Classification of Early Esophageal Cancer Stage using Multi-CNN Networks

CHEMPAK KUMAR A[1,*], D. MUHAMMAD NOORUL MUBARAK [2]

[1,2] Department of Computer Science, University of Kerala, Kariavatom Campus, Kariavatom, Kerala, India, 695581

*Corresponding author: email id: chempakkumar1979@gmail.com

Abstract—*The esophageal cancer is the sixth most lethal cancer with an elevated mortality rate. It is the fastest-growing form of cancer globally. The Early Esophageal Cancer diagnosis is a challenging task for the clinicians. There is a 5- fold increase in Esophageal Cancer patients initially diagnosed with Esophagitis. The persistence of Esophagitis can lead to Barrett's Esophagus and then to Esophageal Cancer. Barrett's Esophagus is one of the key precursors for the development of EC. The Convolution Neural Network has a significant role in the diagnosis of Early Esophageal Cancer. Most of the pre-trained networks can be efficiently used in Computer-Aided Diagnosis of cancer diseases. The proposed model performs the classification of the precancerous stage, Barrett's Esophagus and Esophagitis which is asymptomatic in nature. The proposed work concentrates on the use of the blending of attributes acquired from several Deep Learning networks. A Correlation based Feature Selection is utilized for selecting the relevant attributes using suitable search algorithms. The selected feature set is then classified using a Bayesnet classifier. The proposed multi-CNN model outperforms all the existing methods with an accuracy of 99.61% and an AUC of 100%. We can infer that the multi-CNN model is more efficient than the models with individual pre-trained networks from the experimental analysis.*

Keywords: *Barrett's Esophagus, Bayesnet Classifier, Esophagitis, Feature Selection, Multi -CNN networks*

1. INTRODUCTION

Globally, Esophageal cancer is one of the most widely reported forms of malignancy [1, 2]. In terms of prevalence, this disease ranks seventh and sixth in general mortality. Among the deaths due to cancer, Esophageal cancer accounts for 5% of the deaths [3]. The early detection of Esophageal cancer is vital for decreasing the mortality rate. Even though the early stages are not fatal, the poor prognosis may reduce the survival rate from 95% to 5% [4,5]. So, the diagnosis of the early stages of Esophageal cancer is clinically significant. Also, the nature of many Early Esophageal Cancer lesions is analogous to that of

normal tumour cells, which makes the diagnosis more challenging even for expert clinicians [6]. Various imaging modalities like Computed Tomography, Endoscopy, Positron Emission Tomography, and MRI can be used to evaluate the degree of malignancy [7].

The Esophageal cancer cases are rapidly increasing than any other cancer globally.[8]. The condition of the transformation of indigenous squamous epithelium to the columnar epithelium in the Esophagus due to chronic gastroesophageal reflux disease (GERD) is defined as Barrett's Esophagus [9]. The development starts from Esophagitis to non-dysplastic Barrett's Esophagus, to dysplastic Barrett's Esophagus, and finally to adenocarcinoma [10,11]. In order to reduce the mortality rate, it is better to undergo an endoscopic investigation for those having chronic GERD.[12]. The visible mucosal rifts or erosions of the Esophagus are the preliminary indicators of Reflux Esophagitis.[13]. The diagnosis of Barrett's Esophagus indicates that the patient is having chronic GERD and is more prone to malignancy. Barrett's Esophagus 's investigation was done with endoscopic surveillance of the Z line, Lower Esophageal Sphincter, and top of the gastric folds [14].

Machine Learning's evolution has enhanced the AI system's ability to extract information from scratch and learn from experience [15]. The Deep CNN based networks attained state-of-the-art performance in image classification and segmentation [16]. The ability of a deep neural network to automatically learn significant low-level features and relate them with high-level features enriches deep learning in medical applications [17]. One of the most powerful and generic methods to carry out multiple tasks such as classification, object recognition and segmentation is deep learning with CNN. The deep learning networks require a huge annotated dataset for training, which the medical sectors lack. For overcoming this difficulty, pre-trained networks trained on a large natural dataset can be utilized. The paper proposes a method for detecting Early EC stages using features extracted by different pretrained networks.

The manuscript is structured with section 1.1 describing the existing works, section 1.2 details the contribution of the proposed work. The methodology is described in Section 2 with the dataset used, feature extraction with the multiple CNN network, feature selection method, and classifier. Section 3 provides the results and discussions. Section 4 describes the conclusions of the research work.

1.1 Related works

There are number of literature that deals with the diagnosis of premalignant stages and Esophageal Adenocarcinoma. Van der Putten. et al. [18] combines the Narrow Band Imaging (NBI) and White Light Endoscopy (WLE) images for better localization of early neoplasia in Barrett's Esophagus with a ResNet18 model for accurate biopsy. Canny edge detection with image registration technique is applied for enhancing the image pairs. The

multi-modality-based model has an accuracy of 93% with a sensitivity of 92%, a specificity of 95%, and an 83% accurate biopsy location. The Inception- ResNet model proposed in Liu. et al. [19] classify BE and Esophageal Cancer tissues with an overall accuracy of 85.83%. Ebigbo. et al. [20] performs a patch-based Barrett's Esophagus analysis using a ResNet network with a Leave one Out Cross-validation. Here again, for the WLE images, the sensitivity is 97%, and specificity is 88% and is better than the NBI images. Van Riel. et al. [21] analysed the performance of some of the widely used CNN's with Support Vector Machine and Random Forest for achieving concurrent performance in Early Esophageal Cancer diagnosis. The model achieved an area under the curve ROC value of 92%. Liu. et al. [22] developed a model for early Esophageal Cancer segmentation and classification using DeepLabV3+ network with a Recall of 77.58% and precision of 78.06%. A Gastro Network is developed by Van der Putten [23] with a semi-supervised learning algorithm with Bootstrapping and Ensemble learning to obtain better classification, localization, and semantic segmentation of the Barrett's Esophagus with dysplasia. The model achieved an accuracy, sensitivity, and specificity of 87.5%,92.5%, and 82.5% with ensemble learning. A deep learning system for automated diagnosis and categorization of Esophageal variabilities was proposed by Ghatwary [24]. The automated system with Faster RCNN boosts the performance level through the concatenation of feature maps from the Gabor filters and the DenseNet network. The model obtained a Precision of 92.1% and a Recall rate of 90.2%. Struyvenberg. et al. [25] proposes a ResNet- UNet hybrid model to classify non-dysplastic Barrett's Esophagus and Barrett's Esophagus neoplasia. Here the NBI zoom images were used for the classification with transfer and ensemble learning techniques. The model achieves an accuracy of 84% and AUC of 95% with a sensitivity of 88%and specificity of 78%.

1.2 Contribution of the Proposed Work

The models mentioned above have used only one CNN model to classify and segment various Esophageal cancer and its variants from the dataset available. The combination of the unique features obtained by each pre-trained network can enhance the performance of diagnosis systems. The proposed study experiments with the combination of features from a greater number of CNNs for classification. The proposed work has the following contributions.

- In the proposed work the features are extracted by more than one pre-trained CNNs. Almost all the prevailing approaches use only one pre-trained network for feature extraction.
- The proposed study uses a composition of multiple Deep Learning networks with correlation-based feature selection (CFS) and Bayesnet classifier. No existing methods have employed such a

combination to classify early stages of Esophageal Cancer.

- The majority of the existing methods deal with the segmentation and classification of Esophageal abnormalities and EC.

2 MATERIAL AND METHODS

2.1 Dataset

The proposed work is implemented with the open access Kvasir Dataset [26], which consists of images inside the gastrointestinal (GI) tract. The dataset consists of only 94 images of BE and about 1663 images of esophagitis from various repositories with a resolution of 610x520 to1920x1072. For our analysis, a balanced number of images for training and testing the models were considered. The data augmentation is done to expand the training dataset to achieve a more relevant feature set for more accurate analysis. The dataset is expanded by performing the scaling pixel shifting, flipping, random rotation in different directions (45°, 135°, 225°), stretching vertically and horizontally of the images available. 130 labeled images of the Esophagitis and Barrett's Esophagus images in separate folders are utilized in this work. Fig 1shows the sample endoscopic images of Barrett's Esophagus and Esophagitis.

2.2 Feature Extraction

The multi -CNN network is a combination of the multiple CNN's. These pre-trained networks are generated by training the conventional CNN with the ImageNet dataset [27], a natural image dataset with 1000 class*es.* The network model has four layers viz. convolutional layer, pooling layer, activation, and fully connected layer and a Softmax classifier [28]. A series of convolution kernels matrix convolves with the image pixel values to generate a feature element. The kernel moves over the image in a progressive manner to compute the feature map. If f(m, n) is an image function and h(x, y), the kernel function, then the convolution is given by

$$C = f * h = \sum_{x}\sum_{y} h(x,y)\, f[(m-x),(n-y)] \qquad (1)$$

The feature matrix is then passed to an activation layer, which performs a non-linear operation. These non-linear functions make the network capable of handling complex patterns. The Rectified Linear Unit (ReLu) is one of the most commonly used activation functions [34]. The ReLu function is mathematically expressed as,

$$y(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \qquad (2)$$

where x is the input to the neuron. The ReLu generates only positive values so that the activation layer swipes away the gradients and enhances the computation speed. Sparsity is one of the significant advantages of ReLu. The activation functions like tanh, sigmoid, and Softmax functions are the commonly used ones.

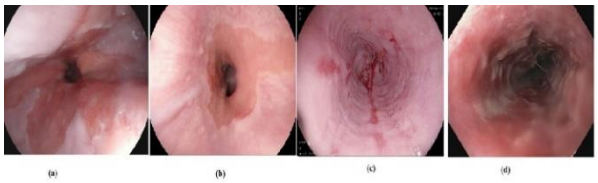

Figure 1 (a) (b)are Barrett's Esophagus images (c) (d) are Esophagitis images

The output of the activation layer is fed to the pooling layer. The pooling layer reduces the dimensions of the feature map through the process of down-sampling. So, the prominent image features are retained, resulting in a reduced number of parameters to be handled by the network. The network will have low computation time and becomes more robust. The Max pooling, mean pooling, and Min pooling are some of the commonly used pooling methods. The max-pooling layer extracts the feature elements having a maximum value in the feature matrix. The pooling layers have two hyperparameters, filter 'f' and stride 'S.' The max filter strides over the feature maps and generates the most relevant feature value. If 'n_h' and 'n_w 'are the height and width of the feature map with 'c' channels, then the dimension of the feature map when a max filter is applied gets reduced $n'_h * n'_w * c$ where;

$$n'_h = \left[\frac{(n_h - f)}{S}\right] + 1 \qquad (3)$$

$$n'_w = \left[\frac{(n_w - f)}{S}\right] + 1 \qquad (4)$$

The pooled attribute matrix is converted into a column feature matrix by concatenating the rows in the fully connected layer. All the neurons in the preceding layer are independently connected to the neurons in the fully connected layer. The CNN model can have a greater number of fully connected layers, based on the level of classification. The fully connected layer's output is applied to the Softmax Classifier, which converts the real-valued feature vectors into the probabilistic values for classification.

In the proposed work, we use eleven individual pre-trained networks. The Residual Network (ResNet) [29] ResNet50, ResNet18, ResNet101, AlexNet [30], DenseNet201[31], InceptionV3[32], Inception-ResNet V2 [33], MobileNet [34], ShuffleNet [35], SqueezeNet [36] and Xception [37] for the analysis. All the pre-trained networks can handle the RGB image but vary in the input size. The input size for the ResNet50, ResNet18, ResNet101, DenseNet201, MobileNet and ShuffleNet is 224*224. The input image size for InceptionV3, Inception-RessNetV2 and Xception is 299*299. For AlexNet and SqueezeNet, the input image size should be 227*227. The images in the dataset have different dimensions. Initially, the image dimensions are made compatible for each of the pre-trained networks for feature extraction. The extracted features are then fused to obtain a larger feature set. The FC layer provides a feature matrix of n x 1000 features where 'n' is the number of images under consideration. A combination of 'm' networks can result in n x1000 x m features for a multi-CNN network. In our proposed work, the best performing multi-CNN network consists of five networks,

DenseNet201, InceptionV3, Inception-RessNetV2, ShuffleNet, and SqueezeNet. This network generates a feature matrix of 260 x 5000.

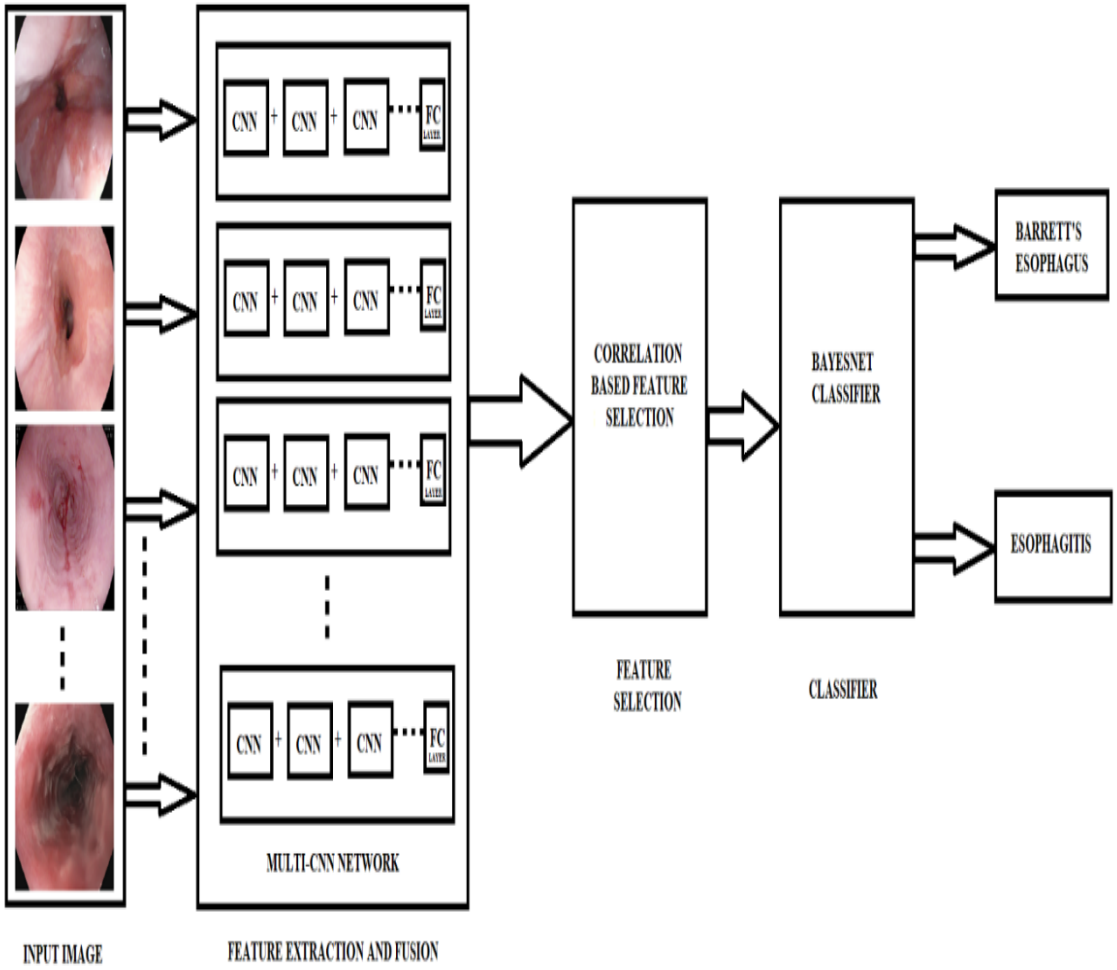

Figure 2: Proposed Architecture

2.3 Feature Selection

Correlation-based Feature Selection is one of the most commonly used filter-based feature selection methods. Correlation-based Feature Selection aims to categorize the subsets with features that are highly correlated and uncorrelated with the class. The algorithm ranks the feature subsets based on the heuristic merit function. Subsets that show high correlation with the class label and less correlation with other features will be ranked at a higher value. Correlation-based Feature Selection evaluates a subset by considering the predictive ability and the degree of redundancy (or correlation) of each one of its features individually. It is necessary to ignore the rest of

the features [38]. The subset evaluation function is given by

$$M_F = \frac{m\,\overline{r_{cf}}}{\sqrt{m+m(m-1)\ \overline{r_{ff}}}} \qquad (5)$$

where M_F is the heuristic "merit" function of a feature subset F with m features, r_{cf} is the mean feature-class correlation, and r_{ff} is the average feature-feature intercorrelation. The numerator indicates the group of features that are predictors of the class, and the denominator indicates the redundancy among them. The irrelevant features are handled by the heuristic function and the surplus features are expelled, as they are strongly associated with one or more other attributes. Correlation-based Feature Selection 's time complexity is very low. In its pure form, the best first search is exhaustive, but using a stopping criterion reduces the probability of searching the entire search space [39]. For a given class, Correlation-based Feature Selection assumes that features are conditionally independent.

The Correlation-based Feature Selection with the best first search algorithm reduces the feature matrix's dimensionality from n x 1000 x m to n x r, where 'r' is the reduced number of features. The feature vector's dimension of the best performing multiple CNN network in this work is reduced to 260 x 203 by the Correlation-based Feature Selection algorithm. The block diagram of the multi-CNN model proposed is shown in the Fig:2

2.4 Bayesnet Classifier

The uncertainty about the relationship being reviewed or observed is represented by a probability function by the Bayesian Classifier. By evaluating these probabilities along with observed data, Bayesian learning can produce the probability distributions of the quantities of interest and make optimal decisions [40]. The Bayesian decision law determines a class with the least conditional risk. The rule selects the class with the highest posterior likelihood in the minimal error rate classification. If '*v*' is an

observation from among '*n*' features such that V= (v_1, v_2….v_n), then for predicting a discrete variable 'C' to the most probable class, a Bayesian network is used. Then,

$$c^* = arg_c \, max \, P(c|v) = arg_c \, max \, P(v, c) \quad (6)$$

A Bayesian network B = [T, *Θ*], where T is a directed acyclic graph (DAG). The DAG encodes the conditional independencies with each node representing a variable in (**V**, *C*). A variable *V* is independent of its non-descendants in T given the values **pa**(*v*) of its parents. The classifier factorizes the joint probability distributions into conditional probability distributions over subsets of variables:

$$P(v, c) = P(c|pa(c)) \prod_{i=1}^{n} P\left(v_i|pa(v_i)\right) \quad (7)$$

The Bayesian network classifies the reduced feature matrix in to Barrett's Esophagus and Esophagitis

3 RESULTS AND DISCUSSIONS

3.1 Experimental Setup

In this work, Matlab2019, b tool is used for the extraction and fusion of the features. The system is configured with a i7 processor with Nvidia GPU 2050. The WEKA3.8 is used for the classification and evaluation of the model generated. The Weka tool is an integrated platform with state-of-the-art machine learning algorithms [42]. The various evaluation metrics considered are Receiver Operating Characteristic's Area Under the curve (AUC), Accuracy, Cohen's Kappa score, F-Measure, Recall, and Precision.

The objective of recall is to measure the percentage of true positives correctly identified.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (8)$$

Precision is a metric that indicates how accurate the true positive found is.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (9)$$

Accuracy is defined as the proportion of successfully identified data to the total number of classifications made by the model.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

On a given dataset, the accuracy of the network model is quantified by the F Score [47]. Most binary classifier models are evaluated using this metric, also called the F1 score. It is the arithematic mean of the network model's recall and precision.

$$\text{F Score} = 2 \text{ x } \frac{precision * recall}{precision + recall} \quad (11)$$

3.2 Search algorithm selection

For the feature extraction process, the pre-trained networks were utilized with default parameters for generating the multi-CNN. After completing the feature extraction process, the model should extract the relevant attributes and avoid redundant ones. For selecting the appropriate attributes, we apply different feature search algorithms along with the

selection method. In the proposed method, the feature selection is made by using the filter-based Correlation-based Feature Selection method. The CFS attribute selector has two search algorithms, best first and greedy stepwise algorithms. For both the search algorithms, the performance of the classifier is similar for the proposed model. The performance of the search algorithm is illustrated in Table: 1

Table 1: Comparison of the performance of Search algorithms

Algorithm	Precision	Recall	F score	AUC	Accuracy
Best Search	99.6	99.6	99.6	100	99.61
Greedy Stepwise	99.6	99.6	99.6	100	99.61

3.3 Initial Parameter Settings

For the Bayesnet classifier in WEKA, default settings are used. A simple estimator algorithm with the alpha value 0.5 is chosen for the calculation of the Bayes network conditional probability tables. The network structure search algorithm is set to the hill-climbing algorithm, K2. Table:2 displays the initial settings.

Table 2: Classifiers Initial settings

Estimator	Alpha	Search algorithm	
Simple estimator	0.5	K2	Hill climbing

3.4 Comparison of the proposed multi - CNN networks and the individual pre-trained networks

The pre-trained networks are used to mine image attributes. The attributes are then fused to obtain a more extensive feature set. Each pre-trained network extracts n x 1000 features, and a combination of 'm' networks can result in n x1000 x m features. In the proposed work, the feature extraction is done by using various combinations of the pre-trained networks. The combination of these multiple networks gives a better performance than the individual networks. We experiment with seven multi-CNN with five pre-trained network, five multi-CNN with three pre-trained networks and eleven individual pre-trained networks. The results of the various network combination are displayed in Table:3. For the individual networks, the accuracy ranged from 84.2% to 93%. The AUC for individual networks varied from 93% to 97.5%. The ResNet50 achieved a maximum accuracy of 92.69% and an AUC of 97.5% among the individual network. The accuracy of multi-CNN with three pre-trained networks varied from 95.3% to 96.5% and AUC about 99.8%. The multi-CNN composed of Resnet101, Resnet50, and Inception- Resnetv2 achieved an accuracy of 96.5% and AUC of 99.8%. For the multi-CNN with five pre-trained networks, the accuracy level reached a maximum value of 99.6 and AUC to 100%. The multi-CNN is composed of Densenet201, Inception-Resnetv2, Inceptionv3, ShuffleNet, and

SqueezeNet. All the other evaluation metrics also varied in a similar pattern, with the single network having lower values and the multi-CNN with five pre-trained networks having the maximum values. Fig: 3 depicts the confusion matrix and ROC for the proposed network.

3.5 Classification results for the proposed network

The Bayesnet Classifier achieved the highest accuracy of 99.61% and an AUC of 100%. Both Naïve Bayes and the Stochastic Gradient Descent classifier achieved an accuracy of 97.69%. But the Naïve Bayes classifier has a better AUC of 99%. When compared, the Support Vector Machine's accuracy is 96.53%, and that of Random Forest is 95.76%. But the Random Forest classifier have a much better AUC of 99.6%. The Logistic Model Tree classifier has an AUC of 98.2% and accuracy of 94.23% , far better than the Support Vector Machine with AUC 96.5%. We performed a Leave One Out Cross Validation (LOOCV) on the dataset for the classification. The results are displayed in Table:4

3.6 Comparison with the existing methods

Most of the existing techniques perform the segmentation and classification of esophageal cancer and its premalignant stage. In the proposed work, the model performs the classification of the early stages of Esophageal cancer and achieves a better result. Most of the methods utilize the same dataset for the analysis with different validation processes and networks. The Table:5 displays the evaluation of the performance of the proposed method with the existing ones. A fair comparison is not possible because of the different validation processes, performance metrics evaluated, and the desired classes. Van der Putten. et al. [18] combines the NBI and WLE images for accurate biopsy of early neoplasia in Barrett's Esophagus with a ResNet18 model. The multi-modality-based model has an accuracy of 93% with a sensitivity of 92%, a 95% specificity, and an 83% precise biopsy location. Liu. et al. [19] proposed an Inception- ResNet model to classify BE and EAC tissues with an overall accuracy of 85.83%. Ebigbo. et al. [20] performs a patch-based Barrett's Esophagus analysis using a ResNet network with a Leave one Out Cross-validation. Here again, for the WLE images, the sensitivity is 97%, and specificity is 88% and is better than the NBI images. In Van Riel. et al. [21] the performance of some of the most widely used CNNs with Support Vector Machine and Random Forest is analysed for achieving instantaneous performance in Early Esophageal Cancer diagnosis. The model achieved an ROC value of 92%. For early Esophageal Cancer segmentation and classification Liu [22] developed a model with DeepLabV3+ having a Recall of 77.58% and a 78.06% precision. proposes A Gastro Network was implemented by Van der Putten [23] with ensemble learning, which achieved an accuracy, sensitivity, and specificity of 87.5%,92.5%,

and 82.5%. A deep learning system for automated diagnosis and categorization of Esophageal variabilities was proposed by Ghatwary [24]. The automated system with Faster RCNN boosts the performance level through the concatenation of feature maps from the Gabor filters and the DenseNet network. The Precision and Recall rate of the model is 92.1%, and a of 90.2% respectively.

Table 3: **Comparison of the performance of the proposed multi -CNN networks and the individual pre-trained networks**

Network Combination	Precision	Recall	F score	AUC	Accuracy	Kappa score
Densenet201+ Inception- Resnetv2+ Inceptionv3+ ShuffleNet+ SqueezeNet	**99.6**	**99.6**	**99.6**	**100**	**99.61**	**0.9923**
Inception- Resnetv2+Resnet50+ Densenet201+Resnet101+ ShuffleNet	99.2	99.2	99.2	100	99.23	0.9846
Resnet101+ Densenet201 + AlexNet + Inception- Resnetv2+ SqueezeNet	98.9	98.8	98.8	100	98.4	0.9769
Resnet101+ AlexNet+ Densenet201 + Inception- Resnetv2+ Resnet18	98.5	98.5	98.5	100	98.46	0.9692
Resnet101+ Resnet50+ AlexNet + Inceptionv3+ Inception- Resnetv2	98.2	98.2	98.2	100	98.25	0.9636
Inceptionv3+ Densenet201+ Inception- Resnetv2 +Resnet101+ MobileNet	97.7	97.7	97.7	99.9	97.69	0.9538
Xception+Resnet50+Resnet101 + AlexNet+ Resnet18	97.4	97.3	97.3	100	97.3	0.9462
Resnet101+ Resnet50+ Inception-Resnetv2	96.7	96.5	96.5	99.8	96.5	0.9265
AlexNet +Inceptionv3+ Resnet50	96.6	96.5	96.5	99.8	96.5	0.9268
Resnet101+ Resnet50+ AlexNet	96	95.9	95.9	99.8	95.91	09144
AlexNet +Inceptionv3+ Inception-Resnetv2	95.9	95.9	95.9	99.3	95.91	0.9152
Resnet101+ Inceptionv3+ Inception-Resnetv2	95.3	95.3	95.3	99.7	95.32	0.9033
Resnet50	93	92.7	92.7	97.5	92.69	0.8538

Resnet101	92.7	91.9	91.9	97.7	91.92	0.8385
Densenet201	92.2	91.9	91.9	97.8	91.92	0.8385
Inception- Resnetv2	91	90.8	90.8	97.4	90.77	0.8154
ShuffleNet	91	90.8	90.8	96.8	90.76	0.8154
Xception	90.8	90.8	90.8	96.3	90.76	0.8154
Resnet18	90.2	90	90	96.9	90	0.8000
AlexNet	88.9	88.5	88.4	94.7	88.46	0.7692
SqueezeNet	88.4	88.1	88.1	93.2	88.08	0.7615
Inceptionv3	86.3	85.8	85.7	92.3	85.77	0.7154
MobileNet	84.2	84.2	84.2	93.2	84.23	0.6846

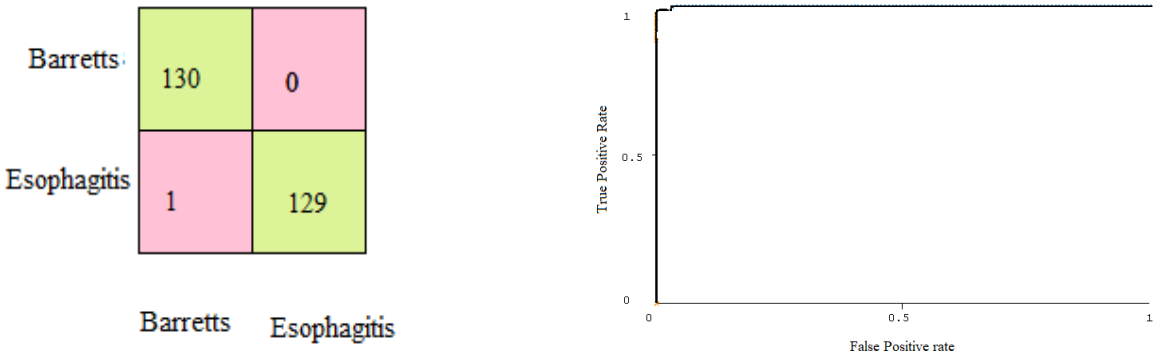

Figure 3: Confusion matrix and ROC of the Proposed Multi-CNN model

Table 4: Comparison of the performance of the various classifiers with the proposed multi -CNN model

Classifier/Metrics	CNN Model	Cross-Validation	Precision	Recall	F Score	AUC	Accuracy	Kappa Score
Bayes Network	Proposed model	LOOCV	**99.6**	**99.6**	**99.6**	**100**	**99.61**	**0.9923**
Naïve Bayes			97.8	97.7	97.7	99.9	97.69	0.9538
Stochastic Gradient Descent (SGD)			97.8	97.7	97.7	97.7	97.69	0.9538

Support Vector Machine (SVM)	96.8	96.5	93.3	96.5	96.53	0.9308
Random Forest	96	95.8	95.8	99.6	95.76	0.9154
Logistic Model Tree (LMT)	94.3	94.2	94.2	98.2	94.23	0.8846

Table 5: Comparison of the performance of the various existing methods with the proposed model

Sl.No.	Reference	Precision %	Recall %	F score%	Accuracy%	AUC %
1	**Proposed model**	**99.6**	**99.6**	**99.6**	**99.61**	**100**
2	Van Riel.et.al 2018.	--	--	--	--	92
3	Struyvenberg.et.al .2019	78	88	--	84	95
4	Van der Putten.et.al 2019,	95	92	97	93	--
5	Liu.et.al 2019	78.06	77.58	--	--	--
6	van der Putten.et.al 2019	82.5	92.5	--	87.5	--
7	Ghatwary.et.al 2019 Kvasir dataset MICCAI Dataset	92.1 91	90.2 95	--	--	--
8	Ebigbo.et.al 2019 NBI WLE	80 88	94 97	-- --	--	--
9	Liu.et.al 2020.	94.67	94.23	--	85.83	--

Struyvenberg. et al. [25] proposes a ResNet-UNet hybrid model to classify non-dysplastic Barrett's Esophagus and Barrett's Esophagus neoplasia. NBI zoom images are used for the classification with transfer and ensemble learning techniques. The model achieves an accuracy of 84% and an AUC of 95% with a sensitivity of 88%and specificity of 78%.

4 CONCLUSIONS

The efficacy of the multi - CNN pre-trained network is analyzed to classify early Esophageal Cancer stages. The fusion of features extracted from the pre-trained networks is considered for the analysis. A Correlation-based Feature Selection algorithm is applied to the feature combination for relevant attribute selection. A Bayesnet classifier is utilized for the classification of the desired classes. The multi-CNN network with combination Densenet201, Inception- Resnetv2, Inceptionv3+, ShuffleNet, and SqueezeNet outperformed all the other networks. The multi-CNN network model with Correlation-based Feature Selection and Bayesnet Classifier achieved an accuracy of 99.61% and an AUC of 1. The results shows the high efficiency of the multi-CNN model for the classification of early stages of Esophageal Cancer. In the future more network combinations with new attribute selectors and classifiers need to be analysed for the diagnosis of Esophageal Cancer stages.

REFERENCES

[1] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman,D., and Bray, F., 2015. Cancer incidence and mortality worldwide: sources, methods, and major patterns in GLOBOCAN 2012. *International journal of cancer*, *136*(5), pp.E359-E386

[2] Wong, M.C., Hamilton, W., Whiteman, D.C., Jiang, J.Y., Qiao, Y., Fung, F.D., Wang, H.H., Chiu, P.W., Ng, E.K., Wu, J.C. and Yu, J., 2018. Global incidence and mortality of oesophageal cancer and their correlation with socioeconomic indicators temporal patterns and trends in 41 countries. *Scientific reports*, *8*(1), pp.1-13.

[3] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, *68*(6), pp.394-424.

[4] Ell, C., May, A., Pech, O., Gossner, L., Guenter, E., Behrens, A., Nachbar, L., Huijsmans, J., Vieth, M. and Stolte, M., 2007. Curative endoscopic resection of early esophageal adenocarcinomas (Barrett's cancer). *Gastrointestinal endoscopy*, *65*(1), pp.3-10.

[5] Whiteman, D.C., 2014. Esophageal cancer: priorities for prevention. *Current*

Epidemiology Reports, *1*(3), pp.138-148.

[6] Liu, D., Jiang, H., Rao, N., Du, W., Luo, C., Li, Z., Zhu, L. and Gan, T., 2020. Depth Information-Based Automatic Annotation of Early Esophageal Cancers in Gastroscopic Images Using Deep Learning Techniques. *IEEE Access*.

[7] Domingues, I., Sampaio, I.L., Duarte, H., Santos, J.A. and Abreu, P.H., 2019. Computer vision in esophageal cancer: a literature review. *IEEE Access*, *7*, pp.103080-103094.

[8] Devesa, S.S., Blot, WJ and Fraumeni Jr, JF, 1998. Changing patterns in the incidence of esophageal and gastric carcinoma in the United States. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, *83*(10), pp.2049-2053.

[9] Spechler SJGoyal, R.K., 1986. Barrett's Esophagus. *N Engl J Med*, *315*, p.362-371.

[10] Reid, B.J., Sanchez, C.A., Blount, P.L. and Levine, D.S., 1993. Barrett's Esophagus: cell cycle abnormalities in advancing stages of neoplastic progression. *Gastroenterology*, *105*(1), pp.119-129.

[11] Altorki, N.K., Oliveria, S. and Schrump, D.S., 1997, July. Epidemiology and molecular biology of Barrett's adenocarcinoma. In *Seminars in surgical oncology* (Vol. 13, No. 4, pp. 270-280). New York: John Wiley & Sons, Inc.

[12] Sampliner, R.E., 1998. The Practice Parameters Committee of the American College of Gastroenterology. Practice guidelines on the diagnosis, surveillance, and therapy of Barrett's Esophagus. *Am J Gastroenterol*, *93*, pp.1028-1032.

[13] Vakil, N., Van Zanten, S.V., Kahrilas, P., Dent, J. and Jones, R., 2006. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. *American Journal of Gastroenterology*, *101*(8), pp.1900-1920.

[14] Hassall, E., 2011. Esophagitis and Barrett esophagus: Unifying the definitions and diagnostic approaches, with special reference to esophageal atresia. *Journal of pediatric gastroenterology and nutrition*, *52*, pp.S23-S26.

[15] McBee, M.P., Awan, O.A., Colucci, A.T., Ghobadi, C.W., Kadom, N., Kansagra, A.P., Tridandapani, S. and Auffermann, W.F., 2018. Deep learning in radiology. *Academic radiology*, *25*(11), pp.1472-1480.

[16] Liu, D., Jiang, H., Rao, N., Du, W., Luo, C., Li, Z., Zhu, L. and Gan, T., 2020. Depth Information-Based Automatic Annotation of Early Esophageal Cancers in Gastroscopic Images Using Deep Learning Techniques. *IEEE Access*.

[17] Xu, J., Jing, M., Wang, S., Yang, C.

and Chen, X., 2019. A review of medical image detection for cancers in digestive system based on artificial intelligence. *Expert review of medical devices*, *16*(10), pp.877-889.

[18] van der Putten, J., Wildeboer, R., de Groof, J., van Sloun, R., Struyvenberg, M., van der Sommen, F., Zinger, S., Curvers, W., Schoon, E., Bergman, J. and de With Peter, H.N., 2019, April. Deep learning biopsy marking of early neoplasia in Barrett's Esophagus by combining WLE and BLI modalities. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (pp. 1127-1131). IEEE

[19] Liu, G., Hua, J., Wu, Z., Meng, T., Sun, M., Huang, P., He, X., Sun, W., Li, X. and Chen, Y., 2020. Automatic classification of esophageal lesions in endoscopic images using a convolutional neural network. *Annals of translational medicine*, *8*(7).

[20] Ebigbo, A., Mendel, R., Probst, A., Manzeneder, J., de Souza Jr, LA, Papa, J.P., Palm, C. and Messmann, H., 2019. Computer-aided diagnosis using deep learning in the evaluation of early oesophageal adenocarcinoma. *Gut*, *68*(7), pp.1143-1145.

[21] Van Riel, S., Van Der Sommen, F., Zinger, S., Schoon, E.J. and de With, P.H., 2018, October. Automatic detection of early esophageal cancer with CNNS using transfer learning. In *2018 25th IEEE International Conference on Image Processing (ICIP)* (pp. 1383-1387). IEEE.

[22] Liu, D.Y., Jiang, H.X., Rao, N.N., Luo, C.S., Du, W.J., Li, ZW and Gan, T., 2019, August. Computer aided annotation of early esophageal cancer in gastroscopic images based on deeplabv3+ network. In *Proceedings of the 2019 4th International Conference on Biomedical Signal and Image Processing (ICBIP 2019)* (pp. 56-61).

[23] van der Putten, J., de Groof, J., van der Sommen, F., Struyvenberg, M., Zinger, S., Curvers, W., Schoon, E. and Bergman, J., 2019, October. Pseudo-labeled bootstrapping and multi-stage transfer learning for the classification and localization of dysplasia in Barrett's Esophagus. In *International Workshop on Machine Learning in Medical Imaging* (pp. 169-177). Springer, Cham.

[24] Ghatwary, N., Ye, X. and Zolgharni, M., 2019. Esophageal abnormality detection using densenet based faster r-cnn with gabor features. *IEEE Access*, *7*, pp.84374-84385

[25] Struyvenberg, M.R., de Groof, J., van der Putten, J., van der Sommen, F., Baldaque-Silva, F., Bisschops, R., Schoon, E., Curvers, W. and Bergman, J., 2019. 297–Deep Learning Algorithm for Characterization of Barrett's Neoplasia Demonstrates High Accuracy on Nbi-Zoom Images. *Gastroenterology*, *156*(6), pp.S-58

[26] Pogorelov, K., Randel, K.R., Griwodz, C., Eskeland, S.L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.T., Lux, M., Schmidt, P.T. and Riegler, M., 2017, June. Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 164-169).

[27] Deng, J., Dong, W., Socher, R., Li, LJ, Li, K. and Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). IEEE

[28] Abraham, B. and Nair, M.S., 2019. Automated grading of prostate cancer using convolutional neural network and ordinal class classifier. *Informatics in Medicine Unlocked*, *17*, p.100256.

[29] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[30] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

[31] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[32] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

[33] Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A., 2016. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint arXiv:1602.07261*.

[34] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

[35] Zhang, X., Zhou, X., Lin, M. and Sun, J., 2018. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6848-6856).

[36] Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J. and Keutzer, K., 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. *arXiv preprint arXiv:1602.07360*.

[37] Chollet, F., 2017. Xception: Deep

learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).

[38] Yildirim, P., 2015. Filter based feature selection methods for prediction of risks in hepatitis disease. *International Journal of Machine Learning and Computing*, *5*(4), p.258.

[39] Hall, M.A., 1999. Correlation-based feature selection for machine learning.

[40] Du, C.J. and Sun, DW, 2008. Object Classification. Computer vision technology for food quality evaluation, p.81.

[41] Mihaljević, B., Bielza, C. and Larrañaga, P., 2020. bnclassify: Learning Bayesian Network Classifiers.

[42] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H., 2009. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, *11*(1), pp.10-18.