Omparative Analysis of Pre-Trained Classifier in Augumented Approach for Ovarian Image

Kavitha Sᵃ, and Dr. VidyaaThulasiramanᵇ

[a]
Research scholar, Department of Computer Science, Periyar University, Salem.
[b]Assistant Professor and Head, Department of Computer Science Government Arts and Science Collegefor Women, Bargur-635 104

_____

Abstract: Ovarian cancer is identified as one of the leading cause for increased mortality rate among women. The early diagnosis of ovarian cancer decreases the mortality rate, which demands for efficient classification technique. Conventional cell classification technique extracts multiple features for recognition of ovarian cancer from complex cell texture with identification of difference between cells. To resolve complexity associated with ovarian cancer cell texture analysis deep neural network exhibits improved performance. In deep learning technique, features are extracted automatically for identification of cell types and texture. However, Annotation based approach exhibits improved classification performance still performance need to be improved for automated cancer diagnosis. To achieve higher accuracy rather than annotation this paper proposed an augmentation of MRI ovarian image. The augmented images are pre-processed with median filtering for contrast enhancement. In next stage, ROI based image segmentation is performed followed by feature extraction. To improve classification performance of augmented images CNN model Inception V3 and Xception model is comparatively examined. The performance of Inception V3 and Xception model is evaluated with Logistics Regression and Random Forest classifiers. The comparative analysis of simulation results expressed that Xception Logistic Regression model provides higher accuracy than the Inception V3 Logistics regression, Inception V3 Random Forest and Xception Random Forest.

Keywords: Ovarian Cancer, Deep learning, Classifier, Inception V3, Xception

1. Introduction

Recently, above two thirds of ovarian cancers has metastasized on the outer surfaceof the actual pelvis [1]. Even for these cancer patients,primary surgery of done with the aim to clear all tumour tissues, macroscopic or microscopic residualinfection is left in at least 50% of the cases. After chemotherapy, second look operations enlarge the morbidity and costs withoutaffecting the results of final treatment [2]. When surgery procedures are discardedin evaluating ovarian cancers to the response of chemotherapy, non-invasive methods are involved. While declining CA 125 levels, clinician is informed that the tumour burden is decreasing. On contrary, the increase imply a tumour reappearance but the information about its location is unknown [3].results while evaluating the location and quantifyingfew residual infection [4].

Based on the classificationmade by World Health Organization (WHO) volumetric data alone is involved to evaluate the responses of chemotherapy in gynaecological masses [5]. Traditionally, for tumour masses, two dimensions are obtained in every computerized tomography (CT)/magnetic resonance imaging (MRI) images which producesthe best representation of the tumour. The foci ofevery tumour are summed up to provide an equivalent numberto the patient's overall tumour bulk [6].

WHO has announced that the death caused by cancer is world's second largest following is heart disease and stroke [7]. According to Ovarian Cancer Research Fund Alliance (OCRFA), Ovarian Cancer (OC) is the 8th most widespread cancer in women [8]. Epithelial Ovarian Cancer (EOC) representing 90% of Ovarian Cancer patients is generallyanalyzedlater yielding five majorhisto subtypes such as high-grade serous (HGSOC), Endometrioid (ENOC), clear cell (CCOC), mucinous (MOC), and low-grade serous (LGSOC) [9]. But, in recent days, 3D MRI images providean useful tool to visualize and analyze intra-abdominal tumours, their volumes, and provides quantitative information about tumour. Usually, 3D MRI is volumetric and is accurately analyzed using segmentation approaches. Gynecologic malignancies are categorized as better responses when 3D volumetric tumourassessment methods are used [10]. While the follow-up CTs in 30 ovarian cancer patients, 12 have been categorized as better response when 3D volumetric analyzing methods are used [11]. Previously, segmentation and volumetric analysis were performed to evaluate ovariancancer [12], but till now not applied to segment ovarian tumours.

Deep learning (DL), a significant artificial intelligence approach, are extensively used in recognizing images [13]. Moreover, convolutional neural network (CNN) provides remarkable results in classifying images. The original image can be processed directly where complex image preprocessing task is avoided [14]. Here, three conceptsnamely local receptive field, pooling, and weight sharing are integrated thereby reducing the count of training parameters in neural networks.

Classifying the subtypes of Ovarian Cancer has been described in several research studies [15], which employed machine learning approaches to classify the subtype and predict it based on the type of tumor cells. In existing, EOC are further classified based on the molecular into two sub-classes ranging from stage II to IV of

International Federation of Gynecology and Obstetrics (FIGO) stage-directed supervised classification approach [16]. Here, it was further independently classified using cell type clinicopathological identification parameters. Further recent research works presents an overview of cancer types, its classification and the way the machine learning approaches are essential part in solving these challenges [17].

1.1 Contribution and Organization

In this paper, image augmentation is performed for increasing accuracy rather than annotation. The developed augumentation technique integrated with filtering for improving accuracy. The augumentation creates own dataset based on rotation, scaling and etc. This provides the normalized view of image at any direction. The augmented image dataset is pre-processed through median filtering followed by image segmentation and feature extraction. The image segmentation is performed through ROI process. The CNN augmented data is comparatively processed with Inception V3 and Xception with logistic regression and Random Forest classifier. The paper is arranges as follows: In section I general description about ovarian cancer and classification is presented. In section II related works for classification is presented followed by research methodology in section III. In section IV and V Stacked approach for feature extraction and classification is presented. In section VI results achieved bythe proposed SLRFC is presented. Finally, in section VII, the overall conclusion about the proposed technique is presented.

2. Related Works

In view of using CNNs for analyzing medical images is continuously increasing. In [18]various CNN architectures were applied in classifying and identifyingovariancells. In [19],transfer learning related to CNN as well as recurrent neural network (RNN) was utilized in the identification and diagnosing lung cancer. Even in [20], growth CNN (GCNN) was involved in segmenting brain tumour from MRI images.

In [21], two approaches were presented for cancerclassification. The first classification approach implemented exhaustive feature selection to identify informative gene features. The next is the generalized approach which worked with several kinds of cancer. Both these approaches was based on machine learning for validating the output. Moreover, the recent researchers introduce few cancer classification techniques with machine learning.

Exhaustive feature selection approach was applied in [22] which explained the way to use supervised learning techniques effectively on classifying cancer or normal tissues. In [22], Support Vector Machine (SVM) classifier was applied to obtain 100% accuracy with leukemia, ovarian, and breastCancer, but obtained only 99.44% of accuracy with Colon Cancer. From the investigation carried out, till now no classification approach has universally outperformed, however, while incorporating data sources such as gene expression and protein-protein interaction, the accuracy of prediction can be increased. Even in [23], ensemble learning was adopted for classifying Breast, Leukemia and Colon Cancer. SVM-recursive feature elimination method was applied which attained 0.84 AUC. In [24], two prognostic classes, metastasis or non-metastasis were considered, and using ensemble classification technique AUC obtained was 1.00. In [25], a strategic classification approach was developed for identifying genes effectively which was used in clinical dataset for the classification of non-small cell lung with Artificial Neural Networks (ANN) and achieved65.71% of accuracy since only a small subset of gene expression data was involved.

Network structure based on Inception v3 along with transfer learning approach attracted the recent researchers, as it performed excellently on a wide range of small datasets. In [26] high-precision classification rate was obtained which used five representative snakes. In [27] the classification of the German Traffic Sign Recognition Standard (GTSRB) was performed. Also, in [28] flowers images taken from Oxford-i7 as well as Oxford-102 flower datasets were classified and obtained better classification accuracy. In [29],gum disease using apical X-ray film was classified successfully. Further, [30] used approaches to identify wild photos which were applied for land-cover classification. In [31], the lymph node metastasis classification in colorectal cancer obtained successful results. In the meantime, [32] realized an effective breast lumpsclassification.

Alternatively, the other approach was more generalized. In [33],Principal Component Analysis (PCA) algorithm was enhanced against binomial classes. Cancer gene expression classification was considered with several cancer types and achieved above 92% of accuracy. Unsupervised clustering was also used to generate multinomial classes; Kernel PCA (KPCA) outperformed PCA with more than 90% of classification accuracy. Yet, on the investigation made, so far no related work exists which covers classification of cancer subtype working on gene expression as well as clinical datasets to address subtype classification of ovarian cancer. This paper introduces an approach which integrates gene expression and clinical dataset for subtype classification of Ovarian Cancer.

3. Proposed Scheme

In this section, some preliminaries for the augmented images for classification are presented. The augmentation consists of creation of own dataset and captures image at different view with respect to rotation, scaling, cropping, brightness, saturation and contrast. This facilitate image at any view can be effectively processed and region can be identified. The augmented images are pre-processed with median filtering for contrast enhancement. The images are segmented with Region of Interest (ROI) based approach followed by extraction of images features based on augmented images. The features selected are size, scale, brightness,

contrast, saturation, entropy and correlation. Finally, feature extracted images are trained and tested with pre-trained models such as Inception V3 and Xception. The pre-trained model performance is comparatively examined with classifiers such as logistic regression and random forest. In figure 1, the overall process adopted in proposed SLRFC is presented.
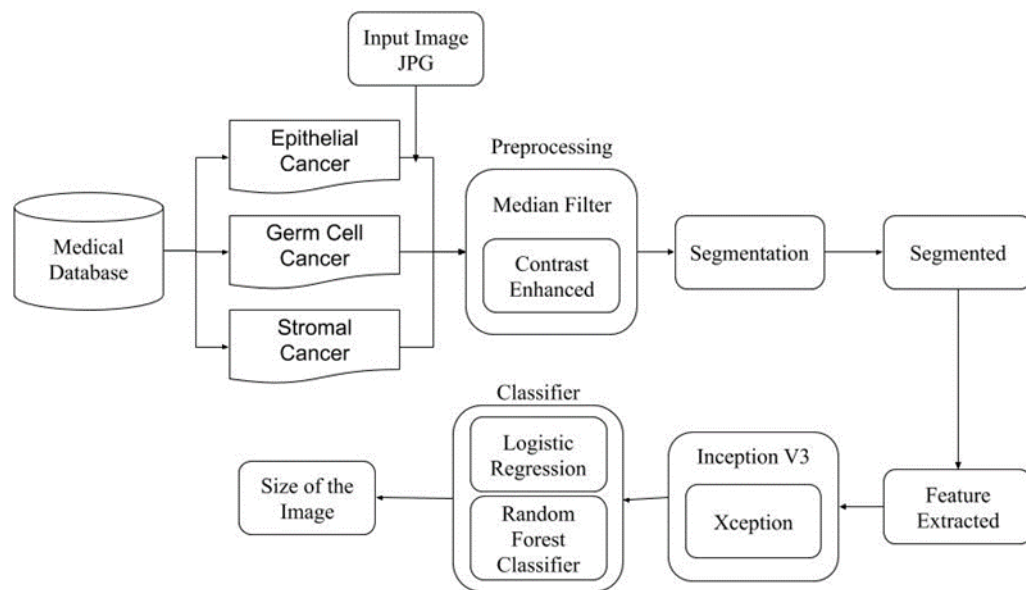

Figure 1: Overview of SLRFC

3.1 Image Augmentation

Image augmentation generatesthe training images artificially by various processing methods or combining several processing techniques like shifts, flips, random rotation, shear etc.For construction of valuable Deep Learning approaches, errors have to be continuously validated so that training error can be decreased which can be effectively achieved by data augmentation. The augmented imageprovidesa complete set of all possible points in an image. Hence, the distance between training and validation set is minimized which even helps testing sets in future. The motivation of developing this approach is to construct a novel MRI ovarian image datasetwhich is the enhance version of the existing dataset; therefore results obtained from segmentation are improved. Segmentation is the process of recognizing the pixels related to the ovarian cancer image and segregating nuclei ignoring the other portions of the image. The real image-mask pair is considered as the input domain. Image augmentationinvolved in either horizontal or vertical flipping of image, the pixels are rearranged completely, howeverpreserving the features. This model is tested with any inverted dog image and hence verticalflips are to be counterproductive. The inverted images can be of different angles since rotated with random degree whose pixel values now differs from those of the actual image.

Even noise can be added to the image which are erroneous pixels distributed randomly all over the image.Traditional augmentation approaches namely flips and rotations are applied to the every training image in the dataset where the images are not processed manually. "ImageDataGenerator" takes several image sets from directory to which transformations like either vertical or horizontal flip or rotation are applied.

Initially, the augmented images undergo pre-processing step to enhance the images earlier to the computation. Under various ways of image investigation, the preprocessing step mainly generates the gathering of images. It converts the applied image to a new one basically identical to the applied image; however, it varies in few dimensions. Some of the preprocessing functions are resizing, masking, segmentation, normalization, elimination of noise, and so on. This study performs the preprocessing operation on the applied product images by resizing the images and filtering the noises present in the image. For resizing the images, every image is converted to a default size of 300x300 pixels. For obtaining better outcome from the product images, resized images are passed to the filtering procedure. It is essential due to the fact that many matrices as well as intensives the sparsity of Convolution Neural Network (CNNs) structure. Then, Logistic Regression (LR) classifier is applied to classify the images. Once the model is created by usingDeep Convolution Neural Network (DCNN) and LR, then the testing of images takes place

Initially, the applied images undergo pre-processing to enhance the images earlier to the computation. Under various ways of image investigation, the preprocessing step mainly generates the gathering of images. It converts the applied image to a new one basically identical to the applied image; however, it varies in few dimensions. Some of the preprocessing functions are resizing, masking, segmentation, normalization, elimination of noise, and so on. This study performs the preprocessing operation on the applied product images by resizing the images

and filtering the noises present in the image. For resizing the images, every image is converted to a default size of 300x300 pixels. For obtaining better outcome from the product images, resized images are passed to the filtering procedure. It is essential due to the fact that many issues arise based on the noise present in the image. An image is assumed as noisy when the value is highly varied from rest of the nearby values.

3.1.1 Median Filtering

To level the noise and preserve the edges of the image, median filter is appliedfrequently. The process of median filtering for an image I is given by equation (1) as follows:

$$I_{mf}(i,j) = median\{I(i+r, j+s), (r,s) \in W\}$$ (1)

where (i, j) [1, 2, …, H] × [1, 2, …, L], H and L represent the width and height of the image. W denotes the set of coordinates within a square window. The rest of the work focuses on filtering with the window of 3×3 and 5×5, which are extensively involved in median filter.

3.1.1.1 Noise Reduction with Median Filtering

The nonlinear median filter has complex mathematical form for the image having random noise. With zero mean noise normally distributed, for the median filter, noise variance is roughly estimatedas in equation (2) as follows:

$$\sigma_{med}^2 = \frac{1}{4nf^2\left(\overline{n}\right)} \approx \frac{\sigma_i^2}{n+\frac{\pi}{2}-1} \cdot \frac{\pi}{2}$$

In above equation, σ_i^2 is defined as input noise power variance, n is denoted as median filter masking, $f^2\left(\overline{n}\right)$ is the noise function density. The average filtering noise is presented in equation (3) as follows:

$$\sigma_0^2 = \frac{1}{n}\sigma_i^2$$

The effects of median filter is based on two aspects namely the mask size and noise distribution. The performance of median filter in reducing random noise is satisfactory than that of the average filtering. However, median filter is more effective with the impulse noise, particularly where narrow pulses are farand pulse width is less than2/n. The performance can be improved when median filtering is integrated with average filtering where mask is resized based on the density of noise. This work involved in processing of MRI images that contains salt and pepper noise those are eliminated with median filtering. In algorithm 1 median filtering process is presented.

Algorithm 1: Median Filtering

1. To compute the gray histogram ihist][(0<i<G, where G denotes the range of gray) of the u nn mask, let u nnN , find the median med and record ltmed (the number of the pixel value which is less than med);

1. To let the left row shift out of the histogram, if the value of the shifting out pixels is less than med , then ltmed-1 ;

2. To let the right row shift in the histogram, if the value of the shifting in pixels is less thanmed , then ltmed _x000E_1;

3.

4. If ltmed< N/2 , then repeat med+1_x000E_, ltmed+hist[med], until ltmedN/2 ;

5. If ltmed> N/2 , then repeat med - 1,ltmed-hist[med] , untilltmedN/2 ;

6. Return the median med.

---

4. ROI Image Segmentation

In this work, segmenting ovarian canceris based on SLRFCwhere the following steps are included: Initially, upgradedSLRFC network framework was created. Next, ovarian cancer imagewas given as input for segmentation. At last, MRI images were the outcome of segmentation. Mostly, methods used to segment images comprises of three phases such as feature extraction, detection and segmentation. The first module identifies the features of the image, the next one realizes the position of the object and classifies them and last module segments the image by generating a binary mask using convolutional networks.

Usually, 'Region of Interest' or ROI is identified based on the intensity values of the pixels or regions determined by the users. The process of sorting outthe required objects from unnecessary ones is called segmentation. The threshold value is defined by the user which is gray scale intensity value above the object and below background; if this is satisfied the image is thresholdedand the process is termed as thresholding. If the object has medium intensity value between 0 and 255 representing black and white respectively, then the image can be segmented ignoring the background. The range of intensity for determining ROI is termed as Density Slice.

After segmentation, the images pixels can be reassigned with intensity either 0 or 1 denoting uninteresting and interesting respectively. During this operation, geometric values remain unchanged, but grayscale image is converted to binary. To determine morphometrics, adjacent pixels with 1 are counted ignoring those with intensity 0.When intensity values are necessary, binary image is utilized to mask the actual image revealing the ROI pixels alone. After masking, intensity values of the image can be obtained.In figure 2, image segmentation adopted for the proposed Inception V3 is illustrated as follows.
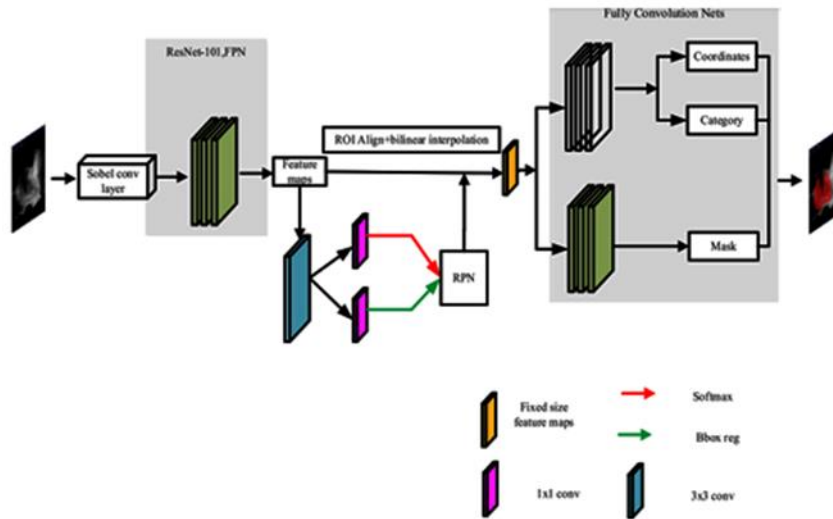

Figure 2: Structure of Segmentation

First, the proposed model takes the ovarian cancer image as input. Next, usingRegion Proposal Network (RPN), candidate region box in the feature map was generated quickly. Through ROI Align, fixed size feature image was provided as output. Further, using the detection module, target box was identified and positioned. In the MRI ovarian cancer image, using CNN in the segmentation module, foreground as well as background was predicted.The related binary mask was employed and the predicted MRI ovarian image was produced as output.

Neighborhood weighted average along with first-order differentiation was performed for detecting the edges. Two matrices were involved to convolve the actual input image. In the X and Y direction, gray difference partial derivatives were estimated. Two matrices involved in median filtering are given in equation (2) and (3):

$$Gx = \left[f(x-1, y-1) + 2 \times f(x-1, y) + f(x-1, y+1) \right] - \\ \left[f(x+1, y-1) + 2 \times f(x+1, y) + f(x+1, y+1) \right] \quad (2)$$

$$Gy = \left[f(x-1, y-1) + 2 \times f(x, y-1) + f(x+1, y-1) \right] - \\ \left[f(x-1, y+1) + 2 \times f(x, y+1) + f(x+1, y+1) \right] \quad (3)$$

whereGx and Gyrepresents the gradients in X and Y directions correspondingly.

For detecting the target image minutely and segmenting MRI ovarian images instantly, ROIpooling is unable to satisfy the necessities of positioning feature points accurately. Hence, bilinear interpolation approach is involved which replaced quantization process on the generated ROI feature graph using ROIAlign layer. Here, preservation of floating point coordinates, reduction of quantization error and realization of accurate mapping of original and feature image pixels was achieved.The principle for bilinear interpolation in x and y direction is given in equations (4) and (5) respectively:

$$f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \quad (4)$$

$$f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22})$$

(5)

The ROI of linear interpolation is stated as follows:

$$f(P) = f(x, y) = \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2)$$

(6)

where f (x, y) represents the values of the pixel of point P that has to be solved, Q11 = (x1, y1), Q12 = (x1, y2), Q21 = (x1, y2), and Q22 = (x2, y2) represents the pixel values whose known points are f (Q11), f (Q12), f (Q21), and f (Q22), and f (R1) and f (R2) denotes the values of the pixel acquired by X interpolation.

5. Pre-trained model with Feature Extraction

This section describes thestructural design of Inception v3 integrated with artificial feature extraction approach to recognize ovariancells with improved accuracy than the conventional recognition approaches. The features considered are size, scale, brightness, contrast, saturation, entropy and correlation. The major motivation is to introduce an effective design for diagnosing ovarian cancer cells with the assistance of computer.

5.1 Extraction of Augmented images

Feature extraction is involved to determine the useful features effectively from the images. Normally, cancer and normal cells differs in color as well as morphology. Typically, featuresextraction is performed. This research work selects nine features which are combined into deep learning. For extracting the color features of normal and abnormal cancer cells, Histogram is employed. The proportion of every color of the cells are described as in equation (7):

$$H(i) = \frac{n_i}{N}, i = 0,1,....,L-1$$

(7)

where i, L, ni and N denotes the gray level of the pixel, total number of gray levels, number of pixels used to represent gray level and total pixels correspondingly.

Features considered for analysis are explained as follows equation (8)-(12):

Size: The image pixel average value

$$\mu = \sum_{i=0}^{L-1} iH(i)$$

(8)

Scale: It exhibits degree of dispersion in image pixel. The increase dispersion leads to higher distribution

$$\sigma^2 = \sum_{i=0}^{L-1} (i - \mu)^2 H(i)$$

(9)

Brightness: It defines whether image is increased or decreased

$$\mu_k = \frac{1}{\sigma^4} \sum_{i=0}^{L-1} (i - \mu)^4 H(i) - 3$$

(10)

Saturation: It defines the image degree of uniformity of image

$$\mu_N = \sum_{i=0}^{L-1} H(i)^2$$

(11)

Entropy: It provides average information for given image

$$\mu_e = -\sum_{i=0}^{L-1} H(i) \log_2 [H(i)]$$

(12)

5.2 Logistic Regression and Random Forest Classifier

Logistic regression, a statistical approach, is employed to analyze the dataset where one or even more independent variables are available to estimate the outcome. Dichotomous variable with only two possibilities is used to measure the outcome. Logistic regression consist of only data coded as 1 representing TRUE or 0 which denotes FALSE.This method involves sigmoid function to classify image.The sigmoidal function $g(Z)$ is stated as in equation (13)

$$g(Z) = \frac{1}{1 + e^{-z}}$$

(13)

For every row in the input image, the feature matrix X is computed with bias value of $x_0^i = 1$, where $i = 1\, to\, m$. The feature matrix is defined as in equation (14)

$$X = \begin{bmatrix} x_0^1 & image1 \\ x_0^2 & image2 \\ x_0^m & image\, m \end{bmatrix} = \begin{bmatrix} 1 & x_j^1 \\ 1 & x_j^2 \\ 1 & x_0^m \end{bmatrix} where j = 1 \text{ to n}$$

(14)

When several parameters are given as input features, overfittingmay occur which provides poor performance of prediction since other noise parameters along withnegligible fluctuations are to be fit into the training data. Overfitting can be prevented by employing regularization function with the objective of minimizing the cost function relating to θ.

Random Forest (RF) classifier,an ensemble learning algorithm, is used for regression and to classify the instances that are re-sampled. Numerous decision trees are generated at the time of training and the class which is the output mode of these trees areprovided as output. In the near past, RF classifier has become a primary tool to classify images used in medical applications. But till now in medical applications, RF integrated with the features extracted using sparse auto encoder was not much investigated. From the results of RF classifier, it was observed that it produced higher performance than approaches such as neural network, bagging and boosting. For classifying objects from images, the performance of RF was equivalent to Support Vector Machine (SVM)but RF was easier for training as well as testing than SVM. With these factors, RF was integrated with sparse auto encoder for significant distinguishing of ovariancancer clinically. In figure 3, the overall architecture of the proposed SLRFC model is presented.
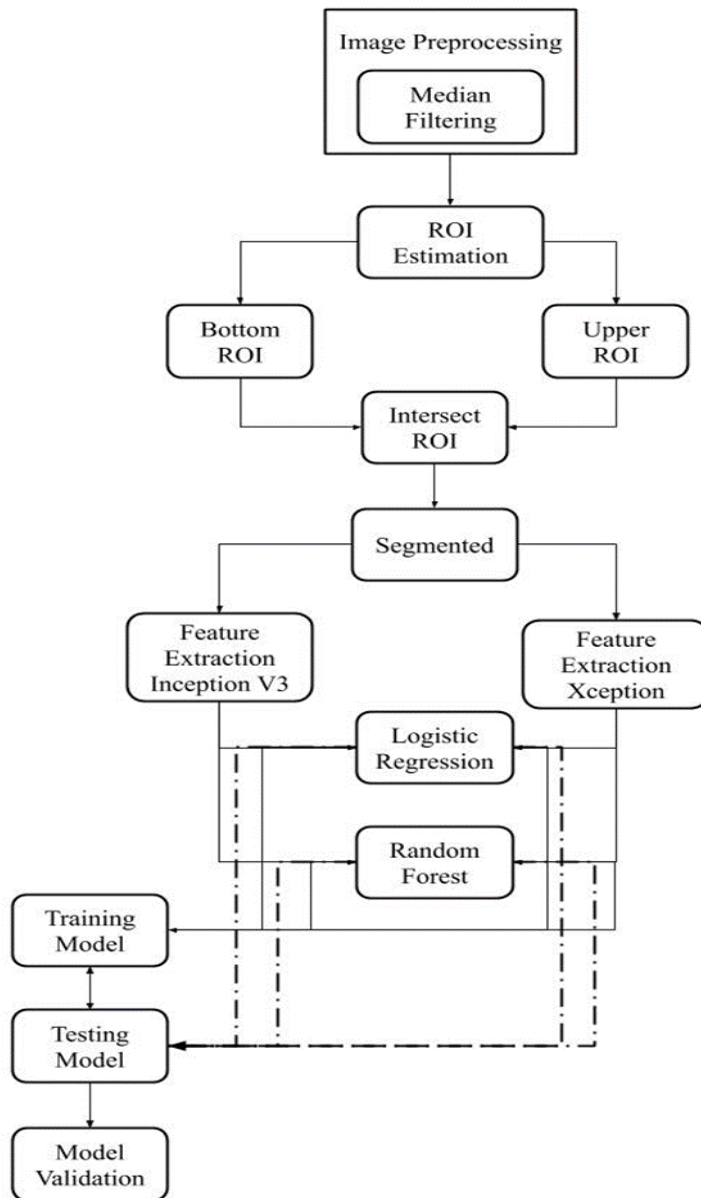
Figure 3: Flow Chart of SLRFC

This network comprises of a single hidden layer sparse auto encoder. This layer is composed 200 neurons. ROI ofApplication Delivery Controller (ADC), Bloomberg Valuation Service (BVAL) and T2 Weighted Image (T2W) are provided as input to the Subcortical arteriosclerotic encephalopathy (SAE). In SAE, the hidden layer encodes the input X to get the high-level representation $h_i^{(1)}$. The higher level features $h_i^{(1)} = \sigma\left(W^{(1)}x_i + b^{(1)}\right)$ extracted from every modality has a dimension of 330x200.The feature matrix of the inputs are combined whose resultant dimension 330×600 is provided to Adaptive Synthesis (ADASYN)to generate a balanced feature matrix. Then, the dimension of the feature matrix is enhanced as 520×600 where the instances of the clinically PCA are significantly oversampled and enhanced from 76 to 266. The balanced feature matrix is passed to a supervised RF classifier along with labels of the corresponding patch. The label of the ith patch is L(i) 2 0, 1, where 1 and 0 repre Inception v3is the enhanced version of previous inception architectures which is computationally low-cost. The primary building blocks of this model are Inception modules which allowed efficient computation along with deepnetworks by reducing the dimensionality with stacked 1×1 convolution. These modules tackled the issue related to cost, overfitting and so on. Here, the basic objective is to generate different filters with various sizes run in parallel instead of serial ones. The networks of Inception modules utilized an additional 1x1 convolution layer stacked prior to 3x3 and 5x5 convolution layers, thereby providing low computational coat and moreover was robust.

In this research work, inceptionv3 model pre-trained on the Cancer Imaging Archive dataset is introduced. In this model, classification segment is replaced using dense layers of 3 x 1 and 128 x 12 x 1 for binary and ternary

classification correspondingly. Further, this model is upgraded with ovarian MRI images to extract features effectively where Inception v3 has an input image of 224 x 224 x 3. This then passes through different inception modules to prevent overfittingand thereby reduced the cost of computation. Then, the obtained features are passes through two dense layers of 128 x 1 and 3 x 1 or 2 x 1 for classification. Then, after several forward and backpropagation iterations, Adam optimizer was used for classifying images and Inception V3 was integrated with SLRFC. In algorithm 2 Inception V3 process is presented.

Algorithm 2: SLRFC classification model with Inception V3 for Ovarian Cancer

---

Input: Input medical image database

Output: Classified Ovarian cancer region

for every image in database do

　　Segmentation of color intensities of medical images img_{seg}

　　Estimate the boundaries of image with ROI estimation of boundaries and local pattern identification $lower_{ROI}$

　　Compute intersection of ROI region $lower_{ROI}$ *and* $upper_{ROI}$.

　　For upper ROI estimation compute upper image boundaries $upper_{ROI}$.

　　Perform segmentation of ROI intersected image $lower_{ROI}.inset.upper_{ROI}$

　　Compute image feature extraction for ROI segmented image databases img_{seg}

end for

　　Generate classification model for IncerptionV3 with estimation of Adam classifiers CNN_{FE}

　　Estimate the logistic value of image features with pixels values of 0's and 1's.

　　Compute the pixel value as i^{th} and j^{th} value of images as LR_{ij}

　　Estimate random forest classification model for feature extraction

　　Compute multitude decision tree for classification as RF_{ij}

　Implement LR_{ij} *and* RF_{ij} *with Inception V3 for classification as* $InLR_{ij}$, $InRF_{ij}$,

　return $InLR_{ij}$, $InRF_{ij}$

Xceptionrepresenting extreme inception is existing from 2016. Xception model is 36 layers deep and does not include fully connected layers at last.Xceptionconsist of depth wise separable layers namelyMobileNet, and also shortcuts in which the output of certain layers and previous layers are summed up. In contrast to inceptionV3, Xceptionpacks the input records to somecompressed lumps.It autonomously maps the spatial connections for every channel, then 1×1 convolution is performed depth wise for capturing cross channel associations. Xceptionis superior to inception v3 in classifying Cancer Imaging Archive dataset. In this research work, Xception model pre-trained on the Cancer Imaging Archive dataset is introduced to detect ovarian cancer. In this model, classification segment is replaced using dense layers of 128 x 1and 3 x 1 and 128 x 1 and 2 x 1 for binary and ternary classification correspondingly. Further, this model is upgraded withovarian MRI images where Xceptionhas an input image of 224 x 224 x 3. This then passes through different depth wise separable layers and shortcuts. Then, the obtained features are passes through two dense layers of 128 x 1 and 3 x 1 or 2 x 1 for classification. The algorithm 3 presents the procedure of SLRFC classifier for ovarian cancer.

Algorithm 3: SLRFC classification model with Xception for Ovarian Cancer

---

Input: Input medical image database

Output: Classified Ovarian cancer region

for every image in database do

    Segmentation of color intensities of medical images img_{seg}

    Estimate the boundaries of image with ROI estimation of boundaries and local pattern identification $lower_{ROI}$

    Compute intersection of ROI region $lower_{ROI}$ and $upper_{ROI}$.

    For upper ROI estimation compute upper image boundaries $upper_{ROI}$.

    Perform segmentation of ROI intersected image $lower_{ROI}.inset.upper_{ROI}$

    Compute image feature extraction for ROI segmented image databases img_{seg}

end for

    Generate classification model for Xception model with estimation of hyperparameters of medical images CNN_{FE}

    Estimate the logistic value of image features with pixels values of 0's and 1's.

    Compute the pixel value as ith and jth value of images as LR_{ij}

    Estimate random forest classification model for feature extraction

---

Compute multitude decision tree for classification as RF_{ij}

Implement LR_{ij} and RF_{ij} withXception for classification as $XcLR_{ij}$ and $XcLR_{ij}$

return $XcLR_{ij}$ and $XcLR_{ij}$

6. Results and Discussion

This section discusses the results achieved by the proposed SLRFC classifier involved for ovarian cancer. The images considered for the experimentation are taken from Cancer Imaging Archive database.

6.1 Experimental setup

Pythontool is used for implementing the proposed SLRFC classifier and the system configuration is PC with Ubuntu, 4GB RAM, and Intel i3 processor.

6.2 Database description

To estimate the effectiveness of the proposed SLRFC classifier, for experimentation, single cell blood smear samples are considered from Cancer Imaging Archive database. This database has numerous collections of cropped portions of epithelial cells, Germ cell and Stromal cell[34]. The gray level properties of the Cancer Imaging Archive database aremostly similar with the Cancer Imaging Archive database, but are largerin dimension.

6.3 Performance Metrics

The confusion matric considered evaluation is based on the evaluation of other parameters like accuracy, precision, recall, and F1 - Score. The stated parameters are evaluated with the estimation of True Positive (TP), False Negative (FN), True Negative (TN), and False Positive (FP).

Accuracy: It defines the number of correctly predicted values to the total predictions made. It is defined in equation (14)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (14)$$

Recall or Sensitivity: It is defined as the correctly predicted value to the total prediction value. It is defined in equation (15)

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \quad (15)$$

Precision: It provides the ratio of true positive values to the total predicted values. It is stated in equation (16)

$$\mathrm{Pr}\,ecision = \frac{TP}{TP + FP} \quad (16)$$

F1 - Score: It provides the ratio between average mean of precision as well as recall. F1-Score is stated in equation (17)

$$F1 - Score = 2 * \frac{\mathrm{Pr}\,ecision * \mathrm{Re}\,call}{\mathrm{Pr}\,ecision + \mathrm{Re}\,call} \quad (17)$$

Confusion Matrix: It presents the performance of the model with a comparative analysis of actual and predicted values. The analysis depends on the estimation of TP, FN, FP, and TN. It is represented as in equation (18)

$$ConfusionMatrix = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \quad (18)$$

Where, True Positive (TP) is stated as forecast value which is anticipated as positive an AI model.

False Positive (FP) is defined as forecast value which is estimated as negative initially and later anticipated as positive in AI model.

True Negative (TN) demonstrated forecast value as negative and anticipated as unfavorable for AI model.

False Negative (FN) is stated as forecast value which is estimated as positive initially and later anticipated as negative in AI model.

6.4 Simulation Results

The simulation results obtained for proposed SLRFC is presented in this section. The simulation is conducted for both logistic and Random Forest classifier separately with Inception V3 and Xception model. In this section,

the results achieved for regression classifier with pre-trained Inception V3 and Xception model is presented. In the next section, Random Forest classifier with Xception model is provided.

6.4.1 Simulation Results for Logistics Regression

The simulation results obtained for logistic regression classifier with Inception V3 and Xception is presented. In figure 4 (a), confusion matrix achieved for proposed SLRFC with logistic regression with inception V3 and in figure 4 (b) xception is presented.
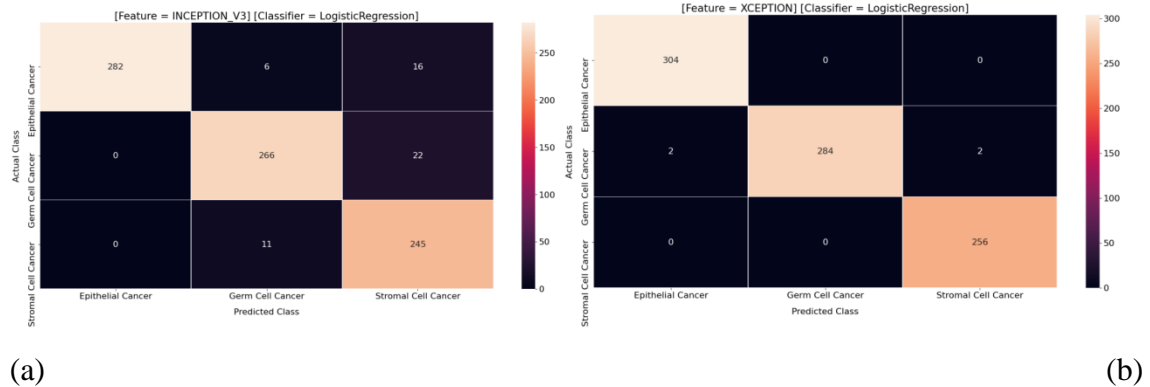

(a)   (b)

Figure 4 (a): Confusion Matrix for Logistics Inception V3

Figure 4 (b): Confusion Matrix for Xception

The evaluation of confusion matrix stated that features are classified based on three categories such as Epithelial cancer, Germ cancer and Stromal cancer. All types of classification is significant with higher TP values. This implies that proposed SLRFC is significant. In figure 5 (a) and 5 (b)performance metrics measured for proposed SLRFC with Logistics Regression Inception V3 and Xception is presented.
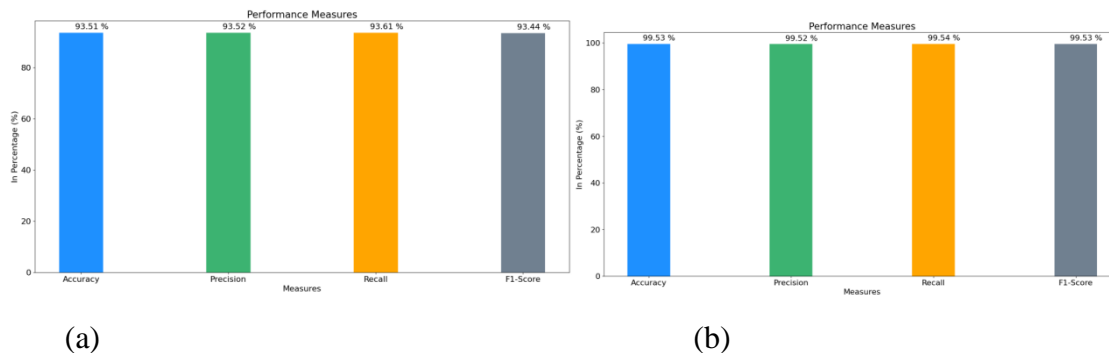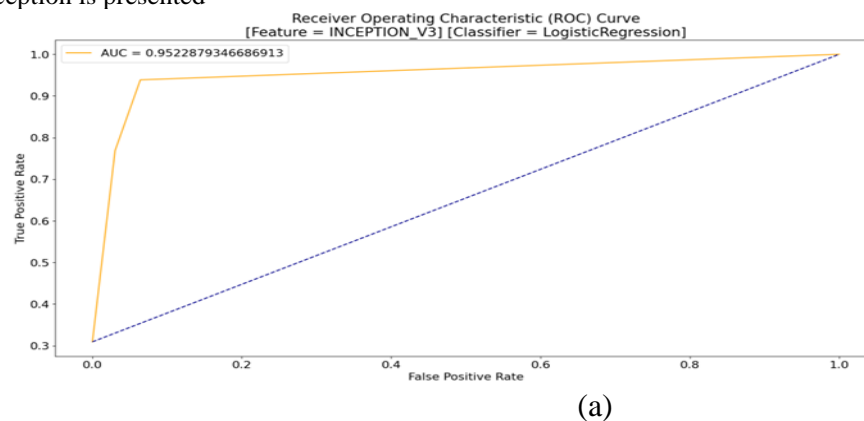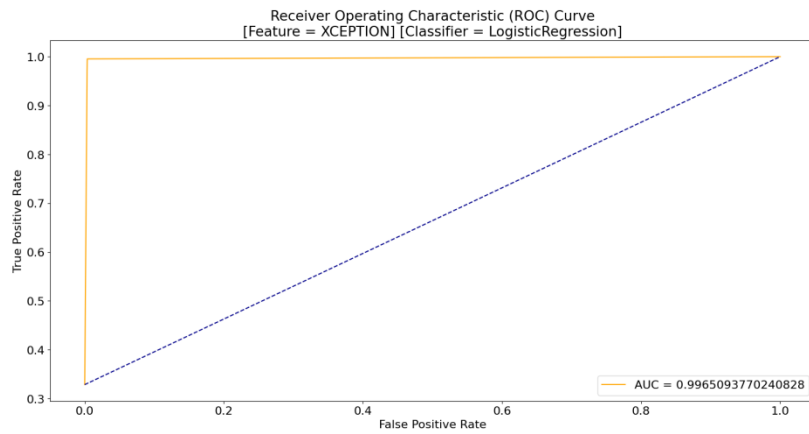

(a)   (b)

Figure 5 (a): Performance Comparison for Logistics Inception V3
Figure 5 (b): Performance Comparison for Xception

The performance metrics measured for proposed SLRFC for logistics regression with Inception V3 and Xception is effective. In figure 6(a) and figure 6(b) ROC obtained for Logistics regression Inception V3 and Xception is presented
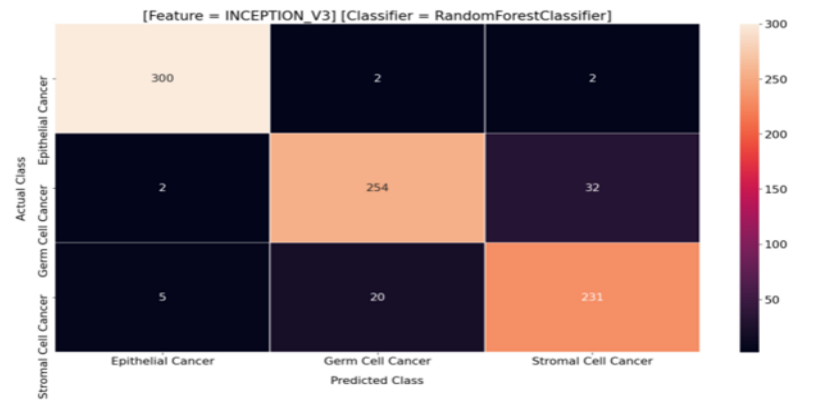

(a)

(b)

Figure 6 (a): ROC for Logistics Inception V3
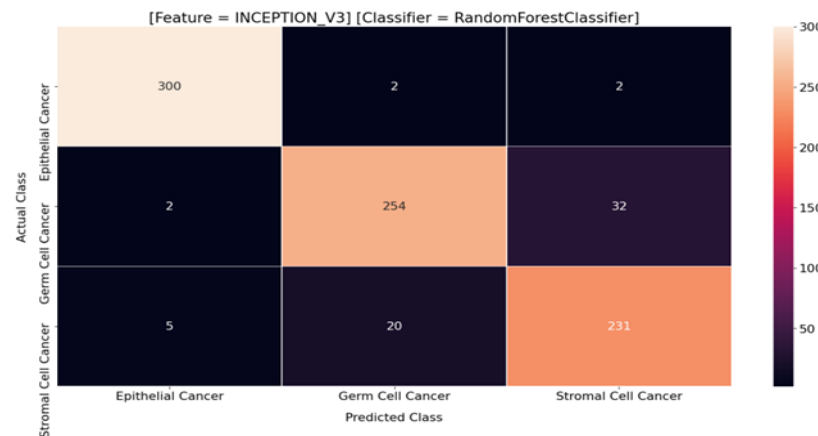
Figure 6 (b):ROC for Logistics Xception

In ROC meaurement ROC values achieved are significant for both pre-trained model implemented with Logistics Regression.

6.4.2 Simulation Results for Random Forest

The simulation results obtained for Random Forest classifier with Inception V3 and Xception is presented. In figure 7 (a) confusion matrix achieved for proposed SLRFC with logistics regression with inception V3 and figure 6 (b) provides xception is presented.
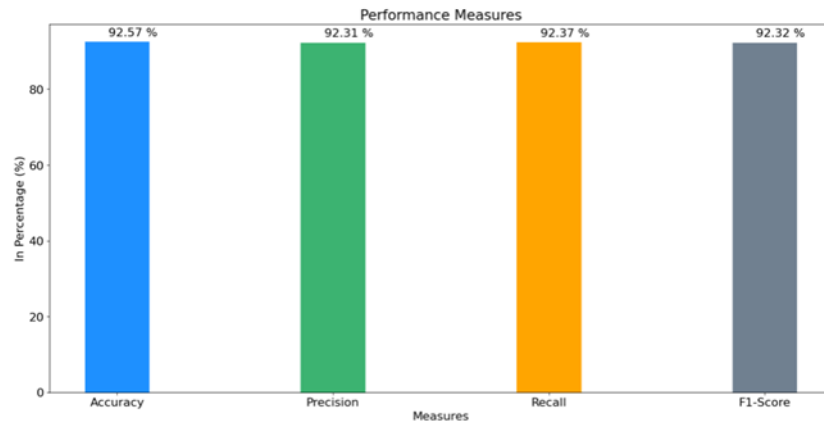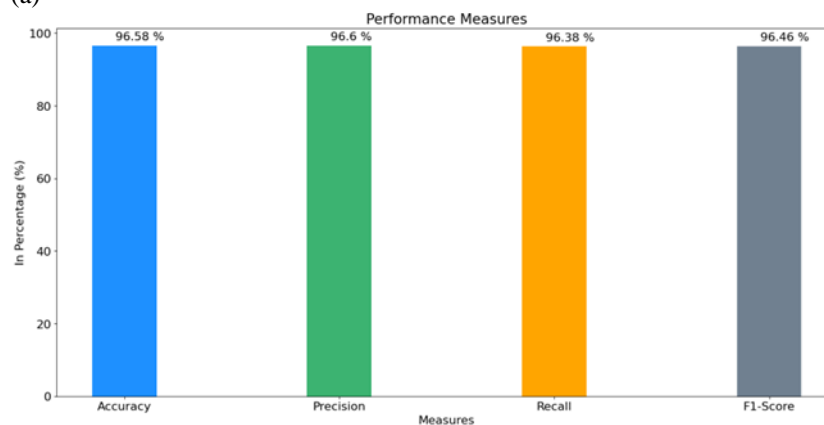

(a)

(b)

Figure 7 (a): Confusion Matrix for Random Forest Inception V3

Figure 7 (b): Confusion Matrix for Random Forest Xception

The evaluation of confusion matrix stated that features are classified based on three categories such as Epithelial cancer, Germ cancer and Stromal cancer. All types of classification is significant with higher TP values. This implies that proposed SLRFC is significant. In figure 8 (a) and 8(b)performance metrics measured for proposed SLRFC with Random Forest Inception V3 and Xception is presented.
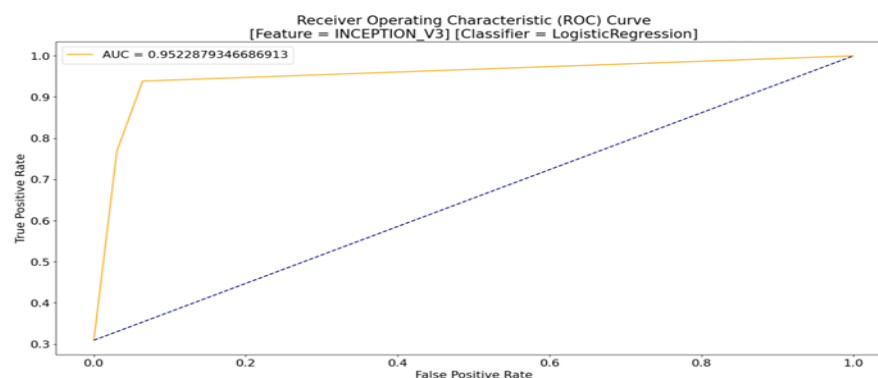

(a)

(b)
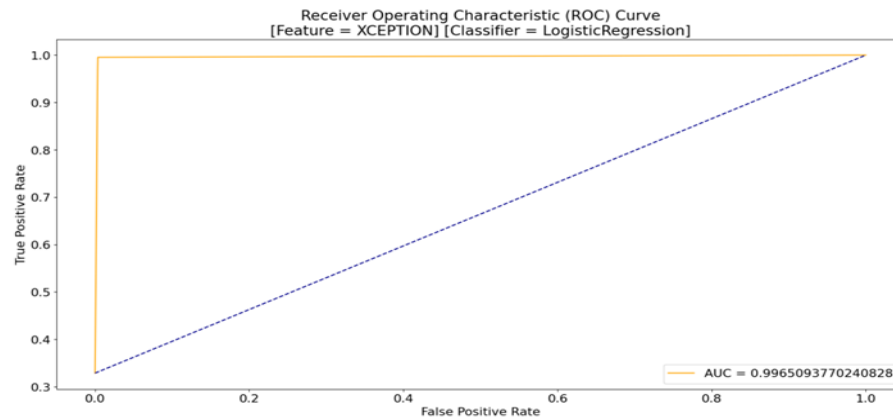
Figure 8 (a): Peformance Comparison for Random Forest Inception V3

Figure 8 (b): Performance Comparison for Random Forest Xception

The performance metrics measured for proposed SLRFC for logistics regression with Inception V3 and Xception is effective. In figure 9(a) and figure 9(b) ROC obtained for Logistics regression Inception V3 and Xception is presented

(a)

(b)

Figure 9 (a): ROC for Logistics Inception V3

Figure 9 (b):ROC for Xception

In ROC measurement ROC values achieved are significant for both pre-trained model implemented with Random Forest Classifier.

6.5 Discussion

This paper presented aaugumentation integrated with median filtering process for ovarian cancer classification stated as SLRFC. The augmented MRI ovarian images collected formCancer Imaging Archive are classified with pre-trained model for improving accuracy. The augmented model is pre-processed with median filtering followed by the ROI based image segmentation. The comparative analysis is conducted for pre-trained model Inception V3 and Xception with classifiers such as Logistics Regression and Random Forest. This research concentrated on augmentation approach to overcome the limitations associated with annotation approach. In existing, annotation subjected to limitations of increased error rate and increased complexity. To overcome the limitation of annotation augmentation is performed for reducing complexity and accuracy is improved with pre-trained models. In table 1 comparative analysis is presented.

Table 1: Overall comparison of performance metrics

Parameters (%)	Annotation		Augmentation			
	SVC	GNB	Inception V3 Logistics Regression	Xception Logistics Regression	Inception V3 Random Forest	Xception Random Forest
Accuracy	-	-	93.51	99.53	92.57	96.58
Precision	95.96	97.7	93.52	99.52	92.31	96.6
Recall	94.31	97.7	93.61	99.54	92.37	96.38
F1 - Score	97.39	98.69	93.44	99.53	92.32	96.46

In above table general comparison is presented with annotation and augmentation technique. In annotation scheme Support vector classifier (SVC) and Gaussian Naive Bayes (GNB) is presented. In augmentation scheme pre-trained models Inception V3 and Xception model is examined with classifier such as Logistics Regression and Random Forest. The overall comparative analysis expressed that Xception logistics regression provides higher precision, accuracy, recall and F1-score.

7. Conclusion

Image classification is necessaryidentifying and classifying the cancer cells in ovary. Likewise,a pre-requisite is that an appropriate classifier has to be selected to attain higher accuracy of recognition. Annotation based approach subjected to error and increased computational complexity. This paper innovatively proposes aaugmentation approach with median filtering for contrast enhancement. The image segmentation is performed based on ROI followed by feature extraction. The processed augmented dataset is classified with pre-trained model Inception V3 and Xception with classifier. The augmented results are obtained for pre-trained model with consideration of logistics regression and random forest classifiers. The comparative analysis of results expressed that Xception Logistics Regression model provides improved performance than the annotation based classifiers and other classifiers integrated with pre-trained models. In future, this research will be extended for medical image registration process for ease of access.

References

1. Torre L. A, Trabert B, DeSantis C. E, Miller K. D, Samimi G, Runowicz C. D and Siegel R. L, "Ovarian cancer statistics", CA: a cancer journal for clinicians, vol.68, no.4, pp.284-296, 2018.

2. Lheureux S, Gourley C, Vergote I and Oza A. M, "Epithelial ovarian cancer", The Lancet, vol.393, no.10177, pp.1240-1253, 2019.

3. Wu M, Yan C, Liu H and Liu Q, "Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks", Bioscience reports, vol.38, no.3, 2018.

4. Tang Z, Yang J, Wang X, Zeng M, Wang J, Wang A and Chen J, "Active DNA end processing in micronuclei of ovarian cancer cells", Bmc Cancer, vol.18, no.1, pp.1-10, 2018.

5. Eckert M. A, Pan S, Hernandez K. M, Loth R. M, Andrade J, Volchenboum S. L and Yamada S. D, "Genomics of ovarian cancer progression reveals diverse metastatic trajectories including intraepithelial metastasis to the fallopian tube", Cancer discovery, vol.6, no,12, pp.1342-1351, 2016.

6. Sawyer T. W, Rice P. F, Sawyer D. M, Koevary J. W and Barton J. K, "Evaluation of segmentation algorithms for optical coherence tomography images of ovarian tissue", Journal of Medical Imaging, vol.6, no.1, 2019.

7. Wang S, Liu Z, Rong Y, Zhou B, Bai Y, Wei W and Tian J, "Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer", Radiotherapy and Oncology, vol.132, pp.171-177, 2019.

8. Zhang L, Huang J and Liu L, "Improved deep learning network based in combination with cost-sensitive learning for early detection of ovarian cancer in color MRI detecting system", Journal of medical systems, vol.43, no.8, 2019.

9. Urase Y, Nishio M, Ueno Y, Kono A. K, Sofue K, Kanda T, Murakami T, "Simulation Study of Low-Dose Sparse-Sampling CT with Deep Learning-Based Reconstruction: Usefulness for Evaluation of Ovarian Cancer Metastasis", Applied Sciences, vol.10, no.13,2020.

10. Sawyer T. W, Rice P. F, Sawyer D. M, Koevary J. W and Barton J. K, "Evaluation of segmentation algorithms for optical coherence tomography images of ovarian tissue", Diagnosis and Treatment of Diseases in the Breast and Reproductive System, 2018.

11. Danala G, Wang Y, Thai T, Gunderson C. C, Moxley K. M, Moore K and Qiu Y, "Improving efficacy of metastatic tumor segmentation to facilitate early prediction of ovarian cancer patients' response to chemotherapy", Biophotonics and Immune Responses, 2017.

12. Zhu Y, Ferri-Borgogno S, Sheng J, Yeung T. L, Burks J, Cappello P and Mok S, "Deep learning on image-omics data in identifying prognostic immune biomarkers for ovarian cancer", 2020.

13. Klein O, Kanter F, Kulbe H, Jank P, Denkert C, Nebrich G and Darb-Esfahani S, "MALDI-Imaging for Classification of Epithelial Ovarian Cancer Histotypes from a Tissue Microarray Using Machine Learning Methods", PROTEOMICS–Clinical Applications, vol.13, no.1, 2019.

14. Yoshida K, Yokoi A, Kato T, Ochiya T and Yamamoto Y, "The clinical impact of intra-and extracellular miRNAs in ovarian cancer", Cancer science, vol.111, no.10, 2020.

15. Bodelon C, Killian J. K, Sampson J. N, Anderson W. F, Matsuno R, Brinton L. A and Ramus S. J, "Molecular classification of epithelial ovarian cancer based on methylation profiling: evidence for survival heterogeneity", Clinical Cancer Research, vol.25, no.19, pp.5937-5946, 2019.

16. Wu M, Yan C, Liu H and Liu Q, "Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks", Bioscience reports, vol.38, no.3, 2018.

17. Wen B, Campbell K. R, Tilbury K, Nadiarnykh O, Brewer M. A, Patankar M and Campagnola P. J, "3D texture analysis for classification of second harmonic generation images of human ovarian cancer", Scientific reports, vol.6, 2016.

18. Klein O, Kanter F, Kulbe H, Jank P, Denkert C, Nebrich G and Darb-Esfahani S, "MALDI-Imaging for Classification of Epithelial Ovarian Cancer Histotypes from a Tissue Microarray Using Machine Learning Methods", PROTEOMICS–Clinical Applications, vol.13, no.1, 2019.

19. Cook D. P and Vanderhyden B. C, "Ovarian cancer and the evolution of subtype classifications using transcriptional profiling", Biology of reproduction, vol.101, no.3, pp.645-658, 2019.

20. Meinhold-Heerlein I, Fotopoulou C, Harter P, Kurzeder C, Mustea A, Wimberger P and Sehouli J, "The new WHO classification of ovarian, fallopian tube, and primary peritoneal cancer and its clinical implications", Archives of gynecology and obstetrics, vol.293, no.4, pp.695-700, 2016.

21. Tseng C. J, Lu C. J, Chang C. C, Chen G. D and Cheewakriangkrai C, "Integration of data mining classification techniques and ensemble learning to identify risk factors and diagnose ovarian cancer recurrence", Artificial intelligence in medicine, vol.78, pp.47-54, 2017.

22. Prabhakar S. K and Lee S. W, "An Integrated Approach for Ovarian Cancer Classification With the Application of Stochastic Optimization", IEEE Access, vol.8, pp.127866-127882, 2020.

23. Rahman M. A, Muniyandi R. C, Islam K. T and Rahman M. M, "Ovarian Cancer Classification Accuracy Analysis Using 15-Neuron Artificial Neural Networks Model", IEEE Student Conference on Research and Development (SCOReD), pp. 33-38, 2019.

24. Amidi E, Mostafa A, Nandy S, Yang G, Middleton W, Siegel C and Zhu Q, "Classification of human ovarian cancer using functional, spectral, and imaging features obtained from in vivo photoacoustic imaging", Biomedical optics express, vol.10, no.5, pp.2303-2317, 2019.

25. Rojas V, Hirshfield K. M, Ganesan S and Rodriguez-Rodriguez L, "Molecular characterization of epithelial ovarian cancer: implications for diagnosis and treatment", International journal of molecular sciences, vol.17, no.12, 2016.

26. Su X, Yuan T, Wang Z, Song K, Li R, Yuan C and Kong B, "Two-Dimensional Light Scattering Anisotropy Cytometry for Label-Free Classification of Ovarian Cancer Cells via Machine Learning", Cytometry Part A, vol.97, no.1, pp.24-30, 2020.

27. A Hambali M and D Gbolagade M, "Ovarian cancer classification using hybrid synthetic minority over-sampling technique and neural network", Journal of Advances in Computer Research, vol.7, no.4, pp.109-124, 2016.

28. Feng Z, Wen H, Bi R, Ju X, Chen X, Yang W and Wu X, "A clinically applicable molecular classification for high-grade serous ovarian cancer based on hormone receptor expression", Scientific reports, vol.6, no.1, pp.1-9, 2016.

29. Rosendahl M, Høgdall C. K and Mosgaard B. J, "Restaging and survival analysis of 4036 ovarian cancer patients according to the 2013 FIGO classification for ovarian, fallopian tube, and primary peritoneal cancer", International Journal of Gynecologic Cancer, vol.26, no.4, 2016.

30. Liu Z, Wu H, Deng J, Wang H, Wang Z, Yang A and Tang X, "Molecular classification and immunologic characteristics of immunoreactive high-grade serous ovarian cancer", Journal of cellular and molecular medicine, vol.24, no.14, pp.8103-8114, 2020.

31. Chen K, Niu Y, Wang S, Fu Z, Lin H, Lu J and Xia D, "Identification of a Novel Prognostic Classification Model in Epithelial Ovarian Cancer by Cluster Analysis", Cancer Management and Research, vol.12, 2020.

32. Winterhoff B, Hamidi H, Wang C, Kalli K. R, Fridley B. L, Dering J and Gostout B. S, "Molecular classification of high grade endometrioid and clear cell ovarian cancer using TCGA gene expression signatures", Gynecologic oncology, vol.141, no.1, pp.95-100, 2016.

33. Arfiani A and Rustam Z, "Ovarian cancer data classification using bagging and random forest", AIP Conference Proceedings, vol. 2168, no.1, 2019.

34. Sampathkumar, A., Murugan, S., Rastogi, R., Mishra, M. K., Malathy, S., & Manikandan, R. (2020). Energy Efficient ACPI and JEHDO Mechanism for IoT Device Energy Management in Healthcare. In Internet of Things in Smart Technologies for Sustainable Urban Development (pp. 131-140). Springer, Cham.