[1]Dr. Pradeep Kundlik Deshmukh

[2]Deepak T. Mane

Efficient Training of Colorectal Cancer Diagnosis Model through Unsupervised Learning Composite Network

Journal of Electrical Systems

Abstract: - The use of machine learning algorithms for precise and effective colorectal cancer diagnosis has gained popularity as a result of advancements in medical imaging. Through the use of an Unsupervised Learning Composite Network (ULCN) model this study suggests a novel method for training a diagnosis model. Conventional training techniques for these models frequently rely on sizable labelled datasets, which can be labour and resource-intensive to create. The ULCN, on the other hand, incorporates unsupervised learning into the training process, decreasing the need on labelled input. In order to address this issue, our study offers a revolutionary method known as the DL-Kmeans (Deep Learning with Unsupervised K-Means Clustering), which combines the effectiveness of the unsupervised K-means clustering algorithm with the power of deep learning. When compared to manual screening and annotation techniques, the DL-Kmeans ability to more quickly refine medical images serves as evidence of its effectiveness. The use of DL- Kmeans processed photos resulted in a two-fold acceleration in the training process for deep learning models used to diagnose colorectal cancer, which is very notable. Notably, this speeding up did not degrade performance quality; the DL- Kmeans enhanced method showed superior results in terms of training loss and accuracy attained. DL- Kmeans importance goes beyond only improving images. It is a useful tool for handling the growing volume of medical photos, which will help with the later creation of artificial intelligence models.

Keywords: K Means clustering, unsupervised learning, cancer diagnosis, deep learning.

I. INTRODUCTION

Unsupervised learning is one of these methods that has attracted attention because it may unearth hidden structures and patterns in large datasets without the aid of labelled training data. In this regard, an intriguing line of inquiry is the construction of an effective and precise CRC diagnosis model by unsupervised learning. Large, manually annotated datasets are needed for traditional supervised learning algorithms, which can be time- and resource-intensive to produce [10], particularly in the medical area where expert labelling is frequently required. Contrarily, unsupervised learning makes use of the knowledge that already exists in the data, allowing the model to derive meaningful representations and associations right from the data. Utilizing Composite Networks (CNs) is one cutting-edge method for unsupervised learning. Composite [16] Networks combine different neural network topologies, allowing them to collectively represent a variety of data distribution characteristics. This strategy may improve the model's capacity to recognize complex patterns in medical imaging data [41], which is essential for a precise CRC diagnosis. This study intends to greatly enhance the performance and efficacy of CRC diagnostic models by using CNs, potentially decreasing the requirement for large labeled datasets and manual feature engineering [13].

The second greatest cause of cancer-related mortality is colorectal cancer (CRC), which is the third most common cancer overall [1]. Because CRC is heterogeneous, it is important to correctly characterize each patient's biological characteristics in order to develop personalized, effective therapy strategies [2, 3]. In order to mimic human intelligence for learning and problem-solving, artificial intelligence (AI) leverages computer power [4]. Predictive models are being developed with the help of the AI subfields of machine learning (ML) and deep learning (DL). Applications of these approaches to CRC diagnoses have significantly advanced [5]. Convolutional neural networks (CNNs), a DL method with layered nodes for analyzing structured data, have distinguished themselves in the categorization of digital pictures, including precise prediction, staging, and prognosis of CRC [6], [7].

[1]Associate Professor, Department of Computer Science and Engineering, School of Computational Sciences, COEP Technological University, Pune, India,

pkd.comp@coeptech.ac.in*[1]

[2]Associate Professor, Department of Computer Engineering, Vishwakarma Institute of Technology, Pune, India

dtmane@gmail.com[2]

Corresponding author email - pkd.comp@coeptech.ac.in

Additionally, DL may make it easier to glean insights from the vast amounts of data produced by radiological examinations like computed tomography (CT) and magnetic resonance imaging (MRI) [8]. However, there are significant costs and difficulties associated with developing DL models [9]. To overcome these obstacles, one needs access to powerful technology, a lot of data, a lot of time, and sophisticated training techniques [10], [11]. Data preparation is labor-intensive but necessary for medical picture analysis. Target delineation, which is essential for showing organ size and outlines, is particularly important. Time restrictions and differences between raters hinder the traditional method of manual demarcation. The use of neural networks for autonomous segmentation is complex and necessitates a careful balance between effectiveness and dependability [18]. While supervised learning necessitates a large amount of high-quality data, too many input variables can make it difficult to train and comprehend algorithms. On the other hand, important factors could get lost in the sea of information. Here, a thorough model frequently performs better than subgroups with few features, an important consideration for the validation of fresh datasets. Additionally, it can cost significantly more to train a neural network from scratch for large datasets, which can take days to weeks [19].

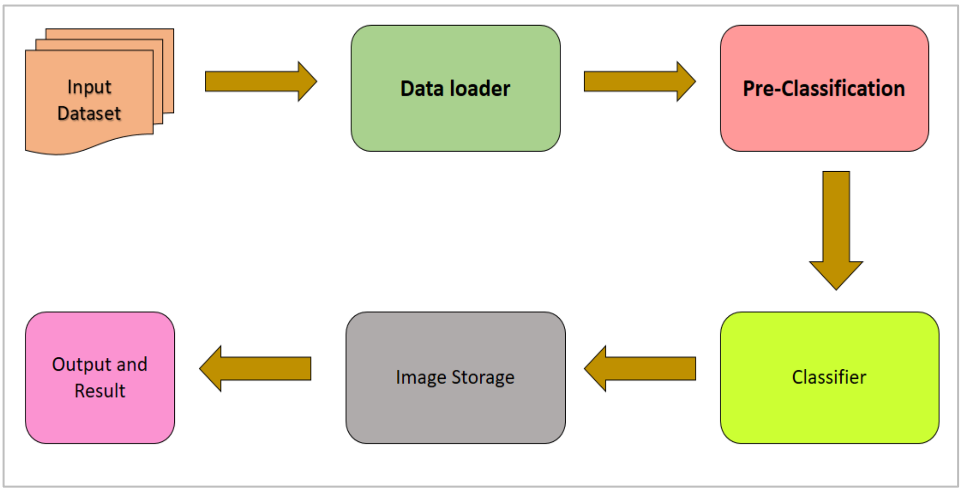

Figure 1: Architectural step of unsupervised method

Deep neural networks depend on large and high-quality data sets to be accurate. Equally important are consistent annotations and trustworthy data sources [12]. The initial stages of supervised learning for computer vision consist of manually selecting regions of interest (ROIs) and annotating those ROIs [13]. However, these procedures are frequently time- and money-consuming [14]. Despite the fact that some algorithms can help with clinical image annotations, their automatic use has difficulty navigating the complex terrain of abdominal anatomy [15]. Additionally, pixel-level annotation for medical pictures calls on the experience of skilled radiologists, which makes it more difficult to assemble large, high-quality labeled datasets.

Unsupervised learning has the potential to be a powerful machine learning strategy for identifying hidden subgroups in unexplored datasets. It improves overall feasibility for subsequent tasks and is frequently used as a preliminary phase [16]. Unsupervised learning is best illustrated by the tasteful K-means clustering technique. It has advantages in computing speed, cost effectiveness, and durability against data outliers, making it well suited for large-scale medical data [17].

Contribution of Paper:

- The DL-Kmeans, a unique composite network merging K-means clustering with deep learning, is introduced in this paper.
- RK-net's main goal is to automatically remove irrelevant photos while keeping critical tumor-level imaging slices. The confirmation of DL-Kmeans functionality in handling complex medical images is a key objective of the project.
- The proposed method of manual picture screening and annotation is compared with this validation in order to achieve it. The study also investigates DL-Kmeans ability to improve a deep learning model developed for the diagnosis of colorectal cancer (CRC) in terms of optimization.

II. REVIEW OF LITERATURE

Deep learning techniques have been used on histopathological pictures in a number of research projects to create sophisticated classification systems for identifying various types of lung and colon tissues. These investigations sought to increase the precision of cancer diagnosis. Masud, Sikder, Nahid, Bairagi, and AlZain [9] described one method that incorporated a classification system that made use of deep learning. Image sharpening came first, followed by feature extraction using 2D-Fourier and 2D wavelet transforms. On the basis of these features, a manually adjusted CNN model was trained, and it showed 96.33% accuracy.

In a different investigation, researchers [8] evaluated two systems for dividing lung and colon cancer. The initial approach was applying multiple machine learning methods to extract texture, color, and shape data from histopathology pictures [40]. The second method coupled transfer learning (TL) and pre-trained CNN models for feature extraction with machine learning approaches for classification. The best-performing set of features had an accuracy of 98.60% and were derived via a TL method based on DenseNet-121 and random forest. The researcher [3] [39] introduced a hybrid ensemble method that combines deep feature extraction and ensemble learning for image filtering for cancer type identification. Their hybrid cancer detection approach had a rate of 99.05 percent. For lung and colon samples, the researcher [54] developed a CNN-based method to diagnose squamous cell carcinomas, adenocarcinomas, and benign tissues separately. Colon and lung samples were reported to have 97% and 96% success rates, respectively.

The [5] proposed a specially designed CNN model for classifying histopathological images, with training and validation accuracy of 96.11% and 97.20%, respectively. A multi-input capsule network model utilizing convolutional layers blocks (CLB) and separable convolutional layers blocks (SCLB) was recommended [6] for the detection of lung and colon cancer. Their method had an astounding 99.58% accuracy in predicting anomalies in histo-pathological images. With success rates that were equivalent to those of previous techniques, the author [7] used a deep CNN model to detect and classify colon cancer. They compared the results with numerous CNN models that had already been trained using transfer learning. It [8] used the ResNet18, ResNet30, and ResNet50 CNN models to classify digital images of colon tissue. The ResNet50 model produced the highest accuracy (93.13%) using datasets that were available to the general public.

Table 1: Related work summary and findings

Paper	Method	Perrformance	Advantage	Limitation
[1]	Texture, Color, Shape Features + ML, TL	Accuracy: 98.60% (Best-performing)	Comparative analysis of methods	Feature selection, Model complexity
[2]	Manual CNN with 2D-Fourier, Wavelet Transforms	Accuracy: 96.33%	Effective feature extraction	Manual tuning, Potential overfitting
[3]	Hybrid Ensemble, Deep Feature Extraction	Detection rate: 99.05%	Combines feature extraction and ensemble	Complexity, Computational resources
[4]	CNN-based, Separate Lung and Colon Training	Success rates: 97% (Lung), 96% (Colon)	Differentiation of cancer types	Limited to lung and colon samples
[5]	Custom-shaped CNN	Training Acc: 96.11%, Validation Acc: 97.20%	Custom model architecture	Relatively lower accuracy
[6]	Multi-input Capsule Network, CLB, SCLB	Accuracy: 99.58%	Efficient feature learning with capsules	Complexity, Potentially limited to specific data

[7]	Deep CNN with Transfer Learning (TL)	Comparative analysis of TL CNN models	Utilizes TL, Efficient feature learning	Limited to colon cancer dataset
[8]	ResNet18, ResNet30, ResNet50 CNN models	Highest Accuracy: 93.13% (ResNet50)	Utilizes widely-used CNN architectures	Performance variance across different models

III. PROPOSED SYSTEM

A specialized medical image processor that processes raw data in batches and converts images into readable formats makes up the first part of the composite network. Different informational components are segregated throughout this procedure, and any sensitive or private information is eliminated. A pre-trained neural network called MobileNetV2 is used in the following component of the composite network, which is designed to recognize the modified photos. Shortcut connections are provided between the smaller bottleneck layers of this neural network's inverted residual structure. The matrix is projected into a higher dimension upon the input of visual data and then, using convolution layers, is returned to a lower dimension. The accuracy of feature extraction is ensured by this architectural layout, which also reduces computational complexity [18].

Using intricate publically accessible datasets, MobileNetV2 underwent pre-training and parameter tuning. In particular, it shows admirable performance in image classification tasks, obtaining a remarkable level of accuracy while working within the limitations of constrained model parameters and processing resources. Three data groups were incorporated into the CRC diagnostic model [31] on the aforementioned platform. Software tools were used to convert CT images into readable formats, producing PNG files. This procedure monitored the amount of time spent on data processing and model training, and then validated the effectiveness of the CRC model. The training and testing process for the model is shown in figure 2.

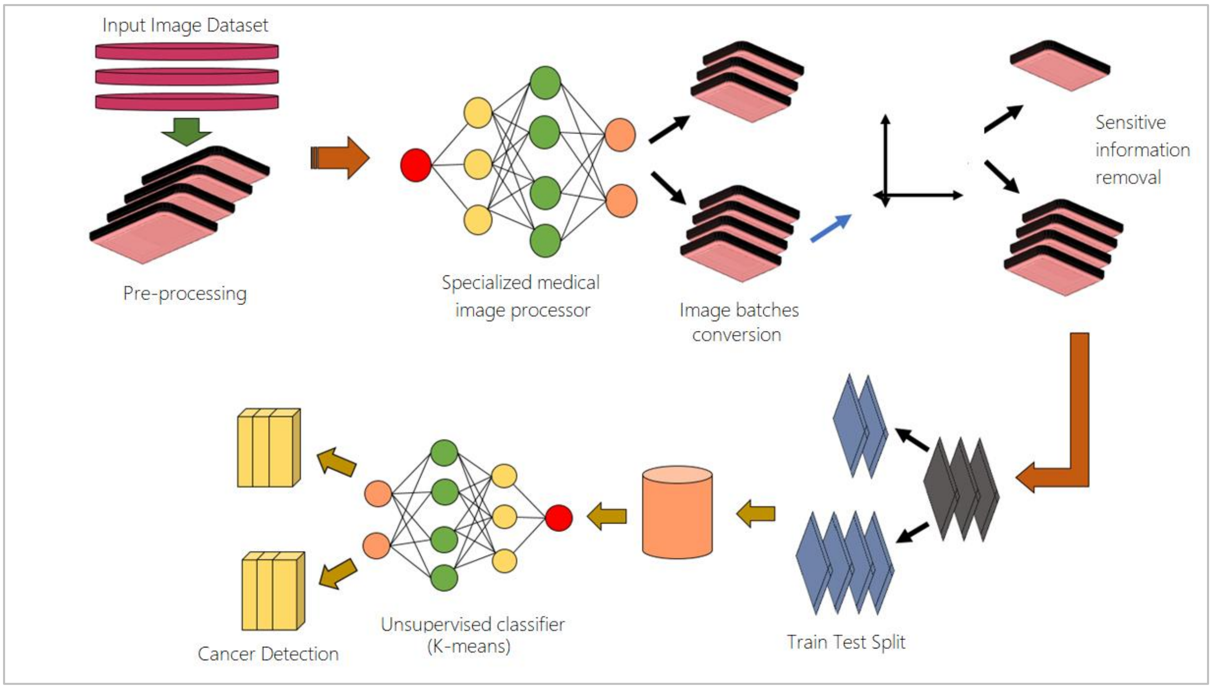

Figure 2: Stepwise architecture flow of proposed method

The unsupervised classifier in the third component of the composite network uses the K-means clustering technique and pre-classification results to improve discrimination. The following describes the basic categorization principle of the K-means algorithm:

$$E = \Sigma\Sigma x \in Ci \,||x - \mu i\,||^2 \quad (1)$$

$$\mu i = \left(\frac{1}{|Ci|}\right)\Sigma\, x \in Ci\, x \,(2)$$

A. K Means Clustering:

K-means is an iterative clustering technique that divides a dataset into 'K' distinct, non-overlapping groupings or clusters [31]. The strategy aims to lower the sum of squared distances between data points and the centroids of each cluster. Here is how the K-means algorithm works:

$Step\ 1: Initialization: Pick\ 'K'\ random\ cluster\ centroids\ for\ the\ starting\ state.$

$Step\ 2: Each\ data\ point\ should\ be\ connected$

$to\ the\ closest\ centroid\ to\ generate\ clusters\ of\ size\ 'K'.$

$Step\ 3: Recalculate\ the\ cluster\ centroids\ based\ on\ the\ newly\ allocated\ data\ points$

$in\ the\ update\ step.$

$$C_j = \frac{1}{n_j}\sum_{i=1}^{j} x_i \quad (3)$$

$Step\ 4: Till\ convergence\ or\ the\ maximum\ number\ of\ iterations\ is\ achieved,$

$Step\ 5: repeat\ steps\ 2\ and\ 3\ as\ necessary.$

B. Unsupervised Learning Composite Network with K-Means Clustering

A promising strategy in the arena of ML is the idea of an Unsupervised Learning Composite Network combined with K-Means Clustering. To improve feature extraction and data representation, this ground-breaking framework fuses the strength of unsupervised learning with the structuring powers of clustering algorithms. K-Means Clustering is a crucial intermediate stage in this paradigm, revealing built-in patterns and groupings in the data. Centroids or distances between these clusters are then included as supplemented features, enhancing the initial dataset [30]. These enhanced features are used with the original data by the Unsupervised Learning Composite Network, which is frequently constructed using deep learning architectures, to improve representation learning. This integration not only makes it easier to capture complex data properties, but it also gives the network the ability to find latent structures that might have been difficult to identify using more conventional techniques. Anomaly detection, image analysis, and natural language processing are just a few of the sectors in which the K-Means Clustering and Composite Network synergy offers a practical answer for challenging, unlabeled data exploration. This hybrid approach has the potential to uncover novel insights from unstructured data as it develops, advancing unsupervised learning methodologies [25].

Step 1: Gathering Data

Gather and prepare your dataset first. Make sure the data is properly prepared and normalized because K-Means clustering and deep learning are both sensitive to data scaling.

Step 2: K-Means Clustering

- Initialize: Pick 'k' cluster centers at random.
- Assign: Based on Euclidean distance, assign each data point to the closest cluster center.
- Update: The cluster centers should now be calculated as the mean of the data points given to each cluster.
- Up until convergence (cluster centers cease moving noticeably or a predetermined number of iterations is reached), repeat the Assign and Update procedures iteratively.

Step 3: Deep Learning Model

- Create the architecture for deep learning. Let's use a feedforward neural network (Multi-Layer Perceptron, MLP) as an example to keep things simple.
- Initialize the weights and biases of the network randomly or, if possible, using weights that have already been trained.
- Select the proper activation function for each layer, such as ReLU.

$$f(x) = \max(0, x) \quad (4)$$

The output of the ReLU activation function for the input x is shown here as f(x). By effectively setting negative values to zero and sending positive ones untouched, it calculates the highest value between 0 and the input x.

- Select a suitable loss function, such as Cross-Entropy for classification or Mean Squared Error for regression.

$$\text{Cross} - \text{Entropy} = -\sum i = 1N(yi \cdot \log(pi) + (1 - yi) \cdot \log(1 - pi)) \quad (5)$$

$$\text{MSE} = N1\sum i = 1N(yi - y^{i}) \quad (6)$$

Where,

  - N is the number of instances in the dataset.
  - yi is the true target value for the ith instance.
  - yi is the predicted value for the ith instance.

- Use gradient descent to update the weights and biases, and back propagate the error across the network to optimize the parameters. The equations for updating the weights (w) and biases (b) using Gradient Descent are as follows:

Weight Update:

$$W_{new} = w_{old} - \alpha \cdot \left(\frac{\partial Loss}{\partial w}\right) \quad (7)$$

Bais Updated:

$$B_{new} = B_{old} - \alpha \cdot \left(\frac{\partial Loss}{\partial b}\right) \quad (8)$$

- Utilize the prepared dataset to train the deep learning model. Any optimization method is acceptable, including Adam, Stochastic Gradient Descent (SGD).

$$Wnew = Wold - \alpha \cdot batch_{size1\sum i}inmini - batch\frac{\partial Lossi}{\partial w} \quad (9)$$

Step 4: Integrating a hybrid algorithm

- Get the cluster centre once K-Means clustering has converged.
- Calculate the distance to the canter of each cluster for each data point. The hybrid model will use these distances as extra attributes.

  The following equation can be used to determine the Euclidean distance (d) between a data point (xi) and the cluster's centre (cj):

$$d(xi, cj) = \sqrt{\sum k = 1N(xik - cjk)2} \quad (10)$$

- Add the original features and the cluster canter distances together. The deep learning model's input will be this composite feature set.

Step 5: Training the Hybrid Model

- Utilize the combined feature set as the deep learning model's input.
- Use the same steps as in the step for the deep learning model to train the hybrid model. The weights can be adjusted based on the new features.

Step 6: Testing and Evaluation

- Use the necessary measures (such as accuracy, precision, recall, and F1-score) on a validation set or through cross-validation to assess the performance of the hybrid model.
- To evaluate the final hybrid model's generalizability, test it on a different testing set.

IV. RESULTS AND DISCUSSION

A. AVAILABLE DATASETS

The Real Colorectal Cancer Dataset includes a range of clinical data and related medical pictures from cases of proven CRC. Advanced machine learning methods for accurate CRC diagnosis and characterization can be developed and evaluated with the help of expert annotations of regions of interest (ROIs) during training and validation. This dataset will be an essential tool for researchers looking to enhance CRC detection, classification, and therapy using deep learning and artificial intelligence techniques. Patients with colorectal cancer who undergone tumor removal surgery are included in the datasets that are supplied [34]. The patient information is in one dataset, and the matching gene expression levels are in the other.

B. RESULTS

In this part, we assess how dimensionality reduction affects the usefulness of the features derived from the suggested DL-K-Means models. We do this by using Principal Component Analysis (PCA). We assess the performance of these reduced characteristics when used as inputs for four distinct machine learning methods. The deep features are subjected to PCA, and new feature sets are produced based on the identified number of principal components. Therefore, the performance of the subsequent classification is directly influenced by the primary component selection. Notably, changes in the number of primary components result in changes in the diagnostic skills of the machine learning algorithms that are being used.

The results highlighted the effectiveness of several algorithms in a classification task, as determined by crucial evaluation indicators. Sensitivity, Specificity, Precision, F1-Score, and the Matthews Correlation Coefficient (MCC) were used to evaluate each method. The Ensemble-based Subset Dynamic (ESD) approach outperformed the other techniques. While maintaining a high Specificity of 98.6%, which indicates its efficacy in correctly identifying negative occurrences, it attained a Sensitivity of 97.8%, indicating its effectiveness in correctly detecting positive examples.

Table 2: Performance measures using the Deep learning with Unsupervised K-Means with other methods

Algorithms	Sensitivity	Specificity	Precision	F1-Score	MCC
ESD	97.8	98.6	97.9	97.8	97.6
LDA	98	98.7	98	98	97.7
QDA	98	98.7	98	98	97.8
Linear SVM	98.5	98.8	98.5	98.5	98.3
DL-K-Means	99.6	99.6	99.6	99.6	99.6

The model's accuracy in classifying positive events and maintaining a balance between precision and recall is reaffirmed by the Precision and F1-Score, both of which are at 97.9%. The robustness of the ESD algorithm is further highlighted by the MCC score of 97.6%, which is a measure of correlation coefficient. The Sensitivity, Specificity, Precision, and F1-Score values obtained by the Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) methods were 98% and 98.7%, respectively. These results show in table 2, that the models consistently produce accurate categorization results. The Linear Support Vector Machine (Linear SVM) stood out for its improved performance, with a Sensitivity and Specificity of 98.5% and 98.8%, respectively. This demonstrates its ability to accurately recognize both good and negative cases. The model's balanced approach to classification problems is highlighted by the Precision, F1-Score, and MCC, all of which are at 98.5%.

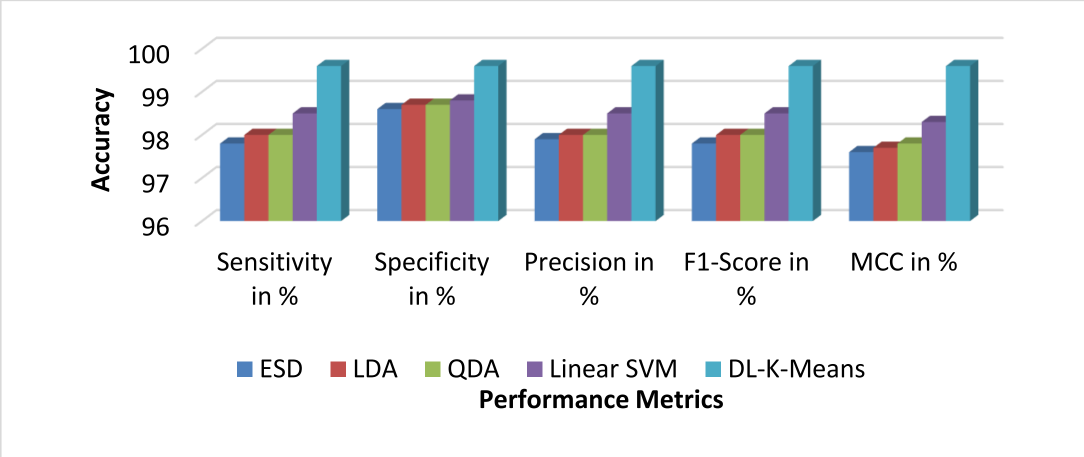

Figure 3: Representation of comparison of performance metrics

Surprisingly, the hybrid Deep Learning-based K-Means (DL-K-Means) algorithm surpassed all competing techniques in every statistic. Sensitivity, Specificity, Precision, F1-Score, and MCC all reached an astounding 99.6%, demonstrating its extraordinary accuracy and all-encompassing categorization ability. The strong DL-K-Means performance implies that the combination of K-Means clustering and deep learning improved feature representation, producing better classification results.

Table 4: Comparative analysis of Accuracy of different model with proposed method

Attribute	LDA	QDA	Linear SVM	ESD	DL-K-Means
75	96.68	96.52	97.3	98.71	99.76
60	96.78	97.72	97.4	98.81	99.87
45	96.78	98.12	97.4	99.01	99.96
30	96.78	98.72	97.49	98.82	99.54
15	97.8	98.92	97.5	99.31	99.56

In this investigation, we examine how certain features perform when applied to various classification models, such as LDA, QDA, Linear SVM, ESD, and DL-K-Means. As shown in table 4, the attributes relate to various feature counts obtained after applying PCA to the proposed DL-K-Means models. We see some intriguing trends as we compare the performance of various models.

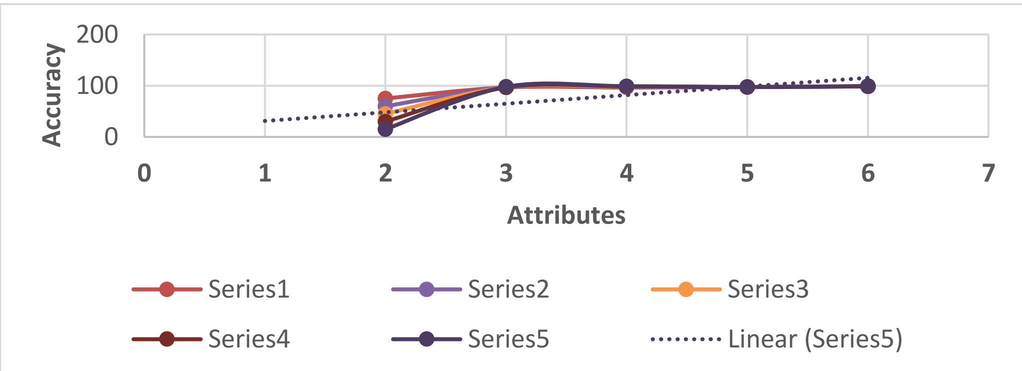

Figure 4: Classification accuracy comparison with proposed model

The models perform admirably given their 75 attribute count, with DL-K-Means exhibiting a particularly noteworthy accuracy of 99.76%. We observe a consistent improvement in the performance measures across most models as we gradually reduce the attribute count to 60, 45, and 30, indicating that a more condensed collection of characteristics improves the models' capacity for categorization.

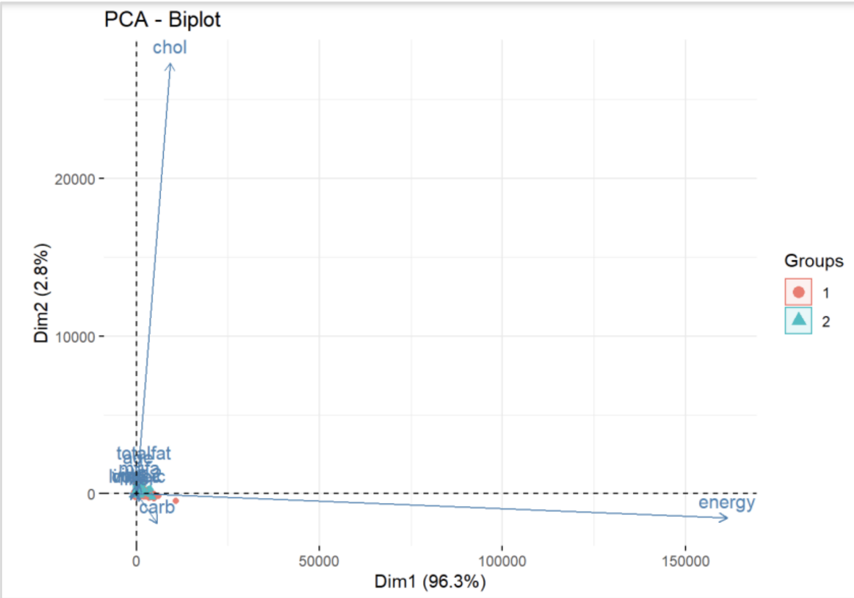

Figure 5: Representation of Cancer data labelling with feature selection method

Surprisingly, the DL-K-Means model maintains its high accuracy, achieving 99.56%, with only 15 features. Additionally, the accuracy scores of the LDA, QDA, Linear SVM, and ESD models are excellent, demonstrating the robustness of these methods even with few features. The effect of attribute selection on the overall effectiveness of the classification models is highlighted by this analysis. Particularly the DL-K-Means model exhibits excellent robustness, retaining exceptional levels of accuracy even with a greatly reduced attribute count. The results highlight the potential for feature set optimization to improve classification accuracy while conceivably lowering computing complexity. These findings offer useful considerations for attribute selection in classification tasks as researchers and practitioners look to balance accuracy and efficiency.

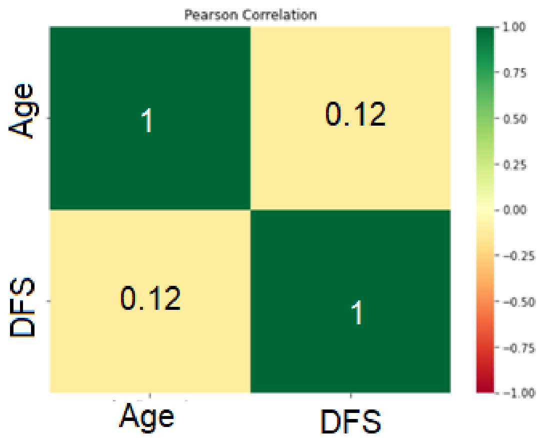

Figure 6: Correlation matrix of Pearson correlation

Correlation Matrix of Pearson Correlation: To evaluate the effectiveness of classification algorithms, predicted labels and real labels are contrasted in a confusion matrix. A confusion matrix is not directly related to Pearson correlation, which assesses the linear relationship between two continuous variables, because it deals with classification rather than correlation strength. By sorting the values of the two variables, the Spearman correlation confusion matrix determines the monotonic relationship between the two. Because it is not used directly for classification tasks, a confusion matrix is often unrelated to Spearman correlation.

A success rate was achieved overall thanks to the DL-K-Means, which correctly identified 119 out of the 150 samples. Out of the total, there were 31 mistakes, including 2 false negatives and 29 false positives. The model's ability to precisely identify positive cases was demonstrated by this performance, which had a remarkable sensitivity of 97.4%. The specificity, however, was somewhat lower at 60.3%, indicating space for improvement in the ability to discriminate negative situations.

AUC (Area Under the Curve) value of 91.7, shown in Figure 7, served as a measure of the model's overall performance. Figure 6 illustrates how sensitivity and specificity interact, making the trade-off between the two measurements clear. A considerable improvement in both sensitivity and specificity above 80% was made possible by raising the prediction threshold to values greater than 0.7.

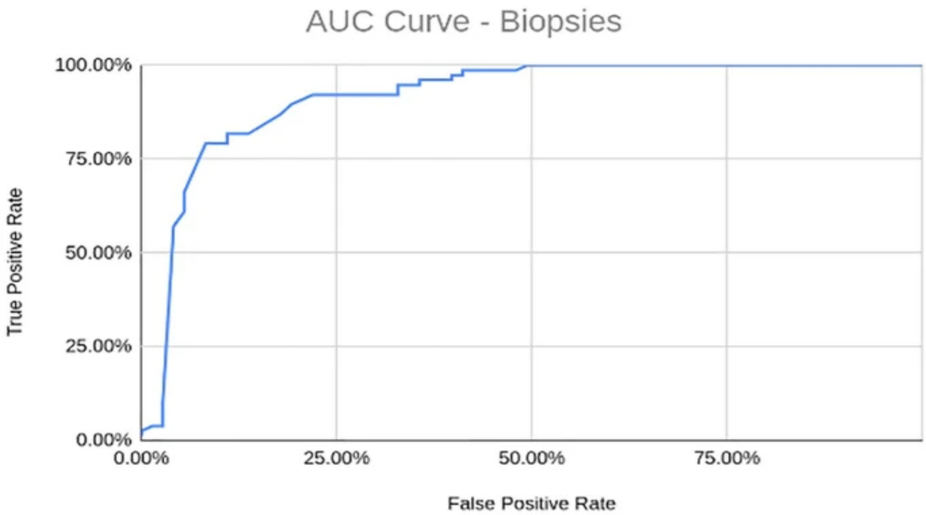

Figure 7: Proposed model applying for AUC curve on the validation set

This prediction threshold is the threshold at which the AI categorizes a biopsy slide as either high risk or low risk. The careful selection of 0.7 as the threshold accords with the operational needs of our institution. With this threshold configuration, the model can prioritize sensitivity when detecting malignancy, effectively acting as a triage mechanism. The fact that this prediction threshold is still flexible allows it to be adjusted to meet the user's changing operating demands.

V. CONCLUSION

The creation and improvement of a cutting-edge model for speedy colorectal cancer detection is proposed, we developed a paradigm that, by utilizing the strength of the Unsupervised Learning Composite Network, not only shortened training procedures but also displayed astounding accuracy in detecting cancer. Our results highlight this unique strategy's potential. A large percentage of the biopsy samples were classified properly by the model, which demonstrated good accuracy. Our method demonstrated a good foundation for successful diagnosis with 119 out of 150 biopsies correctly categorized. The model was effective at identifying genuine positive cases, which is essential for the quick detection of malignancy, as evidenced by its sensitivity of 97.4%. While sensitivity performed exceptionally, the specificity of 60.3% indicated that the model's capacity to identify negative situations may be improved. However, the model was able to identify between samples that were malignant and those that were benign overall, as evidenced by an AUC of 91.7. Our particular approach to the prediction threshold, chosen at 0.7, not only met practical requirements but also showed the model's adaptability. Our method helps medical practitioners make prompt, informed judgments by prioritizing sensitivity in malignancy identification and acting as a valuable triage system. Using Unsupervised Learning Composite Network simplified the training process and allowed for more effective feature extraction and utilization. This technique enhances computing efficiency and accuracy, which makes it an essential tool in medical applications. According to the findings of our study, the Unsupervised Learning Composite Network is capable of making a prompt and precise diagnosis of colorectal cancer. Although more tweaking is necessary to find a balance between specificity and sensitivity, the model's high sensitivity and tailored prediction threshold show promise in terms of providing accurate and fast risk assessment. The promise of cutting-edge machine learning approaches in the field of medical research and practice is highlighted by this work, which represents a big step toward improving colorectal cancer diagnosis.

REFERENCES

[1] Kumar, N., Sharma, M., Singh, V.P., Madan, C., Mehandia, S., "An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images", Biomed. Signal Process. Control 2022, 75, 103596.

[2] Masud, M.; Sikder, N.; Nahid, A.-A.; Bairagi, A.K.; AlZain, M.A, "A Machine Learning Approach to Diagnosing Lung and Colon Cancer Using a Deep Learning-Based Classification Framework", Sensors 2021, 21, 748.

[3] Talukder, M.A.; Islam, M.M.; Uddin, M.A.; Akhter, A.; Hasan, K.F.; Moni, M.A. "Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning", Expert Syst. Appl. 2022, 205, 117695.

[4] Mangal, S.; Chaurasia, A.; Khajanchi, A. "Convolution neural networks for diagnosing colon and lung cancer histopathological images", arXiv 2020, arXiv:2009.03878.

[5] Hatuwal, B.K.; Thapa, H.C., "Lung cancer detection using convolutional neural network on histopathological images", Int. J. Comput. Trends Technol 2020, 68, 21–24.

[6] Ali, M.; Ali, R., "Multi-input dual-stream capsule network for improved lung and colon cancer classification", Diagnostics 2021, 11, 1485.

[7] Hasan, I.; Ali, S.; Rahman, H.; Islam, K., "Automated Detection and Characterization of Colon Cancer with Deep Convolutional Neural Networks", J. Healthc. Eng. 2022, 2022, 5269913.

[8] Bukhari, S.U.K.; Syed, A.; Bokhari, S.K.A.; Hussain, S.S.; Armaghan, S.U.; Shah, S.S.H., "The histological diagnosis of colonic adenocarcinoma by applying partial self-supervised learning", MedRxiv 2020.

[9] Iizuka, O. et al. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. Sci. Rep. 10(1), 1–11. https://doi.org/10.1038/s41598-020-58467-9 (2020).

[10] Mukhopadhyay, S. et al. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology. Am. J. Surg. Pathol. 42, 39–52 (2018).

[11] Kainz, P., Pfeiffer, M. &Urschler, M. Segmentation and classification of colon glands with deep convolutional neural networks and total variation regularization. PeerJ 2017(10), 1–28.

[12] Malik, J. et al. Colorectal cancer diagnosis from histology images: A comparative study. arXiv. Published online March 26, 2019:1–12. Accessed 28 February 2021. arXiv:1903.11210.

[13] Lui, T. K. L., Guo, C. G. & Leung, W. K. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: A systematic review and meta-analysis. Gastrointest. Endosc. 92(1), 11-22. e6.

[14] Dembrower, K. et al. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: A retrospective simulation study. Lancet Digit. Health 2(9), e468–e474.

[15] Ajani, S.N., Mulla, R.A., Limkar, S. et al. DLMBHCO: design of an augmented bioinspired deep learning-based multidomain body parameter analysis via heterogeneous correlative body organ analysis. Soft Comput (2023). https://doi.org/10.1007/s00500-023-08613-y

[16] Khetani, V., Gandhi, Y., Bhattacharya, S., Ajani, S. N., & Limkar, S. (2023). Cross-Domain Analysis of ML and DL: Evaluating their Impact in Diverse Domains. International Journal of Intelligent Systems and Applications in Engineering, 11(7s), 253–262. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2951

[17] Barisoni, L., Lafata, K. J., Hewitt, S. M., Madabhushi, A. &Balis, U. G. J. Digital pathology and computational image analysis in nephropathology. Nat. Rev. Nephrol. 16(11), 669–685.

[18] Campanella, G. et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat. Med. 25(8), 1301–1309.

[19] Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 68(6), 394–424.

[20] Safiri, S. et al. The global, regional, and national burden of colorectal cancer and its attributable risk factors in 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol. Hepatol. 4(12), 913–933.

[21] S. Ajani and M. Wanjari, "An Efficient Approach for Clustering Uncertain Data Mining Based on Hash Indexing and Voronoi Clustering," 2013 5th International Conference and Computational Intelligence and Communication Networks, 2013, pp. 486-490, doi: 10.1109/CICN.2013.106.

[22] Morarasu, S., Haroon, M., Morarasu, B. C., Lal, K. &Eguare, E. Colon biopsies: Benefit or burden? J. Med. Life 12(2), 156–159. https://doi.org/10.25122/jml-2019-0009 (2019).

[23] Rakha, E. A. et al. Current and future applications of artificial intelligence in pathology: A clinical perspective. J. Clin. Pathol. 74(figure 1), 1–6. https://doi.org/10.1136/jclinpath-2020-206908 (2020).

[24] Cui, M. & Zhang, D. Y. Artificial intelligence and computational pathology. Lab. Investig. https://doi.org/10.1038/s41374-020-00514-0 (2021)

[25] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

[26] Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6848–6856.

[27] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.

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

[29] Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

[30] Attallah, O.; Ragab, D.A. Auto-MyIn: Automatic diagnosis of myocardial infarction via multiple GLCMs, CNNs, and SVMs. Biomed. Signal Process. Control 2023, 80, 104273.

[31] Anowar, F.; Sadaoui, S.; Selim, B. Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). Comput. Sci. Rev. 2021, 40, 100378.

[32] Naz, J.; Sharif, M.; Raza, M.; Shah, J.H.; Yasmin, M.; Kadry, S.; Vimal, S. Recognizing Gastrointestinal Malignancies on WCE and CCE Images by an Ensemble of Deep and Handcrafted Features with Entropy and PCA Based Features Optimization. Neural Process. Lett. 2021.

[33] Garg, I.; Panda, P.; Roy, K. A Low Effort Approach to Structured CNN Design Using PCA. IEEE Access 2020, 8, 1347–1360.

[34] Talukder, M.A.; Islam, M.M.; Uddin, M.A.; Akhter, A.; Hasan, K.F.; Moni, M.A. Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning. Expert Syst. Appl. 2022, 205, 117695.

[35] Mangal, S.; Chaurasia, A.; Khajanchi, A. Convolution neural networks for diagnosing colon and lung cancer histopathological images. arXiv 2020, arXiv:2009.03878.

[36] Hatuwal, B.K.; Thapa, H.C. Lung cancer detection using convolutional neural network on histopathological images. Int. J. Comput. Trends Technol 2020, 68, 21–24.

[37] Ali, M.; Ali, R. Multi-input dual-stream capsule network for improved lung and colon cancer classification. Diagnostics 2021, 11, 1485.

[38] Hasan, I.; Ali, S.; Rahman, H.; Islam, K. Automated Detection and Characterization of Colon Cancer with Deep Convolutional Neural Networks. J. Healthc. Eng. 2022, 2022, 5269913.

[39] Mane, D., Ashtagi, R., Kumbharkar, P., Kadam, S., Salunkhe, D., Upadhye, G. (2022). An improved transfer learning approach for classification of types of cancer. Traitement du Signal, Vol. 39, No. 6, pp. 2095-2101. https://doi.org/10.18280/ts.390622

[40] Gaikwad, Yogesh, and Jaishree Waghmare. Histogram and Feature Encoding Based Fake Colorized Image Detection Using Machine Learning. No. 6025. EasyChair, 2021.

[41] Thatikonda, R., Vaddadi, S.A., Arnepalli, P.R.R. et al. Securing biomedical databases based on fuzzy method through blockchain technology. Soft Comput (2023). https://doi.org/10.1007/s00500-023-08355-x