Al-Nahrain Journal of Science
Al-Nahrain J. Sci., Vol. 29, Issue (3), pp. 102 – 120, September, 2026.

Machine Learning Based on Transfer Learning for Predicting Lung and Colon Cancer

Mohammed M. Neamah[1,2,*], Laith A. Al-Ani[1], Loay E. George[3,4]

[1] Department of Physics, College of Science, Al-Nahrain University, Baghdad, 10072, Jadriya, Iraq.
[2] Physics Department, Education College, Mustansiriyah University, Baghdad, 10052, Iraq
[3] University of Information Technology & Communications, Baghdad, 11101, Iraq.
[4] College of Science, University of Baghdad, Baghdad, 10071, Iraq.

Article's Information

Received: 04.05.2025
Accepted: 07.09.2025
Published: 15.09.2026

Keywords:
Machine learning,
Deep learning,
Transfer learning,
Cancer diagnosis,
Lung and colon cancer.

Abstract

Lung and colon (LC) cancer is the leading cause of death from cancer worldwide. Early detection is key in this disease, which at the same time proves to be a challenge because, in its early stages, symptoms are few. Present a method that offers a comprehensive approach using machine learning (ML) and deep learning (DL) for the early detection of LC cancer, achieved through analysis of histopathology images (HPI). Used modified transfer learning (TL) models, which we augmented with average-pooling 2D layers for better feature extraction from the HPI dataset. Then, feed these features into ML classifiers to evaluate their performance. Achieved very impressive results across the board, with the EfficientNet B2 and the SVC classifier reporting an accuracy, recall, F1 score, and precision of 0.9996, and an AUC of 0.9998, which outperforms present methods. Also, we put forth the fact that the use of DL and ML has great promise in health care, in particular in middle- and low-income countries, which have limited access to advanced diagnostic tools. Our proposed system may support early detection, which in turn may reduce mortality rates and improve health care outcomes.

http://doi.org/10.22401/ANJS.29.3.10
*Corresponding author: mohammed.m.neamah@uomustansiriyah.edu.iq
This work is licensed under a Creative Commons Attribution 4.0 International License

1. Introduction

Cancer is a lethal disease that has become a significant issue for the healthcare system due to increasing incidence and mortality rates globally. The International Agency for Research on Cancer (IARC) states there are 53,504,187 five-year prevalent cases, and in 2022, there were 19,976,499 incidence cases and 9,743,832 deaths for both males and females worldwide. The top three leading cancers, ranked by incidence, are breast, lung, and colon cancers, while the leading causes of death are lung, breast, and colon cancer [1]. The human body contains a vast number of cells (trillions of them), and each one has the potential to divide and expand. From time to time, healthy cells proliferate and replace damaged or aged ones. Injured cells can multiply and eventually form either benign or malignant tumors unless they are substituted by healthy cells [2]. A benign tumor is characterized by slowly expanding cells that do not harm the surrounding tissues. In contrast, carcinomas develop when cells grow improperly and rapidly, causing damage to the nearby tissues. The human body can be afflicted with cancerous cells in various locations. Lung cancer can affect both females and males. By the year 2022, the global incidence of LC cancer had surpassed 4.4 million cases (22%), resulting in over 2.72 million recorded deaths (28%) [3]. There are cases where one in seven people will develop both types of LC cancer simultaneously. The research found that smoking increases the risk of lung cancer, which in turn leads to dietary abnormalities that further raise the risk of colon cancer. [4]. When the patient exhibits symptoms, it is often too late, as the tumour has progressed to an advanced stage, and there are few or no indicators of the illness in its early phases. Early LC cancer detection is challenging without a thorough assessment using advanced imaging techniques such as MRI, CT scans, ultrasound, and tissue analysis [5]. Middle- and low-

income states account for 70% of cancer mortalities, according to the World Health Organization (WHO) [3]. Therefore, these countries must construct hospitals fitted with comprehensive diagnostic laboratories at no cost and offer training to qualified healthcare professionals to carry out appropriate diagnostic procedures. Furthermore, the cancer diagnosis process demands a considerable amount of time. It is influenced by various perspectives from medical practitioners, particularly in the early stages of the disease [6]. Several sectors within the healthcare industry may overcome these difficulties. This disparity underscores the urgent need for automated, cost-effective, and scalable early detection methods [7]. ML and DL methods are applied in healthcare for tasks such as early detection of biomedical imaging, health emergencies, disease diagnosis, health crises, and other related applications [8]. ML algorithms are used at various phases of image analysis to minimize human involvement and improve the quality of the results. ML approaches allow for the identification of subtle patterns and correlations among the pixels [9]. DL algorithms can analyze data derived from high-dimensional images [10]. Furthermore, these techniques extract hidden elements and features from medical images that can be used for early cancer diagnosis [11]. The main contribution of this manuscript is:

- Feature extraction: The proposed methods employed modified TL models with added average pooling 2D layers to extract deep features from the HPI dataset. We implemented a system-wide approach to extract high-level features from the modified TL models, enabling memory-efficient processing of large datasets without loading everything into memory simultaneously. Subsequently, the extracted strong features were trained with an ML classifier to evaluate the model's performance.
- High Accuracy Achievement: Our integrated approach scored a remarkable 0.9996 accuracy.

This study uses many TL algorithms (DL models) as feature extractors to extract features from the input HPI of LC cancer (LC25000). The deep features are trained into several ML classifiers for their performance evaluation in predicting the LC cancer. The present work follows this introduction: Section 2 examines the prior work for diagnosing LC cancer. Section 3 introduces our suggested methods, different models, and methodologies for evaluating the efficacy of models by utilizing the HPI LC25000 dataset. Section 4 analyzes the execution performance of all models and provides a succinct overview of their outcomes. Lastly, the article concludes in Section 5 with a conclusion.

2. Previous work

The increasing efficacy of machine learning and DL has rendered them usable in several domains, particularly in healthcare, such as cancer diagnosis. This section provides a comprehensive review of previous relevant studies on LC cancer detection. Numerous research works have reviewed prior studies on AI-based LC cancer detection. This work establishes a Convolutional Neural Network (CNN) architecture for classifying and discriminating between five kinds of LC tissues (benign and malignant) by analyzing their HPI. The findings show that the suggested work can predict cancerous tissues with an accuracy of 0.9633 [12]. This study presents a hybrid ensemble feature extraction approach for effectively identifying LC cancer. For cancer image sets, we present a method that uses High Performance Filtering (HPF) in combination with Ensemble Learning (EL) and deep feature extraction. Evaluated the approach using HPI data for the LC (LC25000). Report that the hybrid model we present achieves accuracy rates of 0.9905, 1.00, and 0.993 for lung, colon, and LC cancer detection, respectively [13]. The study reports on research in which we developed a computer-aided diagnostic method for reliably classifying five types of LC tissues via analysis of their HPI dataset. Used ML, image processing, and feature engineering to that end, and in turn put to use six models: Light Gradient Boosting (LGB), Random Forest (RF), Linear Discriminant Analysis (LDA), Extreme Gradient Boosting (XGB), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP) for the classification of LC tumor images. Our experiments showed that the ML models performed very well, with high precision in classifying LC cancer subtypes. XGB model did the best with an accuracy of 0.99 and an F1 score of 0.988 [14]. This study looked at improving AlexNet by tuning four layers, and then trained the model on the dataset. It achieved a total accuracy of 0.89 with the first set of results. To improve total accuracy while maintaining computational performance, improved image quality in the underperforming class was achieved using a good and efficient contrast enhancement technique; whole-dataset-wide image enhancement proved not to be what was being looked for. The technique presented in this study increased accuracy to 0.984 [15]. SVM, RF, and logistic regression (LR) are used, which aggregate via majority voting to form the ensemble classifier. Used deep features (DF) from the LC

dataset, which were obtained via the use of two different methods: the VGG16 architecture and local binary pattern (LBP). Verified the proposed technique with histopathology (LC25000) datasets for LC cancer. Reported that the proposed method does very well, with an average accuracy of 0.99, a recall of 0.9880, a precision of 0.99, and an F1 score of 0.988 [16]. This paper reports on the MPADL-LC3 approach, which uses DL to classify LC cancers via an algorithm inspired by a marine predator. The MPADL-LC3 approach does an excellent job of differentiating between various types of LC and performs well in HPI. Report first that the MPADL-LC3 approach uses CLAHE-based contrast enhancement as a first step. It uses MobileNet for feature vector creation and also does use of MPA for hyperparameters. Also report the use of deep belief networks (DBN) for LC classification. Analyzed the simulation results of the MPADL-LC3 approach via benchmark datasets, which reported a max accuracy of 0.9927 [17]. This work reports on the use of the EfficientNetB7 TL, which we applied to build a classification model for HPI diagnosis, which in turn differentiates between five types of LC tissues, including three malignant and two benign varieties. Furthermore, precision, F1 score, and recall are utilised to evaluate the model's performance. The EfficientNetB7 architecture achieved an impressive accuracy of 0.98 [18]. The research presented a hybrid technique for the detection of LC cancer. It integrates pre-processing, feature extraction using TL techniques, ML computations, optimal method selection via hyperparameter tuning, and the incorporation of EL approaches. Select the most effective ensemble voting classification technique and the most appropriate feature extraction TL algorithms. MobileNet, DenseNet201, DenseNet169, VGG16, and VGG19 were used in the research. Looked at the performance of six basic ML models, which are MLP, RF, LGB, SVM, XGB, and LR, and we found that LR did the best with an average of 0.9299 [19]. In terms of what the studies looked at, they reported great progress in AI for LC cancer detection, which did very well with accuracy rates of over 0.96 using methods like CNNs and ensemble learning. At the same time, they ignored the issue of computational efficiency for clinical use and also reported using a variety of pre-processing techniques. What we did in our study was to present a more simplified approach which puts forth modified, more efficient TL models (MobileNet, DenseNet201, DenseNet169, VGG16, and VGG19) also we included what we think to be improved feature extraction and classification. This approach utilizes thorough evaluation metrics (ROC-AUC, CM) and standardized preprocessing to boost generalizability and applicability in clinical settings, ultimately aiming to reconcile high performance with real-world application in resource-limited healthcare environments. Table 1 illustrates a summary of related work.

Table 1. Related Work Summary.

Authors	Year	Approaches	Accuracy
Masud, M. et al. [12]	2021	CNN	0.9633
Talukder, M. A. et al. [13]	2022	CNN+EL(ML)	0.9905 1.00 0.9930
Hage Chehade, A. et al. [14]	2022	ML	0.99
Mehmood, S. et al. [15]	2022	CNN+CSIP	0.984
Singh, O. & Singh, K.K. [16]	2023	CNN+EL(ML)	0.99
Mengash HA et al. [17]	2023	CNN	0.9927
Prusty, S., et al. [18]	2023	CNN	0.98
Reddy, S. N., et al. [19]	2024	CNN+ML	0.9299

3. Methodology

This section discusses image preprocessing, deep feature extraction, ML, and the proposed model. It provides a fundamental overview of the ML approaches used to identify LC cancer. It builds our suggested approaches using TL models for feature extraction and ML methods on the LC cancer HPI datasets to predict LC cancer. The block diagram of the proposed models is illustrated in Figure 1. The phases of the proposed model are outlined as follows:

- Phase 1: To perform our experiment, we acquired the HPI datasets (LC25000) for LC cancer. The datasets include three types of lung cancer and two types of colon cancer.
- Phase 2: During the pre-processing phase, the image is resized into 224X224, converted from BGR

into RGB (bgr2rgb), and then transformed into NumPy arrays. In the last steps, implement labeling by appointing a label (0, 1, 2, 3, 4) to each cancer class.

- Phase 3: As part of the deep feature extraction phase, extract deep features from the HPI datasets using four distinct modified TL methods: EfficientNet (EfficientNet B2), VGG19, MobileNet (MobileNetLargeV3), and DenseNet201. Disable the last three layers, add a global average pooling layer, and implement a system-wide approach to extract robust features and collect these features for the next phase.
- Phase 4: To evaluate the performance, successfully extracted features are input into four ML models: RF, SVM, KNN, and LGB.
- Phase 5: Also see to it that accuracy, precision, F1 score, recall, confusion matrix (CM), receiver operating characteristic curve (ROC), and area under the curve (AUC) are used as tools to report on the performance of the proposed model.

Our study looked at tuning hyperparameters through a process of iteration, which included many training sessions with different key parameters like learning rate, batch size, and other ML parameters. We paid close attention to how these changes played out in terms of validation performance, and we kept whichever settings consistently improved performance. Although the tune was done by hand instead of via an algorithm, the selected hyperparameters we reported on had been tested many times, which improved the study's credibility and robustness. By using hold-out validation, which is a technique that puts the data into separate groups usually a training set, an optional validation set, and a hold-out (test) set it is used to do an unbiased assessment of the model's performance on data it had not seen before. During training, the model learns from the training set and may adjust hyperparameters using the validation set if available. The hold-out set is kept strictly separate and only used after training to accurately measure the model's generalization and detect overfitting.

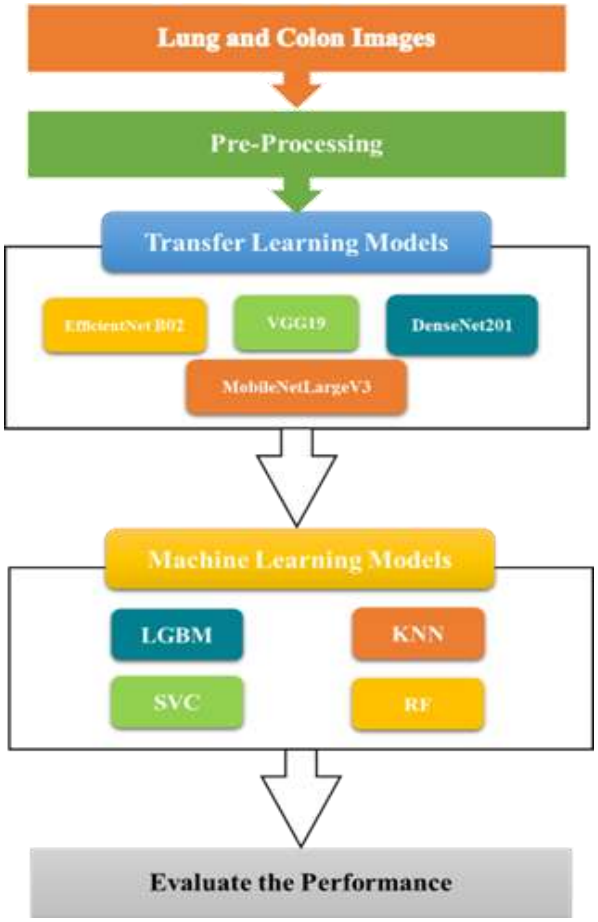


Figure 1. The proposed model block diagram.

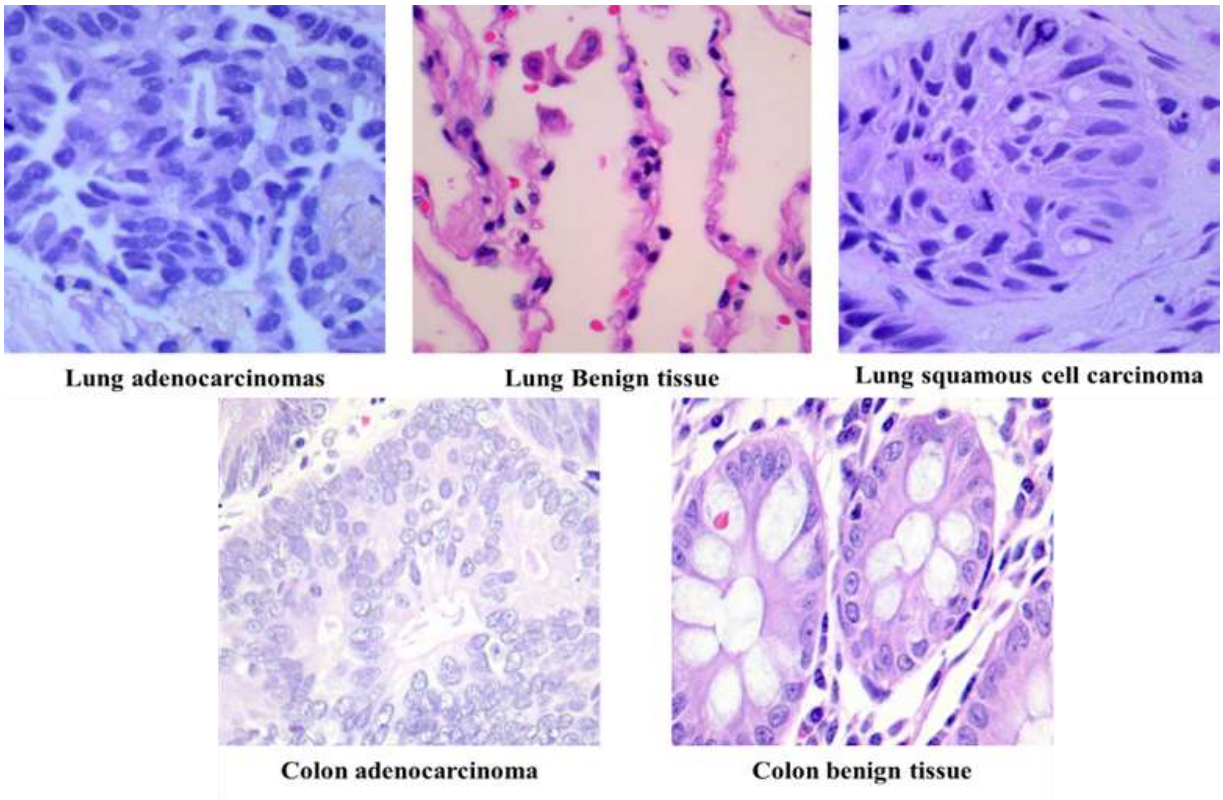


Figure 2. Illustrates an example of an image from the LC cancer dataset.

3.1. Datasets

Borkowski et al. used the new dataset LC25000, released in 2020 [17]. The dataset holds 25,000 colon and lung tissue images categorized into five classes. Lung images are further classified into adenocarcinoma (lungA), benign tissues (lungB), and squamous cell carcinoma (lungS), while colon tissue cancer images are divided into benign (colonB) tissues and adenocarcinoma (colonA). Originally, the dataset combined 1250 colon and lung tissue pathology slide images. By employing augmentation techniques such as flipping and rotating the images under numerous conditions, the dataset has been extended to 25,000 images spread across five classes, with 5000 images for each class. Before augmentation, the images were resized to 768x768 [20]. Table 2 shows the assigned class names for each image class along with the corresponding image numbers and labels. At the same time, Figure 2 presents an example of images from the dataset as well as the distribution of the dataset per class, respectively.

Table 2. The dataset distribution and labelling.

Lung and colon cancer	Class	Lable	No. image
Colon adenocarcinoma	colonA	0	5000
Colon benign tissue	colonB	1	5000
Lung adenocarcinomas	lungA	2	5000
Lung Benign tissue	lungB	3	5000
Lung squamous cell carcinoma	lungS	4	5000

3.2. Preprocessing

In this manuscript, resized the input image to 224x224. Next, to work with the scaled image intensity data, transformed it into bgr2rgb and a NumPy array. At last, performed the image labeling, which is crucial to classify the cancerous images. Figure 3 illustrates the preprocessing steps of our suggested methods.



Figure 3. Pre-Processing steps for the suggested models.

3.3. Transfer Learning (TL)

TL uses characteristics acquired from one issue to address a new, analogous problem [21]. This DL process has attracted considerable attention, prompting the adaptation and modification of various existing models to meet our needs. Rather than training all layers of the algorithms, we focus on a chosen subset while utilising the learnt weights from those frozen layers to extract specific features from the datasets employed [22]. The relevant features of an image significantly influence the efficacy of an identification process. In this study, we analyse HPI datasets using five distinct TL algorithms, including VGG19, EfficientNet (EfficientNet B2), MobileNet (MobileNetLargeV3), and DenseNet201, to extract features and evaluate the performance of each model through four different ML models, namely RF, KNN, LGBM, and SVM.

- EfficientNet is a CNN that, in 2019, was put forth by a group at Google AI; it did, to a large extent, transform the field of DL by putting forward a family of models that achieve state-of-the-art accuracy while at the same time also doing so in a very efficient way. The main innovation of EfficientNet was its approach to model growth. Up until then, typical practices were to independently play with the depth, width, or resolution of the network. What EfficientNet did instead was present a combined scaling approach that modifies all these elements via a single compound coefficient. This is a very broad and at the same time also a very effective approach, which also brings about uniform growth of the network, which in turn improves performance and efficiency [24].
- MobileNet a pre-trained convolutional network which was put forth by Howard et al. in 2017 [25]. What sets MobileNet apart from other models is the reduced number of parameters. It is designed to reduce computational effort in the initial layers [26]. In this, they use pointwise and depthwise convolutions. By input filtering, which does not produce additional features, we mean depthwise convolution. To that end, they include pointwise convolution, which does introduce new features. Thus, the term depthwise separable convolution is used to describe the combination of the two. They use depthwise convolutions, which apply a separate filter for each input and then add a 1x1 convolution (pointwise) to produce the combined output from the depthwise layer. Also, Batch Normalization is used along with the Rectified Linear Unit (ReLU) at each stage of convolution. We see a trade-off in performance between speed and accuracy [27].
- The VGG, introduced in 2014 by Simonyan and Zisserman, saw the introduction of what has become a fundamental architecture in CNNs for computer vision. That model, which has greatly transformed the space of image classification, object detection, and segmentation, has proven itself to be very impactful [28]. The VGG algorithm is a study in balance between the simple and the complex. It presents a network that, although large in structure, is also very simple in design. The architecture is made up of many layers which use convolutions and pooling, some of which are fully connected. In the convolutional layers, see the use of a small 3x3 filter, which in fact increases the receptive field and improves feature extraction [28]. Also have max-pooling layers, which reduce the spatial dimension of feature maps, which in turn allows the network to collect more broad context info. And finally, fully connected layers, which in their role are that of classification, which they do by putting out a probability score for each category [29].
- DenseNet201 is a model that is put forth and very much in use for a wide range of computer vision tasks [30]. What DenseNet brings to the table is its dense connection architecture, which sees each layer connected to all other layers in a feed-forward way. This contrasts with traditional CNNs, which have layers that only connect in sequence. Enhancing the complexity of the connection structure utilises shared features and ensures that information flows smoothly between layers [31]. The DenseNet201 algorithm consists of numerous

dense blocks, each containing convolutional layers. All prior layers in the block provide feature maps, and their output is concatenated with the input feature maps of each layer. This process is applied iteratively to each layer within the block, resulting in a condensed representation of the input data [30].

These four modified TL models will extract deep features from the LC cancer dataset by freezing the last three layers (FCC) and incorporating a global average pooling layer that samples the feature mapping to reduce the computational load and number of parameters in the subsequent layers. They will also adopt a system-wide approach to extract high-level features. The extracted deep features will then be input into various ML classifiers to evaluate their performance.

3.4. Machine Learning Classifier

Four variant ML classifiers were employed to estimate the models' performance: RF, KNN, LGBM, and SVM.

- RF is an ensemble of supervised learning techniques for regression and classification tasks within predictive modelling and ML methodologies [32]. The method aggregates the outcomes and predictions from multiple decision trees (DT) to select the optimal output, defined as the mode (the most frequently occurring value) of the class results or the mean prediction. It is a basic element of ML, offering robustness, interpretability, and versatility.
- KNN: Here is a paraphrased version of the provided text: The KNN algorithm is a very basic yet very efficient instance-based supervised learning method for regression and classification tasks. Its wide use in many fields, such as medical diagnosis, recommendation systems, and image recognition, is due to its simplicity and performance. It is based on the idea that similar data points tend to fall into the same class or have the same target value [33]. What is different about KNN, compared to parametric models that require in-depth training, is that it is what is known as a lazy learner, which delays computation until the time of prediction, thus making it very flexible for use with dynamic data sets. KNN will determine the distance (for example, Euclidean, Manhattan, or Minkowski) between a new query input and all training samples. Which distance metric is used plays a large role in performance, especially in high-dimensional spaces [34]. The algorithm reports the k closest neighbors, which is a parameter defined by the user. With lower values of k, more noise and also a greater response to local variability, while with higher k, less well-defined boundaries, although at the same time the local picture can be lost [35]. In the case of classification, the query input is assigned the class most represented among the k neighbors.
- LGBM is a very fast and very efficient platform for Gradient Boosting, which in terms outperforms other options in the machine learning space. At this point, are seeing wide adoption of it from practitioners who note the speed and scalability. Also, what excites everyone is its speed and ability to manage big data. This method excels when working with vast amounts of data and has been utilised across numerous fields, from healthcare to finance, including recommendations for films and television shows. As researchers publish their findings, they demonstrate that it performs exceptionally well overall. LGBM was introduced by Microsoft Research in 2017 [36]. This approach is grounded in the gradient boosting framework and incorporates various innovations to enhance efficiency and performance. LGBM utilises Gradient One Side Sampling (GOSS) to reduce the number of data instances, and employs Exclusive Feature Bundling (EFB) also reduces the number of executed features. This significantly speeds up training time a lot [37].
- SVM algorithms are a supervised learning method for regression, outlier detection, and classification. In a high or infinite-dimensional space, an SVM creates a hyperplane or multiple hyperplanes, which can be used for various tasks [38]. It is possible to achieve separation by using the hyperplane that maximizes the distance to the closest training data points from any class, referred to as the functional margin, which is effective for achieving separation. A larger margin correlates with a reduced generalization error of the classifier [39]. It is an effective technique; however, its computational and storage demands increase significantly as the number of training vectors rises. An SVM fundamentally involves a quadratic programming problem (QP) that distinguishes support vectors from the remaining training data [40].

3.5 Performance Metrics

The Variant performance metrics evaluate our suggested model, including the confusion matrix, accuracy, F1-score, recall, precision, and AUC-ROC curve. The metrics for estimating the models are expressed below:

1. The Confusion Matrix is a method to measure ML classification performance. When the model is

developed for classification, four outcomes are predicted these include:

- True Positive (TP): when the classifier predicted sample correctly belongs to the positive class.
- True Negative (TN): When the classifier correctly predicts that the sample belongs to the negative class.
- False Positive (FP): When the classifier wrongly predicts a sample that belongs to the negative class as the positive class.
- False Negative (FN): When the classifier wrongly predicts a sample that belongs to the positive class as the negative class [41].

These four measures are plotted in the confusion matrix. Figures 4 show the confusion matrix.

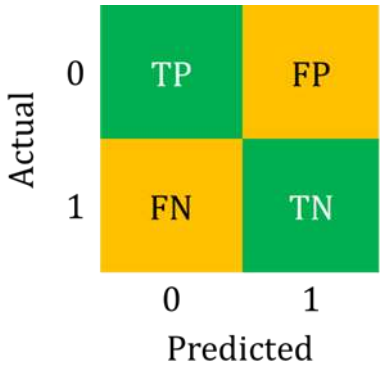


Figure 4. Example of a confusion matrix for the binary classification.

2. Accuracy: The ratio of the true predicted to the total outcomes; it is the most apparent performance metric.

$$Accurac = \frac{TP + TN}{TP + FP + FN + TN} \quad \dots (1)$$

3. Precision: The ratio of the true positive outcomes to the total predicted positive outcomes.

$$Precision = \frac{TP}{FP + TP} \quad \dots (2)$$

4. Recall (Sensitivity): The ratio of the true positive outcomes to the sum of True positive and False-negative outcomes (all outcomes in the true class) [42].

$$Recall = \frac{TP}{TP + FN} \quad \dots (3)$$

5. F1-Score: The weighted average of recall and precision.

$$F1 - Score = \frac{(Recall \times Precision) \times 2}{(Precision + Recall)} \quad \dots (4)$$

6. AUC-ROC curve: ROC curves are two-dimensional graphs often used to assess the efficacy of models [43]. The ROC plot illustrates a model's specificity and sensitivity across all possible prediction thresholds, providing the evaluation and selection of models based on specific operator criteria, often attributed to varying accuracy expectations and error costs [44]. AUC is a measure of discrimination, which ROC is a graph of. It also represents the model's performance in separating out different groups. The y-axis represents the TP rate, whilst the x-axis denotes the FP rate. An AUC nearing one indicates that a predictive model excels in discriminating across class labels, whereas an AUC approaching zero suggests a poor predictive model. In reality, a poor indication of the outcome is being reflected [45]. It is an approach for visualizing the effectiveness of prediction. Models with higher ROC-AUC curves are preferable [46].

$$AUC = \frac{TP\ Rate}{FP\ Rate} \quad \dots (5)$$

4. Results and Discussion

Our in-depth analysis reports very good results we have achieved in predicting LC cancer as compared to what was presented before. Compare results from a number of TL and ML algorithms. Have a set of 25,000 images, which have been broken up into five separate groups. To make the distribution even, we divided the data into 80% for training (20,000 images) and the remaining 20% for validation and testing, split into 2,500 images each. Precision, F1 score, recall, AUC-ROC curve, and accuracy are used as performance metrics. Fig 5 shows the breakdown of the LC cancer dataset:

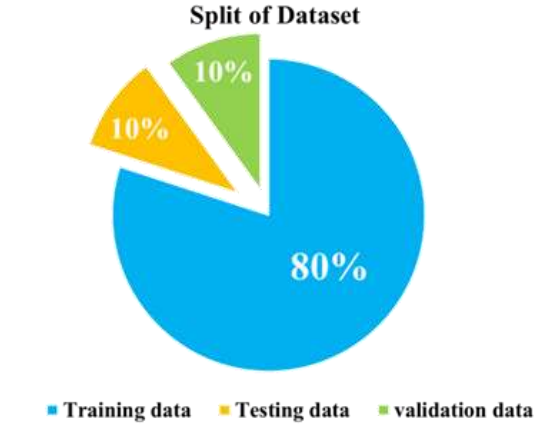


Figure 5. Dataset splitting.

This section evaluates the efficacy of several TL approaches and then examines the results derived

from the chosen high-performance TL algorithms and ML classifiers. Our performance results have been reported. Found that the modified EfficientNet B2 model outperforms all past TL models and that the SVC classifier does better than the other classifiers. Tables 3, 4, 5, 6, and 7 illustrate the ML performance metrics.

Table 3. Model performance in terms of accuracy.

Models	EfficientNet B2	VGG19	DenseNet201	MobileNetLargeV3
KNN	0.9992	0.9996	0.9848	0.9932
RF	0.9888	0.972	0.982	0.9892
LGBM	0.9972	0.9872	0.9932	0.9976
SVC	0.9996	0.9792	0.9816	0.9968

Table 4. Model performance in terms of recall.

Models	EfficientNet B2	VGG19	DenseNet201	MobileNetLargeV3
KNN	0.9992	0.9932	0.9851	0.9996
RF	0.9888	0.972	0.982	0.9893
LGBM	0.9972	0.9872	0.9932	0.9976
SVC	0.9996	0.9792	0.9817	0.9968

Table 5. Model performance in terms of precision.

Models	EfficientNet B2	VGG19	DenseNet201	MobileNetLargeV3
KNN	0.9992	0.9932	0.9851	0.9996
RF	0.9888	0.972	0.982	0.9893
LGBM	0.9972	0.9872	0.9932	0.9976
SVC	0.9996	0.9792	0.9817	0.9968

Table 6. Model performance in terms of F1 Score.

Models	EfficientNet B2	VGG19	DenseNet201	MobileNetLargeV3
KNN	0.9992	0.9932	0.9848	0.9996
RF	0.9888	0.9719	0.982	0.9892
LGBM	0.9971	0.9872	0.9932	0.9976
SVC	0.9996	0.9791	0.9816	0.9968

Table 7. Model performance in terms of AUC.

Models	EfficientNet B2	VGG19	DenseNet201	MobileNetLargeV3
KNN	0.9995	0.9957	0.9905	0.9997
RF	0.993	0.9825	0.9887	0.9932
LGBM	0.9983	0.992	0.9957	0.9985
SVC	0.9998	0.987	0.9885	0.998

Tables 3 – 7 present the results of a wide range of TL algorithm modifications in table form, from which it can be seen that in all models the highest performance is achieved. Among these, the modified EfficientNet B2 model and the SVC classifier deliver the best performance, with an accuracy rate, F1 score, precision, and recall of 0.9996, and a recorded AUC of 0.9998. Additionally, Figures 6 – 10 illustrate that the bar chart highlights the modified EfficientNet B2 and the SVC classifier as having the highest performance metrics compared to other models, indicating that the accuracy and other metrics are superior.

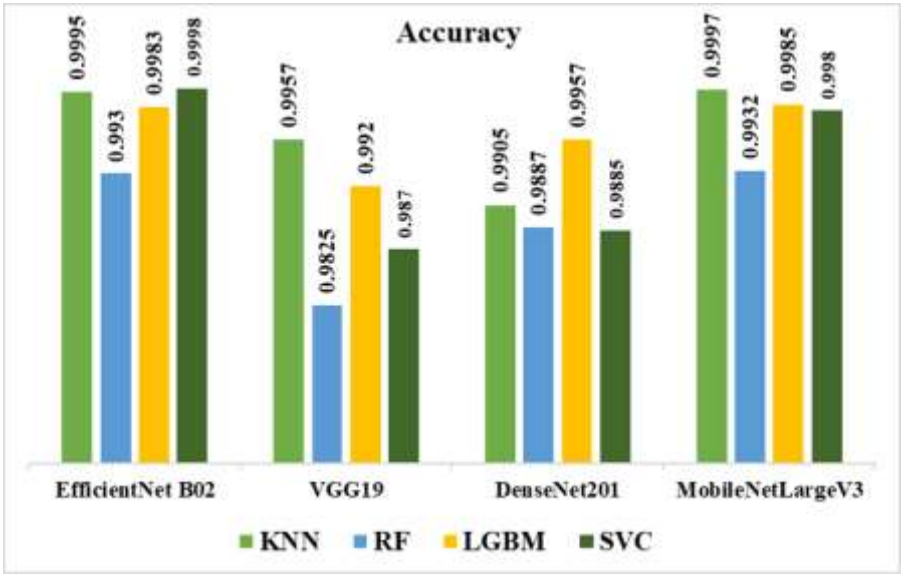


Figure 6. Accuracy analysis for ML models.

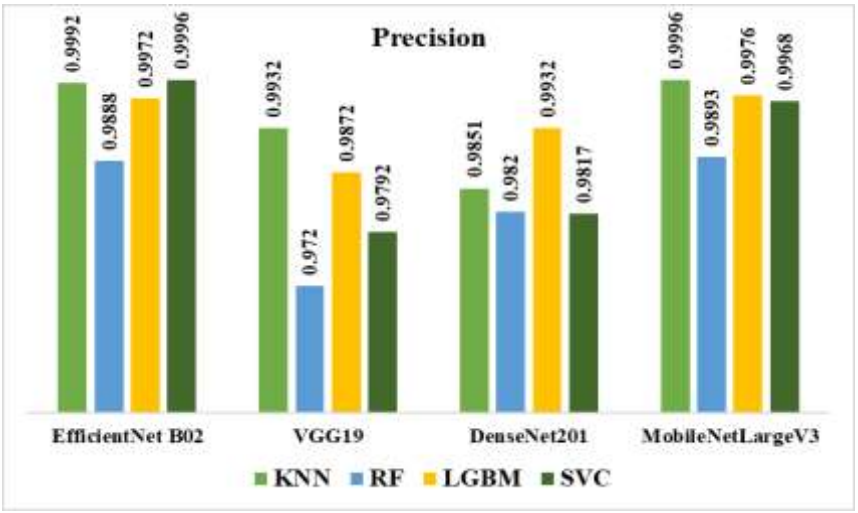


Figure 7. Precision analysis for ML models.

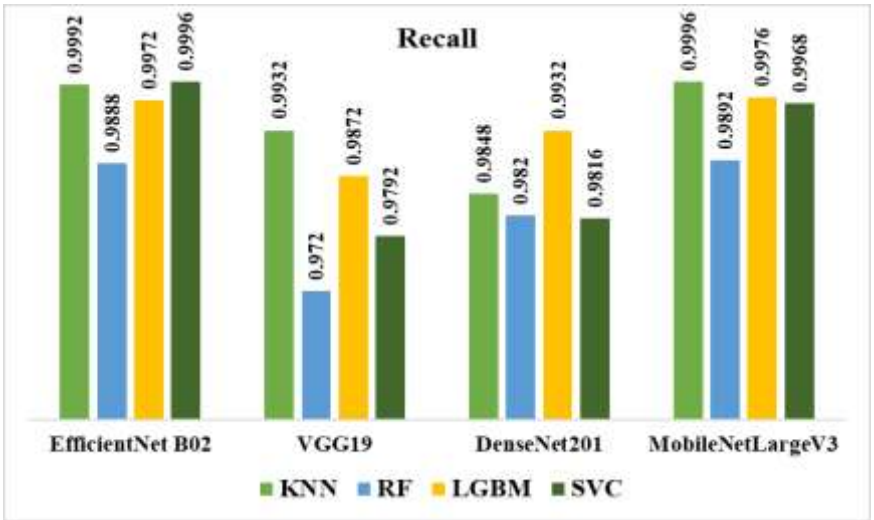


Figure 8. Recall analysis for ML models.

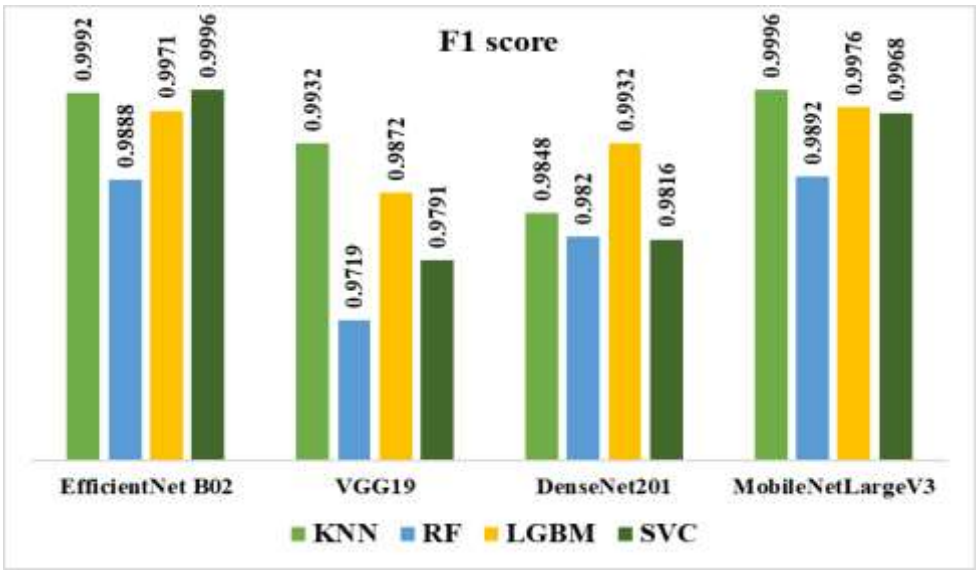


Figure 9. F1 score analysis for ML models.

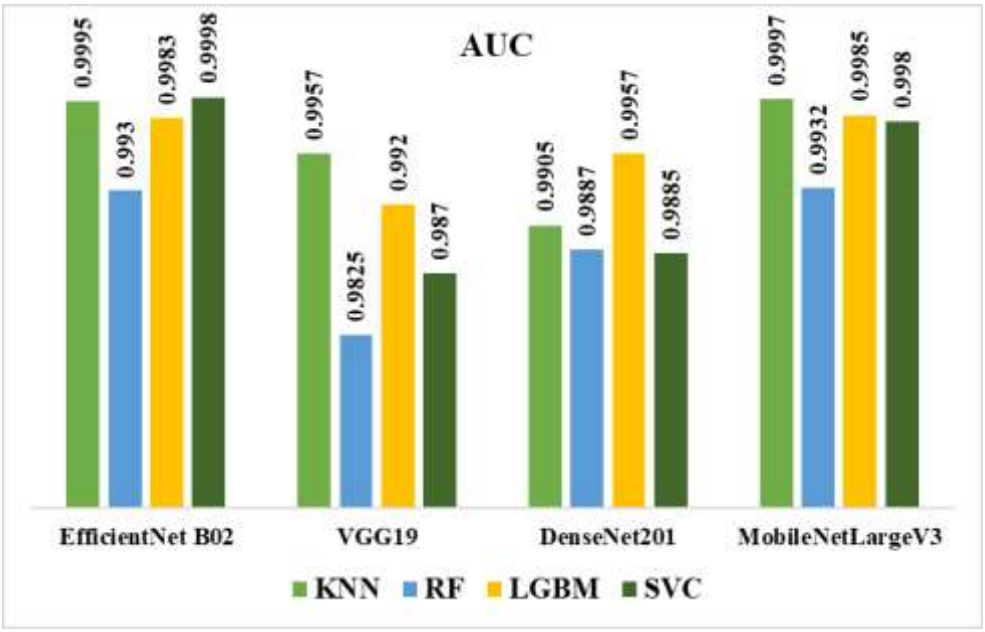
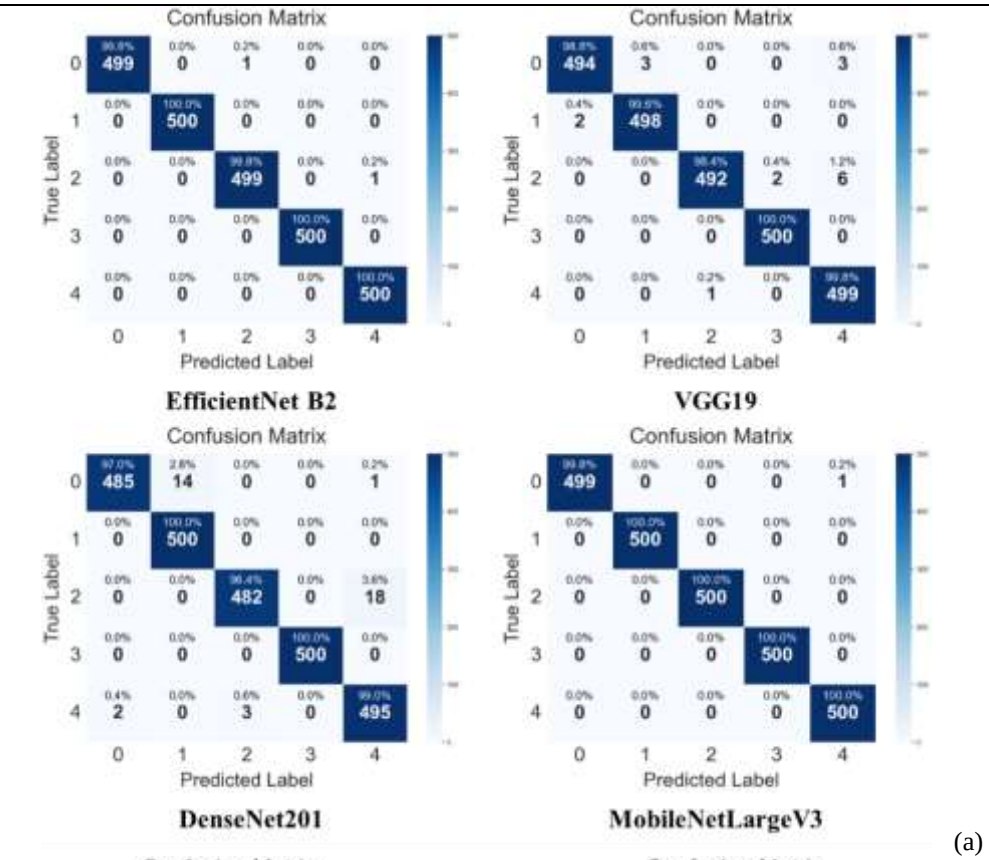
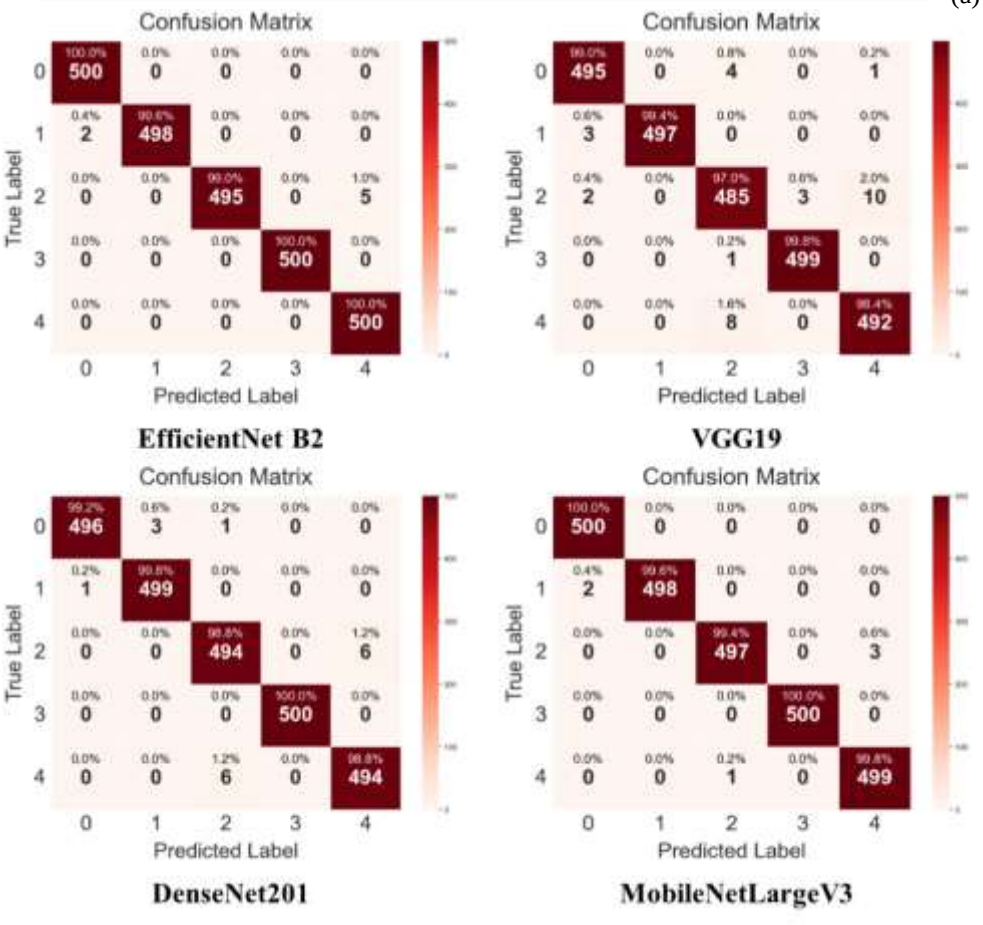


Figure 10. AUC analysis for ML models.

Figures 11 (a) – (d) illustrate the CM of the suggested algorithms. They demonstrate that the TP and TN rates for all modified TL models, particularly for the modified Efficient Net B2 model and the SVC classifier, are exceptionally high. In contrast, the FP and FN rates are very low across all classes. A greater number of TN and TP compared to FN and FP is essential for ML classifiers to predict LC cancer and indicate the highest-performing model.


Confusion Matrix
True Label
Predicted Label
EfficientNet B2
Confusion Matrix
VGG19
Confusion Matrix
DenseNet201
Confusion Matrix
MobileNetLargeV3

(a)


Confusion Matrix
True Label
Predicted Label
EfficientNet B2
Confusion Matrix
VGG19
Confusion Matrix
DenseNet201
Confusion Matrix
MobileNetLargeV3

(b)

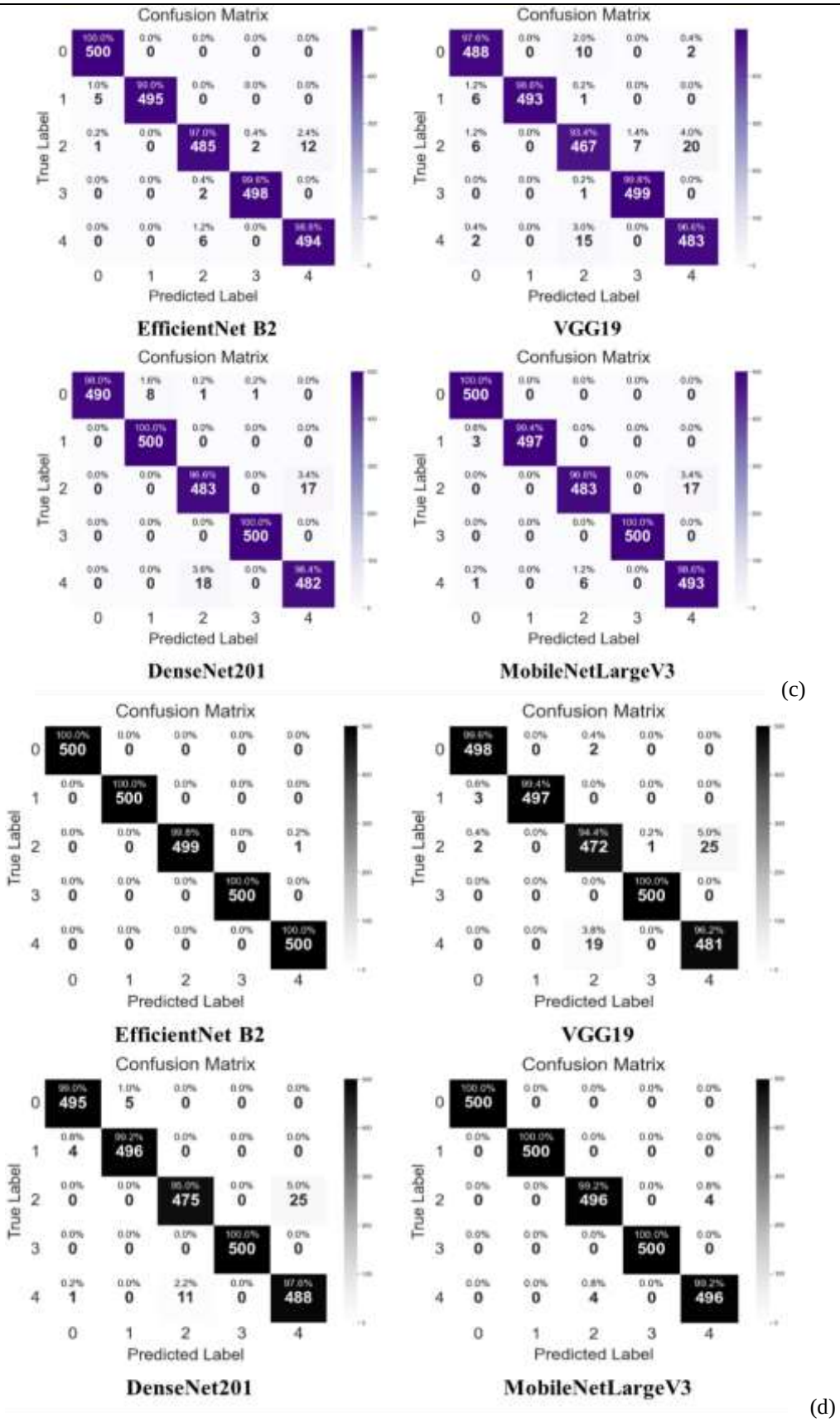
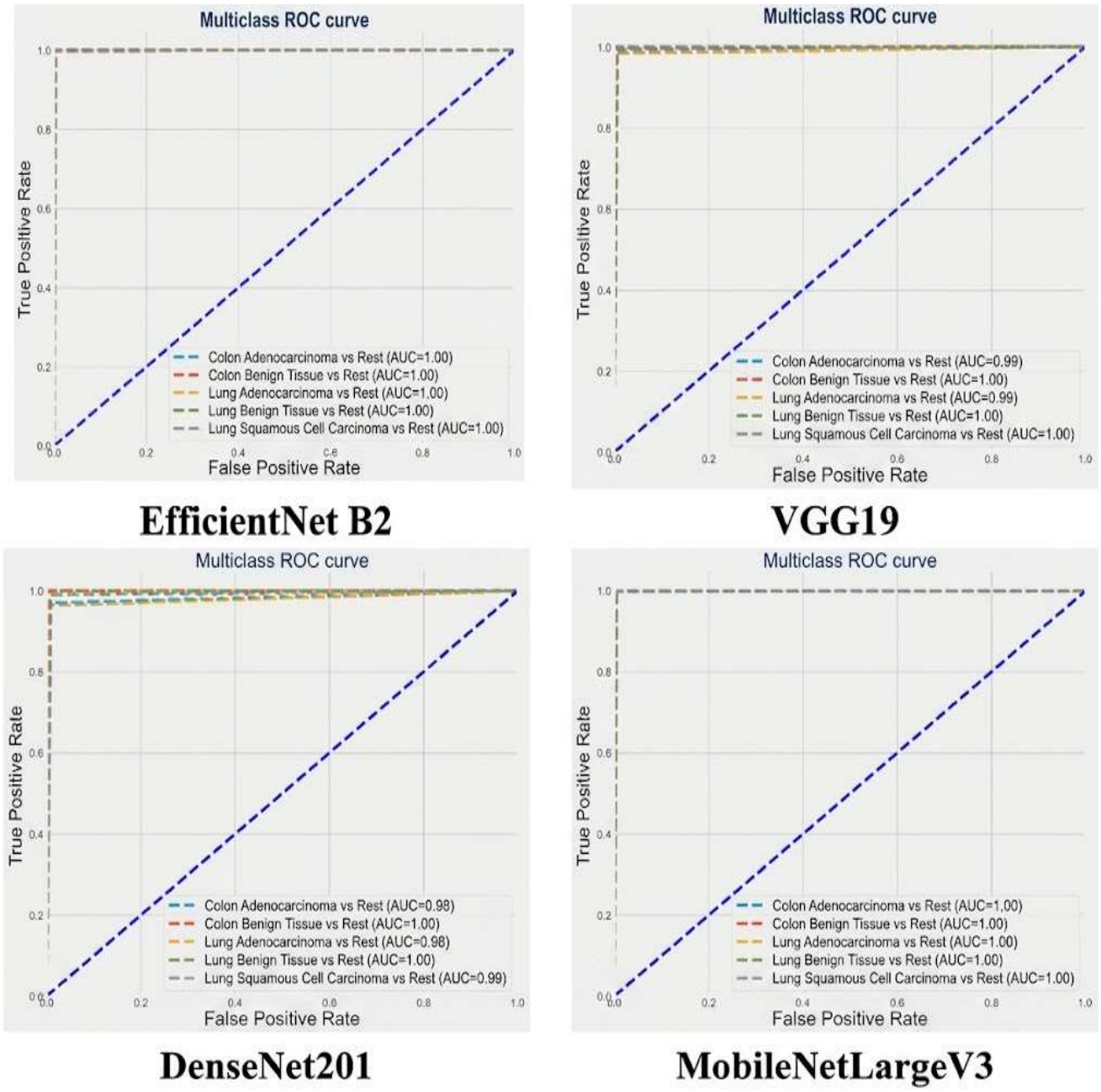
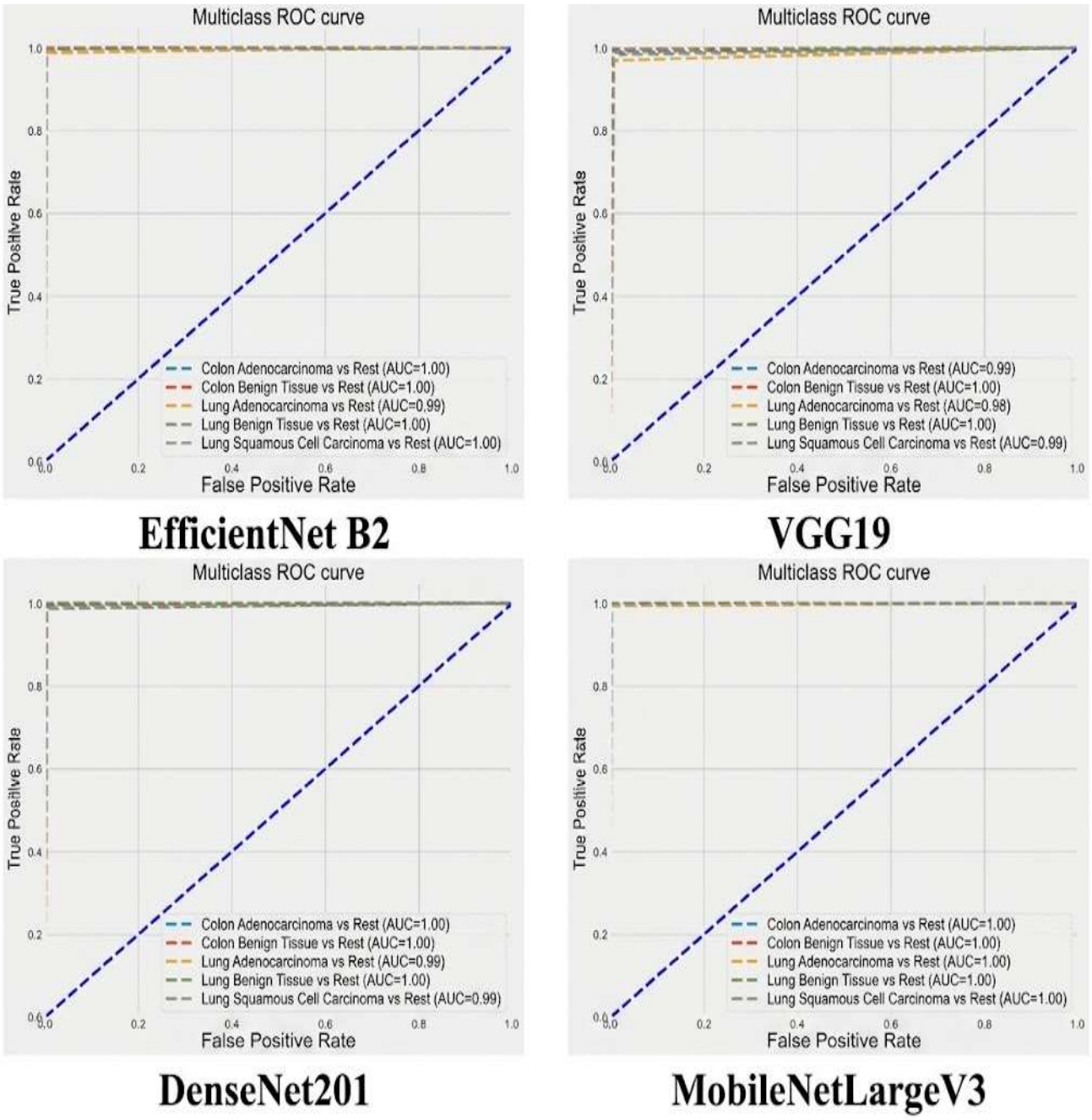


Figure 11. The confusion matrix: a) KNN, b) LGBM, c) RF, and d) SVC.

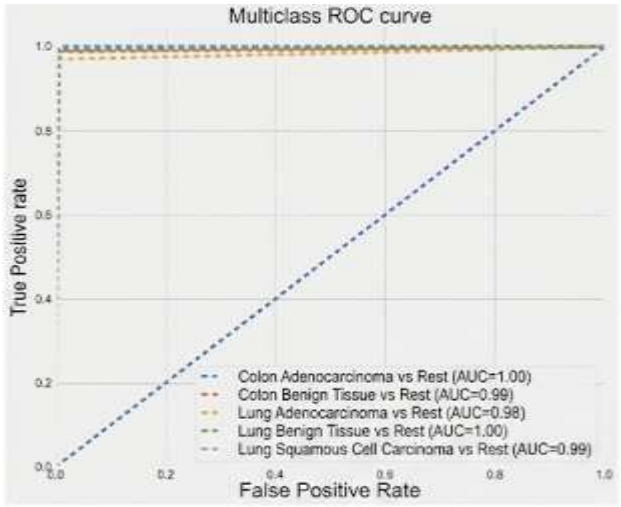
Figures 12 (a) – (d) illustrate the ROC curve for the modified TL models. They demonstrate that the AUC for all modified TL models, particularly the modified EfficientNet B2 model and the SVC classifier, is notably high. The curve closely aligns with the chart and approaches 1, indicating the model's optimal performance. The ROC curves for the proposed models exhibit remarkably high discriminative performance, as indicated by AUC values nearing 1.0. This almost perfect AUC signifies that each model excels at differentiating between positive and negative classes, reliably identifying true positives with a notably low false positive rate. These outcomes imply that the models have established strong decision boundaries and are performing nearly flawlessly on the validation set.



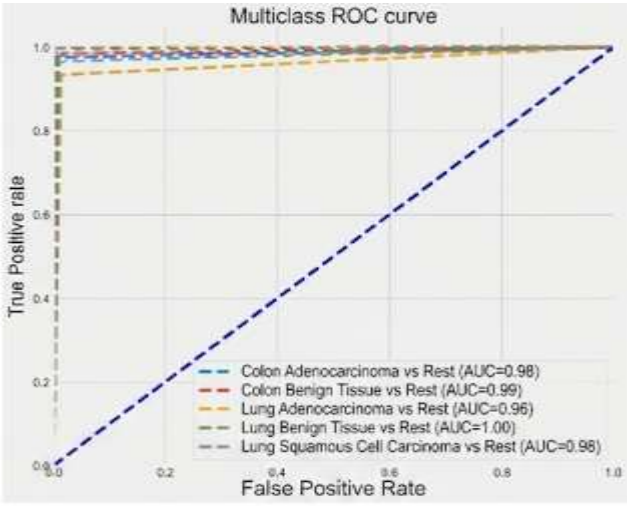
(a)


Multiclass ROC curve
True Positive Rate
False Positive Rate
Colon Adenocarcinoma vs Rest (AUC=1.00)
Colon Benign Tissue vs Rest (AUC=1.00)
Lung Adenocarcinoma vs Rest (AUC=0.99)
Lung Benign Tissue vs Rest (AUC=1.00)
Lung Squamous Cell Carcinoma vs Rest (AUC=1.00)
EfficientNet B2
Multiclass ROC curve
True Positive Rate
False Positive Rate
Colon Adenocarcinoma vs Rest (AUC=0.99)
Colon Benign Tissue vs Rest (AUC=1.00)
Lung Adenocarcinoma vs Rest (AUC=0.98)
Lung Benign Tissue vs Rest (AUC=1.00)
Lung Squamous Cell Carcinoma vs Rest (AUC=0.99)
VGG19
Multiclass ROC curve
True Positive Rate
False Positive Rate
Colon Adenocarcinoma vs Rest (AUC=1.00)
Colon Benign Tissue vs Rest (AUC=1.00)
Lung Adenocarcinoma vs Rest (AUC=0.99)
Lung Benign Tissue vs Rest (AUC=1.00)
Lung Squamous Cell Carcinoma vs Rest (AUC=0.99)
DenseNet201
Multiclass ROC curve
True Positive Rate
False Positive Rate
Colon Adenocarcinoma vs Rest (AUC=1.00)
Colon Benign Tissue vs Rest (AUC=1.00)
Lung Adenocarcinoma vs Rest (AUC=1.00)
Lung Benign Tissue vs Rest (AUC=1.00)
Lung Squamous Cell Carcinoma vs Rest (AUC=1.00)
MobileNetLargeV3

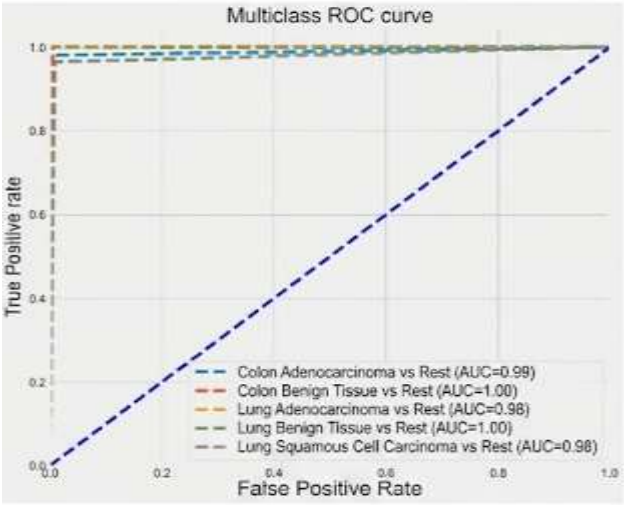
(b)



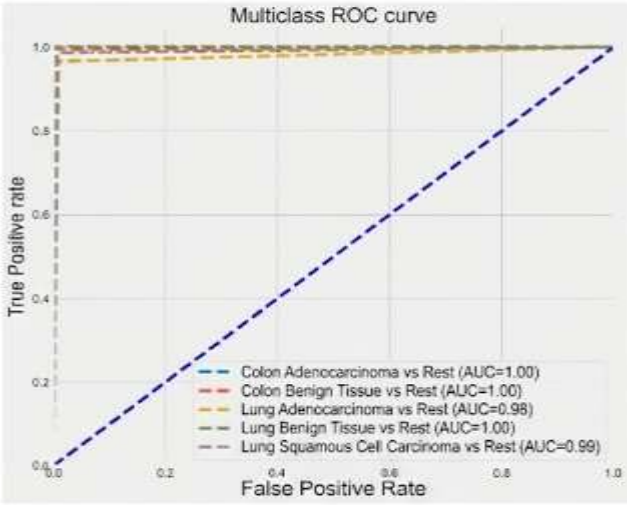
EfficientNet B2



VGG19

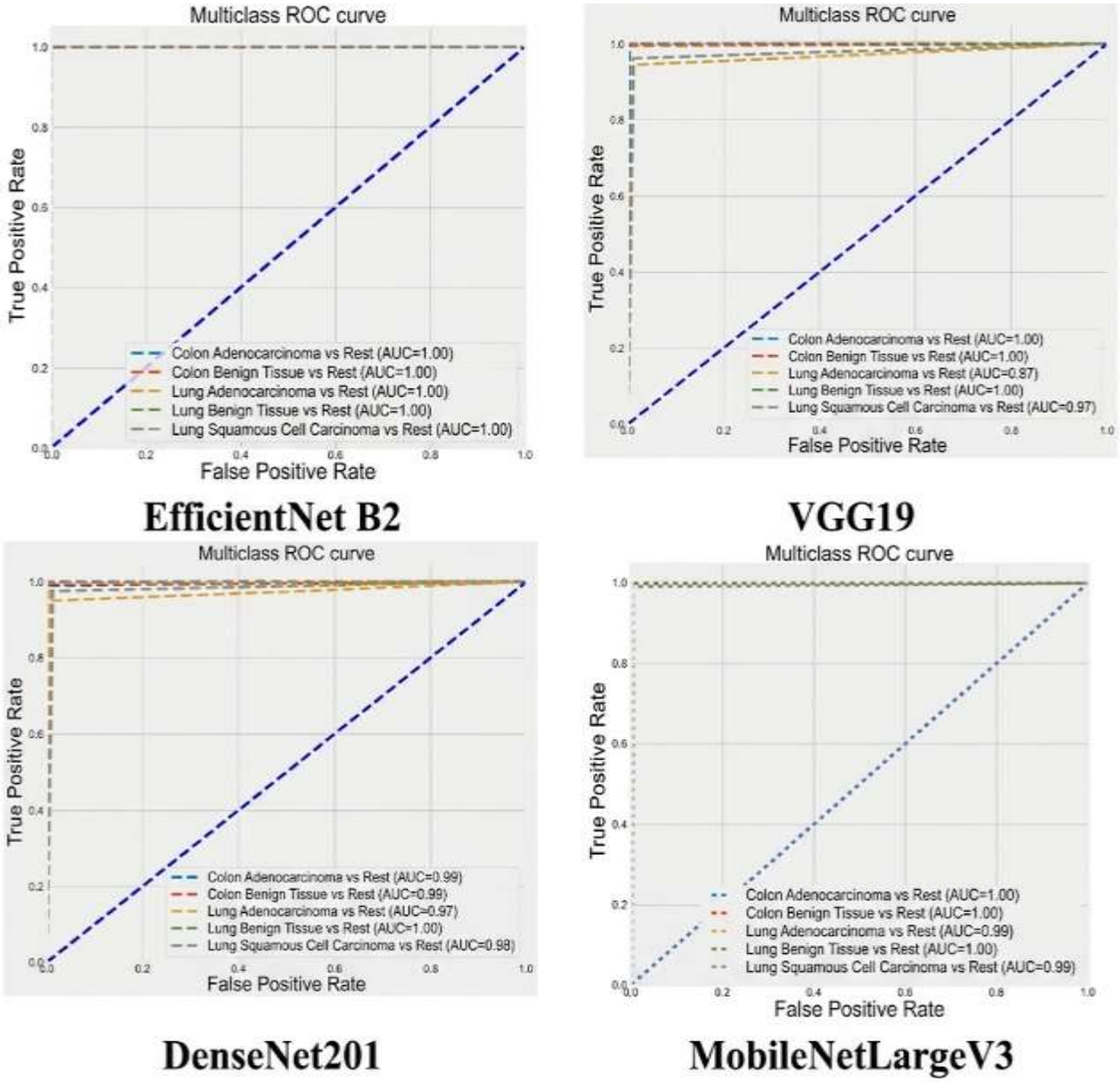


DenseNet201



MobileNetLargeV3

(c)



(d)

Figure 12. The ROC curve: a) KNN, b) LGBM, c) RF, and d) SVC.

5. Conclusions

Our research has put forth an integrated approach that combines DL and ML for the early-stage detection of LC cancer using a high-performance HPI dataset, which we found very promising. We used modified TL models, which we enhanced with average-pooling 2D layers, applied to the adjusted EfficientNet B2 and the SVC, and achieved very impressive results: an accuracy, F1 score, precision, and recall of 0.9996, and an AUC of 0.9998. The results are very accurate and near-perfect, which increases the confidence of novice doctors. Put forth this system as a solution for low- and middle-income countries, which have limited access to advanced diagnostic tools. It supports early detection, which in turn reduces mortality rates and improves health care outcomes. Our results show the great promise of DL and ML in transforming LC diagnosis, especially in resource-constrained settings.

Funding: No funding was obtained for this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Siegel, R. L.; Miller, K. D.; Fuchs, H. E.; Jemal, A.; "Cancer statistics, 2022". CA. Cancer J. Clin., 72 (1): 7–33, 2022.

[2] Mareel, M.; Leroy, A.; "Clinical, Cellular, and Molecular Aspects of Cancer Invasion". Physiol. Rev., 83 (2): 337–376, 2003.

[3] Liu, P.; Lv, H.; Li, Y.; Liu, Z.; Yang, X.; Hu, H.; "Global bone cancer incidence and death rate analysis at 40 years". Discov. Oncol., 16 (1): 1087–1111, 2025.

[4] Kurishima, K.; Miyazaki, K.; Watanabe, H.; Shiozawa, T.; Ishikawa, H.; Satoh, H.; Hizawa, N.; "Lung cancer patients with synchronous colon cancer". Mol. Clin. Oncol., 8 (1): 137–140, 2017.

[5] Del Re, M.; Rofi, E.; Restante, G.; Crucitta, S.; Arrigoni, E.; Fogli, S.; Di Maio, M.; Petrini, I.; Danesi, R.; "Implications of KRAS mutations in acquired resistance to treatment in NSCLC". Oncotarget, 9 (5): 6630–6643, 2017.

[6] Baqir A. Altimmime; Farah A. Rashid; "View of CXCL12 as a Metastasis Inducer Chemokine of Breast Cancer". Al-Nahrain J. Sci., 27 (5): 49–57, 2024.

[7] A. Maryoosh, A.; Pashazadeh, S.; "Leukemia Detection using Machine Learning Algorithms: Current trends and future directions". Mustansiriyah J. Pure Appl. Sci., 2 (3): 33–49, 2024.

[8] van der Velden, B. H. M.; Kuijf, H. J.; Gilhuijs, K. G. A.; Viergever, M. A.; "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis". Med. Image Anal., 79: 102470, 2022.

[9] Madabhushi, A.; Lee, G.; "Image analysis and machine learning in digital pathology: Challenges and opportunities". Med. Image Anal., 33: 170–175, 2016.

[10] Barragán-Montero, A.; Javaid, U.; Valdés, G.; Nguyen, D.; Desbordes, P.; Macq, B.; Willems, S.; Vandewinckele, L.; Holmström, M.; Löfman, F.; Michiels, S.; Souris, K.; Sterpin, E.; Lee, J. A.; "Artificial intelligence and machine learning for medical imaging: A technology review". Phys. Medica, 83: 242–256, 2021.

[11] Alaa Hussein; Mohammed Sahib Mahdi Altaei; "View of Deep Learning Techniques for Detecting and Segmenting Text in Natural Scene Images_ Review". Al-Nahrain J. Sci., 27 (2): 133–144, 2024.

[12] 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, 21 (3): 748, 2021.

[13] 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., 205: 117695, 2022.

[14] Hage Chehade, A.; Abdallah, N.; Marion, J.-M.; Oueidat, M.; Chauvet, P.; "Lung and colon cancer classification using medical imaging: a feature engineering approach". Phys. Eng. Sci. Med., 45 (3): 729–746, 2022.

[15] Mehmood, S.; Ghazal, T. M.; Khan, M. A.; Zubair, M.; Naseem, M. T.; Faiz, T.; Ahmad, M.; "Malignancy Detection in Lung and Colon Histopathology Images Using Transfer Learning With Class Selective Image Processing". IEEE Access, 10: 25657–25668, 2022.

[16] Singh, O.; Singh, K. K.; "An approach to classify lung and colon cancer of histopathology images using deep feature extraction and an ensemble method". Int. J. Inf. Technol., 15 (8): 4149–4160, 2023.

[17] Mengash, H. A.; Alamgeer, M.; Maashi, M.; Othman, M.; Hamza, M. A.; Ibrahim, S. S.; Zamani, A. S.; Yaseen, I.; "Leveraging Marine Predators Algorithm with Deep Learning for Lung and Colon Cancer Diagnosis". Cancers (Basel)., 15 (5): 1591, 2023.

[18] Prusty, S.; Patnaik, S.; Dash, S. K.; "EfficientNetB7: To predict lung and colon cancer at early stages". in RECENT TRENDS IN APPLIED MATHEMATICS IN SCIENCE AND ENGINEERING, 2819(1) 030013, 2023.

[19] Reddy, N.; Parasa, G.; Janani, R.; Thiruveedula, S.; Sujatha, K.; Varaprasad Rao, B.; N, S. M.; "Modernizing Cancer Diagnosis with an Intelligent Model for Lung and Colon Cancer". J. Electr. Syst., 20 (6): 1218–1226, 2024.

[20] Aitazaz, T.; Tubaishat, A.; Al-Obeidat, F.; Shah, B.; Zia, T.; Tariq, A.; "Transfer learning for histopathology images: an empirical study". Neural Comput. Appl., 35 (11): 7963–7974, 2023.

[21] Niu, S.; Liu, Y.; Wang, J.; Song, H.; "A Decade Survey of Transfer Learning (2010–2020)". IEEE Trans. Artif. Intell., 1 (2): 151–166, 2020.

[22] Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q.; "A Comprehensive Survey on Transfer Learning". Proc. IEEE, 109 (1): 43–76, 2021.

[23] Marques, G.; Agarwal, D.; de la Torre Díez, I.; "Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural

network". Appl. Soft Comput., 96: 106691–106702, 2020.
[24] Petrini, D. G. P.; Shimizu, C.; Roela, R. A.; Valente, G. V.; Folgueira, M. A. A. K.; Kim, H. Y.; "Breast Cancer Diagnosis in Two-View Mammography Using End-to-End Trained EfficientNet-Based Convolutional Network". IEEE Access, 10: 77723–77731, 2022.
[25] Srinivasu, P. N.; SivaSai, J. G.; Ijaz, M. F.; Bhoi, A. K.; Kim, W.; Kang, J. J.; "Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM". Sensors, 21 (8): 2852–2879, 2021.
[26] Wang, W.; Li, Y.; Zou, T.; Wang, X.; You, J.; Luo, Y.; "A Novel Image Classification Approach via Dense-MobileNet Models". Mob. Inf. Syst., 2020 (1): 1–8, 2020.
[27] Hsu, C.-C.; Jian, C.-Y.; Tu, E.-S.; Lee, C.-M.; Chen, G.-L.; "Real-Time Compressed Sensing for Joint Hyperspectral Image Transmission and Restoration for CubeSat". IEEE Trans. Geosci. Remote Sens., 62: 1–16, 2024.
[28] Singh, R.; Rana, R.; Singh, S. K.; "Performance Evaluation of VGG models in Detection of Wheat Rust". Asian J. Comput. Sci. Technol., 7 (3): 76–81, 2018.
[29] Tabrizchi, H.; Parvizpour, S.; Razmara, J.; "An Improved VGG Model for Skin Cancer Detection". Neural Process. Lett., 55 (4): 3715–3732, 2023.
[30] Zhao, C.; Shuai, R.; Ma, L.; Liu, W.; Hu, D.; Wu, M.; "Dermoscopy Image Classification Based on StyleGAN and DenseNet201". IEEE Access, 9: 8659–8679, 2021.
[31] Zhang, J.; Lu, C.; Li, X.; Kim, H.-J.; Wang, J.; "A full convolutional network based on DenseNet for remote sensing scene classification". Math. Biosci. Eng., 16 (5): 3345–3367, 2019.
[32] Breiman, L.; "Random Forests". Mach. Learn., 45 (1): 5–32, 2001.
[33] Cover, T.; Hart, P.; "Nearest neighbor pattern classification". IEEE Trans. Inf. Theory, 13 (1): 21–27, 1967.
[34] Zhang, S.; "Challenges in KNN Classification". IEEE Trans. Knowl. Data Eng., 34 (10): 4663–4675, 2022.
[35] Hastie, T.; Tibshirani, R.; Friedman, J.; 'The Elements of Statistical Learning Data Mining, Inference, and Prediction'. 2nd ed.; Springer New York:, New York, NY, USA, 2009.
[36] Liao, H.; Zhang, X.; Zhao, C.; Chen, Y.; Zeng, X.; Li, H.; "LightGBM: an efficient and accurate method for predicting pregnancy diseases". J. Obstet. Gynaecol. (Lahore)., 42 (4): 620–629, 2022.
[37] Zhang, Z.; Wang, S.; Lin, N.; Wang, Z.; Liu, P.; "State of Health Estimation of Lithium-Ion Batteries in Electric Vehicles Based on Regional Capacity and LGBM". Sustainability, 15 (3): 2052, 2023.
[38] Hawwal jasim, A.; Haider k. Hoomod; "Integrative Analysis of Facial and Bodily Expressions for Enhanced Emotion Recognition Using SVM and CNN in Python". Mustansiriyah J. Pure Appl. Sci., 3 (2): 1–11, 2025.
[39] Platt, J. C.; "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods". Adv. large margin Classif. 10.3, 61–74, 1999.
[40] Goh, K.-S.; Chang, E. Y.; Li, B.; "Using one-class and two-class SVMs for multiclass image annotation". IEEE Trans. Knowl. Data Eng., 17 (10): 1333–1346, 2005.
[41] Gong, M.; "A Novel Performance Measure for Machine Learning Classification". Int. J. Manag. Inf. Technol., 13 (1): 11–19, 2021.
[42] Baldi, P.; "Deep Learning in Biomedical Data Science". Annu. Rev. Biomed. Data Sci. Annu. Rev. Biomed. Data Sci. 2018, 8: 181–205, 2024.
[43] Fawcett, T.; "ROC Graphs: Notes and Practical Considerations for Researchers". Mach. Learn. 31.1, 1–38, 2004.
[44] Vergara, I. A.; Norambuena, T.; Ferrada, E.; Slater, A. W.; Melo, F.; "StAR: a simple tool for the statistical comparison of ROC curves". BMC Bioinformatics, 9 (1): 265, 2008.
[45] Fawcett, T.; "An introduction to ROC analysis". Pattern Recognit. Lett., 27 (8): 861–874, 2006.
[46] Yulianto, A.; Sukarno, P.; Suwastika, N. A.; "Improving AdaBoost-based Intrusion Detection System (IDS) Performance on CIC IDS 2017 Dataset". J. Phys. Conf. Ser., 1192: 012018, 2019.