Tech Science Press

Cervical Cancer Detection Based on Novel Decision Tree Approach

S. R. Sylaja Vallee Narayan[1,*] and R. Jemila Rose[2]

[1]Department of Computer Science and Engineering, Infant Jesus College of Engineering, Thoothukkudi, Tamilnadu, 628851, India
[2]Department of Computer Science and Engineering, St. Xavier's Catholic College of Engineering, Nagercoil, Tamilnadu, 629003, India
*Corresponding Author: S. R. Sylaja Vallee Narayan. Email: sylajavalleenarayananpapers@gmail.com

Abstract: Cervical cancer is a disease that develops in the cervix's tissue. Cervical cancer mortality is being reduced due to the growth of screening programmers. Cervical cancer screening is a big issue because the majority of cervical cancer screening treatments are invasive. Hence, there is apprehension about standard screening procedures, as well as the time it takes to learn the results. There are different methods for detecting problems in the cervix using Pap (Papanicolaou-stained) test, colposcopy, Computed Tomography (CT), Magnetic Resonance Image (MRI) and ultrasound. To obtain a clear sketch of the infected regions, using a decision tree approach, the captured image has to be segmented and analyzed. The goal of creating a decision tree is to establish prediction model that anticipate the feature vector based on the input variable. This paper deals with investigating various techniques of segmentation for detecting the cervical cancer. It proposes a novel method to develop an assistance system for the detection diagnosis of cervical cancer, based on work of Martin, Byriel and Norup. The analysis is focused on Pap smear pictures of single cells. Smear testing is a method of detecting abnormalities in the blood. Image processing is an effective method for extracting data. It is used to determine the size of cervical carcinoma and the length of the uterus. Martin's database, which is open source and utilised for analysis and validation, is obtainable for research purposes. Cervical malignancy information utilizing three grouping strategies to anticipate the disease and afterward analyzed the outcomes showed that choice tree is the best classifier indicator with the test dataset. Further investigations ought to be led to improve execution.

Keywords: Cervical cancer; image segmentation; level set; threshold; watershed

1 Introduction

Cervical cancer is one of the most frequent types of cancer [1] in women all over the world. It causes early mortality and long-term impairment in women, resulting in a loss of productive life. The lack of knowledge of the condition, as well as access to screening and health care [2], is the major factor. Many cancer control programmers have been undertaken by the Indian government, but owing to challenges with skilled professionals, facilities, transport, regulatory compliance and scanning frequencies, these strategies have struggled to reach remote communities, causing an increase in death rates [3]. Cervical

cancer is currently the globe's 2nd most common malignancy among ladies. The cervix or gateway to the uterus is where this cancer begins. The nucleus of cancer cells is frequently difficult to detect with the bare eyes, which makes it tough for medical professionals to diagnose them. The nuclei [4] of healthy cells are thinner than those of aberrant ones. Abnormal nuclei are bigger, making it difficult to distinguish between stages of cervical cancer with the bare eyes. This is due to the fact that each professional has a unique view on cancer stage categorization by observing the nucleus without exactly decreasing the size of the class. Based on the current report, one woman dies of cervical cancer every seven minutes, and by 2025, one woman will die every 4.6 min [5]. For the last 30 years, although India has maintained a continuous campaign against cervical cancer, it has had little impact on the disease's mortality and morbidity, with the country ranking 4th globally. If detected early, cervical cancer has a very good prognosis. Genetic testing [6] with Pap smear images is one of the most effective ways of detecting and diagnosing precancerous development, even at an early stage. As a result, a system to assist in sample analysis is necessary. Byriel (5), Martin (6) and Norup (7) focused on categorising Pap smear images. Martin and Norup utilised Dimac Imaging's champ programme for segmentation and classification of cervical cells [7]. Martin's primary categorization approaches were hard c-means (hcm), fuzzy c-means (fcm) and ustafson-kessel clustering (gk). Byriel used a neuro-fuzzy classification technique to classify the cervical cells. Byriel used clustering algorithm (fcm), logistic model kessel clustering (gk) and anfis to categorise cervix cells (adaptive network based fuzzy inference system). Byriel classified cervical cells using fuzzy c-means (fcm), gustafsonkessel clustering (gk) and anfis (adaptive network based fuzzy inference system). The nucleus border has indeed been segmented using techniques rely on the watershed transform [8]. We propose a novel method to develop an assistance system for the detection diagnosis of cervical cancer, based on work of Martin, Byriel and Norup. The analysis is focused on Pap smear [9] pictures of single cells. Martin's database, which is open source and utilised for analysis and validation, is obtainable for research purposes.

The cervix, or cervix uteri, is the bottom portion of the uterus. In less developed nations, around 14,000 women have a serious health problem such as cervical cancer as a result of inadequate immunization efforts. It is a serious illness that affects women. It's a volumetric and anthropomorphic radiation [10] treatment technique. Smear testing is a method of detecting abnormalities in the blood. Image processing is an effective method for extracting data. It is used to determine the size of cervical carcinoma [11] and the length of the uterus.

1.1 Cancer

Cancer is a term used to describe a collection of ailments characterised by abnormal cell development that has the tendency to infiltrate or spread to other regions of the body.

Cervical cancer is a kind of malignant tumor in the cervix and spreads throughout the body. It is caused by abnormal cell growth that has the ability to penetrate or disseminate to other cells in the body.

1.2 Stages of Cervical Cancer

Staging is a strategy for bettering cancer treatment and detecting the stages of cancer. Staging is given a higher importance. It is used to distinguish between cancer and non-cancerous conditions, as well as to aid in therapy selection. It is considered non-invasive cancer at stage 0 and metastatic cancer at stage 4 [12].

Stage 0: Cancer cells are only seen on the cervix's epithelium since that stage, not in the underlying tissue.

Stage IA: This is an aggressive cervical cancer that can only be detected using a microscope, not by looking at it.

Stage IA1: The invasion region is no deeper than 3 mm and no wider than 7 mm.

Stage IA2: The invasion region is more than 3 mm in diameter but less than 5 mm deep and 7 mm broad.

Stage IB: It is a somewhat larger tumor than stage IA. Microscopes are unable to view it.

Stage IB1: Only under a microscope can these tumors be noticed. The invasion is no more than 5 mm deep and 7 mm broad, or it may be seen under a microscope but is no more than 4 cm.

Stage IB2: These tumors are greater than 4 cm in diameter and may be seen without the use of a microscope.

Stage II: Beyond the cervix, the cancer has progressed to the pelvic wall and the lower portion of the birth wall, but not to the pelvic wall or the genital canal itself. Malignancy has spread beyond the cervix to the top two-thirds of the vaginal opening but not to the uterus at this stage. Stage IIA1-The tumor is visible without magnification [13] and has a diameter of less than 4 cm. Stage IIA2-The tumor is bigger than 4 cm in diameter and visible without the use of a microscope.

Stage IIB: The malignancy has progressed to the soft tissue the uterus and the upper two-thirds of the urethra, but not to the uterus itself.

Stage III: These malignancies may have progressed to the bottom 1/3 of the genitalia, the hip wall, and/ or the kidneys, in furthermore to the above 2/3 of the genitals and the tissues surrounding the uterus, as mentioned above.

Stage IIIA: The malignancy has progressed to the lower third of the vaginal wall although not to the parietal peritoneum.

Stage IIIB: Cervical cancer can still be defined as stage IIIB for a variety of causes. One of those is if it has spread to the pelvic wall. The latter is if it has obstructed one or perhaps both uteruses (tubes that connect the kidneys to the bladder), causing the kidneys to expand or cease functioning properly.

Stage IV: The tumor has progressed well beyond the cervix, involving the bladder or sphincter membrane, or has migrated to other tissues of the plant in stage IV cervical cancer.

Stage IVA: These malignancies have progressed to the point where they have colonized the bladder, rectum, or both (and have spread to other pelvic organs).

Many academics have recently suggested approaches for detecting and categorizing Pap smear pictures [14] to diagnose cervical cancer. This method can enhance system analysis and design reliability, leading to more efficient information alignment and data and improve results. Some individuals are first diagnosed as being in Stage 2, but following further testing, they are shown to be in Stage 4, with very slim possibilities of recovery. Uneven striping, exaggerated or dead blood characterize depiction on a single cell, and they also overlapped. One of the most amazing capabilities of screening pictures is computerized pelvic ultrasound tracking [15,16] and data collection. The phases of cervical cancer are depicted in Fig. 1.

2 Literature Review

In 2019, Robert P, Celine Kavida, the main purpose of this paper is to classify the microscopic cervical images in order to identify the true impact of cancer that helps the patient to be treated properly. The Pap smear test is most efficient medical test, but it generates problem at the time of interpretation under the microscope. In order to unravel this drawback, automatic cancer detection is developed. This detection process includes few techniques of the image processing such as segmentation, and enhanced SVM (Support Vector Machine) classification algorithm. The final outcome of this proposed technique is compared to previous classification techniques such as ANN (Artificial Neural Network), KNN (K-Nearest Neighbor). The proposed algorithm is found to yield a good result from the experimental results & performance evaluation.
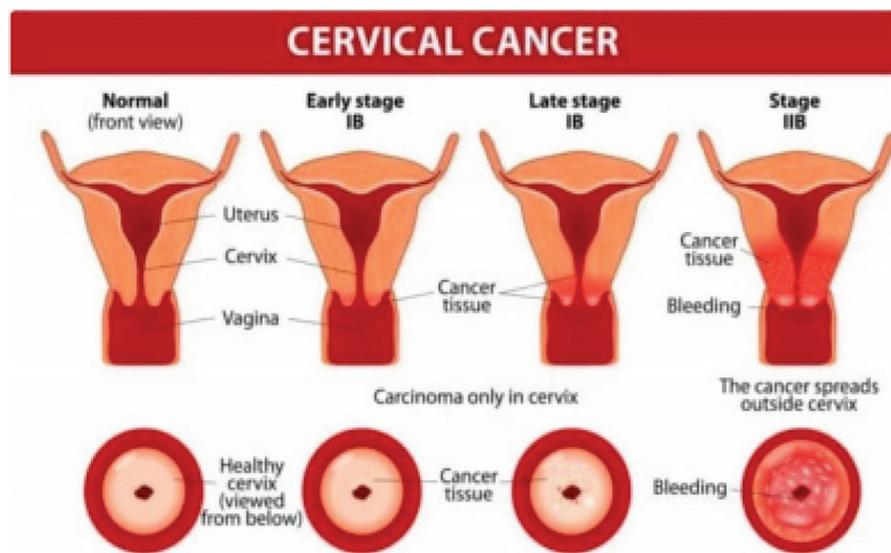
Figure 1: Stages of cervical cancer

Johnes Obungoloch and Wasswa William, Andrew Ware, Annabella Habinka Basaza Ejiri and Andrew Ware will be featured in 2019. On the basis of Pap-smear pictures, a method for the automated detection and categorization of cervical cancer is described in this study. The technique that was employed Image categorization was done using a Trainable Weka Integration classifier and debris rejection was accomplished using a sequential elimination approach. Simulink model integrated with a wrapper filter was used to pick features, while a fuzzy C-means technique was used to classify them. Lastly, they gave outcomes and evaluations of their efforts. The classifier was evaluated using three various algorithms (single cell images, multiple cell images and Pap-smear slide images from a pathology lab). '98.88 percent', '99.28 percent' and '97.47 percent' were achieved for each dataset, respectively. Defuzzification and classifications performance were enhanced due to a rigorous features extraction approach that properly picked cell characteristics that increased segmentation results and the amount of groups employed. Once implemented to the Herlev benchmark Pap smear dataset, the approach surpasses several current schemes in terms of sensitivity (99.28 percent), specificity (97.47 percent) and accuracy (98.88 percent). These results were achieved when applied to Pap-smear slides from the clinical laboratory.

During the year 2018, B Karthiga Jaya, S Senthil Kumar, we provide a cervical cancer identification and classification approach that uses an ANFIS (Adaptive Network-based Fuzzy Inference System) classifier. Enrollment of images, edge detection, classifying, and segmenting are the steps of the proposed system. Images are registered using FFTs (Fast Fourier Transforms). The approved cervical image is then used to obtain the Grey Level Co-occurrence Matrix (GLCM), Grey Level and Trinary Features. Next these extracted features are trained and classified using ANFIS classifier. The ANFIS classifier is then used to train and classify these retrieved features. Cancer is detected and segmented by implement changes procedures on the categorised cervical picture. It has been shown in studies that the suggested technique surpasses existing methodologies in terms of sensitivity, specificity and accuracy.

When S. Yamuna Devi employed a model to solve cervical cancer in 2018, the results were impressive. The picture is subjected to noise reduction to eliminate the noise. Image refinement is used to determine the image's boundaries. Regions of a picture are divided via segmentation. Watershed segmentation and threshold segmentation are the two methods of segmentation utilized in this study. In grayscale pictures, the threshold value transforms grayscale to binary. The threshold value is the primary selection in this

case. This approach of pixel grouping is the most effective. After watershed segmentation, perform thresholding on the watershed segments. Tumor components are separated from a picture using this method. These characteristics are used in the multiclass classification phase of the proposed system. To categorise the tumor, this approach employs a multiclass SVM. It is estimated that 80–90 percent of tumors may be accurately classified using SVM classifier. In medical imaging science, the fine identification of tumorous areas is a difficult problem. Affected areas can be detected using the suggested technique.

In 2018, Lidiya Thampi, Varghese Paul, this article reviews and summarizes various approaches used in digital cervical images and discusses its challenges. The papillomavirus is found everywhere and is the most prevalent unprotected sex viral illness in both male and females. Cervical cancer is typically the most dangerous illness among females in poor nations like India, accounting for up to 25% of all female cancers. Several sophisticated approaches and tools for the early diagnosis of cervical cancer have evolved in modern healthcare. With the advent of the Papanicolaou-stained (Pap) smear or Pap test, the incidence rate tolls of cervical cancer have been reduced by about half to two-thirds. The difficulties stemming from stain fluctuation, cell duplication and lower-level picture characteristics make segmentation of cervical cell nuclei difficult.

Ismaliza Ramli, NurIllani Binti Aziz, Yin Mon Myint, Eko Supriyanto and Christina Pahl were among those honored in 2017. The use of an active contour level set approach is discussed in this work. The cervix can be successfully removed in this way, although the procedure is semi-automatic. In many parts of the globe, cervical cancer is serious harm to women and girls, yet its early signs are not readily apparent. A method for identifying cervical cancer at an early stage is crucial for women to defeat this threat to their lives. To evaluate the cervix, physicians will use three-dimensional (3D) imaging. The goal of this study is to retrieve the cervix from an acoustic picture. This procedure, however, can be difficult since the cervix; bony segment, and accompanying vaginal tissue are not separated on sonar.

Identifying suspicious cells in whole slide cervical cell images using computer-assisted analytical methods was proposed in 2016 by Meng Zhao, Aiguo Wu, Jingjing Song, Xuguo Sun and Na Dong, in this work (WSCCI). The main difference between our technique and the usual computation: instead of segmented cells, the image is split into blocks of a particular size, which reduces computing complexity dramatically. Some statistical characteristics, such as texture and color, indicate substantial changes between frames either with or without problematic cells when analyzed using predictive analytics. As a result, support vector machine classifiers will use these characteristics as input. To create a model, 1100 quasi blocks (110 suspicious blocks) are trained, whereas 1040 blocks (491 quasi blocks) from 12 additional WSCCIs are evaluated to ensure the algorithm's viability. Our technique's reliability is around 98.98 percent, according to the results. More importantly, according to the images analyzed in the study, the susceptibility, which is more dangerous in cancer screening, is 95.0 percent, while the accuracy is 99.33 percent. Finally, unlike previous approaches, the computation analysis is based on block images. Whereas other analytical tasks must be completed ahead of time, the model's creation will significantly enhance processing performance. Furthermore, because the methodology is based on the real WSCCI, the technique will be of obvious diagnostic value.

S. S. Dhumal and S. S. Agrawal were among the top performers in 2015. The need for amazing precision when compared to human life motivates automated categorization and identification of cancers in various medical pictures. Because of the nature of the Tumor cells, detecting a Cervical Tumor is a difficult task. This system proposes a segmentation approach based on a spatial fuzzy clustering algorithm for segmenting Magnetic Resonance pictures in order to discover and evaluate anatomical features in Cervical Tumors in their early stages. The vector machine will be utilised to determine if the Cervical MRI test image is normal or abnormal. To analyze texture, Dual Tree DWT (Discrete wavelet transforms)

multi scale decomposition is utilised. The classification findings will be utilised to diagnose Cervical Tumors earlier, increasing the patient's chances of survival. The decision-making process for an automated Cervical Tumor classification was split into two stages: feature extraction using GLCM (Gray-Level Co-occurrence Matrix) and classification using SVM. This classifier's performance was assessed in terms of training efficiency and classification accuracy.

Cervical cancer, according to N. Sakthi Priya, is one of the biggest killers known, as well as a major study topic in image processing. The primary issue with this malignancy is that it is impossible to diagnose even though it does not manifest symptoms until it is advanced. This is due to the disease itself, as well as a scarcity of pathologists to test the malignancy. Using ultrasonic contouring, a novel technique to classifying different cancers in cervical imaging was presented. We utilized an SVM classifier for categorization, which helped us categorize the stages of cancer and let the physician diagnose the disease more accurately. The picture was put to the test using a variety of photos and proven to be effective.

Sreedevi M. T., Usha B. S. and Sandya S. suggested a new technique for both the early diagnosis of cervical cancer utilising Pap smear scans in a study published in 2012. The most effective effort of clinical research and practise again for early identification of cervical cancer is regular Pap smear examination. Cervical cell analysis by hand is time-consuming, tedious and error-prone. The goal of this research is to create a tool for identifying whether cervical cells are normal or abnormal. It is tested on 80 Pap smear images, and the results show that the approach is equivalent to prior work and offers good results in terms of susceptibility (100%) and accuracy (100%).

3 Overview of Cervical Cancer

Cervical tumor is a kind of malignancy that originates in the cells of the cervix or the uterus's neck. These cell types do not develop into cancer instantly. Conversely, the cervix's healthy cells progressively acquire neoplastic alterations that eventually lead to cancer. When compared to normal cells, cancerous cells have a larger nucleus. This distinguishing trait can be utilized to classify cervical cells as normal or aberrant at the first level. The Fig. 2 depicts normal cell and abnormal cell [Normal cells have a tiny nucleus but a big intracellular region, whereas defective cells have a wider nucleus but a smaller intracellular area.
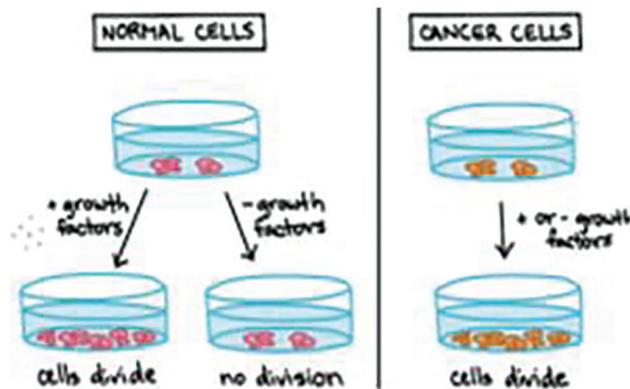

Figure 2: Illustration of normal and abnormal cancer cells

According to a WHO (World Health Organization) publication, the main cause of squamous cell carcinoma is persistent infection caused by pathogenic forms of human papillomavirus (HPV). HPV is a sexually transmitted disease (STD) that is quite prevalent in densely populated areas. It creates a tiny,

solid and innocent growing on the epidermis of non-Hispanics, although this innocuous growth often does not indicate that a patient has HPV. There are many different kinds of HPV, but only a handful of them end up causing cervical cancer. Plenty of the lifestyle factors for STDs also apply to cervical cancer. Cervical cancer is more common in smokers, for example. If cancer is detected early enough, 90 percent of cases can be, treated.

The change that happens in cervix cells in response of HPV infection is seen in Fig. 3. Low-grade alterations called low-grade squamous intraepithelial lesions (LSIL) or cervical intraepithelial neoplasia 1 (CIN1) occur in the early stages. These alterations tend to lessen over time, but they might also increase. Further screening findings show it's not like these alterations are transformed into substantial changes. High-Grade Squamous Intraepithelial Lesions (HSIL) or CIN2 are severe alterations. If not monitored closely, HSIL generally develops into cervical cancer.
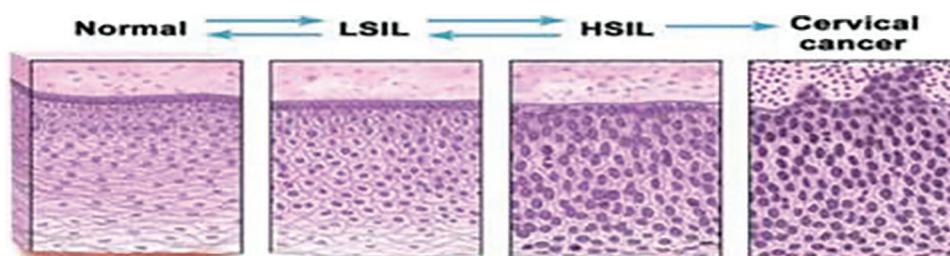

Figure 3: HPV infected cervix cells

As per recent research, developing nations account for roughly 85 percent of new occurrences of cervical cancer. In underdeveloped countries, about 80%–90 percent of cervical cancer cases are diagnosed. Women over the age 35 and up are more likely to be in nations. Cervical cancer is a kind of cancer that affects the female. From precancerous lesion to advanced cancer, it develops slowly. Across the board, females under the age of 25 have a relatively low risk of cancer. However, between the ages of 35 and 40, the incidence rises and reaches its peak. Women in their 50s and 60s have the highest levels of estrogen. Tab. 1 gives the estimation of infected cases and death rate worldwide.

Table 1: Estimated cases & death rate worldwide

Estimated numbers	Cases	%	Deaths	%	5-Years prevalence	%
India	134	25.3	72	26.2	338	21.7
China	75	14.8	33	12	232	15
Europe	31	5.8	13	4.7	1.6	6.8
IARC-membership countries (22)	193	36.4	96	34.8	546	35.2
WHO Western pacific region	105	19.8	46	16.7	332	21.4

4 Classification Types of Cervical Cancer

There are different types of cervical cancer and the Classification types of cervical canceris shown in Fig. 4. Almost any patient with cervical cancer is unique. Cancer Treatment Centers of America® (CTCA) has significant expertise in appropriately assessing and confirming the illness, as well as designing a diagnosis unique to each patient's form of cervical cancer.
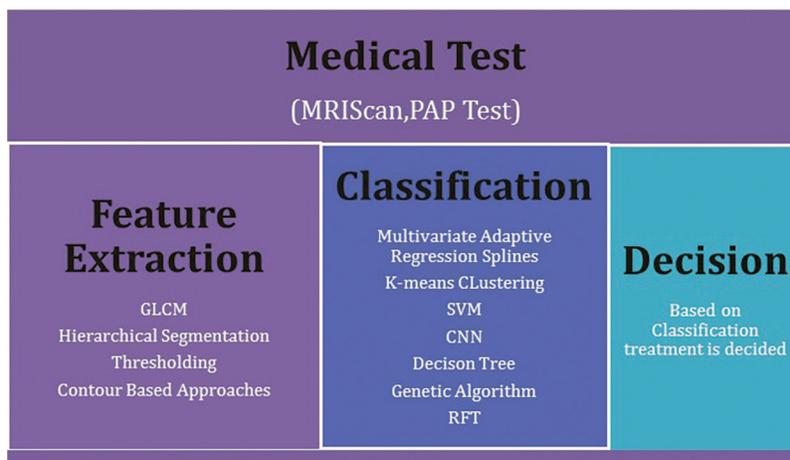
Figure 4: Classification types of cervical cancer

Cervical cancer develops when the cells lining the cervix start to alter abnormally. These aberrant cells may develop malignantly or revert to normal over time. Melanoma and denocarcinoma are the two most common forms of cervical cancer. As seen under light microscopy, each one is differentiated by its cellular composition.

The slender, smooth cells that border the underside of the cervix are the origin of squamous cell carcinoma. According to the National Cancer Institute, this kind of cancer accounts for about 80% of all cervical malignancies. Cervical adenocarcinomas develop from the seminiferous tubules that coat the cervix's top surface. The adenocarcinomas of the cervix account for around 20 percent of all cervical cancer cases. Cervical cancer may involve both types of cells. Cervical cancers of other kinds, however, are extremely rare. Metastatic cervical cancer, for example, begins in the cervix.

The Pap smear test, also known as the Papanicolaou test, is a cervix screening procedure that would be used to diagnose cells in the cervix that are cancerous or on their way to becoming carcinogenic. In a Pap smear test, cells infected with the HPV (Human Papillomavirus) are studied under a camera. All or several of the cells were obtained from the cervix may be malignant, or only one cell may be abnormal; in any scenario, cervical cancer has occurred. The first step in preparing a slide for a Pap-smear test is to fix and stain the cells. This created slide is next examined with a lens under a microscopy and digital pictures are recorded. The digital pictures are then magnified while preserving correct pixel intensity, and then segmented and classified using a variety of machine learning and deep learning techniques. The nuclei image is separated from the cytosol image during the segmentation stage. The picture generated in this phase is known as a segmented image and it is supplied as input into the classification phase, which determines whether the cell is normal or aberrant.

Nodal involvement has a predictive value for cervical cancer and so impacts treatment options. The anatomical parameters for nodal evaluation and categorization in cervical cancer are listed in Tab. 2. The most significant parameter used to distinguish benign from cancerous nodes is nodal size. The Nodes will be deemed suspicious if this axis is more than 1 cm. Nodes bigger than 8 mm are considered suspicious in cervical cancer and have a higher diagnostic accuracy than conventional imaging. The node is shaped like a blob. Benevolent nodes are more likely to be ovoid, while malignant infiltration causes them to become more round. Additionally, if the long-axis diameter to short-axis diameter ratio is less than 2, then it is more likely to be malignant. In ordinary medical practice, the nodal boundary itself can be relied on to identify normal from pathological nodes. Nonetheless, the presence of an ill-defined border might help in the treatment of a patient. The GLCM radiation measure is used to determine boundary

values. In ordinary medical practice, the nodal boundary itself can be relied on to identify normal from pathological nodes. Nonetheless, the presence of an ill-defined border might help in the treatment of a patient. The GLCM radiation measure is used to determine boundary values.

Table 2: Anatomical nodal criteria for cervical cancer cell

Nodal criteria	Normal	Abnormal
Size	<8 mm	>8 mm
Shape	Elliptical	Round
Contour	Regular	Irregular
Appearance	homogeneous	heterogeneous

Sterilization techniques, such as high-risk human papillomavirus (HPV) testing, Pap smear cytology testing, colposcopy and visual inspection of the cervix with acetic acid (*VIA*), are now extensively used, each having their own set of advantages and disadvantages.

- A bimanual pelvic examination: The practitioner does a sensory examination. Optical exams using a biopsy and tactile inspections with fingers are also part of the process. This test is generally followed by a Pap test because it is insufficient on its own.
- Cervical cytopathology: The Papanicolaou Smear (Pap smear), also known as liquid-based cytology, is a technique in which cervical cells are gently scraped and examined under a microscope. Computers can also be used to digitally examine it.
- An HPV testing: Cervical cancer is caused by a chronic infection of the cervix with carcinogenic human papillomavirus (HPV) strains including HPV16 and HPV18. It's generally done in conjunction with a Pap test or if a Pap test reveals abnormal cervical alterations. The presence of HPV does not imply the presence of malignancy.
- Colposcopy: Colposcopy is a procedure that involves using a specific tool called a colposcope to examine the cervix visually. Like a magnifying glass, the instrument magnifies the cervix region under examination. It is safe for pregnant women to use.

5 Methods and Algorithms

One of the heuristic strategies for strategic thinking is adaptive threshold metric. It is mostly used to determine which splitting standard separates the provided information tests the best. Supervised Learning, Entropy and Gain Value are three well-known metrics. The paper presents a novel decision tree approach for deciding the type of cervical cancer based on features extracted from the preprocessed image.

5.1 Attribute Selection Measure

The goal of creating a decision tree is to establish prediction model that anticipate the feature vector based on the input variable. Decision trees are widely used for data supervised learning. Decision tree models produce a categorical range of target variable values, whereas regression tree models produce a continuous set of target variable values. To create a decision tree from the supplied data, each node's attribute must be divided. The associated property will have a potential value for each branch. The splitting attribute is the most informative of all the characteristics. An algorithm uses a factor known as Entropy to choose the most informative characteristic. The quality of a split is defined by the amount of information gained. The attribute with the greatest information gain is thought to be divided. The dataset

is divided into sections based on the values of all the characteristics. An excellent attribute is one that separates information data in such a way that each succeeding node is as good as it looks, that is, each node's distribution of dataset instances must contain only instances of a single class.

To choose the splitting standard that "best" differentiates a given description of the various tests, R, of a C class for turning data points into singleton classes. It determines how the item sets at a specific hub will be combined. These measures beautifully depict the higher-positioning for each quality of the provided tuples. At each hub in the tree, the information gain metric is used to determine the test quality. The attribute with most noteworthy data gain is picked as the test quality for the hub.

5.2 Algorithm

Assume J is a collection of j data sets. The class label attribute contains n different values, indicating that there are n different classes. C_i(for i = 1,2…n) s_i is the count of J samples in the C_i class.

- The information required to categorise a sample is provided by

$I(s_1,s_2….s_n) = -\sum L_i \log_2(L_i)$ where i = 1….n

L_i- the likelihood (probability) that every given sample belongs to the C_i class

- The entropy or expected information based on the partitioning into subsets by H, is given by

$E(H) = \sum(s_{ij}+….+s_{mj}/s).I(s_{ij}…..s_{mj})$ where j = 1….r

- Calculate gain value

Gain (H) = $I(s_1,s_2….s_m)$-E(H)

Evey attribute's knowledge gain is calculated using the technique. For the provided set J, the attribute with the largest training samples is chosen as the test attributes and the comparison chart of different phases of processing are shown in Fig. 5.
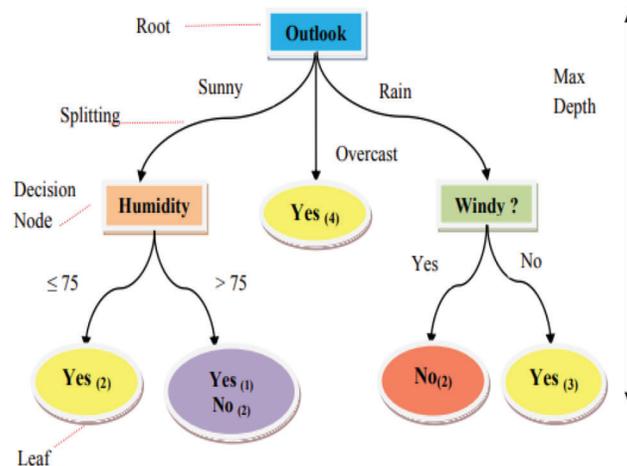

Figure 5: A simple decision tree

5.3 Decision Tree Approach

A supervised classifier with a tree structure made up of nodes is known as a decision tree. Internal nodes and leaves are the two types of nodes, [17]. Internal nodes are also known as decision nodes since they test qualities before making a decision and partitioning the data set. Predicted or decision classes are represented

by leaves. Splitting criteria are used in decision tree construction to partition attributes. Depending on the size of the dataset and the number of attributes, building a tree takes polynomial time.

The dataset is divided into training and testing datasets using decision tree techniques. To build trees, it employs a recursive technique. It creates a trained classifier by recursively partitioning a training set. When the subset at a node has the same prediction value as the target variable, the recursion process is complete. A training dataset is made up of a set of instances, each of which has attributes and a class label, [18]. An attribute's values can be ordinal, real, or Boolean. An example of a decision tree is shown in Fig. 5.
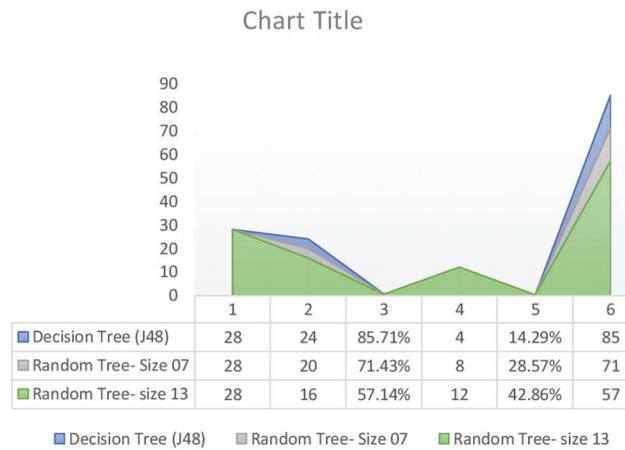
6 Results and Discussion

On a Windows 7 Intel Core i7 CPU (Central processing unit) with a 2.53 GHz operating frequency and 8GB of accessible memory, the implementation is carried out in MATLAB (MATrix LABoratory) R2018. In past findings, single cells did the most of the job, while the Pap smear slides received many and overlapping cells. We obtained cytological pictures from pathology labs in Jaipur. We have 50 shots in total; all of them are in JPEG format. There are 15 normal cells, 20 CIN1 cells, and the rest are CIN2/CIN3. The image is made up of numerous cells that are overlapping. The attribute is labeled on a node, branches are formed for each value of the attribute and the data are subdivided properly and the predictor variables are tabulated in Tab. 3.
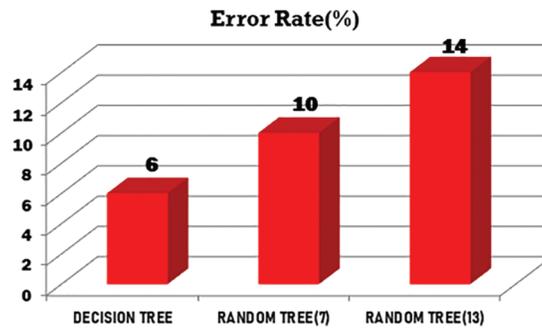
Table 3: Predictor variables

| Predictor variables	Total number of instances	Correctly classified instances	Percentage	Incorrectly classified instances	Percentage	Detailed accuracy by class
Decision tree (J48)	28	24	85.7143%	4	14.2857%	85
Random tree Size of the tree: 07	28	20	71.4286%	8	28.5714%	71
Random tree Size of the tree: 13	28	16	57.1429%	12	42.8571%	57

Fig. 6 shows the comparison of different phases of processing. The error rate of the decision tree and random tree algorithms of sizes 7 and 13 is shown in Fig. 7. The decision tree algorithm has a minimal error rate. Fig. 8 shows the performance data for the texturing feature. It is evident that the accuracy of the proposed method has shown a tremendous improvement regarding detection of cervical cancer based on GLCM features. Fig. 9 depicts the accuracy measurements for the various categorization techniques presented. The goal of this study is to take a comprehensive and in-depth look at the latest state of the art in automated cervical cytopathology cell segmentation and categorization. We searched databases such as PubMed, arXiv, Google Scholar, IEEE (Institute of Electrical and Electronics Engineers), ACM (Association for Computing Machinery), Springer and Elsevier to create this complete review. We also double-checked the references in all of the articles we chose.

Chart Title

	1	2	3	4	5	6
■ Decision Tree (J48)	28	24	85.71%	4	14.29%	85
■ Random Tree- Size 07	28	20	71.43%	8	28.57%	71
■ Random Tree- size 13	28	16	57.14%	12	42.86%	57

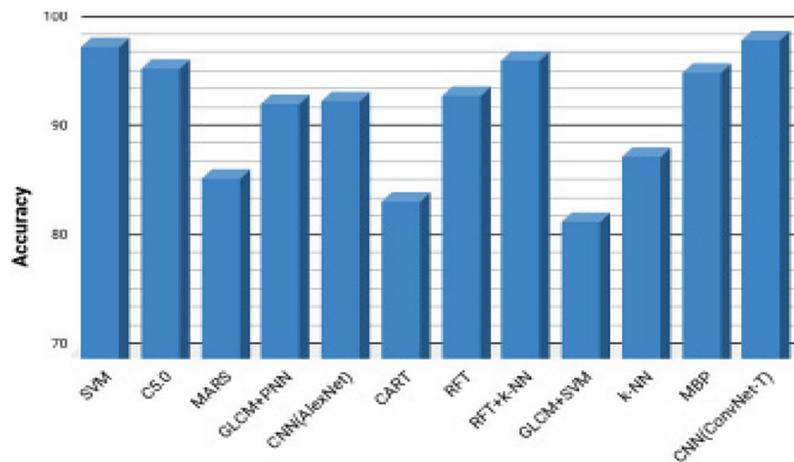■ Decision Tree (J48)   ■ Random Tree- Size 07   ■ Random Tree- size 13

Figure 6: Comparison of different phases of processing

Figure 7: Graphical representation error rate
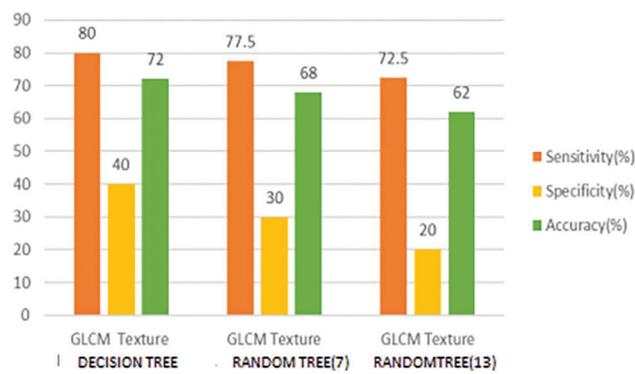
Figure 8: Graphical representation of texture for various techniques

Figure 9: Graphical representation of accuracy for various algorithms

7 Conclusion

In this study, we looked at several cervical cancer screening techniques. In comparison to medical diagnostics, a combination of image preparation and careful computation techniques produced more exact findings for detecting cervical cancer. Finding is done through phases such as preprocessing, division, highlight extraction and arranging, all of which use complex technologies to produce precise findings. According to the evaluation, the highest quality results will be obtained with greater precision when a medical image is equipped with middle sifting/Gaussian channel systems in the preprocessing stage, with the adaptive division method in the division stage, by selecting shading as a component alongside island evacuation post handling strategy and with choice tree as classifier. When these techniques are combined and conducted on a predetermined clinical picture, the underlying stage of cancer can be determined. There is different information mining procedures that can be utilized for the forecast of cervical malignancy. In this paper, broke down cervical malignancy information utilizing three grouping strategies to anticipate the disease and afterward analyzed the outcomes. The paper likewise gives a thought of the trait choice measure utilized by different choice trees calculations like irregular tree utilizes data gain, the J48 calculation utilizes entropy and choice tree calculation utilizes most noteworthy increase value as the property determination measure. The paper likewise gives the techniques for figuring of these property choice measures. Taking all things together, we find that these calculations for choice tree enlistment are to be utilized at various occasions as per the circumstance with less computational time. The outcomes showed that choice tree is the best classifier indicator with the test dataset. Further investigations ought to be led to improve execution of these order methods by utilizing more factors and deciding for a more drawn out subsequent span. In future more number of traits can be gathered and recognize the irregularities without any problem.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] K. A. Dinshaw, S. S. Shastri and S. S. Patil, "Cancer control programme in India: Challenges for the new millennium," *Health Administrator*, vol. 7, no. 1, pp. 10–13, 2005.

[2] M. E. Plissiti, C. Nikou and A. Charchanti, "Combining shape, texture and intensity features for cell nuclei extraction in pap smear images," *Pattern Recognition Letters*, vol. 32, no. 6, pp. 838–853, 2011.

[3] P. Robert and A. C. Kavida, "Classification of microscopic cervical cancer images using regional features and HSI model," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 8, pp. 24–28, 2019.

[4] W. William, A. Ware, A. H. B. Ejiri and J. Obungoloch, "A pap-smear analysis tool (PAT) for detection of cervical cancer from pap-smear images," *Biomedical Engineering Online*, vol. 18, no. 1, pp. 1–22, 2019.

[5] Z. Meng, A. Wu, J. Song, X. Sun and N. Dong, "Automatic screening of cervical cells using block image processing," *Biomedical Engineering Online*, vol. 15, no. 1, pp. 1–20, 2016.

[6] N. S. Priya, "Cervical cancer screening and classification using acoustic shadowing," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 1, no. 8, pp. 1676–1679, 2013.

[7] M. T. Sreedevi, B. S. Usha and S. Sandya, "Pap smear image based detection of cervical cancer," *International Journal of Computer Applications*, vol. 45, no. 20, pp. 35–40, 2012.

[8] A. Nithya, A. Ahilan, N. Venkatadri, D. R. Ramji and C. A. Palagan, "Kidney disease detection and segmentation using artificial neural network and multi-kernel k-means clustering for ultrasound images," *Measurement*, vol. 149, no. 2, pp. 106952, 2020.

[9] R. Sundarasekar and A. Appathurai, "Efficient brain tumor detection and classification using magnetic resonance imaging," *Biomedical Physics & Engineering Express*, vol. 7, no. 5, pp. 055007, 2021.

[10] L. Thampi and V. Paul, "Automatic segmentation and classification in cervical cancer images: Evaluation and challenges," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 12, pp. 12549–12560, 2018.

[11] L. Zhang, H. Kong, C. T. Chin, S. Liu, X. Fan *et al.,* "Automation-assisted cervical cancer screening in manual liquid-based cytology with hematoxylin and eosin staining," *Cytometry Part A*, vol. 85, no. 3, pp. 214–230, 2014.

[12] D. R. Ramji, C. A. Palagan, A. Nithya, A. Appathurai and E. J. Alex, "Soft computing-based color image demosaicing for medical image processing," *Multimedia Tools and Applications*, vol. 79, no. 15, pp. 10047–10063, 2020.

[13] W. William, A. Ware, A. H. B. Ejiri and J. Obungoloch, "Cervical cancer classification from pap-smears using an enhanced fuzzy C-means algorithm," *Informatics in Medicine Unlocked*, vol. 1, no. 14, pp. 23–33, 2019.

[14] K. Subrata and D. D. Majumder, "A novel approach of mathematical theory of shape and neuro-fuzzy based diagnostic analysis of cervical cancer," *Pathology & Oncology Research*, vol. 25, no. 2, pp. 777–790, 2019.

[15] S. Manas, C. Sukkasem, S. Sasivimolkul, P. Suvarnaphaet, S. Pechprasarn *et al.,* "Automated screening of cervical cancer cell images," in *Proc. Biomedical Engineering International Conf. (BMEiCON)*, Chiang Mai, Thailand, pp. 1–4, 2018.

[16] R. Sundarasekar and A. Appathurai, "Efficient brain tumor detection and classification using magnetic resonance imaging," *Biomedical Physics & Engineering Express*, vol. 7, no. 5, pp. 055007, 2021.

[17] P. A. Harrison, R. Dunford, D. N. Barton, E. Kelemen, B. Martín-López *et al.,* "Selecting methods for ecosystem service assessment: A decision tree approach," *Ecosystem Services*, vol. 29, no. 2, pp. 481–498, 2018.

[18] F. Y. Osisanwo, J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi *et al.,* "Supervised machine learning algorithms: Classification and comparison," *International Journal of Computer Trends and Technology*, vol. 48, no. 3, pp. 128–138, 2017.