Sustainable Machine Intelligence Journal

Journal Homepage: sciencesforce.com/smij

Sustain. Mach. Intell. J. Vol. 9 (2024) 57–78

Paper Type: Original Article

Revolutionizing Early Detection: The Influence of Machine Learning in Esophageal Adenocarcinoma Diagnosis

Rabia Younas [1], Hafiz Burhan Ul Haq [1,*] and Hafsa Naveed [2]

[1] Department of Information Technology, Faculty of Computer Science, Lahore Garrison University Pakistan.
Emails: rabia.younas@lgu.edu.pk; burhanhashmi64@lgu.edu.pk

[2] Department of Software Engineering, Faculty of Science, University of Lahore, Pakistan; hafsa.naveed@se.uol.edu.pk.

Received: 05 Mar 2024 **Revised**: 24 Aug 2024 **Accepted**: 25 Sep 2024 **Published**: 01 Oct 2024

Abstract

Machine learning is an analysis technique of big data that uses computers and mathematical models to develop algorithms. These algorithms are applied in many situations and one of the most common uses is in developing prediction models. Machine learning is the activity of making prognosis of other data points which are related to the self using key factors from a dataset. It is designed to effectively predict the dysplasia within a given setting using machine learning algorithms on a dataset with primary focus on Barrett's Esophagus. The study investigates the use of five classification methods: As per the classification, the techniques identified include Naïve Bayes, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbors, and Logistic Regression. Barrett's Esophagus Below are the methods applied using and without Principal Component Analysis (PCA) on the special dataset of Barrett's esophagus. The primary objective is the identification of the effectiveness of each algorithm by comparing their functional characteristics, such as F1 score, recall value, accuracy percentage, precision ratio, and specificity coefficient. From the results obtained from the experiments, two algorithms that yield good accuracy for predicting Barrett's esophagus in this dataset are the SVM algorithm with PCA and Logistic Regression algorithm with PCA and both algorithms returned review ratings of 1. undefined This career has earned a maximum Score of 0.1000, Naïve Bayes perform well in the assessment with 0.964 of accuracy, it followed by K-Nearest Neighbors with 0.9349 and Logistic Regression with 0.923. While working on it, it is found out that SVM has an accuracy of 0%. undefined the south region performing with an average of 0. In total, 834 is the sum of the recognition rate of decision tree with the smallest recognition rate of all the five algorithms.

Keywords: Machine Learning Algorithms; SVM; Principal Component Analysis; Esophageal Adenocarcinoma; Early Detection; K-Nearest Neighbour; Actual Negative; Actual Positive; Predicted Negative; Predicted Positive.

1 | Introduction

The development of esophageal adenocarcinoma is preceded by Barrett's esophagus. Since esophageal disorders have only recently been identified, the appearance of dysplasia in Barrett's esophagus offers a chance for possibly curative therapy. Before becoming malignant, normal cells undergo several profound alterations that lead to their transformation. Notable phases in this process are high hyperplasia and dysplasia. An increase in the number of cells inside an organ or tissue is known as hyperplasia, and it may be seen under a

Corresponding Author: burhanhashmi64@lgu.edu.pk

https://doi.org/10.61356/SMIJ.2024.9383

microscope. Dysplasia, which is observable under a microscope as well, denotes aberrant cellular alterations that are not yet malignant. Cancer may or may not result from either dysplasia or hyperplasia. Regrettably, individuals frequently receive advanced diagnosis, which makes it difficult to receive adequate therapy and frequently comes too late. Figure 1 shows a diagram of the human upper digestive system, specifically highlighting the esophagus and the stomach. The esophagus is labeled as the tube that connects the throat to the stomach, while the stomach is shown as the organ where digestion begins after food passes through the esophagus.

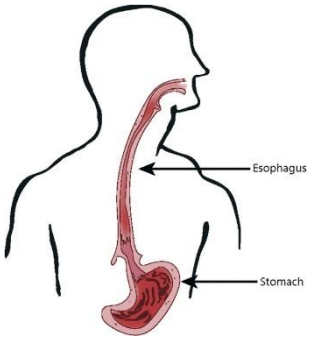

Figure 1. Esophagus in human body.

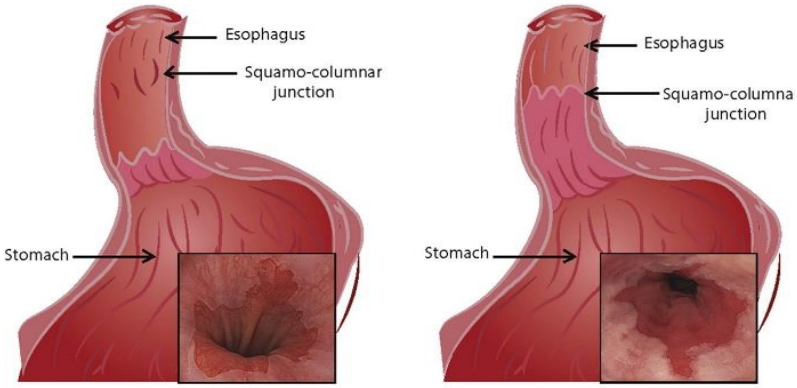

Figure 2. Endoscopic view of Esophagus.

Cancer is prone to accurate risk assessment and estimation of prognosis, tumor node metastasis (TNM) staging independent indications have become an essential accessory in cancer research where the main element could be approximate portrayal of outcomes for patients with esophageal cancer. The classification of cancer patients into high-risk and low-risk groups, And the increasing demand for application in biomedical and bioinformatics various research teams have started to investigate this phenomenon by using Artificial Intelligence (AI). Figure 2 provides a more detailed view of the junction between the esophagus and the stomach, specifically highlighting the squamo-columnar junction (also known as the Z-line). The image on the left shows a normal esophagus and stomach with a clear and healthy squamo-columnar junction. The image on the right shows a condition where the squamo-columnar junction appears abnormal, which could indicate a medical issue such as gastroesophageal reflux disease (GERD) or Barrett's esophagus.The insets in both images show close-up views of the esophagus-stomach junction.

In effect, these have become a de facto template upon which the decipherment and amelioration of numerous disease states has been based. In addition, the capability of AI tools to identify important elements from complex data sets highlights their importance. A novel image enhancement technology, I-Scan is used for optimized visualization of the gastro esophageal junction to facilitate detection abnormalities in this area while using ulcerative findings reflecting underlying interobserver and intra-obsever agreement. The principle of I-Scan is the modification of light wavelengths aimed at microstructures and microvasculature visualization [1]. A consistent issue is the diagnosis of Barrett's esophagus (BE). The endoscopic surveillance of Barrett's esophagus is now done in the diagnostic phase to identify regions at high risk for Esophageal

Adenocarcinoma. Current surveillance is based on white light endoscopy sampling of evenly spaced quadrants every 2 cm over the Barrett's esophagus region [15]. Machine learning (ML), is the name of a practice series that involves computerized processing and prediction under mathematical computations, applied at one Diagram concerning dysplasia [3] as it constitutes Particular Data science. We employ diverse ML based approaches in this study for predicting affected individuals from the disease.

2 | Related Work

A dysplasia delegate against degree. The cell layer of inner human progress is present in cells with light dysplasia. 66% of its thickness and severe dysplasia are features of moderate disappointments. when disposable cells comprise the whole layer of cells. When cells penetrate the basement layer, intestinal cancer develops. That means that the injuries disappear since the Barrett instructor's perspective is less than 5%. The workshop members hope to clarify that Barrett's esophageal reflux disease (GERD) screening and observation are not inherently harmful, but rather that the quality of the currently available evidence does not yet demonstrate that Barrett's GERD screening and observation are beneficial or practical. It is anticipated that professionals would find the workshop's outcomes useful when dealing with challenging administrative problems. Furthermore, these findings could also help clinical scientists identify areas in which additional information is needed and provide guidance to physicians on clinical matters [4]. There is also growing evidence that using the I-filter at the end of colorectal polyps increases the accuracy of dysplasia detection [5, 6].

A variety of information rating computations are available to assign a rating to each of the two or more kinds of information. Professionals use the decision tree as soon as time permits, most of the time. In three endoscopic symptomatic endoscopic assessments, enter the "WEKA" ML bundle to help us with the expectation model development [7]. machine learning approach was used to predict when the helicopter would be ready [8]. Using Bayesian systems, expectation persistence in skin diseases is frequently shown by the hereditary computation [9]. For example, the paper employed computerized reasoning in the field of wellbeing to conclude using inductive and Bayesian learning [10]. An intuitive electronic learning methodology that may be applied to enhance the structure and size of Barrett's throat-related [11]. Assess the importance of student performance in continuing endoscopy: That in the BB part is the primary question. The client's structure has to be made clear by this framework since fear or warnings alone are not enough. In order to detect dysplasia in Barrett's esophageal patients, recommend standard definition as opposed to white-white light endoscopy during plan It is possible to fully disregard Barrett's dietary checks utilizing HD and SD endoscopic structures rather than microvascular models. The customer's structure has to be explained by this system because just alerts or nodularity are not seen. It may be significant since, in the course of developing our I-Scan extraordinary strategy for treating Barrett's throat dysplasia, we have discovered that authorities frequently have more typical microvascular structure than test level examples [12].

The application of machine learning in the healthcare domain has grown rapidly. For example, machine learning has recently been used to enhance composed discovery with medications and boost cancerous development [15-16]. It is expected that the information is necessary in order to integrate a specific reaction variable, even if the prediction and clarification objectives have the potential to undermine the major differentiation. The proper answer might be numerical or corresponding, therefore classifying predictive data mining into temperature and relapse separately. Thus, in essence, the algorithms make use of quiet explicit data to assist experts in forecasting or locating clinical reviews [17]. A notable model that incorporates IBM's "Watson Path Project," which enhances clinical understudies at the Cleveland Clinic's preparation and symptomatic abilities. A number of models have been established, including AI for melanoma, affectability rate for CVD, risk of recurrent breast cancer, and thyroid disease detection [18].

Led an evaluation with endoscopic ultrasound (EUS) to determine textural memory for a pixel-by-pixel analysis to provide a measurement of the early tissue of oesophageal cancer. The usual final results of

correctness, affectability, specifiability, and negative estimate were solely 89.4% and 94% when just the early oesophageal cancer want was taken into consideration [19].

The SVM classifier was carried out on the data set of 30 images with dysplastic BE and 30 images without dysplastic BE. Three features associated with the remarkable Heraldic characteristics were designed more The three features based on the Heraldic traits were Minibar, Red Carpet and the VIP section of 'White Wedding'. On the other hand, research in this paper lingers at 0. 81 attained by the clinical want mannequin, the makers received a area under the Receiver Operating Characteristic (ROC) twist of 0. 95 using a 10-overlay cross-endorsement display[20]. An approach for small-scale evaluation of the stomach-related tract motility has presented that does not require the identification of characteristics for every motility event by using Deep CNN. Namely, the particular technique described in the work achieved a mean accuracy of 96% of six intestinal motility events [21].

Employing an ignore some cross-endorsement (LSOCV) approach, it was further validated that perfect models can be achieved concerning the production attributes of BE recognised evidence for CNN categorization. Thus, sensitivity and specificity were 82 and 80%, respectively whereas accuracy of PDS, AUC and overall accuracy were 81%, 79% and 81% [22]. Four arrangements have been identified from the datasets (125 WLE photos, 122 NBI pictures, and 150 Chromo endoscopy views): The Good group contains samples of Normal Squamous, Gastric Mucosa, BE and High-grade dysplasia which is also known as adenocarcinoma. The accuracy ranged from 36:SVM and KNN are most widely used classifiers and according to the endoscopic approach images and classifiers used it varied from 36% to 89:17% [23]. Twenty-one patients were identified by their respective practitioners as having a squamous cell danger and recommended for endoscopic therapy; examinations of mucosal sites were attained by confocal imaging out of which 43 images were taken of injuries. Based on the k-bits of information, the total hits marked a success rate of 95 %, the affectability was at 100 %, while the specificity at 87 %. Inter-observer agreement regarding the stages involved was acceptable (kappa 0. 79), and intra-observer agreement almost perfect (kappa 0. 95) [24]. Furthermore, decision trees might be able to introduce fundamental principles for the dysplasia's termination according to the creators. In a dataset of 47 HD endoscopic records of the throat (23 dysplastic and 24 non-dysplastic), the researchers employed two decision models: while reviewing study 2, the author described one group as the ace decisions about dysplasia and no-dysplasia group and another group as the non-ace decisions group. As for the cited models, the total hit rate was approximately 79 % (when the decision was made by the authorities) and 77 % (when the decision was made without the experts' influence) [25].

Another learning technique that utilised many event perspectives was suggested to employ the computed tomography to diagnose the tumor interruption of the gastric risk. To achieve affirmation speeds of approximately 76, the developers proposed improving the Citation k-NN (Citation-kNN) [26]. In accordance with the nine VLE volumetric datasets of seven patients, relative to the extreme BE length at the pre-RFA SGS characteristics, alterations in the most extreme BE length at the follow-up phase post-RFA were found to be significantly correlated. Challenges and their importance towards recognition of SGS [27]. As mentioned in the previous paragraphs, several algorithms have been used to classify the dysplasia and the dataset was used to capture this accurately. The rationale of this project is to help to coordinate and develop standard training, as well as performing qualitative analysis protocols for the staff that performs endoscopy in safely resected Barrett esophagus. In this study, five classification methods were used such as KNN, SVM, and implementing PCA. The first aim was to detect the actual amount of details of output and feasibilities with regards to vale of information that can be accurately recalled and correctly specified. The exploratory result further proves the use of vector analysis along with PCA technique for analysis. With regards to classification efficiency, KNN performed well and had a score of 97% and SVM also scored well with a value of 91% [31].

Table 1. Feature-based algorithms applied to the Dysplasia Database.

Author	Classifier	Year	Specificity (%)	Accuracy (%)
Rosenfeld et al. [25]	D-Tree	2014		79%
Serpa-Andrade et al.[28]	KNN and Random Forest(RF)	2015	72%	81%
Boschetto et al. [29]	RF	2016	87.3%	83.9%
Zhang et al. [26]	SVM	2016	95%	89.4%
Swager et al. [20]	SVM and KNN	2017	93%	95%
V. Sehgal et al. [3]	DT	2018	88%	80%
S. Hayat et al. [31]	SVM and KNN	2022	88%	91%

Recent studies by Saqlain et al. [32-34] emphasize the increasing use of artificial intelligence in multiple disciplines, including education software engineering and foreign trade. For example, AI with deep learning is used to more accurately score the readability of English teaching material, which allows for a better design of curriculum and therefore greater benefit for student comprehension in education. Software Engineering: Machine learning models are used to predict software development efforts, helping project managers with resource allocation and timeline planning. Moreover, at a time of global trade and investment uncertainty AI offers autonomous risk assessment capabilities (through the use edge cloud computing and deep learning) that empower more informed action by safely ensuring market integrity behind international borders. This highlights how AI is changing the way decisions are taken, removing efficiency and accuracy improvements from a variety of domains.

3 | Methodology

The suggested techniques' functional sequence and implementation are examined in detail. A comprehensive overview outlines how various processes will operate and collaborate to achieve the fundamental goals. Figure 3 presents a visual representation of the proposed methodology's complete sequential flow. The specifics of the suggested approach are as follows:

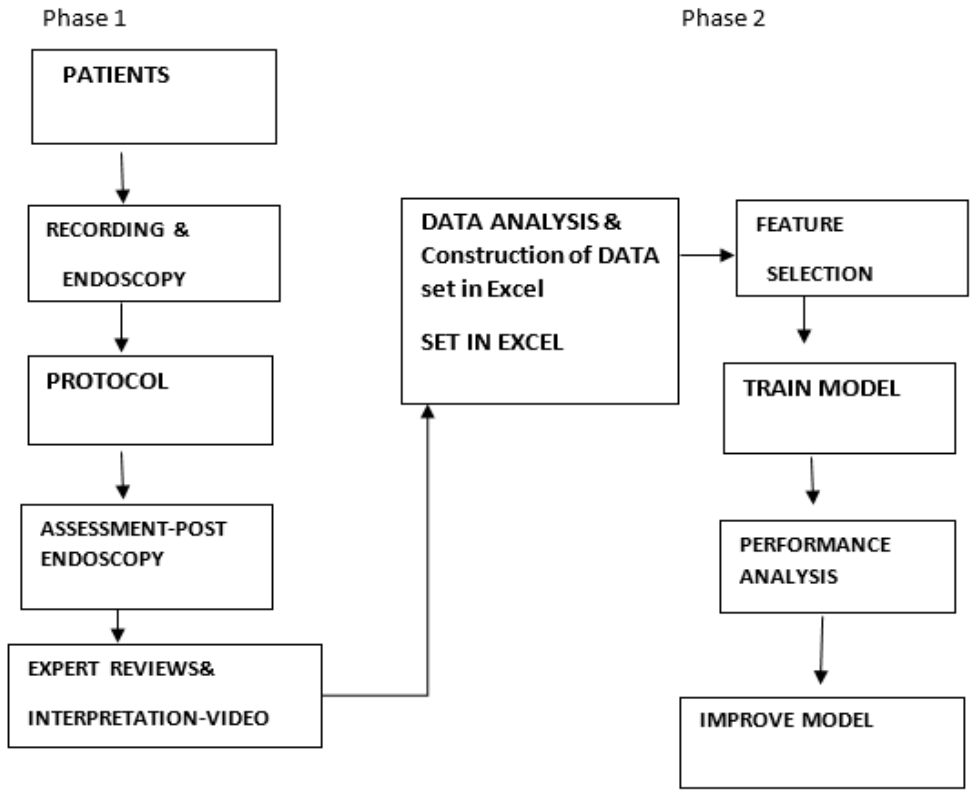

Figure 3. Visual representation of proposed methodology.

3.1 | Patients

The patient information utilized in this study was collected under the direction of Dr. Irfan from the liver clinic in Shadman, Lahore. For this study, 200 patients were chosen, with a mean age of 43 and 130 men and 70 females.

Table 2. Patient's details.

Parameter	Count
Total Patients	200
Gender	
Female Participants	70
Male Participants	130
Mean age(in year)	43.4
Gut preparation	
Fair	110
Excellent	40
Good	50

3.2 | Recording and Endoscopy

Barrett's Esophagus is a condition that has waned attention from the medical fraternity despite affecting a significant number of patients especially in the liver clinics where the condition often manifests itself. During a period of 12 weeks, 200 patients attending Liver Clinic Shadman were considered for a more extensive study of Barrett's Esophagus within that population. Barrett's esophagus is a premalignant disease, in which the lower esophageal squamous mucosa is changed for metaplastic columnar epithelium, which is a precondition for malignant neoplasms of the esophagus, predominantly adenocarcinoma.

These 200 patients were all taking proton pump inhibitors, a class of medications commonly prescribed to manage gastroesophageal reflux disease (GERD) and its associated symptoms. To thoroughly examine the esophageal mucosa, the patients underwent high-definition (HD) video gastroscopy using state-of-the-art Pentax endoscopy equipment, specifically the EG-2990i or EG29-i10 models. The procedures were performed by one of three highly experienced and trained endoscopists, all of whom had a specialized focus on the detection and management of Barrett's esophagus.

The endoscopic procedures entail general patient relaxation sometimes under anesthesia due to the nature of the procedures Both midazolam which is a benzodiazepine and fentanyl which is an opioid analgesic are used in intravenous sedation. In some cases, when strong contractions of the muscular tissue of the esophagus were registered, an antisymioticy agent – hyoscine-N butylbromide or buscopan was given intravenously to decrease the contraction of muscles of the esophagus. Before the examination by the endoscopic technique, the sites which were going to be inspected were rinsed with clean water in order to avoid formation of mucus or debris in the way hence enhancing the endoscopic view. To improve the visibility of the esophageal mucosa and potentially seen less visible sign of BE, N-acetylcysteine was administered as mucolytic agent. It helps in removal of debris and thick secretions which in turn helps in observing the colour and nature of the epithelium more easily. The endoscopist said the whole procedure would be recorded at intervals with the practice mode of WLE, and another type of mode called I-examine, which would probably have provided more concrete results. In particular, endoscopic biopsies were taken from each of the zones regarded to have BE and HPE was performed with the aid of special stains which is today considered definitive for BE. ACA was applied to the mucosa by the endoscopist before taking biopsy samples in 12 of the patients, i.e., 3% of the total number of suspected Barrett's esophagus patients. Such an approach can facilitate differentiation between normal and dysplastic or neoplastic changes in the, esophageal epithelium because of the increase in ACA fluorescence.

Using this intensive, approaches to endoscopic examination and biopsy sampling, the purpose of this research was to provide detailed description of prevalence and characteristics of Barrett's esophagus in the this treated

PPI patients which may potentially benefit better understanding of this significant precancerous lesion. Endoscopy is a difficult technique of examination that is quite time consuming and demands the full concentration of the endoscopist, which is the doctor or healthcare personnel who undergoes the examination. During these procedures the endoscopist patient safety and welfare is paramount, this is together with the correct diagnosis of any form of disease in the patient.

Due to the complexity of endoscopic procedures, the endoscopist lacks a formal process of recording the examination it is considered a secondary task. The endoscopist's hands are usually engaged in manipulating the endoscope, a long slender tube containing a camera and light at the end, through the patient's body and using it to accomplish a view of the internal structures. A lot of care has to be taken while handling the endoscope and maneuvering it around, to ensure that the patient is not in any way harmed or uncomfortable during the process. Also, the endoscopist has to focus on the displayed images, to assess, and to decide during the procedure what actions should or should not be further undertaken.

It is here important to note that if the endoscopist were to shift concentration to recording the examination, then the above crucial tasks may be interfered with to the detriment of the health of the patient. Sometimes the endoscopist himself may decide to capture the event for teaching or study, yet this is normally allowed with patient's permission and in a way that does not impede the goal of the assessment. The endoscopist's main accountability is the overall well being of the patient focusing more closely and diligently than any to make the endoscopy as safe and successful as it could be.

3.3 | Protocol

Two highly skilled gastrointestinal pathologists, Drs. Irfan and Mehreen, closely evaluated each biopsy sample. To guarantee neutral judgments, these pathologists were not informed of the endoscopist's opinion. An revised Vienna Classification, a standardized approach for identifying gastrointestinal lesions, was used to the biopsy samples in order to analyze them [26]. When dysplasia was suspected, both pathologists examined the samples separately and came to a mutually agreed upon conclusion.

3.4 | Assessment Post Endoscopy

After the procedure, the data and video recordings were transferred to a secure memory drive in order to ensure data integrity and secrecy. Subsequently, the video clips were edited using Apple Inc.'s iMovie software to create a short, sixty-second movie that summarized the key points of the second "I-scan" mode stories.

After the relevant histology findings were acquired and examined, the data was anonymized to protect patient privacy and transferred to a structured database. The expert reviews and the films were uploaded to the database for future study and reference.

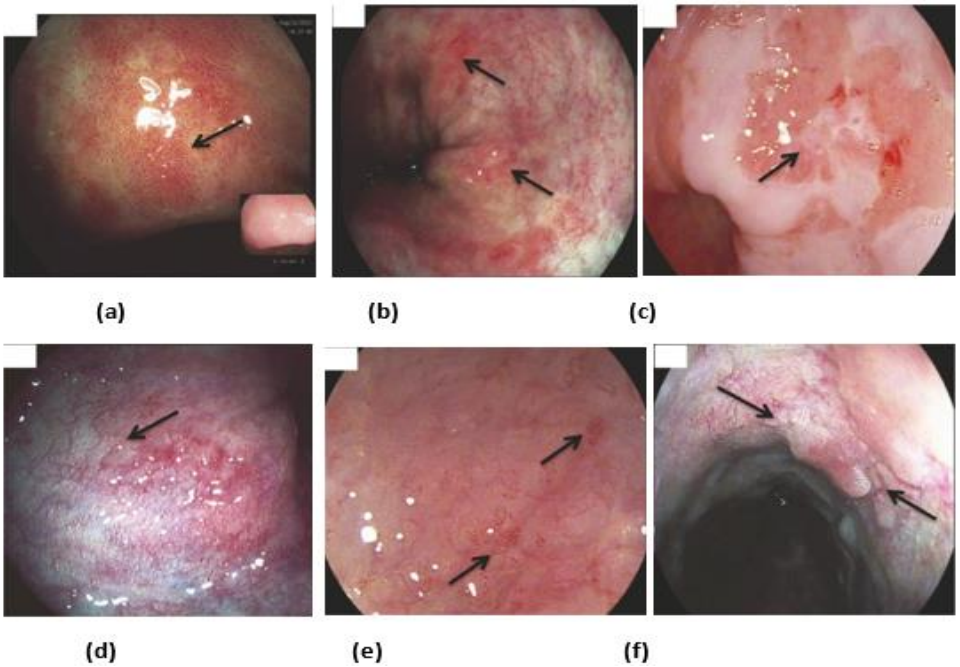

Figure 4. Endoscopic images showing various mucosal and microvascular patterns, and the design of Barrett's fragment using I-Scan. Arrows indicate specific areas of interest. (a) M1: typical round or oval pits; (b) M1: villous/gyrus grooved pits; (c) M2: distorted, featureless mucosa with surface ulceration; (d) V1: normal uniform vessels; (e) V2: irregular, "corkscrew-like" vessels; (f) nodularity with surrounding distortion of the normal mucosal surface pattern.

All of the video recordings were inspected by four endoscopists: Drs. Irfan, Aftab, Mehreen, and Hussain. An endoscopist was mandated to provide data for each film, which comprised specific observations, M and V scores, and the results of histological analysis (ND-BE or D-BE)[26].

When the endoscopists were contacted for the video analysis, they were asked to rate how confident they were in their assessment. This score was based on a 0–1 scale, where 1 represented complete certainty and 0 represented absolute ambiguity in their evaluation.

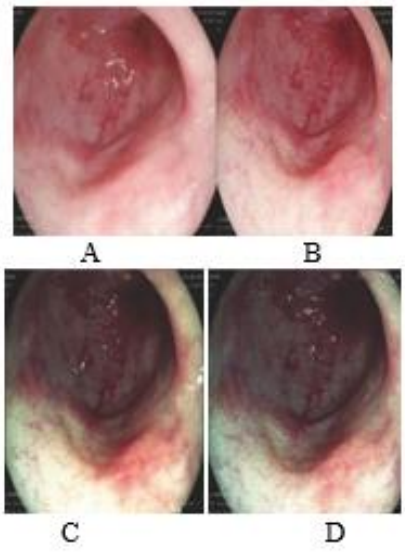

Figure 5. Vascular and Mucosal.

Image A:

This image likely shows a segment of the esophagus with visible mucosal patterns. The mucosa appears slightly inflamed with a regular pit pattern, possibly indicative of M1 characteristics, such as regular circular or villous pits.

Image B:

This image can look like Image A, although the mucosal pattern could look a little different or could be of a different focus. Perhaps it could be M1 features but magnified, or the pit pattern might be more obvious or there could be the slightest alteration to the coloration.

Image C:

It is still another pattern that seems to represent the segment of the esophagus. The pits may look less symmetrical or, in some cases, appear as small craters, which could indicate M2. The mucosa might look more irregular or ulcerated, but this would depend on the type of ulcer.

Image D:

There is also an indication that this image displays an esophageal segment with mucosal and vascular patterns. There might be better defined, smooth, and regular margined vessels, seen as an equivalent of V1 vessels. On the other hand, if the 'vessels appear dilated and tortuous it could be just an appearance of V2 features. Collectively, these endoscopic views show choice appearances of the esophagus and demonstrate different mucosal and vascular patterns that are important in diagnosis such as those found in Barrett's esophagus and the identification of dysplasia [30].

Table 3. Example Dataset.

Id	Diagnosis	Radius_ mean	Texture_ mean	esophageal _regular circular	esophageal _irregular pits	esophageal_ uniform for vessels	esophageal_dilated tortuous vessels
12668	Dysplasia	22.99	20.38	0.56	0.76	0.86	0.55
12669	Dysplasia	33.57	18.77	0.77	0.88	0.67	0.84
12670	No-dysplasia	29.69	21.25	0.35	0.46	0.45	0.34
12671	No-dysplasia	15.42	20.38	0.45	0.38	0.52	0.43
12672	Dysplasia	34.29	14.34	0.69	0.74	0.82	0.61

3.5 |Evaluation Measures

In this research, the effectiveness of the machine learning algorithms is evaluated using performance metrics, with a focus on classification tasks. The performance metrics discussed here are specifically targeted at binary classification problems. The objective variable is the resultant variable used to forecast Barrett's esophagus (BE). A value of 1 for the target variable (DYSP) indicates a positive case, meaning the patient has dysplasia. On the other hand, a negative case—one in which the patient is not dysplastic—occurs if the target variable has a value of 0 (Non-DYSP).

3.5.1 |Confusion Matrix

A confusion matrix involves comparing the numbers of 'positive' and 'negative' events classified and misclassified so that the efficiency of a system can easily be determined. Using the intended and achieved labels it offers a graphical presentation of how the model works.

Confusion matrix is a 2 x n table where the first row with n columns contains the predicted or the classes and the first column contains the real or the classes. The confusion matrix helps us analyze how effective the model is in terms of:

Table 4. Confusion matrix.

	PN	**PP**
AN	TN	FP
AP	FN	TP

3.5.2 | Accuracy

The level of accuracy of the model's prediction is measured by the percentage of the total number of cases predicted correctly among the total number of cases. Therefore, this gives the following formula for accuracy:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP+FP+FN+TN}} \tag{3.1}$$

3.5.3 | Recall

It expresses the model's accuracy in identifying positive instances as the ratio of true positive predictions to real positive cases. One way to test recall is as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.2}$$

3.5.4 | Precision

It measures how effectively the model accurately detects positive situations by quantifying the percentage of genuine positive predictions among all positive predictions. The following is the Precision technique:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.3}$$

3.5.5 | Specificity

The ability of the model to accurately detect negative situations is demonstrated by calculating the fraction of true negative predictions among all real negative cases. Analysis of specificity is quantified as follows:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3.4}$$

3.5.6 | F1 Score

It gives a fair assessment of the accuracy recall of the model by ensuring that the true positives given by the model are achieved. It is the middle point of precision and recall, the overall measure is the harmonic mean of the two. It is computed in this manner:

$$\text{F1} - \text{Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.5}$$

3.6 | Principal Component Analysis

The aim and scope of the PCA is to decrease the number of features in a feature vector or in a dataset while converting the given features to a number of new features known as the principal components. This transformation is used to try and reduce the degree of dispersion in the data set as much as possible while storing the most relevant information.

Regardless of the fact, PCA reduces the variation or diversity of the set of variables by transforming the set of original variables into the principal components. These are actual vectors derived from the covariance matrix and can be seen as the direction in which most of the variation in the data is achieved.

While undertaking PCA, one should possibly confirm that the dataset is normalized correctly. The scaling of the dataset also has a significant purpose of diminishing bias in the results while making sure that the contribution of each variable in the analysis is proportional. This operation should be relative to the dataset because the results depend upon the method of scaling chosen.

For instance, suppose, there is a set of wines and each wine has attributes like quality, age, etc., then for the small set, PCA can be used to tell which of these features, contribute the most to the variation in the set of wines. In another point, it shrinks the wines by doing Generalizations in which they minimize the features of the wines into a lower dimensional format.

In general, PCA enables users to have a bird's eye view of the dataset in the reduced dimension or projection so as to grasp the information in a different perspective manner. It is useful in offering a basis for selection of the most relevant features and simplify the given set of data.

3.7 | Train and Split

The dataset is usually split into training and testing sets for machine learning, with the validation set being utilized occasionally as a third set. The model is trained using the training set, which enables it to learn from the data and modify its parameters. The testing set is used to examine the correctness of the model objectively and to gauge how well it performs on unknown data.

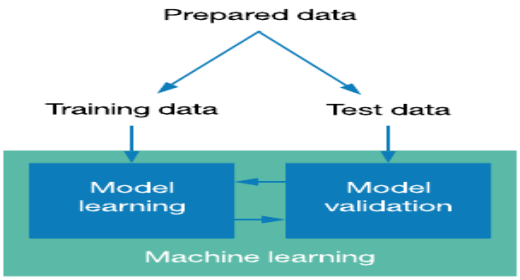

Figure 6. Train and split data.

For different projects based on their overall number of samples and complexity of the model, training, test, and validation samples can be different in size. Thus, while models with a small number of hyperparameters might require only smaller validation sets to tune them, others may require larger training sets to be learned from.

The dataset used in this work is split into two parts: Where training consumed the larger percentage of 70 percent while testing consumed the remaining 30 percent of the student's time. Precisely 63 % of the entire data set is used in training while 7% of the entire data set is as a cross-validation set for the model. Cross-validation is a process of data splitting where in the analysis is performed on one set known as the training set as the validity of the result is checked with the other set known as the validation or testing set.

Different data subsets can be used in numerous rounds of cross-validation to minimize variability and provide more robust findings. An estimate of the model's prediction performance is obtained by averaging or combining the validation findings from each iteration.

The thesis attempts to ensure a trustworthy assessment of the model's prediction skills by splitting the dataset and use cross-validation to train the model efficiently on a significant chunk of the data and evaluate its performance properly on unseen data.

3.8 | Support Vector Machine

In predictive models, supervised learning technique; SVM may be used in classification problems for instance; prediction of Barrett's esophagus (BE).

Input:

Patient Data set

Output:

Chosen attributes and the SVM classifier

i). The first step In data analysis is to import the dataset into the computer for analysis.

ii). Split the collected data into training set that will contain seventy percent of the data, and testing set that will contain thirty percent of the data.

iii). Setting of the target variable.

iv). The classifier should be generated based on the selected training dataset.

v). Teach the classifier with a linear kernel function (similarity function).

vi). Using the classifier that has been trained, use it to predict the testing dataset.

vii). Assess the classifier.

viii). Perform recursion on the selected features and use the weight as the structure for the new selection.

In the setting of Barrett's esophagus, SVM work in a manner where features in the planes or hyperplanes provide a discrete set of outcomes depending on the characteristics that are measured. Thus, the maximum separation between the classes is attained in accordance with the margin or distance between them paying attention to the hyperplane obtained.

Because kernel functions are used, SVMs may handle data that is both linearly and non-linearly separable. By using these kernel functions, the SVM may convert the input data into a feature space with larger dimensions, potentially leading to the separability of the classes. Kernels using polynomial, radial basis function (RBF), and linear functions are frequently utilized.

The SVM algorithm is trained on a labeled dataset including characteristics pertaining to patients' medical records, test results, or any other pertinent data before being applied to the prediction of Barrett's esophagus. In order to build a decision boundary that can categorize fresh, unknown occurrences into the proper category (such as dysplasia or non-dysplasia), the SVM learns the patterns within the data.

3.9 | Naïve Bayes

One method that is frequently used in classification problems, including forecasts of Barrett's Esophagus (BE), is the Naïve Bayes method.

Naive Bayes works on the generalized assumption that all the features of the given data are independent of each other and it uses the theorem of Bayes in the context of Barrett's esophagus. Naive Bayes is nevertheless applied in numerous applications of real life such as the diagnosis of diseases even with the assumption it makes.

For predicting Barrett's esophagus using the Naive Bayes classification technique, cleaning and labeling of the dataset that contains characteristic features of patient medical records, tests or any relevant data is necessary. In the training phase, the different characteristics are related to a target class by computing statistical measures such as for example dysplasia or no-dysplasia.

In the prediction stage, Naive Bayes determines the posterior probability of a novel, unseen instance belonging to each class using Bayes' theorem Eq. (4.1) The procedure makes the assumption that the characteristics are conditionally independent given the class, which means that the existence or lack of one feature does not influence the existence or lack of another. Although it makes the math easier, this assumption might not always be accurate.

Using the training set of data, Naive Bayes determines the probability of each characteristic happening in each class. The posterior probabilities are then calculated by combining these likelihoods with each class's prior probability. The projected class for the new instance is designated as the one with the highest posterior probability.

$$P(C|A) = \frac{P(A|C)P(C)}{P(A)} \quad (4.1)$$

Input:

Patient Data set

Output:

Selected features and Naive Bayes classifier

i). The first step In data analysis is to import the dataset into the computer for analysis.

ii). Split the collected data into a training set that will contain seventy percent of the data and a testing set that will contain thirty percent of the data.

iii). Setting of the target variable.

iv). The classifier should be generated based on the selected training dataset.

v). Teach the classifier with a linear kernel function (similarity function).

vi). Using the classifier that has been trained, use it to predict the testing dataset.

vii). Assess the classifier.

viii). Perform recursion on the selected features and use the weight as the structure for the new selection.

4 | Results

Five different classification techniques are used for the dataset SVM, LR, NB, DT, and KNN Classifier. Accuracy(F1), Precision(F2), Recall(F3), F1 Score(F4), and Specificity(F5) are used to calculate each classifier's performance as mentioned below in Table 5:

Table 5. Scores of Specificity, Recall, F1 Score, Accuracy, Precision.

	F5	F4	F3	F2	F1
LR	0.783	0.949	0.976	0.923	0.923
NB	0.902	0.977	0.984	0.969	0.964
SVM	0.745	0.944	0.991	0.9	0.917
DT	0.585	0.88	0.99	0.732	0.834
KNN	0.804	0.957	0.984	0.931	0.935

The performance scores of many algorithms to predict Barrett's esophagus using PCA are displayed in Table 6 below:

Table 6. Scores of F1 Score, Precision, Accuracy, Recall, and Specificity with PCA.

	F5	F4	F3	F2	F1
LR	0.65	0.912	1	0.838	0.876
NB	0.74	0.94	0.983	0.9	0.911
SVM	0.696	0.93	1.000	0.869	0.899
DT	0.655	0.91	0.974	0.854	0.869
KNN	0.692	0.923	0.974	0.877	0.887

Table 7 shows the Mean, Standard Deviation(SD), and Standard error of Mean (SEM on the test dataset with/without the dimensionality reduction method PCA.

Table 7. Dataset test.

	Without PCA	With PCA
Mean	0.91460	0.88840
SD	0.04855	0.01699
SEM	0.02171	0.007660

4.1 | Comparing the Decision Tree Model's Performance Before and After Applying PCA

The normalized and confusion matrix of the Decision Tree classifier is shown in the images below, both with and without the application of PCA (Figure 7 and Figure 8). The results demonstrate that, without PCA, Decision Tree performed remarkably well for this dataset, with accuracy, precision, and recall scores of 0.834, 0.792, and 0.99, respectively. The decision tree's precision, accuracy, and F1 Score all rise when PCA is applied, with increases observed for each of the three performance metrics. In any case, once PCA is used, there is an issue with the recall. Without PCA, we may assume that the Decision tree performs better. The number of actual diagnoses in Table 8 that were diagnosed as dysplasia and that the algorithm predicted as dysplasia (True Negatives) is 103. Similarly, there are 27 data that have been classified with dysplasia but that the algorithm predicts to have no dysplasia (False Positives). Furthermore, 38 data points have been diagnosed as dysplasia and are classified as such by the algorithm (also known as True Positives). On the other hand, 1 data point has been classified as dysplasia and marked as no-dysplasia by the algorithm (also known as False Negatives).

Table 8. Confusion Matrix of Decision Tree without PCA.

	Predicted 0	**Predicted 1**
Actual 0	TN=103	FP=27
Actual 1	FN=1	TP=38

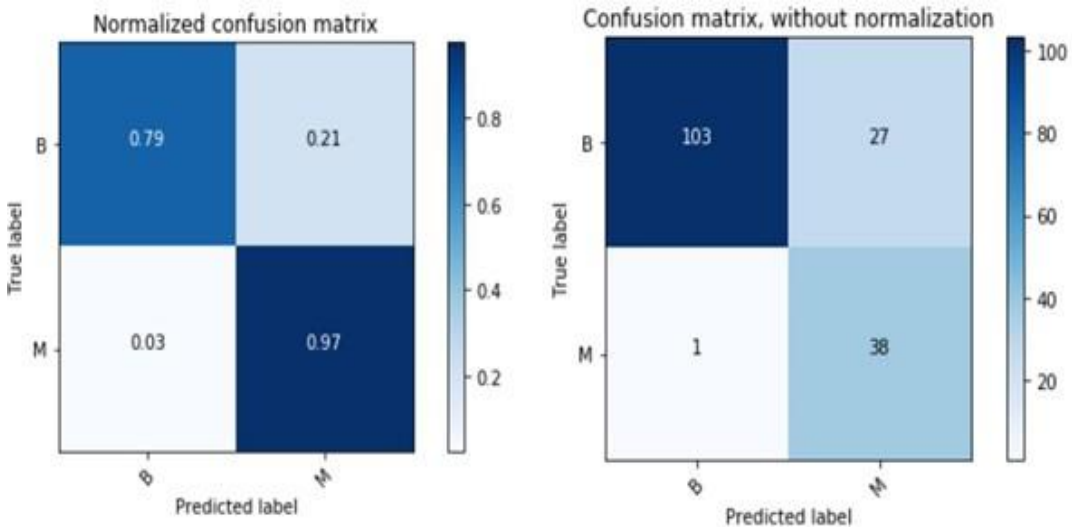

Figure 7. Normalized & Non-Normalized Confusion Matrix of Decision Tree Model without PCA.

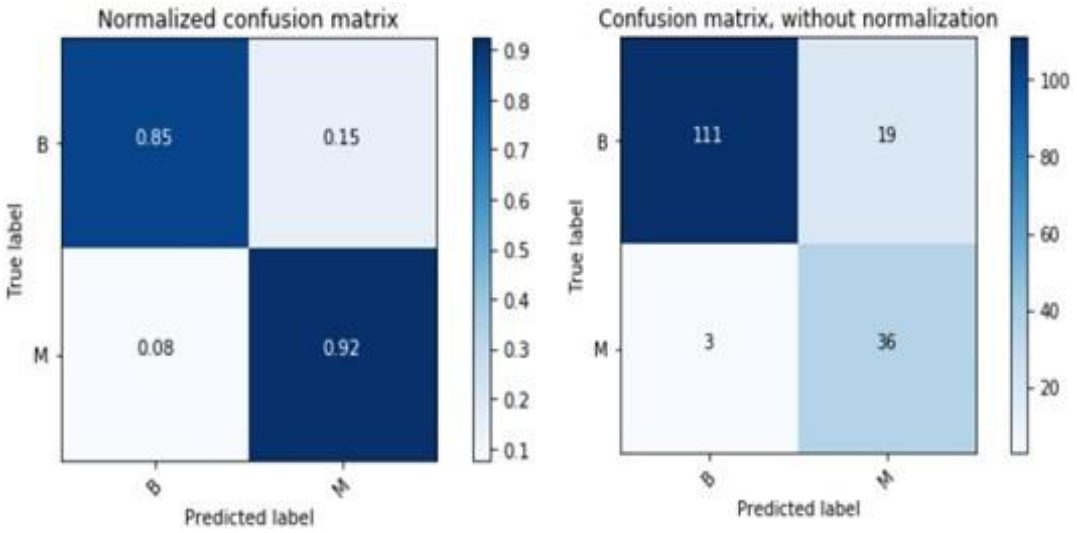

Figure 8. Normalized & Non-Normalized Confusion Matrix of Decision Tree model with PCA.

4.2 |Comparing the KNN Model's Performance Before and After Applying PCA

Figure 9 and Figure 10 depict the normalized and the confusion matrix of the K-Neighbor on this data set with/without applying PCA. The values shown above prove that K-Neighbour has high accuracy with a coefficient of 0. 935. Also, without the help of PCA, Accuracy, Recall, and F1-Score receive good results equal to 0. 931, 0. 984, and 0. 957 separately. Introduction: As it can be derived from the above PCA results the performance decline on all four aspects of exactness (from 0. 935 to 0. 887), accuracy (from 0. 931 to 0. 877), and recall (from 0. 984 to 0. 974 or 0. 973 to 0. 907 depending on the utilized model). When inspecting a given data set, it is preferred to apply K-neighbor without illustrating it in a manner such as the given PCA.

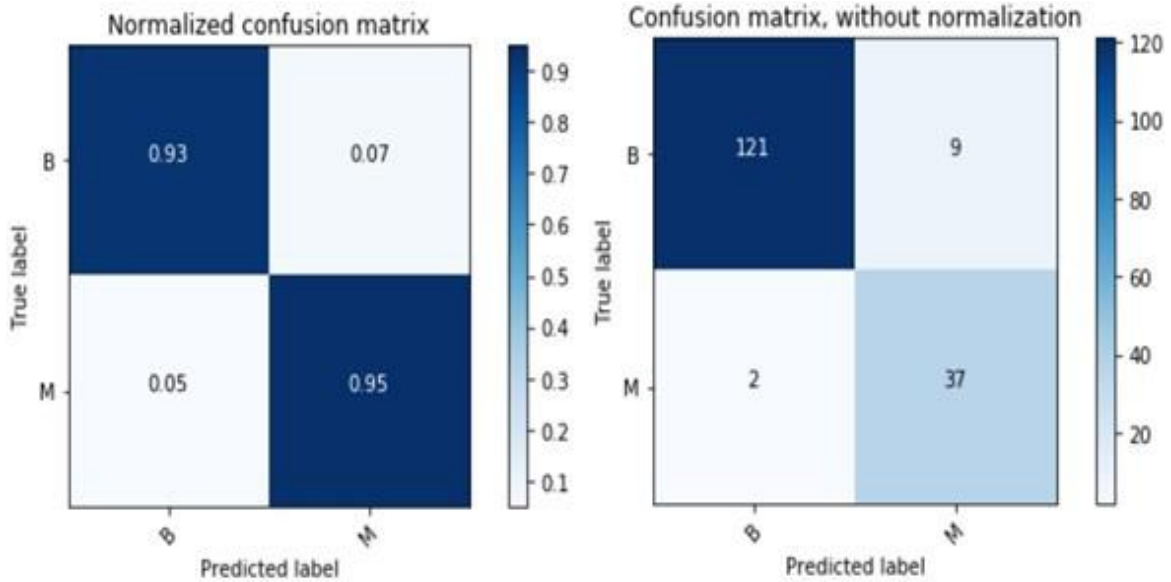

Figure 9. Normalized & Non-Normalized confusion matrix of KNN model without PCA.

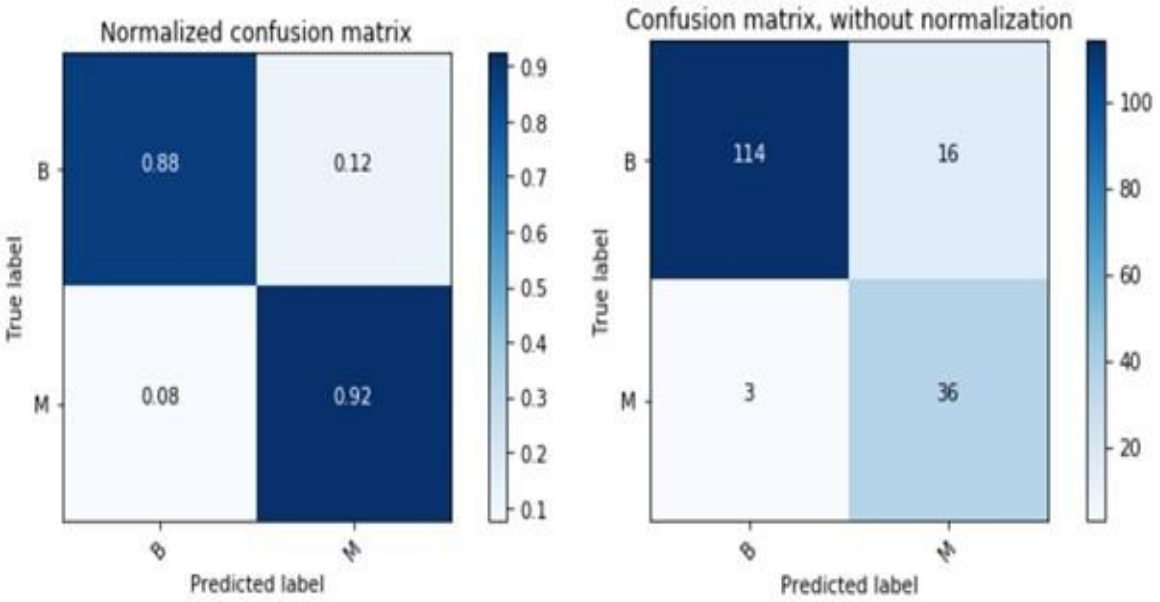

Figure 10. Normalized & Non-Normalized confusion matrix of KNN model with PCA.

4.3 | Comparing the Logistic Regression Model's Performance Before and After Applying PCA

The normalized and confusion matrix of Logistic Regression is shown in the images below, both with and without the application of PCA (Figure 11 and Figure 12). The results demonstrate that Logistic Regression without PCA yields a respectable exactness of 0.923 in addition to numbers for accuracy, review, and F1 Score that are, respectively, 0.923, 0.976, and 0.949. As the computation's exactness, accuracy, and F1 Score decrease, the PCA's display of the dataset yields mixed results. Regardless, recall achieves an optimal score of 1.00 with the use of PCA; as a result, strategic relapse may be employed in conjunction with PCA to estimate bosom malignant development.

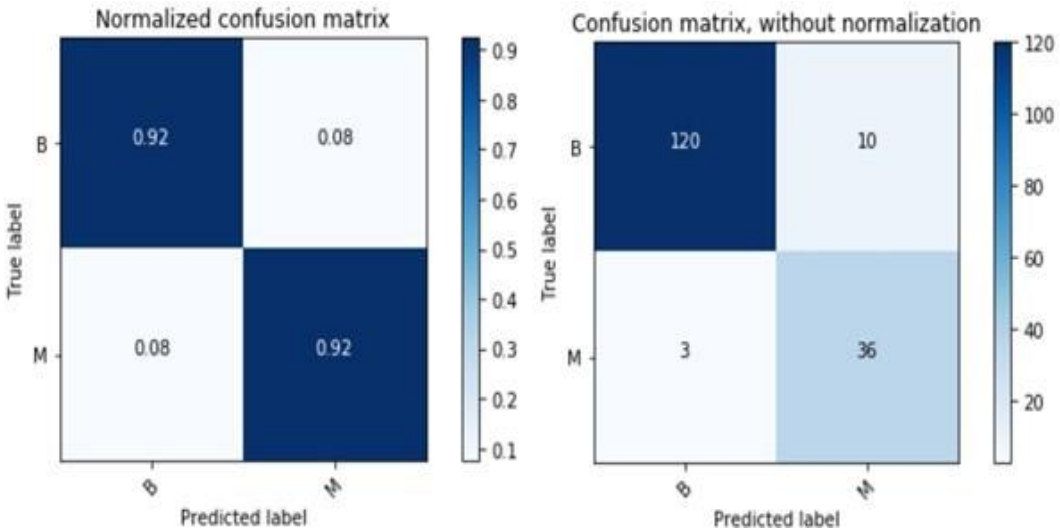

Figure 11. Normalized & Non-Normalized confusion matrix of Logistic Regression model without PCA.

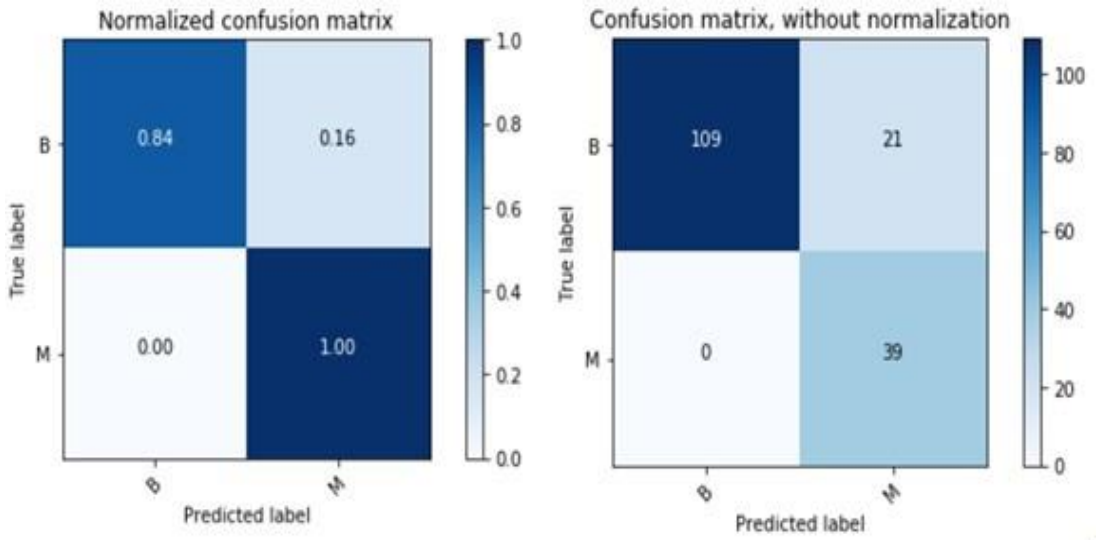

Figure 12. Normalized & Non-Normalized confusion matrix of Logistic Regression model with PCA.

4.4 | Comparing the Naive Bayes Model's Performance Before and After Applying PCA

The normalized and confusion matrix of Naïve Bayes is displayed in the images below, both with and without PCA application (Figure 13 and Figure 14). Moreover, Innocent Bayes scores 0.969 in accuracy, 0.984 in review, and 0.977 in F1Score, resulting in a respectable exactness score of 0.964. When PCA is presented, the estimations of exactness, accuracy, and F1 score decrease, and when Naïve Bayes is used, the estimation of review decreases by 0.001. Therefore, using this method without using PCA will be ideal.

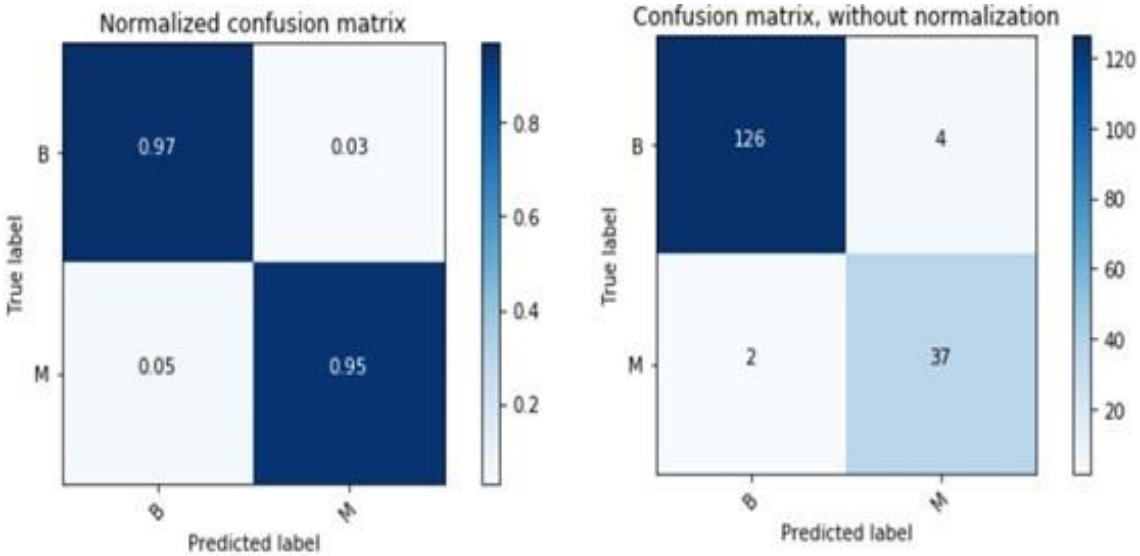

Figure 13. Normalized & Non-Normalized confusion matrix of Naive Bayes model without PCA.

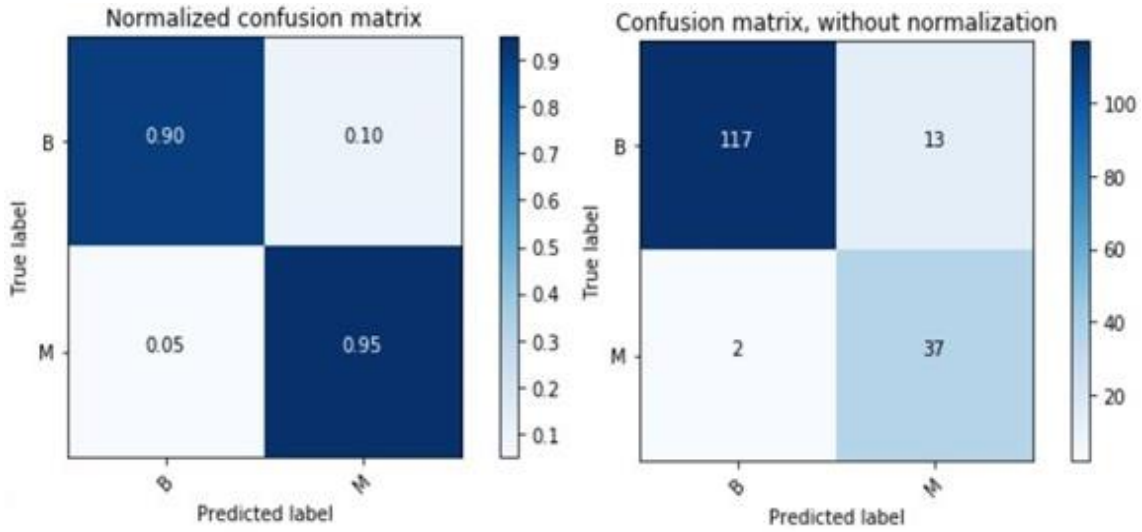

Figure 14. Normalized & Non-Normalized confusion matrix of Naive Bayes model with PCA.

4.5 | Comparing the SVM Model's Performance Before and After Applying PCA

The SVM normalized and confusion matrix are shown in Figures 15 and 16, respectively, both with and without PCA applied to the dataset. The SVM results were rather satisfying because the model predicted a precision of 0.917. For this issue, scores of 0.944 for the F1 Score, 0.991 for the review, and 0.991 for correctness are also noted. If there should be any instances of exactness (0.917 to 0.899), accuracy (0.9 to 0.869), or F1 Score (0.944 to 0.93), the PCA presentation has witnessed a decline. Review nonetheless receives a concentrate score of 1.000 with the application of PCA; hence, the poor accuracy score may be disregarded when SVM is used in conjunction with PCA since the review becomes increasingly relevant in predicting illness.

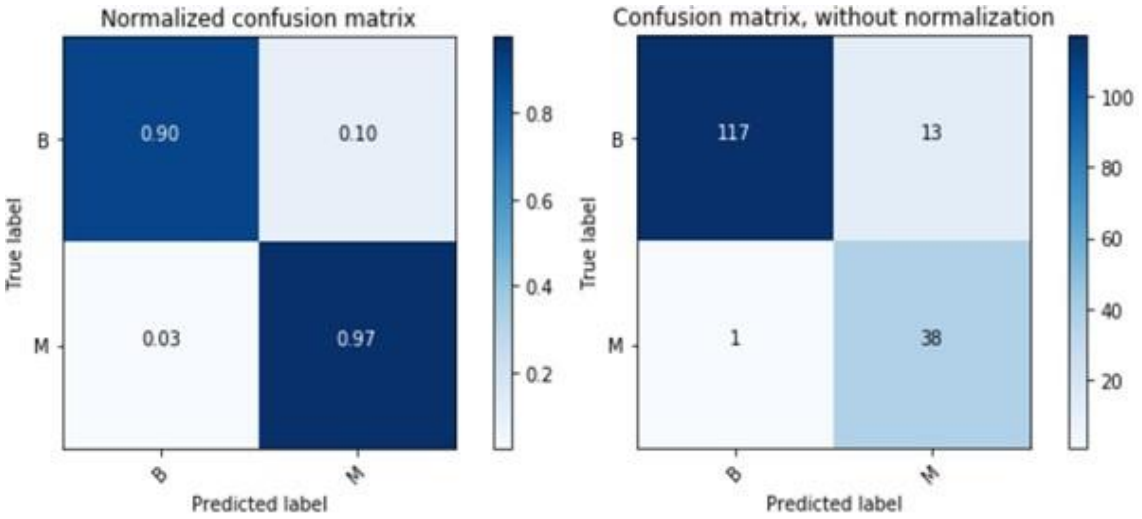

Figure 15. Normalized & Non-Normalized Confusion Matrix of SVM without PCA.

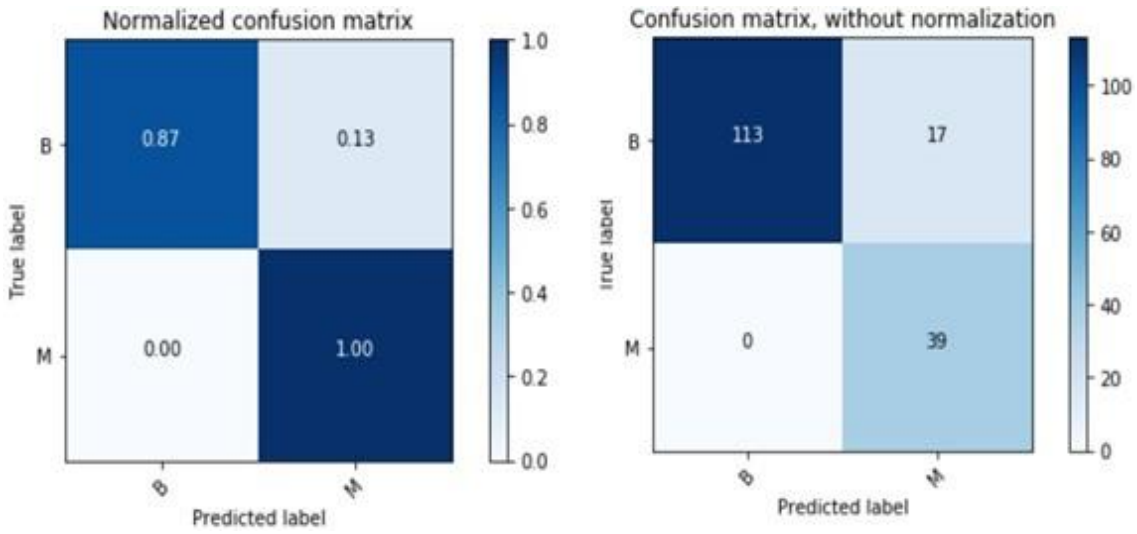

Figure 16. Normalized & Non-Normalized Confusion Matrix of SVM with PCA.

4.6 | Result Summary

Naïve Bayes on its own proved to be quite efficient when it did not incorporate PCA, achieving an accuracy of 0. undefined While they trailed slightly, K-Neighbors were at 0. 9349 and Logistic Regression was at 0. 923, which indicates a high accuracy. Moreover, the accuracy score on SVM was 0. undefined Their average was recorded to be 0. In sample 834, the decision tree worked ineffectively among the five models. The use of PCA led to the recognition reduction in all algorithms with the exception of decision trees.

Reflecting on the whole package of performances, one could state the following conclusions. When PCA is applied to the dataset it is observed that the performance of the Decision Tree, K-Neighbors, and Naïve Bayes increases albeit slightly, whereas Logistic Regression and SVM's performances are already higher. Importantly, after using PCA, both the models of SVM and Logistic Regression receive an accuracy of one. Many of them experience drops in another important performance indicator such as Recall which measures the number of cases that are predicted and recalled at a rate of 0.001, although both of these algorithms have decreased in the other measures of performance.

Based on the analysis of the given dataset, it can be concluded that the better performance of the Logistic Regression and Support Vector Analysis with PCA for the prediction of Barrett's esophagus (BE) is because of the impact of the PCA in the sense that it reduces the number of features by which runtime increases considerably for both small and large data sets and the important thing is that by using the PCA authors maintain the 100 % recall score.

5 | Conclusion

Our goal in conducting the proposed research is to provide a standardized method to increase the accuracy of doctors' prognostications in instances of cancer of the Barrett's esophagus. We have used classification methods to improve the Barrett's esophagus dataset's accuracy. We have gathered up-to-date information about Barrett's esophagus patients in order to guarantee the dataset's dependability. This information was painstakingly gathered at the Lever Clinic Shadman, with the cautious guidance of Drs. Ghulam Mustafa and Irfan.

Moreover, Naive Bayes is the best classifier for preliminary predictions, as per our thorough analysis and comparison of five different classifiers. The automated approach and dataset we have suggested provide

reliable results every time, removing the need for patients to undertake numerous procedures (MRI, ultrasound, and mammography), which can cause them substantial discomfort and radiation exposure.

Looking at things more broadly, we think that the models that were used in this study are quite helpful for medical practitioners to use when making judgments. Several performance indicators, including Accuracy, Recall, Precision, specificity, and the F1 Score, have been utilized to provide a comprehensive evaluation and identification of the optimal method.

5.1 | Future Work

Although the six methods we used yielded accurate and exact answers, we still want to make sure that the limited size of our dataset did not introduce bias into our results. To see if the results hold true, we want to look for a bigger dataset and run a parallel study. Furthermore, because our dataset is the result of human efforts, it is possible to collect even more exact numerical data by utilizing modern technologies and extra prediction criteria.

Acknowledgments

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Author Contributions

All authors contributed equally to this work.

Funding

This research was conducted without external funding support.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Lipman, G., Bisschops, R., Sehgal, V., Ortiz-Fernández-Sordo, J., Sweis, R., Esteban, J. M., ... & Haidry, R. J. (2017). Systematic assessment with I-SCAN magnification endoscopy and acetic acid improves dysplasia detection in patients with Barrett's esophagus. Endoscopy, 49(12), 1219-1228. https://doi.org/10.1055/s-0043-113441

[2] Wang, K. K., & Sampliner, R. E. (2008). Updated guidelines 2008 for the diagnosis, surveillance, and therapy of Barrett's esophagus. Official journal of the American College of Gastroenterology| ACG, 103(3), 788-797. https://doi.org/10.1111/j.1572-0241.2008.01835.x

[3] Sehgal, V., Rosenfeld, A., Graham, D. G., Lipman, G., Bisschops, R., Ragunath, K., ... & Lovat, L. B. (2018). Machine Learning Creates a Simple Endoscopic Classification System that Improves Dysplasia Detection in Barrett's Oesophagus amongst Non-expert Endoscopists. Gastroenterology Research and Practice, 2018(1), 1872437. https://doi.org/10.1155/2018/1872437

[4] Sharma, P., McQuaid, K., Dent, J., Fennerty, M. B., Sampliner, R., Spechler, S., ... & Weinstein, W. (2004). A critical review of the diagnosis and management of Barrett's esophagus: the AGA Chicago Workshop. Gastroenterology, 127(1), 310-330. https://doi.org/10.1053/j.gastro.2004.04.010

[5] Banks, M. R., Haidry, R., Butt, M. A., Whitley, L., Stein, J., Langmead, L., ... & Lovat, L. B. (2011). High resolution colonoscopy in a bowel cancer screening program improves polyp detection. World journal of gastroenterology: WJG, 17(38), 4308. https://doi.org/10.3748/wjg.v17.i38.4308

[6] Pigò, F., Bertani, H., Manno, M., Mirante, V., Caruso, A., Barbera, C., ... & Conigliaro, R. L. (2013). i-Scan high-definition white light endoscopy and colorectal polyps: prediction of histology, interobserver and intraobserver agreement. International journal of colorectal disease, 28, 399-406. https://doi.org/10.1007/s00384-012-1583-7

[7] Bauknecht, H. C., & Klingebiel, R. (2005). Aktuelle Bildgebung des Felsenbeins: CT versus MRT. Radiologie up2date, 5(02), 165-184. https://doi.org/10.1055/s-2005-861241

[8] Richards, R. A. (2002, June). Principle hierarchy based intelligent tutoring system for common cockpit helicopter training. In International Conference on Intelligent Tutoring Systems (pp. 473-483). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-47987-2_50

[9] Sierra, B., & Larranaga, P. (1998). Predicting survival in malignant skin melanoma using Bayesian networks automatically induced by genetic algorithms. An empirical comparison between different approaches. Artificial intelligence in Medicine, 14(1-2), 215-230. https://doi.org/10.1016/S0933-3657(98)00024-4

[10] Kononenko, I. (1993). Inductive and Bayesian learning in medical diagnosis. Applied Artificial Intelligence an International Journal, 7(4), 317-337. https://doi.org/10.1080/08839519308949993

[11] Bergman, J., Pech, O., Ragunath, K., Armstrong, D., Tytgat, G. N. J., Dent, J., ... & Sharma, P. (2017). 59 A Novel, Interactive Web-Based Educational Tool Improves Detection and Delineation of Barrett's Oesophagus Related Neoplasia (Born): The Born Project. Gastrointestinal Endoscopy, 85(5), AB48. https://doi.org/10.1016/j.gie.2017.03.040

[12] Hoffman, A., Sar, F., Goetz, M., Tresch, A., Mudter, J., Biesterfeld, S., ... & Kiesslich, R. (2010). High definition colonoscopy combined with i-Scan is superior in the detection of colorectal neoplasias compared with standard video colonoscopy: a prospective randomized controlled trial. Endoscopy, 42(10), 827-833. https://doi.org/10.1055/s-0030-1255713

[13] Sami, S. S., Subramanian, V., Butt, W. M., Bejkar, G., Coleman, J., Mannath, J., & Ragunath, K. (2015). High definition versus standard definition white light endoscopy for detecting dysplasia in patients with Barrett's esophagus. Diseases of the Esophagus, 28(8), 742-749. https://doi.org/10.1111/dote.12283

[14] Sharma, P., Bergman, J. J., Goda, K., Kato, M., Messmann, H., Alsop, B. R., ... & Waxman, I. (2016). Development and validation of a classification system to identify high-grade dysplasia and esophageal adenocarcinoma in Barrett's esophagus using narrow-band imaging. Gastroenterology, 150(3), 591-598. https://doi.org/10.1053/j.gastro.2015.11.037

[15] Kuo, W. J., Chang, R. F., Chen, D. R., & Lee, C. C. (2001). Data mining with decision trees for diagnosis of breast tumor in medical ultrasonic images. Breast cancer research and treatment, 66, 51-57. https://doi.org/10.1023/A:1010676701382

[16] Vlahou, A., Schorge, J. O., Gregory, B. W., & Coleman, R. L. (2003). Diagnosis of ovarian cancer using decision tree classification of mass spectral data. BioMed Research International, 2003(5), 308-314. https://doi.org/10.1155/S1110724303210032

[17] Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: current issues and guidelines. International journal of medical informatics, 77(2), 81-97. https://doi.org/10.1016/j.ijmedinf.2006.11.006

[18] Yeh, D. Y., Cheng, C. H., & Chen, Y. W. (2011). A predictive model for cerebrovascular disease using data mining. Expert Systems with Applications, 38(7), 8970-8977. https://doi.org/10.1016/j.eswa.2011.01.114

[19] Zhang, Z., Bai, L., Ren, P., & Hancock, E. R. (2016). High-order graph matching kernel for early carcinoma EUS image classification. Multimedia Tools and Applications, 75, 3993-4012. https://doi.org/10.1007/s11042-015-3108-1

[20] Klomp, S., van der Sommen, F., Swager, A. F., Zinger, S., Schoon, E. J., Curvers, W. L., & Bergman, J. J. (2017, March). Evaluation of image features and classification methods for Barrett's cancer detection using VLE imaging. In Medical Imaging 2017: Computer-Aided Diagnosis (Vol. 10134, pp. 84-93). SPIE. https://doi.org/10.1117/12.2253860

[21] Seguí, S., Drozdzal, M., Pascual, G., Radeva, P., Malagelada, C., Azpiroz, F., & Vitrià, J. (2016). Generic feature learning for wireless capsule endoscopy analysis. Computers in biology and medicine, 79, 163-172. https://doi.org/10.1016/j.compbiomed.2016.10.011

[22] Chan, D. K., Zakko, L., Visrodia, K. H., Leggett, C. L., Lutzke, L. S., Clemens, M. A., ... & Wang, K. K. (2017). Breath testing for Barrett's esophagus using exhaled volatile organic compound profiling with an electronic nose device. Gastroenterology, 152(1), 24-26. https://doi.org/10.1053/j.gastro.2016.11.001

[23] Rajan, P., Canto, M., Gorospe, E., Almario, A., Kage, A., Winter, C., ... & Münzenmayer, C. (2010). Automated diagnosis of Barrett's esophagus with endoscopic images. In World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany: Vol. 25/4 Image Processing, Biosignal Processing, Modelling and Simulation, Biomechanics (pp. 2189-2192). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03882-2_581

[24] Pech, O., Rabenstein, T., Manner, H., Petrone, M. C., Pohl, J., Vieth, M., ... & Ell, C. (2008). Confocal laser endomicroscopy for in vivo diagnosis of early squamous cell carcinoma in the esophagus. Clinical Gastroenterology and Hepatology, 6(1), 89-94. https://doi.org/10.1016/j.cgh.2007.10.013

[25] Rosenfeld, A., Sehgal, V., Graham, D. G., Banks, M. R., Haidry, R. J., & Lovat, L. B. (2014, June). Using data mining to help detect dysplasia. In 2014 IEEE international conference on software science, Technology and Engineering (pp. 65-66). IEEE. https://doi.org/10.1109/SWSTE.2014.21

[26] Li, C., Shi, C., Zhang, H., Chen, Y., & Zhang, S. (2015). Multiple instance learning for computer aided detection and diagnosis of gastric cancer with dual-energy CT imaging. Journal of biomedical informatics, 57, 358-368. https://doi.org/10.1016/j.jbi.2015.08.017

[27] Wang, Z., Lee, H. C., Ahsen, O. O., Liang, K., Figueiredo, M., Huang, Q., ... & Mashimo, H. (2016). Sa2030 Novel Optical Coherence Tomography Image Analysis Reveals Subsquamous Glandular Structures as Strong Predictors of Poorer Response to Radiofrequency Ablation in Barrett's Esophagus. Gastroenterology, 150(4), S434.

[28] Serpa-Andrade, L., Robles-Bykbaev, V., González-Delgado, L., & Moreno, J. L. (2015, November). An approach based on Fourier descriptors and decision trees to perform presumptive diagnosis of esophagitis for educational purposes. In 2015 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC) (pp. 1-5). IEEE. https://doi.org/10.1109/ROPEC.2015.7395123.

[29] Boschetto, D., Gambaretto, G., & Grisan, E. (2016, March). Automatic classification of endoscopic images for premalignant conditions of the esophagus. In Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging (Vol. 9788, pp. 65-70). SPIE.

[30] Fitzgerald, R. C., Di Pietro, M., Ragunath, K., Ang, Y., Kang, J. Y., Watson, P., ... & De Caestecker, J. (2014). British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. Gut, 63(1), 7-42. https://doi.org/10.1136/gutjnl-2013-305372

[31] Hayat, S., Rehman, F., Riaz, N., Sharif, H., Irshad, S., & Shareef, S. (2022, May). Using machine learning algorithms to detect dysplasia in Barrett's esophagus. In 2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2) (pp. 1-6). IEEE. https://doi.org/10.1109/ICoDT255437.2022.9787479.

[32] Saqlain, M. (2023). Evaluating the readability of English instructional materials in Pakistani Universities: A deep learning and statistical approach. Education Science and Management, 1(2): 101-110.

[33] Meharunnisa, Saqlain, M., Abid, M., Awais, M., and Stević, Ž. Analysis of software effort estimation by machine learning techniques. Ingénierie des Systèmes d'Information, 28(2023),1445-1457.

[34] Abid, M. & Saqlain, M. (2023). Utilizing Edge Cloud Computing and Deep Learning for Enhanced Risk Assessment in China's International Trade and Investment. Int J. Knowl. Innov Stud., 1(1), 1-9.