Article

Lung and Colon Cancer Classification Using Multiscale Deep Features Integration of Compact Convolutional Neural Networks and Feature Selection

Omneya Attallah [1,2]

1   Department of Electronics and Communications Engineering, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria 21937, Egypt; o.attallah@aast.edu

2   Wearables, Biosensing, and Biosignal Processing Laboratory, Arab Academy for Science, Technology and Maritime Transport, Alexandria 21937, Egypt

Abstract: The automated and precise classification of lung and colon cancer from histopathological photos continues to pose a significant challenge in medical diagnosis, as current computer-aided diagnosis (CAD) systems are frequently constrained by their dependence on singular deep learning architectures, elevated computational complexity, and their ineffectiveness in utilising multiscale features. To this end, the present research introduces a CAD system that integrates several lightweight convolutional neural networks (CNNs) with dual-layer feature extraction and feature selection to overcome the aforementioned constraints. Initially, it extracts deep attributes from two separate layers (pooling and fully connected) of three pre-trained CNNs (MobileNet, ResNet-18, and EfficientNetB0). Second, the system uses the benefits of canonical correlation analysis for dimensionality reduction in pooling layer attributes to reduce complexity. In addition, it integrates the dual-layer features to encapsulate both high- and low-level representations. Finally, to benefit from multiple deep network architectures while reducing classification complexity, the proposed CAD merges dual deep layer variables of the three CNNs and then applies the analysis of variance (ANOVA) and Chi-Squared for the selection of the most discriminative features from the integrated CNN architectures. The CAD is assessed on the LC25000 dataset leveraging eight distinct classifiers, encompassing various Support Vector Machine (SVM) variants, Decision Trees, Linear Discriminant Analysis, and k-nearest neighbours. The experimental results exhibited outstanding performance, attaining 99.8% classification accuracy with cubic SVM classifiers employing merely 50 ANOVA-selected features, exceeding the performance of individual CNNs while markedly diminishing computational complexity. The framework's capacity to sustain exceptional accuracy with a limited feature set renders it especially advantageous for clinical applications where diagnostic precision and efficiency are critical. These findings confirm the efficacy of the multi-CNN, multi-layer methodology in enhancing cancer classification precision while mitigating the computational constraints of current systems.

Keywords: lung and colon classification; deep learning; convolutional neural networks; computer-aided diagnosis; canonical correlation analysis; feature selection

1. Introduction

Lung and colon or colorectal (LC) cancers are important factors contributing to global malignancy-related mortality, highlighting cancer's status as a major and urgent health

issue worldwide [1]. LC cancers represent the most prevalent tumours, constituting approximately 40% of all cancer diagnoses annually [2]. In 2020, LC cancers became the most prevalent causes of cancer-related mortality between both genders worldwide, with 2.21 million new cases of lung cancer and 1.93 million cases of colorectal cancer documented worldwide, and 1.80 million mortalities from lung cancer and approximately 1 million deaths from colorectal cancer [3]. Tobacco consumption elevates the risk of lung cancer, while an unbalanced diet has the potential to heighten the risk of colon cancer [4]. The detrimental impact of lung cancer on the guts may result in colon cancer. A patient may develop LC cancer in this scenario [4]. Consequently, it has become imperative to examine and identify LC cancer among people concurrently. The rapid detection and accurate diagnosis of such cancers are essential for improving patient outcomes, improving strategies for treatment, and thereby lowering death rates [5].

The early identification of LC cancer is made through the analysis of different kinds of imaging techniques, including positron emission tomography, computed tomography, magnetic resonance imaging, and ultrasound, although this is challenging [6]. The manual inspection of medical photos by professionals is a meticulous and challenging endeavour. Consequently, it is time-consuming and necessitates intense concentration [7]. Furthermore, the identification of cases is significantly more challenging during early diagnosis, as the initial symptoms are often ambiguous and hard to ascertain. Once symptoms manifest, it is too late for prompt intervention [8]. Moreover, the dependable subtyping of these cancers may be unattainable through such screening methods [1]. Histopathological imaging is regarded as the most accurate and reliable technique for cancer detection and subtype classification [9], in contrast to other diagnostic modalities. Nonetheless, conventional manual analysis of histological images is susceptible to errors, labour-intensive, expensive, and significantly dependent on the expertise of pathologists [10]. Therefore, autonomous image processing techniques for diagnosing LC cancer subcategories are necessary to alleviate the workload on pathologists [2].

Advances in artificial intelligence (AI) in several industries including health, agriculture, and medicine [11,12] have motivated and enabled computer-aided diagnosis (CAD) tools to function as a decision support system, promoting early diagnosis and aiding physicians in the swift identification of various diseases [13–15]. The application of AI in the analysis of biomedical scans for identifying illnesses has demonstrated significant potential, exhibiting performance that is in line with, and in certain instances superior to, that of medical experts [16–18]. This has heightened the necessity to employ machine learning (ML) and deep learning (DL) methodologies for the detection of cancerous cells in histopathology images [19–22]. CAD systems used to detect illness alleviate the diagnostic workload on physicians and healthcare facilities while facilitating effective and trustworthy identification for patients [23–25]. ML predominantly depends on manually crafted features derived from expert knowledge to enhance interpretability. In ML, feature extraction is conducted prior to image classification [26]. Nevertheless, these methods are hindered by the inadequacy of the feature extraction technique employed and the resultant loss of data throughout the extraction process [27]. Conversely, DL has gained prominence in medical diagnostic fields by mitigating these drawbacks and demonstrating robust discriminative capabilities [28]. Biomedical data typically consist of images; therefore, the Convolutional Neural Network (CNN) is the prominent DL architecture frequently employed for the analysis of medical scans [29]. In recent times, predesigned and pre-trained CNN architectures are often favoured for their convenience and superior effectiveness [29].

CNN models derive high-level attributes from medical images using their deep structured multi-layers. Thus, intricate and challenging data can be effectively comprehended through CNNs. The efficacy of the CNN designs is attributable to the intricacy and pro-

fundity of their construction. As the model's sophistication escalates, the quantity of parameters within the model concurrently rises [30]. Several CNN designs necessitate vast modification of parameters throughout the learning process. Nevertheless, a substantial amount of CNNs parameters would adversely impact the model's generalisation capacity and lead to overfitting [31]. Decreasing the dimension of attributes and employing compact DL models are strategies to mitigate overfitting that may arise from complexities in models [32]. Generally, the large number of attributes from pre-trained CNN structures can lead to overfitting, negatively impacting classification performance. Multiple feature selection and reduction techniques can be employed to prevent overfitting, thereby eliminating superfluous and/or redundant attributes [33]. The literature that has employed CNN to construct CAD systems for LC cancer diagnosis is extensive. Many of these CAD frameworks relied on CNNs with a massive number of deep layers and parameters. Some of the existing CAD frameworks depend on lightweight CNN but extract features of huge dimensions and do not employ feature selection or reduction methods. Even those which use a feature selection or reduction approach still have feature vectors which are lengthy and need further reduction. Several current CAD systems utilise a single CNN model either customised or pre-trained, whereas fusing features of multiple CNNs can improve diagnostic performance. Moreover, transfer learning and pre-trained CNN can boost performance compared to building a CNN from scratch. Additionally, most of the existing CAD systems obtain a single-scale feature vector from one deep layer of a DL model. However, acquiring multiscale features could enhance classification performance.

To overcome the previously mentioned limitations, this study proposes a CAD framework based on multiple pre-trained compact CNNs of different architectures. It employs transfer learning to extract features from each CNN. In addition, it acquires multiscale variables from two distinct deep layers of each deep neural network. Furthermore, it adopts feature reduction and selection approaches, including canonical correlation analysis (CCA), analysis of variance (ANOVA), and Chi-Squared methods to reduce feature dimensionality and select the most impactful features. The following is a list of the research's novelties and contributions:

- Establishing an efficient CAD leveraging lightweight deep neural network architectures.
- Acquiring deep variables from dual separate layers of each deep neural network model labelled as Layer I and Layer II features.
- Applying CCA to lower the dimension of Layer I features, thus diminishing classification complexity.
- Integrating reduced Layer II with Layer I deep variables for each DL model to produce multiscale features instead of using single-scale deep variables obtained from one layer.
- Merging the dual-layer deep features of the three DL models to incorporate the benefits of their distinct architectures.
- Applying the ANOVA and Chi-Squared feature selection methods to pick the most meaningful deep variables and diminishing the dimensionality of feature space, thus reducing complexity.

The remaining sections of the paper are organised as follows: Section 2 represents previous works that used deep learning techniques for LC cancer diagnosis. Section 3 demonstrates the methods and materials including the dataset used and the proposed CAD system, Section 4 illustrates the performance metrics used to evaluate the proposed CAD and the hyper-parameters finetuning. Section 5 presents the experimental results. Section 6 discusses the key findings of the proposed CAD. Lastly, Section 7 gives a brief conclusion to the study.

2. Previous Works

The research of [34] presented a marine predator's algorithm (MPA) integrated with DL for the classification of LC cancer. The suggested framework sought to precisely differentiate various categories of LC cancer by employing histopathological images. The framework exploited contrast-limited adaptive histogram equalisation (CLAHE) for contrast enhancement. Furthermore, the framework utilised MobileNet for the production of feature vectors. Simultaneously, the framework employed MPA for optimisation. Additionally, deep belief networks (DBN) were employed to classify LC tissues. The framework accomplished an accuracy of 99.28%. Another study [7] established a hybrid classification method by integrating three separate models: random forest (RF), support vector machine (SVM), and logistic regression (LR). The ultimate estimates were ascertained through a majority voting methodology. Deep attributes were derived from LC histopathological photographs employing VGG16 and local binary pattern (LBP) methods, with the incorporated attributes functioning as inputs for the hybrid model. The approach was verified on the LC25000 dataset, exhibiting strong performance with a mean accuracy and precision of 99.00%, and a recall and F1-score of 98.80%.

The research of [35] established a robust and efficient model for the classification of LC cancer by exploiting advanced deep learning methodologies. ResNet50 and EfficientNetB0 have been adjusted through layer modifications and subsequent manually operated hyperparameter modifications. Subsequently, deep transfer learning was employed for training the models. Features were derived from both deep models and fused with a priority-based serial integration technique. The normal distribution-based grey wolf optimisation (GWO) method was employed to improve feature reliability by choosing the most pertinent features as inputs for five classifiers. The ultimate prediction was produced via a soft voting process, attaining an accuracy of 98.73% on the LC25000 dataset. Another automatic framework [1] employed EfficientNetV2 models, including large, medium, and small architectures for LC cancer classification. EfficientNetV2-L demonstrated outstanding performance, attaining an accuracy of 99.97%. Gradient-weighted class activation mapping (Grad-CAM) was applied to produce visual saliency maps, emphasising significant areas in test set photos where the model focused its attention throughout estimations.

One paper [2] introduced a sophisticated structure that integrated DL and meta-heuristic optimisation methods for precisely predicting colon, lung, or LC malignancies from histopathological photographs. ResNet-18 first served for binary classification, whereas EfficientNet-b4-wide was applied to multi-class situations with three and five classes, employing the LC25000 dataset for training. Deep attributes obtained from such models were subsequently enhanced using a hybrid meta-heuristic technique, AdBet-WOA, which combines the whale optimisation algorithm (WOA) with adaptive β-Hill Climbing for efficient feature selection. The chosen attributes were categorised employing SVM, attaining accuracies of 99.96% for LC combined datasets. Another work [36] introduced an innovative CNN structure to detect LC cancer. The model that was suggested is compact. The algorithm efficiently used multiscale feature extraction. Explainable AI methods, involving Grad-CAM and Shapley additive explanation (SHAP), were used to emphasise significant data regions, thereby improving transparency while recognising possible flaws. The experimental findings demonstrated that the model attained an accuracy of 99.20% in multi-class predictions across five categories, exceeding rival methodologies.

One study [37] presented a technique based on DL for the automated categorisation of LC cancer. A CNN employing a VGG16 construction, in conjunction with CLAHE, was implemented for classifying the histopathological photos involved in the LC25000 dataset. The findings indicated that the suggested method attained a peak classification accuracy of 98.96%. The integration of CLAHE markedly improved identification accuracy relative

to the approach lacking CLAHE. The study of [38] introduced a hybrid segmentation and classification method called ColonNet for the precise identification of mitotic nuclei in histopathological images of LC tumours. The method enhanced classification robustness and generalisation by integrating two separate CNN designs. These models were developed to represent the morphological and textural features of mitotic cells, tackling issues related to weak annotations via asymmetrical split transform-merge and label optimisation techniques. A global–local pyramid pattern (GLPP) facilitated feature extraction and model cooperation, while deep residual blocks improved performance. The suggested model, ColonNet, surpassed current CNN algorithms, including VGG, ResNet, and DenseNet, in classification accuracy.

One article [39] delineated a pipeline based on an optimisation method using Al-Biruni earth radius (BER) for the diagnosis of LC cancer. This pipeline employed a progressed ShuffleNet structure for obtaining features, with its hyperparameters refined by the BER method. A deep convolutional recurrent neural network (DCRNN) architecture was adopted for the accurate identification of LC cancer, whereas the coati optimisation algorithm (COA) was used for choosing the parameters of the DCRNN. Comprehensive experiments on the LC25000 dataset revealed that the integration of BER and COA techniques markedly enhanced cancer identification efficacy relative to current approaches. The study [40] devised three approaches, each incorporating two CNNs and Artificial neural networks (ANN) for the detection of LC cancer from the LC25000 dataset. The GoogLeNet and VGG19 models generated highly dimensional attributes; therefore, irrelevant and superfluous attributes were eliminated to decrease dimensions and preserve vital attributes using the principal component analysis (PCA) technique. The initial approach for detecting cancer via ANN employs essential features from the GoogLeNet and VGG19 models independently. The subsequent approach employs an ANN that integrates the attributes of GoogLeNet and VGG-19. Two systems were employed; one system diminished dimensions and integrated features while the other merged the large dimension of features and subsequently decreased large dimensions. The final approach employs an ANN that integrates features from GoogLeNet and VGG19, along with handcrafted features. The ANN achieved a sensitivity of 99.85%, precision of 100%, accuracy of 99.64%, specificity of 100%, and an AUC of 99.86% by integrating VGG19 fusion variables with handcrafted characteristics.

The work [41] combined ResNet-50, InceptionV3, and DenseNet, with Kernel Extreme Learning Machine (KELM) to facilitate the swift and precise identification of LC cancer with histopathology scans. A feature fusion technique merged the advantages of such structures, increasing the technique's capacity to identify varied attributes and enhance classification efficacy. KELM effectively handled the multidimensional feature set produced by the models, facilitating rapid and accurate classification. This holistic methodology attained remarkable diagnostic efficacy, resulting in an accuracy of 99.0%. The article [42] employed three pre-trained models including MobileNet, ShuffleNet, and SqueezeNet as feature extractors followed by different transformation techniques including PCA and fast Walsh–Hadamard transform (FWHT) to reduce their dimensionality. Next, these reduced features of the three CNNs were fused using discrete wavelet transform (DWT) and then classified using an SVM classifier reaching 99.6% accuracy with 510 features. The study [43] introduced a framework based on an Inception-ResNetV2 network coupled with LBP attributes to improve the diagnostic accuracy of LC cancer diagnosis. The framework exhibited exceptional performance with an accuracy of 99.98%. Furthermore, XAI methodologies, particularly SHAP, were employed to clarify the model's decision-making procedure, thereby enhancing the transparency and comprehension of the DL system.

This study employs established CNN and ML models; however, its novelty and importance stem from the suggested framework's novel methodology to address the

limitations typically found in relevant CAD systems for lung and colon cancer classification. The primary contributions of this study are as follows:

- This research develops an effective CAD framework, exploiting compact deep neural network architectures to mitigate issues related to computational complexity and overfitting and to achieve a balance between efficiency and diagnostic performance, rendering it appropriate for practical clinical applications.
- This framework differentiates itself from numerous existing CAD systems by employing multiscale feature extraction, obtaining variables from two separate deep layers (Layer I and Layer II) of each CNN. This method improves the model's capacity of capturing both low- and high-level representations, thereby enhancing classification accuracy.
- The system integrates dual-layer features from three disparate pre-trained CNN architectures—MobileNet, EfficientNetB0, and ResNet-18. By incorporating attributes from these models, the framework leverages their distinct architectural advantages, thus improving diagnostic efficacy.
- To overcome the challenges associated with extensive feature dimensions and the potential for overfitting, canonical correlation analysis (CCA) is utilised to diminish the dimensionality of Layer I features. The diminished Layer I features are subsequently combined with Layer II features to generate multiscale feature sets. Feature selection is applied to identify the most significant features, thereby substantially decreasing the dimensionality of the feature space and enhancing classification efficiency.

These contributions resolve significant deficiencies observed in current CAD systems, including a dependence on singular CNN architectures, the extraction of single-scale features, and inadequate feature selection or dimensionality reduction techniques. This study advances the establishment of efficient and effective CAD systems for lung and colon cancer by integrating multiscale features, utilising robust feature selection methods, and harnessing the complementary strengths of various CNN architectures.

3. Materials and Methods

3.1. Lung and Colon Cancer Histopathology Dataset

The LC25000 database developed by Borkowski, et al. [44] serves as a prominent benchmark in histopathological image analysis and the classification of LC cancers in medical imaging. The dataset consists of 25,000 high-resolution histopathological photographs categorised into five distinct classes: colon adenocarcinoma, colon benign tissue, lung adenocarcinoma, lung benign tissue, and lung squamous cell carcinoma. Each category comprises an equal total of 5000 images, thereby guaranteeing equitable distribution across classifications. Such database photos were cropped and scaled to 768×768 pixels, preserving adequate detail for cellular and morphological feature analysis. The data collection is annotated and organised to support supervised learning tasks, rendering it appropriate for the development and assessment of machine learning and DL models. Figure 1 presents an exemplar for each category of LC photos.

3.2. Canonical Correlation Analysis

The canonical correlation procedure is a statistical approach employed to investigate the connections among two separate variable sets by determining the linear combinations of each set that exhibit the greatest correlation with each other. Such a technique finds the largest correlation between pairs of canonical attributes, which are linear combinations of the original variables from each set. In mathematical terms, given a pair sets of variables $X = [x_1, x_2, \ldots, x_p]$ and $Y = [y_1, y_2, \ldots, y_q]$, canonical correlation analysis (CCA) aims to

identify linear combinations $u = X W_x$ and $v = Y W_y$, where W_x and W_y are weight vectors, such that the correlation ρ between u and v is maximised using the following formula [45].

$$\rho = \max_{w_x, w_y} \frac{w_x^T \sum_{XY} w_y}{\sqrt{w_x^T \sum_{XX} w_x} \sqrt{w_y^T \sum_{YY} w_y}} \tag{1}$$

where $\sum_{XX}$ and $\sum_{YY}$ are the covariance matrices of X and Y, respectively, and $\sum_{XY}$ is the cross-covariance matrix between X and Y.
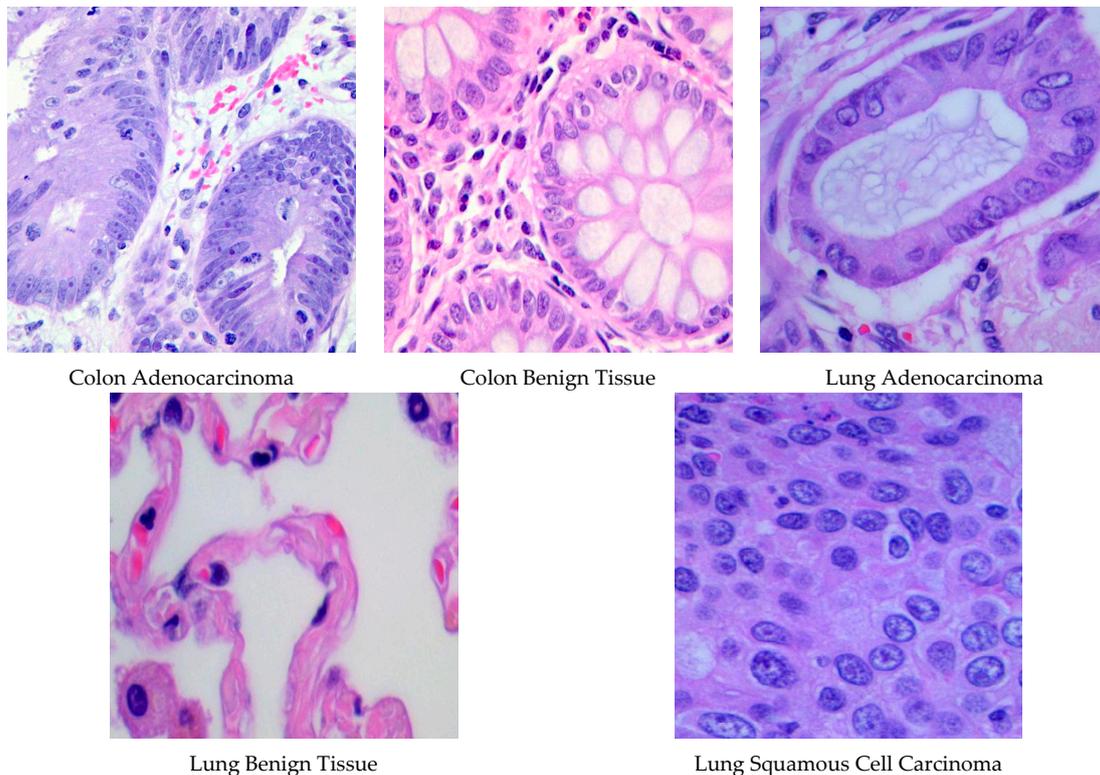

Figure 1. Examples of pictures taken from the LC25000 dataset.

CCA reduces features by transforming the original set into a smaller collection of canonical variables, thereby minimising redundancy and maintaining the interrelationships among features. By choosing the most significant canonical attributes, CCA diminishes the data's dimensionality, thus streamlining the framework while preserving essential interdependencies. This method has been deployed in numerous ML applications to alleviate overfitting, decrease computational complexity, and improve model generalisation by removing unnecessary or duplicate characteristics [46]. CCA is especially useful in multidimensional datasets where maintaining the latent correlations among feature sets is essential.

3.3. Suggested System

The suggested research presents a CAD framework aimed at overcoming the deficiencies identified in current systems for identifying types of lung and colon cancer. This framework employs compact deep neural network designs to attain an acceptable compromise between computational efficiency and diagnostic accuracy, rendering it appropriate for clinical applications. This framework leverages multiscale feature extraction, collecting variables from two separate layers, referred to as Layer I and Layer II, of each CNN, in contrast to many current CAD systems that depend on a singular CNN architecture and single-scale feature extraction. This method allows the model to acquire both low- and

high-level representations, thus improving classification accuracy. The introduced CAD system consists of five stages: LC photo preliminary processing, training compact deep networks and dual-layer feature extraction, dimensionality reduction and dual-layer features merging, multi-deep networks feature incorporation and selection, and ultimately LC cancer classification. Figure 2 presents a summary of each of those phases. First, the photos are scaled-down and augmented. Next, compact deep neural network models—MobileNet, EfficientNetB0, and ResNet-18—are utilised for obtaining attributes from two separate layers, designated as Layer I and Layer II. Layer I length is large; thus, the dimensionality of these attributes is diminished through CCA. Afterward, the diminished features of Layer I are integrated with variables obtained from Layer II.
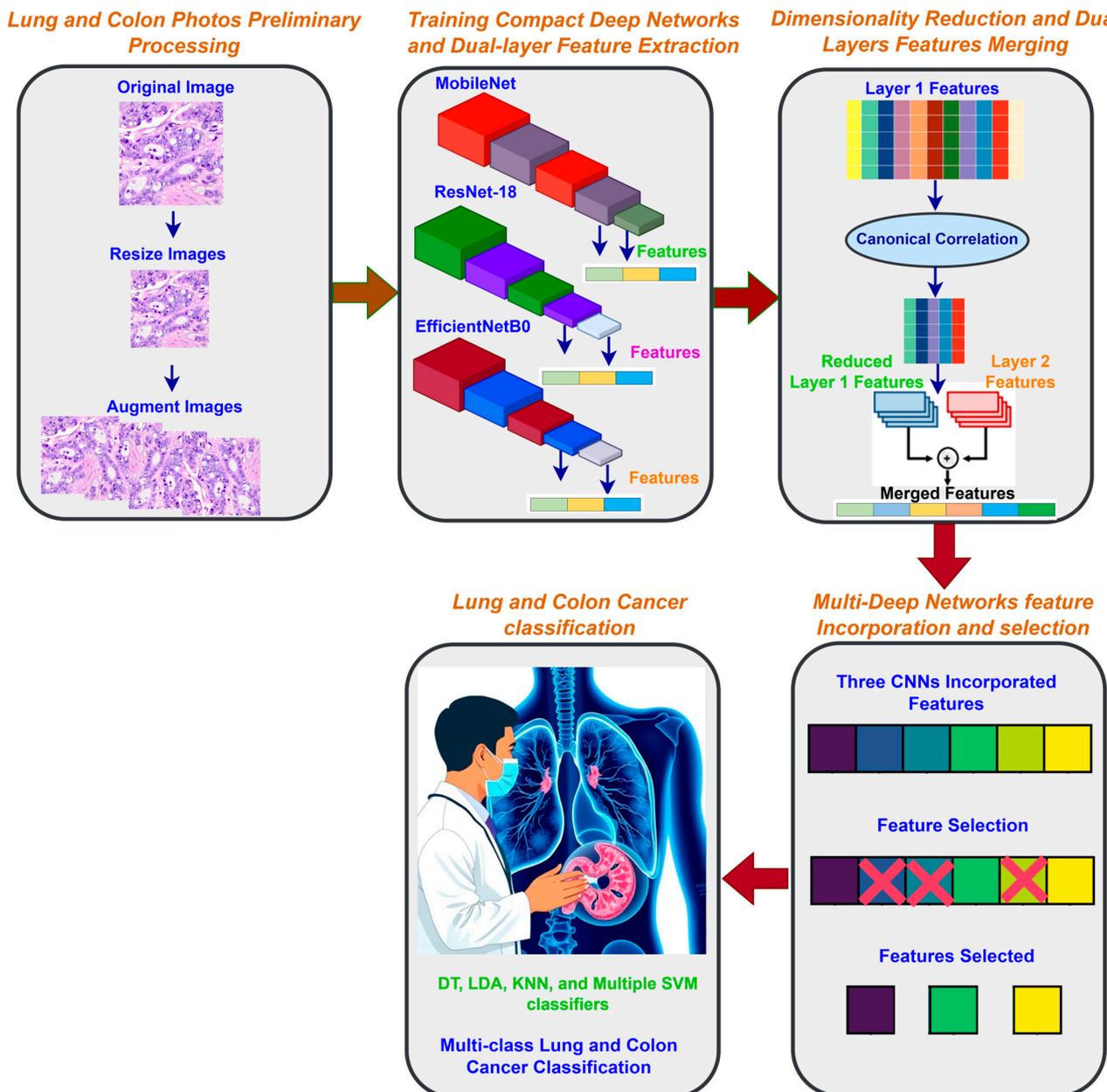

Figure 2. A summary of the phases of the proposed CAD.

Following this, the framework incorporates dual-layer attributes from the three pre-trained CNN structures. The framework combines features from these models to leverage their unique architectural benefits, thereby enhancing diagnostic efficacy. CCA helps to

mitigate the difficulties posed by substantial feature dimensions and the risk of overfitting by reducing the dimensionality of Layer I features. Afterward, the diminished attributes are subsequently integrated with Layer II features to produce multiscale feature sets, facilitating a more resilient representation for classification. Subsequent to the aggregation of features gathered from the dual-layers of the three CNN models, further processing is conducted, adopting ANOVA and Chi-Squared feature selection methodologies. These approaches discern the most essential attributes, significantly reducing the dimensionality of the feature space while maintaining classification efficacy. The chosen features are then employed to train various machine learning classifiers, guaranteeing that the system provides precise and dependable diagnostic outcomes.

3.3.1. Lung and Colon Photos Preliminary Processing

The training procedure for each DL model commences with the input layer necessitating photos of a specified dimension. Accordingly, the sizes of the histopathological images for LC cancer have been altered to meet the input layer requirements of the three selected deep learning models, specifically $224 \times 224 \times 3$. Furthermore, data augmentation strategies are exploited to improve the training effectiveness of such models and reduce overfitting. Data augmentation enhances the diversity of images within the dataset, facilitating the improved generalisation of models throughout learning. This study employs various augmentation techniques, including scaling in both x and y dimensions between an interval of [0.5, 2], flipping along both axes, translating photos in x and y using angles within [−20, 20], and shearing between a range of [−45, 45].

3.3.2. Training Lightweight Deep Networks and Dual-Layer Feature Extraction

The present step employs transfer learning, retaining three lightweight CNNs that have been pre-trained on the ImageNet database. The choice of MobileNet, EfficientNetB0, and ResNet-18 for the study was determined by critical factors such as network structure, computational effectiveness, and performance. The aforementioned designs were chosen for their synergistic advantages, rendering them appropriate for diverse DL applications. MobileNet [47] enhances computational efficiency through depthwise separable convolutions, markedly decreasing model complexity while maintaining accuracy. EfficientNetB0 [48] utilises a hybrid scaling approach that optimises the network's complexity, width, and resolution, yielding outstanding performance with minimal computational expense. ResNet-18 [49], recognised for its deep residual learning architecture, mitigates the vanishing gradient issue, facilitating the training process of deeper networks and improving feature extraction. Collectively, these networks provide versatility, accommodating both resource-constrained applications and high-performance tasks, exemplifying their adaptability in practical scenarios. The architecture of each model is modified to incorporate five fully connected layers to align with the five categories in the LC25000 database. Hyperparameters are meticulously adjusted based on the experimental configuration, with each model trained autonomously on histopathology images, enabling each network to capitalise on its distinct advantages while guaranteeing consistent learning. Subsequent to retraining, deep feature extraction is executed. The present research differentiates itself from traditional methods by extracting attributes from two separate layers of each network: the pooling layer (Layer I) and the fully connected layer (Layer II), and their lengths are illustrated in Table 1.

Table 1. The length of the deep variables acquired from each layer of the three deep neural network models.

Deep Neural Network Model	Layer I	Layer II
MobileNet	1280	5
ResNet-18	512	5
EfficientNetB0	1280	5

In a CNN, layers are structured to incrementally acquire more intricate features. The initial layers detect fundamental features like edges and textures, whereas the subsequent deep layers concentrate on recognising complex, disease-specific structures linked to lung and colon cancer [50,51]. The last layers of a CNN are essential for feature extraction, as they yield the most important data for classification tasks. Therefore, the present study employs the final pooling and fully connected layers as feature extractors. The pooling layer diminishes the spatial dimensions of feature maps produced by convolutional layers while preserving the most vital attributes. Pooling methods including average pooling are used to compress the essential activations within designated areas of the feature map [52]. This decrease in dimension streamlines the feature set, highlighting the most significant attributes for classification. Subsequent to the pooling layer, the fully connected layers function as highly sophisticated feature extractors. In contrast to convolutional layers that examine localised areas, fully connected layers integrate information from the entire feature map, allowing for the identification of global patterns as well as relations within features [53]. This method is especially beneficial in the classification of lung and colon cancer, where comprehending the fundamental characteristics of the disease is crucial. By integrating attributes from the final pooling layer, which encapsulates spatial details, with those from the fully connected layers, which denote global patterns and relationships, the framework attains hierarchical and thorough representations of the input data.

Studies demonstrate that the incorporation of features from various layers typically enhances classification performance relative to dependence on a singular layer [50,54]. Each layer captures distinct characteristics of the input picture, and the combination of these attributes yields a more resilient representation for the classification task. This method improves the system's capacity to deliver a comprehensive and precise understanding of the characteristics of lung and colon cancer.

3.3.3. Dimensionality Reduction and Dual-Layers Merging

Due to the substantial size of the features from Layer I as shown in Table 1, dimensionality reduction is implemented through CCA to efficiently decrease the size of such attributes. An ablation analysis is carried out to study the effect of varying the number of canonical variables on the classification performance. Subsequently, the reduced canonical features of Layer I of each CNN are integrated with the deep features obtained from Layer II. This step is accomplished to determine whether fusing deep variables from different layers of each CNN could enhance performance and is superior to using deep attributes from a single layer of each deep neural network.

CCA was used as it identifies linear combinations of variables from a pair of feature sets that optimise their mutual correlation. In contrast to principal component analysis (PCA), which prioritises variance maximisation, CCA explicitly enhances the correlation among distinct feature sets, rendering it especially appropriate for the proposed method that integrates dual-layer features from multiple CNNs. Although t-SNE is proficient in visualisation, it does not maintain the global patterns with the same efficiency as CCA. Moreover, t-SNE has a non-deterministic nature which limits its use in real-time applications.

In the experiments, CCA exhibited strong efficacy in dimensionality reduction while preserving classification accuracy. However, CCA has certain limitations, such as its susceptibility to noise and dependence on linear correlations, which could inadequately represent intricate relationships within the data. To address these issues, CCA was integrated with another robust method, ANOVA, to guarantee that the attributes extracted are both discriminatory and computationally effective.

3.3.4. Multi-Deep Networks Feature Incorporation and Selection

At this point, the feature vectors obtained from three separate deep neural networks are methodically combined to create a holistic multidimensional representation. This combined strategy seeks to investigate the prospect of attaining performance enhancements by integrating features from various CNN topologies and to evaluate the efficacy of different combination sets. Nevertheless, managing high-dimensional data presents considerable computational difficulties, requiring the implementation of an advanced feature selection (FS) technique. FS is an essential procedure that identifies and preserves the most informative variables while discarding irrelevant or redundant ones to improve model efficiency and accuracy. This study employs analysis of variance (ANOVA), a form of statistics utilised to assess the relative importance of individual features by assessing their variance among distinct classes. For a feature x_i, the ANOVA F-statistic is calculated as follows:

$$F = \frac{Variance\ between\ groups}{Variance\ within\ group} = \frac{\frac{1}{c-1}\sum_{k=1}^{c} n_k (\mu_k - \mu)^2}{\frac{1}{n-c}\sum_{k=1}^{c} \sum_{j=1}^{n_k} n_k \left(x_{k,j} - \mu_k \right)^2} \tag{2}$$

where c represents the sum of categories, n_k denotes the number of observations in category k, μ_k is the mean of attribute x_i in class k, μ is the overall mean of x_i, and n signifies the whole number of observations. Features exhibiting elevated F-statistics are deemed more pertinent to the classification task and are prioritised for incorporation into the model. The present research employs ANOVA-based feature selection in order to minimise the dimensionality of the fused feature set while retaining the most discriminative variables, thus enhancing the computational effectiveness and predictive ability.

ANOVA was selected for its efficacy in detecting statistically important features by assessing variance among classes, rendering it especially appropriate for datasets where class separation is essential. ANOVA is computationally effective and relatively easy to implement, which aligns with the lightweight and resource-constrained emphasis of the suggested approach. To substantiate this selection, a comparative analysis with alternative FS methodologies was performed, including Chi-Squared.

The Chi-Squared FS approach [55] ranks variables according to their significance to the classification process by calculating the Chi-Squared statistic for each attribute. This allows it to assess the dependence among variables and target categories. This technique compares the actual and projected frequencies of a feature's values throughout different categories to determine its statistical independence. Larger Chi-Squared score variables are ranked for training models since they are regarded to be more significant. The Chi-Squared method presumes independence among attributes; thus, its efficacy may be restricted, even if it is good at lowering feature dimensions where interactions across features and classes are nonlinear.

3.3.5. Lung and Colon Cancer Classification

During the classification phase of LC tissue, eight ML classifiers are applied to categorise the five different categories in the LC25000 database. The classifiers comprise decision tree (DT), linear discriminant analysis (LDA), k-nearest neighbour (KNN), and five

variants of support vector machines (SVMs) employing distinct kernel functions: linear, cubic, medium Gaussian, coarse Gaussian, and quadratic. Every single classifier has been developed to employ distinct computational approaches to enhance performance for the multi-class classification task. Fivefold cross-validation is employed to guarantee rigorous assessment and reduce overfitting. This validation method partitions the database into five equal segments, utilising four segments for training and one for testing in each round of validation, with the test set rotating among the folds. The cross-validation procedure guarantees that each data instance is employed for both training and testing, yielding a dependable evaluation of classification performance across all eight models. The comparative evaluation of the above classifiers underlines their efficacy in precisely differentiating between LC histopathological classes, providing insights into their appropriateness for medical image analysis tasks.

4. Hyperparameter Refinement and Performance Assessment

By adjusting the learning rate to 0.0001, the batch size to 4, and the quantity of epochs to 5, the deep neural networks are optimised for the training procedure. Validation was performed after every 130 iterations to assess the learning error at the completion of each epoch. The deep models are trained by employing the stochastic gradient descent optimiser with momentum, while all the other hyperparameters remain at their default settings. The system is executed utilising MATLAB R2024a.

For assessing the efficacy of the suggested system on the categories included in the LC25000 dataset, various metrics are employed, offering a thorough evaluation of its performance. The metrics encompass precision, accuracy, F1-score, sensitivity, specificity, and Matthew's correlation coefficient (MCC), all calculated using their corresponding Formulas (3)–(8). Precision quantifies how many true positive estimations there are out of all the positive estimations, whereas sensitivity computes the classification algorithm's capacity to accurately detect positive cases. Specificity, conversely, denotes the model's precision in accurately recognising negative instances. The F1-score delivers a harmonic mean of precision and sensitivity, serving as a balanced metric for imbalanced datasets. The MCC assesses the comprehensive quality of classification by evaluating true and false positives and negatives and is especially valuable in multi-class situations. Furthermore, the confusion matrix is produced to illustrate the model's prediction distribution among the classes, and the receiver operating characteristic (ROC) curve is calculated to evaluate the trade-off between true positive and false positive rates. Collectively, these metrics guarantee a comprehensive assessment of the model's efficacy.

$$Accuracy = \frac{TP + TN}{TN + FP + FN + TP} \tag{3}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{6}$$

$$F1 - Score = \frac{2 \times TP}{(2 \times TP) + FP + FN} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

In ML classification, true positives (TP) denote occurrences where the ML algorithm accurately identifies the positive class, such as detecting cancer when it exists, whereas true negatives (TN) signify instances in which the model correctly recognises the negative class, such as affirming the absence of cancer when it is not present. Conversely, false positives (FP) arise when the model erroneously identifies an instance that is negative as belonging to the positive class, exemplified by diagnosing cancer in the absence of the disease. False negatives (FN) occur when the model erroneously classifies a positive instance as negative, exemplified by the failure to identify the presence of cancer.

5. Results

This section will discuss the results of the LC cancer classification step of the presented system. Initially, it will present the performance measure values of the eight ML classifiers trained with Layer II deep variables acquired from each deep neural network separately. Next, it will demonstrate the results of the ablation study, which examines the impact of changing the number of canonical variables obtained after applying CCA to Layer I deep features, on the classification performance. After that, the results achieved for each deep neural network, where the deep variables gathered from Layer II and reduced variables of Layer II are combined, are presented. The findings demonstrate whether integrating features from different layers of a CNN could improve performance. Afterwards, the results after the incorporation of the dual-layer features of the three DL models and the application of the ANOVA FS approach used to select the most important variables are presented.

5.1. Layer II Deep Features Results

This section presents the results of the eight ML algorithms constructed with deep variables of Layer II of each CNN model which corresponds to the fully connected layer. Table 2 illustrates the classification accuracy attained by different ML classifiers using Layer II features from each CNN model—EfficientNetB0, ResNet-18, and MobileNet. Of the three models, MobileNet continually displays superior performance, attaining a maximum accuracy of 99.3% with the LSVM, CGSVM, and MGSVM classifiers, followed by KNN and QSVM, which achieved 99.2% accuracy, and LDA and DT, which attained an accuracy of 99.1% and 98.9%, respectively. ResNet-18 closely follows MobileNet, reaching a peak accuracy of 99.2% across various classifiers, including LSVM, QSVM, CSVM, and MGSVM. Subsequently, KNN and CGSVM accomplished 99.1% accuracy, and DT and LDA achieved 98.8% accuracy. On the other hand, EfficientNetB0, although marginally inferior to its counterparts, still achieves commendable results, reaching a maximum accuracy of 98.8% using LSVM and CSVM followed by 98.7% with MGSVM, CGSVM, QSVM, and KNN. The results presented highlight the efficacy of Layer II features, obtained from the fully connected layers, in acquiring discriminatory information for classification tasks. The uniform performance among classifiers emphasises the dependability and scalability of the features, with MobileNet's efficient structure and feature extraction capacities proving to be the most beneficial in this analysis. These findings validate the feasibility of lightweight CNNs and their incorporation with ML classifiers for the accurate classification of LC cancer.

Table 2. The classification accuracy (%) of the classifiers fed with Layer II features.

	DT	LDA	KNN	LSVM	QSVM	CSVM	MGSVM	CGSVM
EfficientNet	98.4	98.6	98.7	98.8	98.7	98.8	98.7	98.7
ResNet-18	98.8	98.8	99.1	99.2	99.2	99.2	99.2	99.1
MobileNet	98.9	99.1	99.2	99.3	99.2	99.2	99.3	99.3

5.2. Layer I Deep Features Results

This section shows the results of the ablation study, which investigates the effect of altering the number of reduced canonical variables obtained after applying CCA on the Layer I features of each CNN on the classification performance. Table 3 displays a detailed ablation study illustrating classification accuracy outcomes across varying quantities of canonical features (10–50) for three CNN architectures (EfficientNet, ResNet-18, and MobileNet) employing eight distinct classifiers. The findings indicate that MobileNet attained the highest overall performance, achieving 99.4% accuracy with QSVM, CSVM, and MGSVM classifiers utilising 50 canonical features. For all three CNNs, enlarging the number of canonical variables typically resulted in enhanced classification accuracy, with the most substantial improvements typically observed between 10 and 30 variables. EfficientNet achieved its optimal performance of 99.1% using both CSVM and MGSVM with 50 features, whereas ResNet-18 reached a peak of 99.3% with QSVM at 50 features. Among the classifiers, SVM variants (notably QSVM, CSVM, and MGSVM) consistently surpassed DT and LDA across all three networks. The KNN classifier demonstrated robust performance, particularly with MobileNet features, attaining 99.1% accuracy using 50 features. LSVM and CGSVM exhibited commendable performance, albeit marginally inferior to their more intricate SVM variants. LDA exhibited the highest sensitivity to the quantity of canonical features, with its performance significantly enhancing as the number of attributes raised especially for ResNet-18, where it rose from 94.6% at 10 attributes to 97.4% at 50 variables.

Table 3. The ablation study which shows the classification accuracy (%) of the classifiers fed versus altering the number of Layer I features after applying CCA.

Number of Canonical Features	DT	LDA	KNN	LSVM	QSVM	CSVM	MGSVM	CGSVM
EfficientNet								
10	97.9	97.8	98.2	98.2	98.3	98.3	98.3	98.2
20	97.9	97.9	98.5	98.4	98.6	98.7	98.7	98.4
30	97.9	97.9	98.6	98.4	98.9	99.0	98.9	98.3
40	97.8	98.2	98.7	98.5	98.9	98.9	99.0	98.5
50	97.8	98.3	98.8	98.6	99.0	99.1	99.1	98.5
ResNet-18								
10	97.2	94.6	98.2	98.1	98.4	98.3	98.3	97.9
20	97.4	95.6	98.8	98.7	98.9	99.0	98.9	98.7
30	97.3	96.6	98.9	99.0	99.2	99.1	99.1	98.9
40	97.4	97.4	98.7	98.9	99.1	99.1	99.2	98.9
50	97.5	97.4	98.7	99.1	99.3	99.2	99.2	98.9
MobileNet								
10	98.1	97.2	98.7	98.7	98.8	98.8	98.9	98.6
20	98.0	97.8	99	98.9	99.2	99.1	99.2	98.9
30	98.2	97.7	99.0	99.1	99.2	99.3	99.3	99.0
40	98.0	97.9	99.0	99.1	99.3	99.3	99.3	99.0
50	98.0	98.1	99.1	99.2	99.4	99.4	99.4	99.0

5.3. Dual-Layer Deep Features Results

This section illustrates the outcomes of the classifiers fed with the combined Layer I and Layer II deep features. Table 4 provides an in-depth analysis of classification accu-

racies obtained by integrating Layer I and Layer II attributes from three neural networks (EfficientNet, ResNet-18, and MobileNet), indicating that the incorporation of attributes typically enhanced performance across all models. MobileNet attained the most outstanding overall performance, achieving 99.7% accuracy with its integrated features employing QSVM, CSVM, and MGSVM classifiers, surpassing the individual outcomes of Layer I (99.4%) and Layer II (99.3%). Similarly, ResNet-18 demonstrated notable enhancement with integrated features, attaining 99.6% accuracy with CSVM, in contrast to 99.3% and 99.2% for Layer I and Layer II, respectively. Furthermore, EfficientNet exhibited incremental improvements, achieving 99.3% accuracy with combined features utilising CSVM and MGSVM, surpassing its Layer I (99.1%) and Layer II (98.8%) accuracy. Across the classifiers, SVM variants (notably QSVM, CSVM, and MGSVM) consistently surpassed other methodologies throughout all networks. The DT classifier exhibited the greatest variability in performance, ranging from 97.8% to 99.0%, whereas LDA demonstrated relatively competent performance, though mostly inferior to the SVM variants. KNN exhibited exceptional performance, especially with the integrated features of MobileNet (99.5%). LSVM and CGSVM exhibited remarkable performance yet were marginally inferior to their more intricate SVM counterparts. The integration of attributes typically resulted in enhanced stability in performance across all classifiers, indicating that the multi-layer methodology yields more resilient feature representations for classification.

Table 4. Classification accuracy (%) of the classifiers fed with the combined Layer I and Layer II deep variables for each deep neural network.

Features	DT	LDA	KNN	LSVM	QSVM	CSVM	MGSVM	CGSVM
EfficientNet								
Layer I	97.8	98.3	98.8	98.6	99.0	99.1	99.1	98.5
Layer II	98.4	98.6	98.7	98.8	98.7	98.8	98.7	98.7
Combined	98.4	98.7	99.1	99.0	99.2	99.3	99.3	98.8
ResNet-18								
Layer I	97.5	97.4	98.7	99.1	99.3	99.2	99.2	98.9
Layer II	98.8	98.8	99.1	99.2	99.2	99.2	99.2	99.1
Combined	98.8	98.6	99.2	99.4	99.5	99.6	99.5	99.3
MobileNet								
Layer I	98.0	98.1	99.1	99.2	99.4	99.4	99.4	99.0
Layer II	98.9	99.1	99.2	99.3	99.2	99.2	99.3	99.3
Combined	99.0	98.8	99.5	99.5	99.7	99.7	99.7	99.4

5.4. Feature Selection Results

This section demonstrates the results of the ML classifiers trained with the selected deep variables after applying the ANOVA FS method to the integrated dual-layer deep variables of the three deep networks. Table 5 displays the classification accuracy outcomes following the execution of ANOVA and Chi-Squared FS on the aggregated dual-layer features from all three CNNs (MobileNet, ResNet-18, and EfficientNet), indicating substantial enhancements compared to the individual CNN results presented in Table 4. The peak classification accuracy increased from 99.7% (attained by MobileNet's integrated features with QSVM, CSVM, and MGSVM) to 99.8% through the utilisation of ANOVA-selected features, concurrently significantly diminishing feature dimensionality. The optimal performance was attained by QSVM, CSVM, and MGSVM classifiers utilising only 50 or 60 selected features, illustrating the efficacy of FS in enhancing performance. Furthermore,

while deploying only 20 features, the majority of classifiers attained remarkable accuracy exceeding 99.7%, demonstrating that ANOVA effectively discerned the most distinguishing features, whereas CSVM and CGSVM achieved 99.6% accuracy and DT and LDA.

Table 5. Classification accuracy (%) of the classifiers fed with the selected features after applying ANOVA and Chi-Squared FS on the combined dual-layer variables of the three deep networks.

Number of Features	DT	LDA	KNN	LSVM	QSVM	CSVM	MGSVM	CGSVM
				ANOVA				
10	99.0	98.7	99.2	99.5	99.4	99.4	99.5	99.4
20	99.2	99.1	99.7	99.7	99.7	99.6	99.7	99.6
30	99.2	99.1	99.7	99.7	99.7	99.7	99.7	99.7
40	99.2	99.0	99.7	99.7	99.7	99.7	99.7	99.6
50	99.2	99.0	99.7	99.7	99.7	99.8	99.7	99.6
60	99.1	99.0	99.7	99.7	99.8	99.8	99.8	99.6
70	99.1	99.1	99.7	99.7	99.8	99.8	99.8	99.6
80	99.1	99.2	99.7	99.7	99.8	99.8	99.8	99.6
90	99.1	99.1	99.6	99.7	99.8	99.8	99.8	99.6
100	99.1	99.1	99.7	99.7	99.8	99.8	99.8	99.6
				Chi-Squared				
10	97.7	98.0	98.3	98.3	98.5	98.4	98.5	98.3
20	98.5	98.7	98.9	98.8	99.0	99.1	99.1	98.9
30	98.7	98.8	99.2	99.2	99.3	99.4	99.4	99.0
40	98.6	98.8	99.2	99.4	99.4	99.5	99.4	99.2
50	98.6	98.8	99.4	99.3	99.5	99.5	99.4	99.2
60	98.6	98.8	99.4	99.3	99.4	99.5	99.4	99.2
70	98.9	98.9	99.5	99.5	99.6	99.7	99.6	99.4
80	99.0	99.0	99.6	99.6	99.7	99.7	99.7	99.5
90	98.9	99.0	99.7	99.6	99.7	99.8	99.7	99.5
100	99.1	99.1	99.7	99.7	99.8	99.8	99.8	99.6

Additionally, the enhancement via ANOVA FS was especially significant across all classifiers in comparison to the individual CNN outcomes presented in Table 4. The DT classifier enhanced from 99.0% (the optimal performance of MobileNet) to 99.2%, KNN from 99.5% to 99.7%, and LDA from 98.8% to 99.2%. The LSVM exhibited an enhancement from 99.5% to 99.7%, sustaining this performance with few features, whereas the CGSVM attained a maximum of 99.7%, surpassing its prior best of 99.4%. The performance peak noted after 20 features indicates that this may be the optimal feature set size, as the inclusion of additional features did not produce substantial enhancements. On the other hand, all classifiers produced somewhat smaller classification accuracies when using the Chi-Squared FS technique. The CSVM attained a peak accuracy of 99.5% using Chi-Squared with feature sizes of 50–60, in contrast to 99.8% with ANOVA. Comparable patterns were noted for classifiers such as DT, LDA, and KNN, wherein Chi-Squared attained accuracies 0.2–0.5% inferior to ANOVA throughout the majority of feature numbers. These distinctions highlight ANOVA's superiority in improving classifier efficacy by concentrating on features with greater discriminative strength.

The findings also demonstrate the influence of feature count on classification accuracy. Both ANOVA and Chi-Squared demonstrated consistent or marginally enhanced accuracy

trends as the quantity of selected features escalated from 10 to 100. Nonetheless, ANOVA optimised this enhancement more efficiently, consistently attaining superior accuracies across all classifiers. This suggests that although both FS techniques gain from expanded feature sets, ANOVA optimises the use of the supplementary information more effectively.

More performance metrics are calculated for the eight classifiers fed with the optimal set of features. These measures involve sensitivity, precision, F1-score, MCC, and specificity. Table 6 delineates the assessment measures for the eight classifiers trained with optimally chosen attributes subsequent to the deployment of the ANOVA FS technique on the integrated dual-layer deep variables of the three CNN architectures (EfficientNet, ResNet-18, MobileNet). The findings indicate that CSVM achieved the highest performance, exhibiting a sensitivity of 0.9978, specificity of 0.9995, precision of 0.9978, F1-score of 0.9978, and MCC of 0.9973. QSVM and MGSVM exhibited outstanding performance, attaining a sensitivity of 0.9975 and 0.9976, specificity of 0.9994, precision of 0.9975 and 0.9976, F1-score of 0.9975 and 0.9976, and MCC of 0.9969 and 0.9970, respectively. KNN and LSVM demonstrated robust performance, with KNN realising a sensitivity of 0.9968, specificity of 0.9992, precision of 0.9968, F1-score of 0.9968, and MCC of 0.9960. LSVM exhibited comparable metrics, with a sensitivity of 0.9967, specificity of 0.9992, precision of 0.9967, F1-score of 0.9967, and MCC of 0.9959. The results highlight the efficacy of the ANOVA feature selection method in improving classification performance across all classifiers, with CSVM attaining the highest overall performance. The findings illustrate the accuracy and reliability of the suggested CAD framework to identify LC cancer.

Table 6. Evaluation metrics of the classifiers fed with the optimal selected features after applying the ANOVA feature selection.

| Classifier	Sensitivity	Specificity	Precision	F1-Score	MCC
DT	0.9916	0.9979	0.9916	0.9916	0.9896
LDA	0.9911	0.9978	0.9911	0.9911	0.9889
KNN	0.9968	0.9992	0.9968	0.9968	0.9960
LSVM	0.9967	0.9992	0.9967	0.9967	0.9959
QSVM	0.9975	0.9994	0.9975	0.9975	0.9969
CSVM	0.9978	0.9995	0.9978	0.9978	0.9973
MGSVM	0.9976	0.9994	0.9976	0.9976	0.9970
CQSVM	0.9965	0.9991	0.9965	0.9965	0.9957

The confusion matrices of the highest-performing classifiers including QSVM, CSVM, and MGSVM are also determined. Figure 3 presents confusion matrices illustrating the proportions of the right and wrong classifications for each class. The confusion matrices indicate that the categories of colon adenocarcinoma, colon benign tissue, and lung benign tissue have been accurately identified with sensitivity values equal to 100% using CSVM, QSVM, and MGSVM algorithms. The lung squamous cancer subtype is the most frequently incorrectly categorised among these three classifiers. Figure 3 illustrates that QSVM is the classifier with the highest rate of incorrect classifications among the three classifiers for photos in the lung squamous category.

In contrast to these metrics, ROC curves are also illustrated in Figure 4 for QSVM, CSVM, and MGSVM, which attained the highest performance. This curve is generated by graphing the true positive rate in relation to the false positive rate. For effective classification, it is preferable for the Area Under the Curve (AUC) to approach 1 [56]. Figure 4 illustrates that the AUC value is determined to be 1 for all curves. The findings indicate that

the suggested CAD offers a cost-effective, impartial, and considerably precise identification of LC cancer.
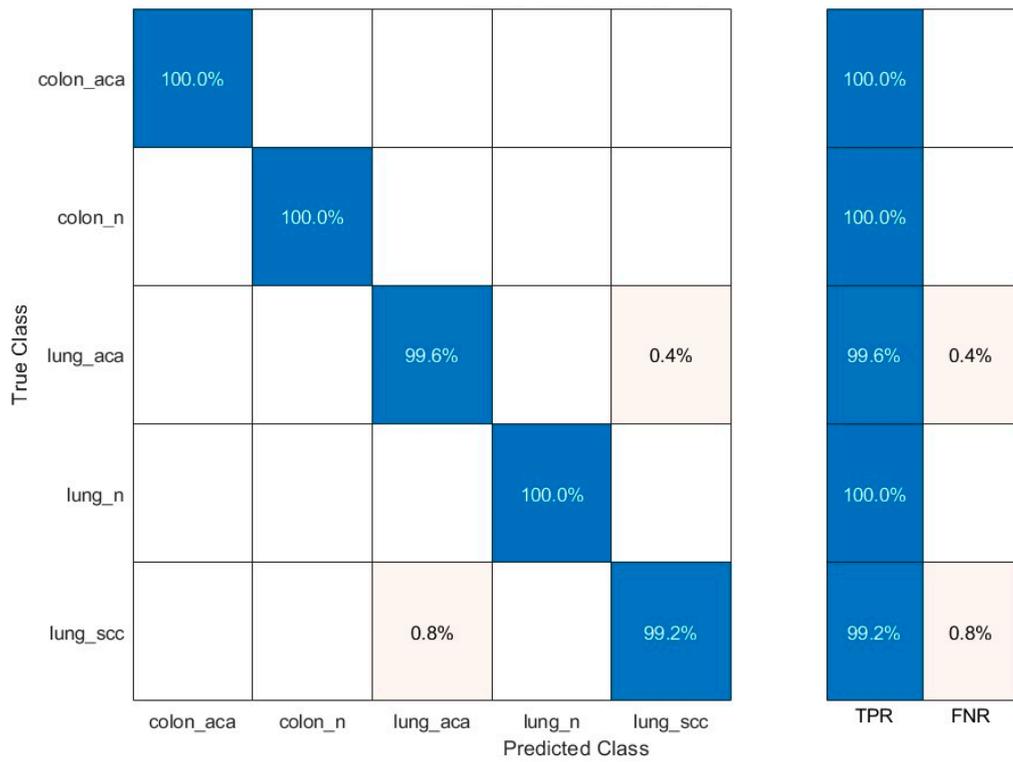
(**a**) QSVM
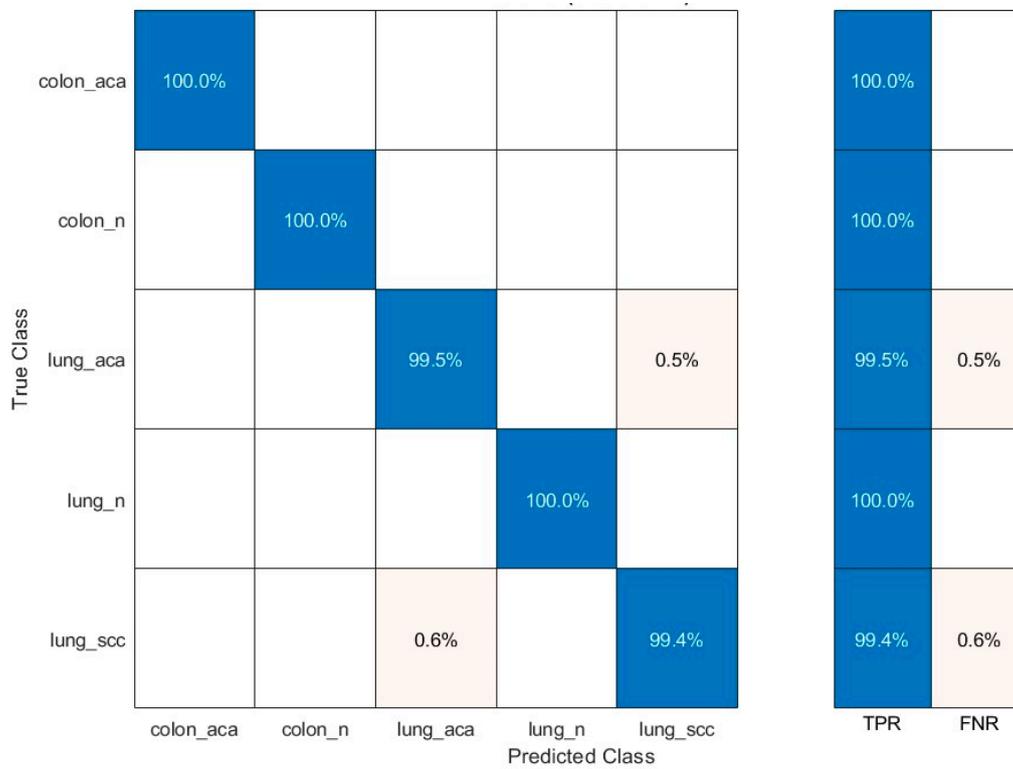
(**b**) CSVM

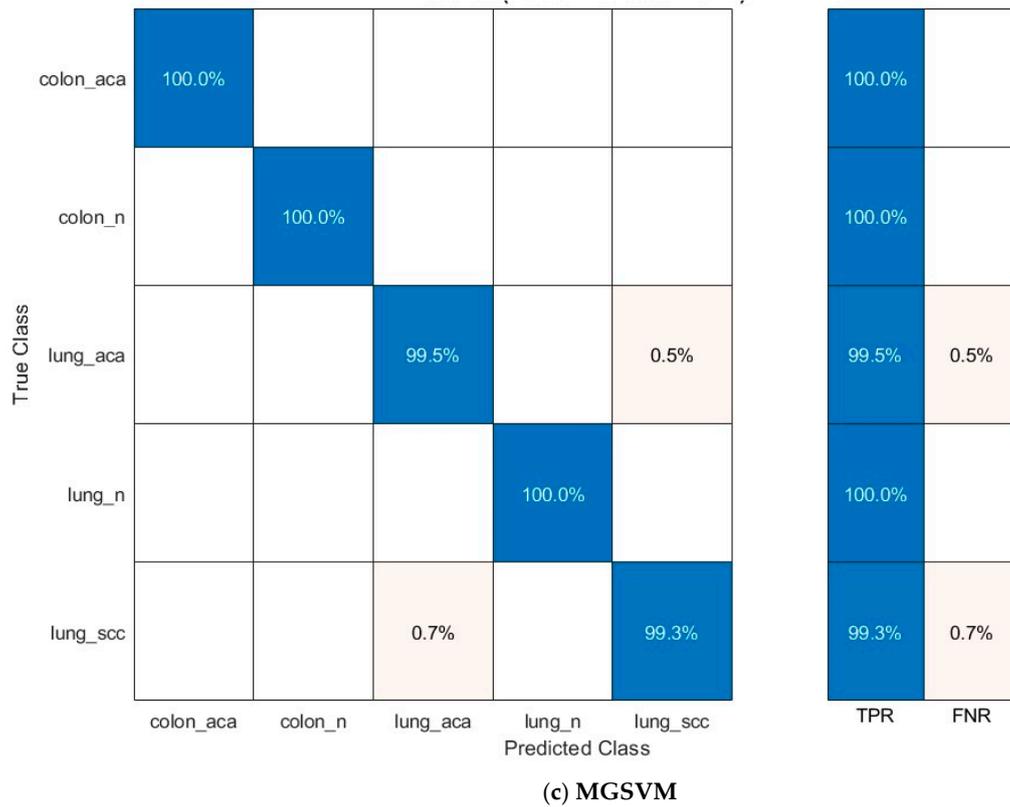**Figure 3.** *Cont.*

(**c**) **MGSVM**

Figure 3. Confusion matrices of the SVM classifiers fed with the optimal selected deep variables after applying the ANOVA method on the combined dual-layer features of the three networks; (**a**) QSVM, (**b**) CSVM, (**c**) MGSVM.
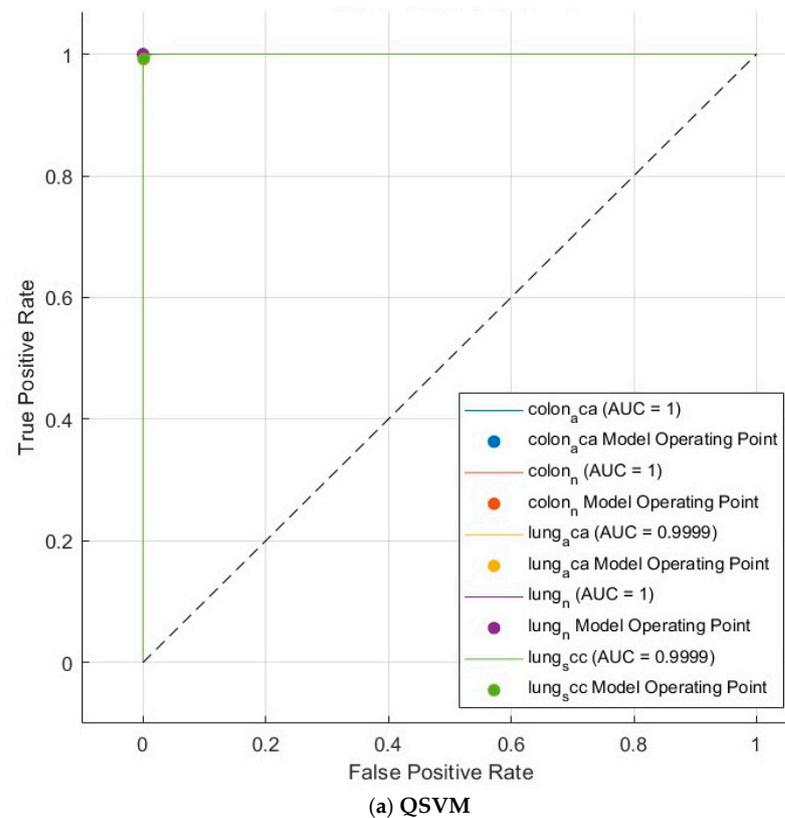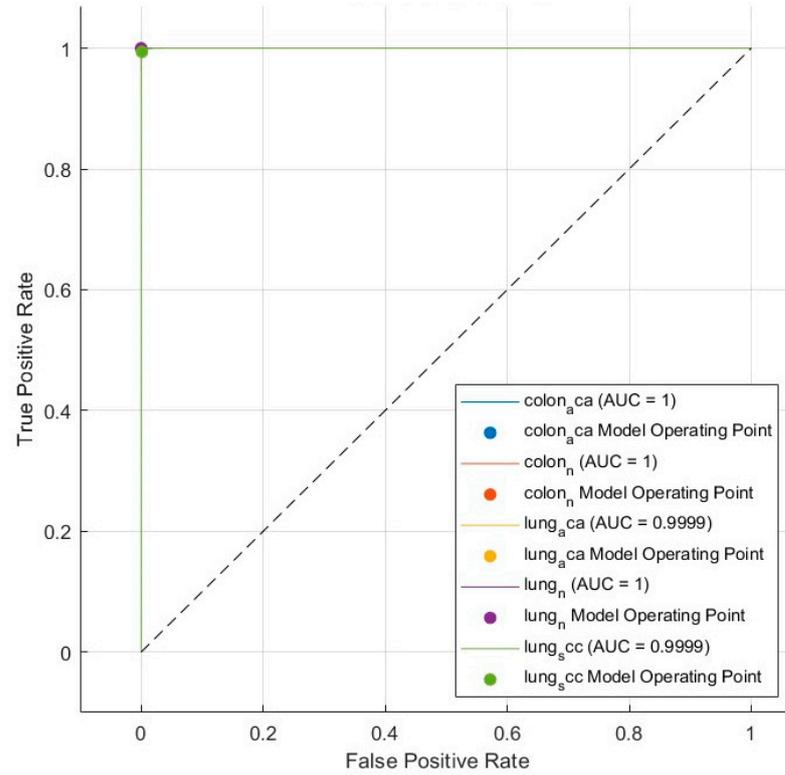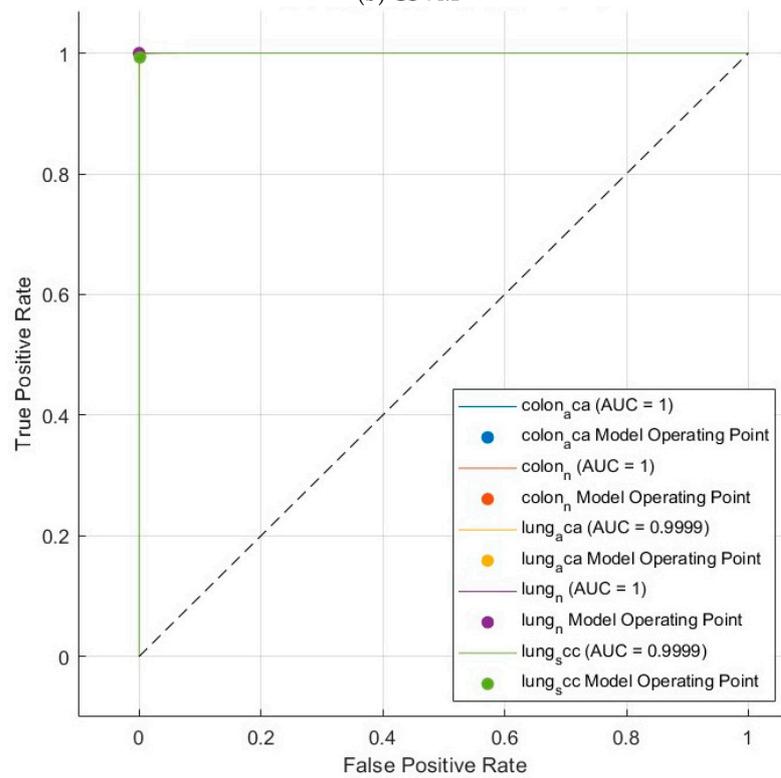

(**a**) QSVM

Figure 4. *Cont.*

(**b**) **CSVM**

(**c**) **MGSVM**

Figure 4. The ROC curves of the SVM classifiers inputted using the optimal selected deep variables after applying the ANOVA method on the combined dual-layer features of the three networks; (**a**) QSVM, (**b**) CSVM, (**c**) MGSVM.

6. Discussion

The experimental findings illustrate the efficacy of the proposed multi-CNN framework, which incorporates dual-layer feature extraction, reduction, and selection for identifying different types of LC cancers. The examination of distinct layers and their integration, along with FS, uncovers numerous critical insights regarding classification efficacy. Preliminary experiments with Layer II variables (Table 2) demonstrated encouraging outcomes, with MobileNet attaining 99.3% accuracy using LSVM and MGSVM classifiers. The ablation study on Layer I features (Table 3) demonstrated that expanding the quantity of canonical variables typically enhanced classification performance, with MobileNet achieving the highest accuracy score of 99.4% employing QSVM, CSVM, and MGSVM classifiers when employing 50 canonical variables. This indicates that the elevated attributes from both layers possess comparable discriminatory capabilities.

The integration of Layer I and Layer II attributes (Table 4) illustrated the efficacy of multiscale feature representation, with MobileNet's aggregated features attaining 99.7% accuracy leveraging QSVM, CSVM, and MGSVM classifiers. This significant enhancement compared to single-layer performance (99.4% for Layer I and 99.3% for Layer II) substantiates the assumption that attributes from various network depths encapsulate complementary information beneficial for classification. ResNet-18 demonstrated comparable advantages from feature incorporation, attaining 99.6% accuracy with CSVM, whereas EfficientNet achieved 99.3% using both CSVM and MGSVM.

Notable performance improvements were evident following the deployment of ANOVA feature selection on the aggregated dual-layer features from all three CNNs (Table 5). This method attained an overall accuracy of 99.8%, utilising QSVM, CSVM, and MGSVM classifiers with only 50 and 60 selected features. Remarkably, despite employing merely 20 selected features, the system achieved outstanding performance (99.7%), illustrating the efficacy of ANOVA in pinpointing the most discriminating attributes while significantly diminishing computational complexity.

Among the evaluated classifiers, SVM variants, especially the QSVM, CSVM, and MGSVM, consistently exhibited outstanding results owing to their capacity to adeptly manage nonlinear relationships within the data. Distinct patterns emerged during the assessment. The classifiers employing diminished feature sets derived from CCA and ANOVA demonstrated enhanced computational efficiency and generalisation, with negligible loss in accuracy. The QSVM, CSVM, and MGSVM classifiers performed exceptionally well in situations where the dimensionality of the feature vectors was markedly diminished, as it could take advantage of the most distinctive attributes for optimal classification. This pattern highlights the significance of effective feature selection in optimising classifier performance while considering the trade-offs between accuracy and computational expense. These results demonstrate the versatility of SVMs in accommodating diverse feature sets and their appropriateness for resource-limited contexts.

The outcomes of the proposed CAD underline numerous pivotal findings: (1) the beneficial nature of features across various network layers and architectures, (2) the efficacy of FS in sustaining or enhancing performance while minimising dimensionality, and (3) the improved accuracy of SVM variants, especially with nonlinear kernels, in managing the intricate feature spaces generated by deep networks. The uniform enhancement observed in all classifiers exploiting the chosen attributes indicates that the suggested system effectively harnesses the advantages of various CNN structures while reducing redundant data. The performance peak noted after 20 chosen attributes signifies an optimal equilibrium among feature dimension and classification accuracy, implying that additional feature inclusion may not produce substantial enhancements. This discovery has significant

implications for practical applications, enabling efficient implementation while preserving high diagnostic accuracy.

A complexity analysis has been conducted to demonstrate the computational efficiency of the proposed CAD. As shown in Table 7, by selecting merely 50 features through sophisticated feature selection technique, our model substantially diminishes the parameter count in comparison to complex deep learning models. It includes compact CNN structures such as ResNet-18 (11 million parameters) and EfficientNetB0 (5.3 million parameters). The decreased number of parameters of such deep models, combined with a reduced number of layers (82 layers for EfficientNetB0, 28 Layers for MobileNet, and 18 for ResNet-18), leads to diminished memory and storage requirements. The proposed system facilitates quicker inference times which is especially beneficial for real-time applications. The computational complexity of the classification phase of the proposed model, denoted as $O(m^2p + m^3)$, illustrates its efficiency, particularly in systems with constrained computational resources. This is because the classification complexity signifies that the computational expense correlates with the quantity of features and samples yet stays feasible owing to the diminished feature set (50 features). This design guarantees adaptability as well as effectiveness in situations with constrained computing resources, guaranteeing expedited inference times without sacrificing performance.

Table 7. A complexity study of the proposed CAD.

Model	Input Data Size to Classifier	Amount of Deep Network Parameters	Amount of Layers	Classification Complexity (O)
ResNet-18	$224 \times 224 \times 3$	11 M	18	$O(k \cdot n \cdot d^2)$ [57]
MobileNet	$224 \times 224 \times 3$	3.5M	28	k: kernel length n: the overall length of the pattern (the amount of input entries)
EfficientNetB0	$224 \times 224 \times 3$	5.3 M	82	d: dimensionality of presentation
Classification Phase of the CAD	50 Attributes	**CSV**M C: regularisation constraint *Gamma:* kernel's length	-	$O(m^2p + m^3)$ [58] p: number of features m: number of input samples

6.1. State-of-the-Art Comparisons

Table 8 delineates a comprehensive comparison comparing the suggested CAD system and various cutting-edge frameworks for identifying different types of LC cancer employing the LC25000 dataset. The suggested framework integrates ResNet-18, MobileNet, and EfficientNetB0 with CCA for dimensionality reduction and ANOVA for FS, attaining an accuracy of 99.78%, sensitivity of 99.78%, specificity of 99.95%, precision of 99.78%, and an F1-score of 99.78%. This performance illustrates the system's resilience and efficacy in classification tasks. Compared to current CAD systems, EfficientNet with GradCAM [1] attains the highest accuracy and sensitivity at 99.97%, marginally exceeding the proposed system. This approach does not incorporate the FS and reduction techniques employed in the suggested CAD system, which diminishes the feature set to merely 50 attributes, thereby substantially diminishing computational requirements. Likewise, EfficientNet with AdBet-WOA [2] attains an accuracy of 99.96% and a specificity of 99.96%, employing 445 features. The dependence on an extensive feature set, although effective, heightens complexity in contrast to the streamlined nature of the proposed system. Although the proposed system's performance on specific metrics is marginally inferior to the method cited in reference [2], these improvements render the proposed approach more versatile and

efficient, especially in contexts with constrained computational resources. The proposed system utilises merely 50 features, whereas reference [2] employs 445 features to attain a comparable level of performance. These factors offer a significant trade-off and a crucial contribution to the domain of CAD systems.

Table 8. State of art comparisons with existing CAD frameworks for LC cancer identification based on the LC25000 dataset.

Article	Methods	Feature Selection/ Reduction	Accuracy	Sensitivity	Specificity	Precision	F1-Score
[34]	CLAHE+ MobileNet + DBN	No	0.9927	0.9817		0.9818	0.9817
[7]	VGG16 + LBP + Majority Voting (SVM + RF + LR)	No	0.9900	0.9880		0.9900	0.9880
[35]	ResNet + EfficientNetB0 + GWO + Soft Voting (LR + SVM + DT + KNN + RF)	Yes	0.9873	0.9873		0.9873	0.9873
[1]	EfficientNet Large + GradCAM	No	0.9997				0.9997
[2]	EfficientNet + AdBet-WOA	Yes (445 features)	0.9996	0.9997		0.9996	0.9996
[36]	Customized CNN + GradCAM and SHAP	No	0.9920	0.9936		0.9916	0.9916
[37]	CLAHE + VGG16	No	0.9896				
[38]	ColonNet + GLPP	No	0.9631	0.9567	0.9497	0.9611	0.9488
[39]	ShuffleNet + DCRNN + BER + COA	No	0.9922	0.9806		0.9807	0.9806
[40]	VGG19 +PCA + Handcrafted Features (DWT, LBP, FCH, GLCM)	Yes (699 features)	0.9964	0.9985	1.000	1.00	
[41]	ResNet-50, InceptionV3, DenseNet + KELM	No	0.9900	0.9650	0.9670	0.9770	0.9820
[42]	MobileNet + ShuffleNet + SqueezeNet + FWHT + DWT +SVM	Yes (510 features)	0.9960	0.9960	0.9990	0.9960	0.9960
[43]	InceptionResNet + LBP + SHAP	No	0.9988	0.9942		0.9946	0.9976
Presented	ResNet-18 + MobileNet + EfficinetB0 + CCA + ANOVA + CSVM	Yes (50)	0.9978	0.9978	0.9995	0.9978	0.9978

The VGG19 integrated with PCA and manually crafted features (e.g., DWT, LBP, GLCM) [40] accomplishes an accuracy of 99.64%, sensitivity of 99.85%, and flawless specificity and precision of 100%. This approach depends on manually crafted features, which can be labour-intensive and less flexible for varied datasets than DL methods. Inception-ResNet, utilising LBP and SHAP [43], realises an accuracy of 99.88%, sensitivity of 99.42%, and specificity of 99.46%, demonstrating robust performance, albeit with diminished specificity relative to the proposed method. Alternative frameworks, including MobileNet and DBN (99.27% accuracy) [34] and ShuffleNet with DCRNN + BER (99.22% accuracy) [39], demonstrate adequate performance but are inferior to the proposed system regarding sensitivity, specificity, and F1-score. Moreover, approaches leveraging multiple classifiers, such as VGG16 + LBP with majority voting (accuracy 99.00%) [7] or ResNet + EfficientNetB0 with soft voting (accuracy 98.73%) [35], demonstrate inferior overall results relative to the proposed system.

The main benefit of the presented CAD system is the combination of multiscale features from different CNN structures (ResNet-18, MobileNet, and EfficientNetB0), along with effective feature reduction and selection methods (CCA and ANOVA), resulting in a concise and informative feature set comprising 50 attributes. This yields a lightweight, computationally efficient model whose performance matches or surpasses that of most current frameworks. The system attains an exceptional specificity of 99.95%, thereby reducing false positives, which is crucial in medical diagnostics to prevent unneeded treatments. The implementation of compact CNNs guarantees adaptability in resource-limited settings, rendering the proposed system a viable solution for practical medical image analysis tasks.

The research study emphasises computational effectiveness via feature dimensionality reduction, resulting in a minor compromise in classification accuracy. Lowering the feature set to 50 attributes yielded an accuracy around 0.002 less than the approach described in reference [2], which employs 445 features. Nonetheless, this trade-off is negligible and is surpassed by the substantial enhancement in computational effectiveness and the diminished risk of overfitting. Furthermore, the proposed CAD achieved higher performance with only 50 features than other studies [40,42] which employed 699 and 510 features, respectively. The chosen 50 features were refined to preserve the most distinguishing information, guaranteeing that the system sustains competitive performance while being appropriately adapted for implementation in resource-limited settings. This equilibrium demonstrates the significance of the methodology for practical applications constrained by computational resources.

6.2. Constraints and Upcoming Areas of Improvement

Considering the promising findings of the presented CAD system, various constraints and prospective avenues for future research must be admitted. Initially, while the system exhibited outstanding performance on the LC25000 dataset, additional validation on varied, multi-centre datasets is necessary to confirm its generalisability across different patient populations and imaging protocols. The present study was confined to histopathological photos, and the system's efficacy for alternative medical imaging modalities, such as CT scans or MRI, has yet to be investigated. A further limitation is that although the ANOVA-based FS strategy proved effective, alternative approaches, such as mutual information or correlation-based methodologies, could possibly improve performance. Moreover, the present system comprises distinct processing phases for feature extraction, dimensionality reduction, and FS. Upcoming efforts may concentrate on creating a unified, comprehensive trainable architecture to optimise these procedures. Additionally, investigating the interpretability and biological relevance of the chosen features is crucial, as this may yield significant insights for clinical applications. Enhancing the framework to accommodate multi-modal data integration, encompassing clinical metadata, involving different patient populations, and establishing real-time processing capabilities for medical decision support represent additional promising avenues for future research. Integrating explainable AI methodologies may enhance the system's clinical acceptability by providing interpretable results for healthcare professionals. Furthermore, expanding the framework to encompass additional cancer types and medical conditions could enhance its clinical relevance and influence. Future clinical trials would be crucial in evaluating the system's influence on clinical decision-making and patient outcomes in practical environments.

7. Conclusions

This research introduced an innovative CAD system for identifying types of LC cancer, leveraging numerous compact CNNs with dual-layer feature extraction, reduction, and selection. The suggested framework harnesses the complementary advantages of MobileNet, ResNet-18, and EfficientNetB0 structures, employing variables from both pooling and fully connected layers to obtain multiscale representations. The experimental findings evidenced the efficacy of the proposed methodology, attaining an exceptional classification accuracy of 99.8% with QSVM, CSVM, and MGSVM classifiers employing ANOVA-selected attributes, surpassing the individual CNN performances of 99.7% (MobileNet), 99.6% (ResNet-18), and 99.3% (EfficientNet). This performance was sustained with only 50 chosen attributes, emphasising the efficacy of the feature selection methodology. The systematic assessment of feature combination and selection techniques demonstrated that the integration of multi-layer attributes from multiple CNN structures, coupled with feature selection, yields enhanced classification performance relative to single-network or single-layer methods. The framework's capacity to attain high accuracy with a diminished feature set renders it especially appropriate for medical applications where diagnostic precision and computational efficiency are essential. Although the suggested framework shows promise in creating a powerful and effective CAD system for the identification of colon and lung cancer, some drawbacks should be noted. Since the study mainly assesses the framework's effectiveness on particular datasets, one possible drawback is its generalisability to other datasets. Future research may concentrate on verifying the methodology across varied datasets to strengthen its resilience. Furthermore, while the framework leverages compact CNN designs to diminish computational complexity, deploying the system in resource-limited environments may still present difficulties. Examining these factors in subsequent research could enhance the system's relevance and efficacy in practical clinical settings.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data utilised in the current research can be found at https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images (accessed on 15 September 2022).

Conflicts of Interest: The author declares no conflicts of interests.

References

1. Tummala, S.; Kadry, S.; Nadeem, A.; Rauf, H.T.; Gul, N. An explainable classification method based on complex scaling in histopathology images for lung and colon cancer. *Diagnostics* **2023**, *13*, 1594. [CrossRef]
2. Bhattacharya, A.; Saha, B.; Chattopadhyay, S.; Sarkar, R. Deep feature selection using adaptive β-Hill Climbing aided whale optimization algorithm for lung and colon cancer detection. *Biomed. Signal Process. Control* **2023**, *83*, 104692. [CrossRef]
3. Provath, M.A.-M.; Deb, K.; Dhar, P.K.; Shimamura, T. Classification of lung and colon cancer histopathological images using global context attention based convolutional neural network. *IEEE Access* **2023**, *11*, 110164–110183. [CrossRef]
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.* **2018**, *8*, 137–140. [CrossRef]
5. Sanchez-Peralta, L.F.; Bote-Curiel, L.; Picon, A.; Sanchez-Margallo, F.M.; Pagador, J.B. Deep learning to find colorectal polyps in colonoscopy: A systematic literature review. *Artif. Intell. Med.* **2020**, *108*, 101923. [CrossRef]
6. Adu, K.; Yu, Y.; Cai, J.; Owusu-Agyemang, K.; Twumasi, B.A.; Wang, X. DHS-CapsNet: Dual horizontal squash capsule networks for lung and colon cancer classification from whole slide histopathological images. *Int. J. Imaging Syst. Technol.* **2021**, *31*, 2075–2092. [CrossRef]
7. 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.* **2023**, *15*, 4149–4160. [CrossRef]

8. 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. [CrossRef]

9. Li, M.; Ma, X.; Chen, C.; Yuan, Y.; Zhang, S.; Yan, Z.; Chen, C.; Chen, F.; Bai, Y.; Zhou, P. Research on the auxiliary classification and diagnosis of lung cancer subtypes based on histopathological images. *IEEE Access* **2021**, *9*, 53687–53707. [CrossRef]

10. Pacal, I.; Karaboga, D.; Basturk, A.; Akay, B.; Nalbantoglu, U. A comprehensive review of deep learning in colon cancer. *Comput. Biol. Med.* **2020**, *126*, 104003. [CrossRef]

11. Attallah, O.; Zaghlool, S. AI-Based Pipeline for Classifying Pediatric Medulloblastoma Using Histopathological and Textural Images. *Life* **2022**, *12*, 232. [CrossRef]

12. Hamamoto, R.; Suvarna, K.; Yamada, M.; Kobayashi, K.; Shinkai, N.; Miyake, M.; Takahashi, M.; Jinnai, S.; Shimoyama, R.; Sakai, A. Application of Artificial Intelligence Technology in Oncology: Towards the Establishment of Precision Medicine. *Cancers* **2020**, *12*, 3532. [CrossRef]

13. Attallah, O. Acute lymphocytic leukemia detection and subtype classification via extended wavelet pooling based-CNNs and statistical-texture features. *Image Vis. Comput.* **2024**, *147*, 105064. [CrossRef]

14. Attallah, O. ADHD-AID: Aiding Tool for Detecting Children's Attention Deficit Hyperactivity Disorder via EEG-Based Multi-Resolution Analysis and Feature Selection. *Biomimetics* **2024**, *9*, 188. [CrossRef]

15. Karaman, A.; Karaboga, D.; Pacal, I.; Akay, B.; Basturk, A.; Nalbantoglu, U.; Coskun, S.; Sahin, O. Hyper-parameter optimization of deep learning architectures using artificial bee colony (ABC) algorithm for high performance real-time automatic colorectal cancer (CRC) polyp detection. *Appl. Intell.* **2023**, *53*, 15603–15620. [CrossRef]

16. Bi, W.L.; Hosny, A.; Schabath, M.B.; Giger, M.L.; Birkbak, N.J.; Mehrtash, A.; Allison, T.; Arnaout, O.; Abbosh, C.; Dunn, I.F. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J. Clin.* **2019**, *69*, 127–157. [CrossRef]

17. Kohli, M.; Prevedello, L.M.; Filice, R.W.; Geis, J.R. Implementing machine learning in radiology practice and research. *Am. J. Roentgenol.* **2017**, *208*, 754–760. [CrossRef]

18. Attallah, O.; Samir, A. A wavelet-based deep learning pipeline for efficient COVID-19 diagnosis via CT slices. *Appl. Soft Comput.* **2022**, *128*, 109401. [CrossRef]

19. Pacal, I. MaxCerVixT: A novel lightweight vision transformer-based Approach for precise cervical cancer detection. *Knowl.-Based Syst.* **2024**, *289*, 111482. [CrossRef]

20. Attallah, O. CerCan· Net: Cervical cancer classification model via multi-layer feature ensembles of lightweight CNNs and transfer learning. *Expert Syst. Appl.* **2023**, *229*, 120624. [CrossRef]

21. Attallah, O. MB-AI-His: Histopathological diagnosis of pediatric medulloblastoma and its subtypes via AI. *Diagnostics* **2021**, *11*, 359. [CrossRef]

22. Sharkas, M.; Attallah, O. Color-CADx: A deep learning approach for colorectal cancer classification through triple convolutional neural networks and discrete cosine transform. *Sci. Rep.* **2024**, *14*, 6914. [CrossRef]

23. Lee, J.-G.; Jun, S.; Cho, Y.-W.; Lee, H.; Kim, G.B.; Seo, J.B.; Kim, N. Deep learning in medical imaging: General overview. *Korean J. Radiol.* **2017**, *18*, 570–584. [CrossRef]

24. Attallah, O. A Hybrid Trio-Deep Feature Fusion Model for Improved Skin Cancer Classification: Merging Dermoscopic and DCT Images. *Technologies* **2024**, *12*, 190. [CrossRef]

25. Coudray, N.; Ocampo, P.S.; Sakellaropoulos, T.; Narula, N.; Snuderl, M.; Fenyö, D.; Moreira, A.L.; Razavian, N.; Tsirigos, A. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nat. Med.* **2018**, *24*, 1559–1567. [CrossRef]

26. Chhillar, I.; Singh, A. A feature engineering-based machine learning technique to detect and classify lung and colon cancer from histopathological images. *Med. Biol. Eng. Comput.* **2024**, *62*, 913–924. [CrossRef]

27. Attallah, O. Cervical cancer diagnosis based on multi-domain features using deep learning enhanced by handcrafted descriptors. *Appl. Sci.* **2023**, *13*, 1916. [CrossRef]

28. Aslan, M.F.; Sabanci, K.; Durdu, A. A CNN-based novel solution for determining the survival status of heart failure patients with clinical record data: Numeric to image. *Biomed. Signal Process. Control* **2021**, *68*, 102716. [CrossRef]

29. Anwar, S.M.; Majid, M.; Qayyum, A.; Awais, M.; Alnowami, M.; Khan, M.K. Medical image analysis using convolutional neural networks: A review. *J. Med. Syst.* **2018**, *42*, 226. [CrossRef]

30. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [CrossRef]

31. Xu, Q.; Zhang, M.; Gu, Z.; Pan, G. Overfitting remedy by sparsifying regularization on fully-connected layers of CNNs. *Neurocomputing* **2019**, *328*, 69–74. [CrossRef]

32. 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. [CrossRef]

33. 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**, *55*, 115–140. [CrossRef]

34. 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* **2023**, *15*, 1591. [CrossRef]

35. Ijaz, M.; Ashraf, I.; Zahid, U.; Yasin, A.; Ali, S.; Attique Khan, M.; Alqahtani, S.A.; Zhang, Y.-D. A Decision Support System for Lung Colon Cancer Classification using Fusion of Deep Neural Networks and Normal Distribution based Gray Wolf Optimization. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2023**. [CrossRef]

36. Hasan, M.A.; Haque, F.; Sabuj, S.R.; Sarker, H.; Goni, M.O.F.; Rahman, F.; Rashid, M.M. An End-to-End Lightweight Multi-Scale CNN for the Classification of Lung and Colon Cancer with XAI Integration. *Technologies* **2024**, *12*, 56. [CrossRef]

37. Hadiyoso, S.; Aulia, S.; Irawati, I.D. Diagnosis of lung and colon cancer based on clinical pathology images using convolutional neural network and CLAHE framework. *Int. J. Appl. Sci. Eng.* **2023**, *20*, 1–7. [CrossRef]

38. Iqbal, S.; Qureshi, A.N.; Alhussein, M.; Aurangzeb, K.; Kadry, S. A novel Heteromorphous convolutional neural network for automated assessment of tumors in colon and lung histopathology images. *Biomimetics* **2023**, *8*, 370. [CrossRef]

39. AlGhamdi, R.; Asar, T.O.; Assiri, F.Y.; Mansouri, R.A.; Ragab, M. Al-biruni Earth radius optimization with transfer learning based histopathological image analysis for lung and colon cancer detection. *Cancers* **2023**, *15*, 3300. [CrossRef]

40. Al-Jabbar, M.; Alshahrani, M.; Senan, E.M.; Ahmed, I.A. Histopathological Analysis for Detecting Lung and Colon Cancer Malignancies Using Hybrid Systems with Fused Features. *Bioengineering* **2023**, *10*, 383. [CrossRef]

41. Gowthamy, J.; Ramesh, S. A novel hybrid model for lung and colon cancer detection using pre-trained deep learning and KELM. *Expert Syst. Appl.* **2024**, *252*, 124114. [CrossRef]

42. Attallah, O.; Aslan, M.F.; Sabanci, K. A framework for lung and colon cancer diagnosis via lightweight deep learning models and transformation methods. *Diagnostics* **2022**, *12*, 2926. [CrossRef]

43. Alsubai, S. Transfer learning based approach for lung and colon cancer detection using local binary pattern features and explainable artificial intelligence (AI) techniques. *PeerJ Comput. Sci.* **2024**, *10*, e1996. [CrossRef]

44. Borkowski, A.A.; Bui, M.M.; Thomas, L.B.; Wilson, C.P.; DeLand, L.A.; Mastorides, S.M. Lung and colon cancer histopathological image dataset (lc25000). *arXiv* **2019**, arXiv:1912.12142.

45. Hardoon, D.R.; Szedmak, S.; Shawe-Taylor, J. Canonical correlation analysis: An overview with application to learning methods. *Neural Comput.* **2004**, *16*, 2639–2664. [CrossRef]

46. Fan, X.; Konold, T.R. Canonical correlation analysis. In *The Reviewer's Guide to Quantitative Methods in the Social Sciences*; Routledge: London, UK, 2018; pp. 29–41.

47. Howard, A.G. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.

48. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 15–19 June 2019; pp. 6105–6114.

49. 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.

50. Pan, Z.; Wang, J.; Shen, Z.; Chen, X.; Li, M. Multi-layer convolutional features concatenation with semantic feature selector for vein recognition. *IEEE Access* **2019**, *7*, 90608–90619. [CrossRef]

51. Attallah, O. Skin-CAD: Explainable deep learning classification of skin cancer from dermoscopic images by feature selection of dual high-level CNNs features and transfer learning. *Comput. Biol. Med.* **2024**, *178*, 108798. [CrossRef]

52. Goodfellow, I. Deep Learning. In *Healthcare Informatics Research*; The MIT Press: Cambridge, MA, USA, 2016. [CrossRef]

53. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Into Imaging* **2018**, *9*, 611–629. [CrossRef]

54. Attallah, O. ECG-BiCoNet: An ECG-based pipeline for COVID-19 diagnosis using Bi-Layers of deep features integration. *Comput. Biol. Med.* **2022**, *142*, 105210. [CrossRef]

55. Liu, H.; Setiono, R. Chi2: Feature selection and discretization of numeric attributes. In Proceedings of the 7th IEEE International Conference on Tools with Artificial Intelligence, Herndon, VA, USA, 5–8 November 1995; pp. 388–391.

56. Aslan, M.F. A robust semantic lung segmentation study for CNN-based COVID-19 diagnosis. *Chemom. Intell. Lab. Syst.* **2022**, *231*, 104695. [CrossRef]

57. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5999–6009.

58. Younes, H.; Alameh, M.; Ibrahim, A.; Rizk, M.; Valle, M. Efficient algorithms for embedded tactile data processing. In *Electronic Skin*; Taylor and Francis: Oxfordshire, UK, 2022; pp. 113–138.