Preprints.org

Article

Leveraging the Novel MSHA Model: A Focus on Adrenocortical Carcinoma

Mubarak Taiwo Mustapha [*] , Berna Uzun , Dilber Uzun Ozsahin

Article

Leveraging the Novel MSHA Model: A Focus on Adrenocortical Carcinoma

Mubarak Taiwo Mustapha [1,*], Berna Uzun [1,2] and Dilber Uzun Ozsahin [1,3,4,]

[1]   Operational Research Centre in Healthcare, Near East University, TRNC Mersin 10, Nicosia, 99138, Turkey
[2]   Department of Mathematics, Near East University, TRNC Mersin 10, Nicosia, 99138, Turkey
[3]   Department of Medical Diagnostic Imaging, College of Health Science, University of Sharjah, Sharjah, UAE
[4]   Research Institute for Medical and Health Sciences, University of Sharjah, Sharjah, UAE
*   Correspondence: mubarak.mustapha@neu.edu.tr

Highlight:

- Development of a novel model incorporating mixed-scale dense convolution, self-attention mechanism, hierarchical feature fusion, and attention-based contextual information techniques.
- Implementation of the model for ACC detection using CT images.
- MSHA model aids surgical planning and postoperative monitoring for ACC.
- Performance evaluation of the MSHA model with ResNet50, VGG16, VGG19, and inceptionV3.

Abstract: This study aims to explore the utilization of deep learning models, specifically the innovative Multi-Modal Contextual Fusion Convolutional Neural Network (MSHA), for the detection and diagnosis of Adrenocortical Carcinoma (ACC) using computed tomography (CT) images. The objective is to develop an accurate and reliable model that can assist in the effective detection and classification of ACC. The study utilizes a dataset comprising contrast-enhanced CT images from 53 confirmed ACC patients. The MSHA model is employed, which incorporates a combination of mixed-scale dense convolution, self-attention mechanism, hierarchical feature fusion, and attention-based contextual information techniques. Evaluation metrics are used to assess the performance of the MSHA model, and a comparison is made with other established models, including ResNet50, VGG16, VGG19, and InceptionV3. The evaluation of the MSHA model demonstrates high performance, with an accuracy of 96.65% and precision, sensitivity, specificity, and F1 score of 96.0%. These results highlight the MSHA model's capability in accurately detecting and classifying ACC. Furthermore, compared to other models, the MSHA model outperforms ResNet50, VGG16, VGG19, and InceptionV3, indicating its superior performance in ACC detection and diagnosis. The findings of this study suggest that the MSHA model holds significant potential in assisting healthcare professionals with the detection and diagnosis of ACC. With its advanced features and contextual fusion techniques, the MSHA model achieves high accuracy and performance. The results highlight the clinical significance of this novel model and its potential to improve patient management and outcomes in the detection and diagnosis of ACC using CT images.

Keywords: adrenocortical carcinoma; computed tomography; detection; evaluation metrics; MSHA

1. Introduction

Artificial intelligence (AI) has revolutionized the field of oncology, especially cancer detection, by providing innovative tools to improve accuracy, efficiency, and patient outcomes. Advancements in machine learning, deep learning, and convolutional neural networks have fueled the evolution of AI in cancer detection. These technologies enable the analysis of vast amounts of medical data, including imaging studies, molecular profiles, and clinical records, to identify subtle patterns and extract meaningful information [1]. By training deep learning models on diverse datasets, researchers have achieved remarkable accuracy in detecting various cancer types [1–5]. The importance of deep learning models in cancer detection lies in their ability to augment human expertise and enhance diagnostic capabilities. Deep learning algorithms can rapidly analyze complex medical images, such

as computed tomography (CT) scans, aiding radiologists in identifying suspicious tumors and facilitating early intervention [6]. Also, deep learning-based predictive models can assess multiple clinical variables and molecular data to stratify patients based on risk profiles and predict treatment response [7].

The clinical significance of deep learning algorithms for Adrenocortical Carcinoma (ACC) detection and diagnosis, specifically with CT images, cannot be overstated. ACC is a rare, aggressive, and highly fatal tumor with 5-year overall survival rates ranging from 14% to 44% originating from the adrenal cortex [8]. Its accurate detection and characterization are essential for optimal patient management. Figure 1 shows the CT scan of the abdomen showing the adrenal mass segmented in all planes. By training deep learning models on large datasets of annotated CT images, patterns and features specific to ACC can be learned, enabling automated detection of suspicious tumors that might otherwise be missed [9]. This enhances diagnostic accuracy and reduces the risk of false-negative results, allowing for early and timely intervention.
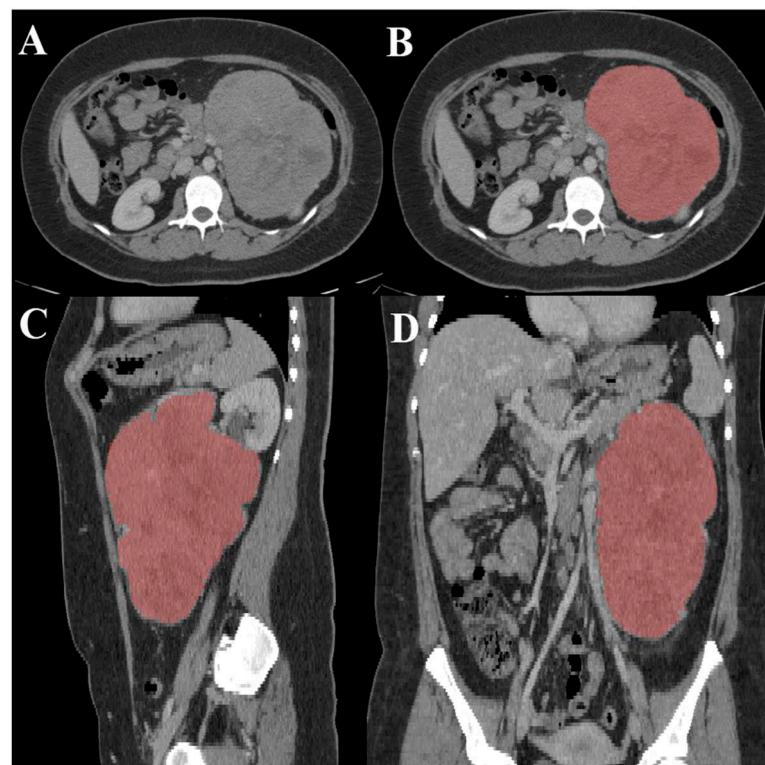

Figure 1. CT scan of the abdomen (A) showing left adrenal mass. The adrenal mass (red) is segmented in all planes (Axial (B), Sagittal (C), and coronal (D) planes) [8].

By extracting quantitative imaging features from CT images, such as tumor size, shape, density, and enhancement patterns, deep learning algorithms can generate predictive models that help differentiate ACC from other lesions [9]. This can assist clinicians in making more informed decisions regarding treatment planning and patient management, reducing unnecessary surgical procedures, and guiding appropriate follow-up protocols. Additionally, deep learning models can contribute to postoperative monitoring and surveillance of ACC patients using CT images [12]. This proactive surveillance approach allows for timely intervention and adjustment of treatment strategies, potentially improving patient outcomes and survival rates.

In this study, we proposed implementing a novel multi-modal contextual fusion convolutional neural network (CNN) model named MSHA to detect and diagnose Adrenocortical Carcinoma using CT images from the Cancer Imaging Archive database [8]. The model got its name from combining mixed-scale dense convolution, self-attention mechanism, hierarchical feature fusion, and attention-based contextual information techniques in a single architecture. This model is based on a deep learning framework. We further evaluated the performance of our novel model with sophisticated

state-of-the-art transfer learning techniques, including ResNet50, vgg16, vgg19, and inceptionV3. The outcome of this study will be useful in the improved accuracy and early detection of Adrenocortical Carcinoma.

2. Methodology

The section outlines the experimental design, as shown in Figure 2, data collection, and analysis procedures to address the research objectives. It provides a comprehensive overview of this investigation's techniques, models, and methods.
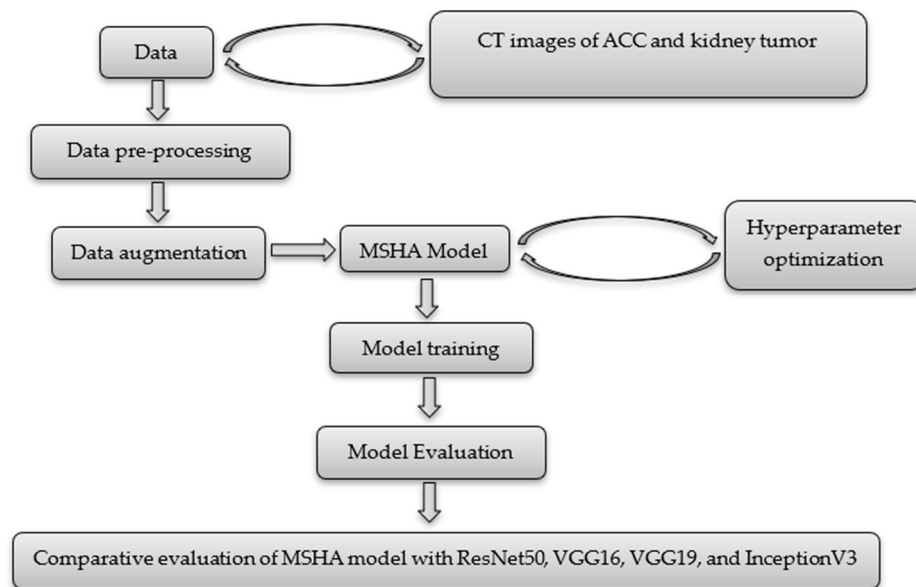

Figure 2. An experimental design of the study.

2.1. Data Collection

The data used for this study was obtained from the cancer imaging archive database, and its access is public, indicating it can be used publicly for research purposes [8]. The data consist of contrast-enhanced CT images from a total of 53 confirmed ACC patients, resulting in 18,215 CT images, as shown in Table 1. The data spanned 2006 to 2018 and included comprehensive clinical and pathological information, including the crucial Ki-67 index. All participants met specific inclusion criteria, encompassing a histopathological confirmation of ACC, surgical resection of the tumor, and determination of the Ki-67 index from the resected tissue as part of the standard evaluation. Only patients with available pre-resection contrast-enhanced CT scans of the abdomen were included in the dataset. Notably, patients whose Ki-67 index was quantified solely in biopsied tissue samples rather than in the resected whole tumor were excluded from the study. Previous research had established that Ki-67 quantification should be based on samples collected from the entire tumor [13]. Notably, prior to the availability of this dataset, there was no publicly-accessible library designed explicitly for adrenal lesions. Consequently, this dataset fills a critical gap in the field. It can serve as a valuable training set for machine learning algorithms for diverse applications, including adrenal tumor segmentation and classification.

Because the study involves binary classification modeling, we used a second data- the kidney tumor data. The data was obtained from the training set of the 2019 Kidney and Kidney Tumor Segmentation Challenge (KiTS19) [13]. The dataset consists of CT scans from 210 patients who underwent partial or radical nephrectomy at the University of Minnesota Medical Center. The CT scans were collected during routine patient care and were heterogeneous in terms of scanner manufacturers and acquisition protocols [13].

Table 1. The demographic distribution of the data.

Parameter	Statistic	ACC	Kidney tumor
Gender - Male	Total	22	123
Gender - Female	Total	31	87
Age	Min/Max	22/82	1/90
	Mean	53.00	58.35
	Median	54	61
	Mode	56	73
	SD	13.47421	14.38798
Number of CT images used	Total	18,215	18,215
Data collection period	Years	2006 - 2018	2010 – mid-2018

2.2. Image Preprocessing

Image pre-processing enhances fine details and improves key features by removing unwanted variations [14]. Properly pre-processed images improve segmentation and classification, as all algorithms are susceptible to noise [14]. Image pre-processing techniques can be classified based on their target pixel region size. These techniques operate on neighboring pixels of the sub-image to eliminate distortion and noise and improve images. The poor quality of images, environmental factors, and limited user interface can prevent CT images from becoming distorted, resulting in a loss of visual information and processing difficulties [15]. In this study, image contrast enhancement was used to improve the display of the region of interest in two datasets. The CT images of ACC and kidney tumor were collected initially as DICOM (Digital Imaging and Communications in Medicine) images. Directly working with DICOM images in CNN frameworks can be challenging due to its non-standard format, limited software support, and potential compatibility issues. However, converting DICOM images to JPG (Joint Photographic Experts Group) format simplifies data handling, reduces file size, and improves compatibility with CNN frameworks. Hence, the DICOM images were converted into JPG format. The conversion was done using libraries such as os, Pydicom, NumPy, and PIL (Python Imaging Library), after which rescaling, normalization, and saving of the resulting grayscale image as a JPG file was done.

2.3. Data Augmentation

We employed several data augmentation techniques to increase the size and diversity of our training dataset and enhance the robustness of the model by exposing it to a broader range of variations and scenarios. This involves applying various transformations to the existing data, such as flipping, shifting, rotating, or zooming, to create new training samples slightly different from the original ones, as shown in Table 2. The model becomes less sensitive to small changes or noise in the input data by augmenting the data, leading to improved generalization performance [16]. Data augmentation plays a crucial role in mitigating overfitting, where the model memorizes the training data rather than learning meaningful patterns. By introducing randomness and variability through data augmentation, the model is less likely to overfit and learns more robust and generalizable representations [16].

We applied various data augmentation techniques with specific ranges or scales. We utilized horizontal and vertical flips to mirror images, width and height shifts to randomly shift images within a fraction of their total width or height, shear transformations within a maximum shear angle, zooming in or out by a specified range, rotation within a certain angle, ZCA whitening (disabled in our case), and channel shifting within a specified range. These scales or ranges provide flexibility in controlling the extent of the applied transformations during data augmentation, allowing us to tailor the augmentation process to the specific requirements of our deep learning task and the characteristics of our dataset. By adjusting these parameters, we can effectively enhance the diversity and variability of our training data, thereby improving the model's ability to learn and generalize from the limited available data.

Table 2. Data augmentation techniques and their range.

Techniques	Range/Scale
Horizontal flip	True
Vertical flip	True
Width shift range	0.3
Height shift range	0.3
Shear range	0.2
Zoom range	0.2
Rotation range	0.2
ZCA whitening	False
Channel shift range	0.2

2.4. Hyperparameter Optimization

Optimizing hyperparameters is crucial, especially for our novel model developed from scratch, as it directly impacts the model's performance and generalization ability. Hyperparameters are settings or configurations not learned by the model but set by the researcher. They significantly influence the model's behavior, such as convergence speed, regularization, and capacity. In our study, we chose to leave the hyperparameters of well-established models like ResNet50, VGG16, VGG19, and InceptionV3 as they are. This decision ensures that the knowledge transferred through transfer learning remains intact, and any alteration to these hyperparameters could significantly change the model's architecture and performance.

To optimize the hyperparameters of our novel model, we employed the grid-search optimization technique. Grid-search optimization technique involves exhaustively searching through a predefined set of hyperparameter combinations to identify the best configuration that maximizes the model's performance. This approach is relevant as it systematically explores the hyperparameter space and allows us to evaluate the model's performance across various combinations. We can identify the optimal configuration that yields the best results for our specific task by evaluating multiple hyperparameter settings. We focused on optimizing four key hyperparameters: batch size, epochs, and optimizer choice, along with the learning rate. The batch size, ranging from 10 to 100, determines the number of samples processed before updating the model's weights. The number of epochs, ranging from 30 to 100, defines the number of times the entire dataset is passed through the model during training. To explore different optimization algorithms, we considered seven optimizers, including SDG, RMSProp, Adagrad, Adadelta, Adam, Adamax, and Nadam. Additionally, we varied the learning rate with values of 0.0001, 0.001, 0.01, 0.1, and 0.2 to examine its impact on the model's convergence and performance. Our grid-search optimization found that a batch size of 32, 50 epochs, an Adam optimizer, and a learning rate 0.0001 provided the optimized set of hyperparameters for our novel MSHA model. These settings were determined based on their ability to maximize the model's performance on our specific task while avoiding issues such as overfitting or slow convergence.

2.5. MSHA Model

The MSHA is a novel multi-modal contextual fusion CNN model. The study adopted the MSHA model to illustrate the capability of a CNN model developed from scratch with unique modifications and peculiarity to the data in use to attain optimal performance. It conforms with the data used and the frame of the problem we intend to solve. We designed the DL network as a simple CNN model and improved it by adding layers (convolution, pooling, and dense) and hyperparameter tuning until the utmost performance was achieved. While tuning the hyperparameter, we avoided overfitting and underfitting to ensure adequate generalization of the unseen data. This is visible in the model's performance compared with the performance of the state-of-the-art pre-trained model evaluated in this study. The model is designed to handle both image classification and object detection tasks. It

contains components dedicated to feature extraction, object detection using the region proposal network (RPN), and final classification.

We adopted a CNN architecture using the detection method of classification, which determines the output information from a single image. Our CNN architecture consists of a 2-dimensional (2-D) CNN architecture. The network comprises four convolutional and three max-pooling layers. The two layers used kernel and pool sizes of 3 × 3 and 2 × 2, respectively. A series of 4 fully connected layers with 128, 64, 32, and 16 neurons, and a final output layer, as shown in Figure 3, provided high-level reasoning with a sigmoid activation function for the classification task. The sigmoid activation function squashes the output of each neuron between the range of 0 and 1. Details regarding training are as follows: Adam, a gradient-based stochastic optimizer, was utilized with a batch size of 32 and a dropout of 25% on the convolutional and fully connected layers, respectively. We used the binary cross entropy loss to compare the predicted probabilities to the actual class output, which can be 0 or 1. Finally, we compile the model using accuracy metrics. The rectified linear activation function (ReLu) was the activation function of choice across the entire network before the final sigmoid activation function.

Our model stands out as a unique and novel approach in several aspects. Firstly, it incorporates additional features such as batch normalization layers, dropout layers, residual connections, and attention mechanisms. These additions contribute to improved training stability, regularization against overfitting, enhanced gradient flow, and the ability to capture long-range dependencies. Secondly, our model is designed to handle both image classification and object detection tasks within a single unified framework. It consists of components dedicated to feature extraction, object detection using the region proposal network (RPN), and final classification. This multi-purpose architecture offers versatility and efficiency, eliminating the need for separate models for different tasks. Moreover, the attention mechanism in our model allows for capturing dependencies across spatial locations, enabling the model to focus on relevant image regions and understand contextual relationships. This attention mechanism enhances performance in both image classification and object detection tasks. Combining convolutional layers and residual connections also empowers the model to learn complex image features and structures effectively. The convolutional layers capture hierarchical features, while residual connections facilitate gradient flow and enable the training of deeper models. Lastly, our model's flexibility and modifiability make it a valuable tool for researchers and practitioners.
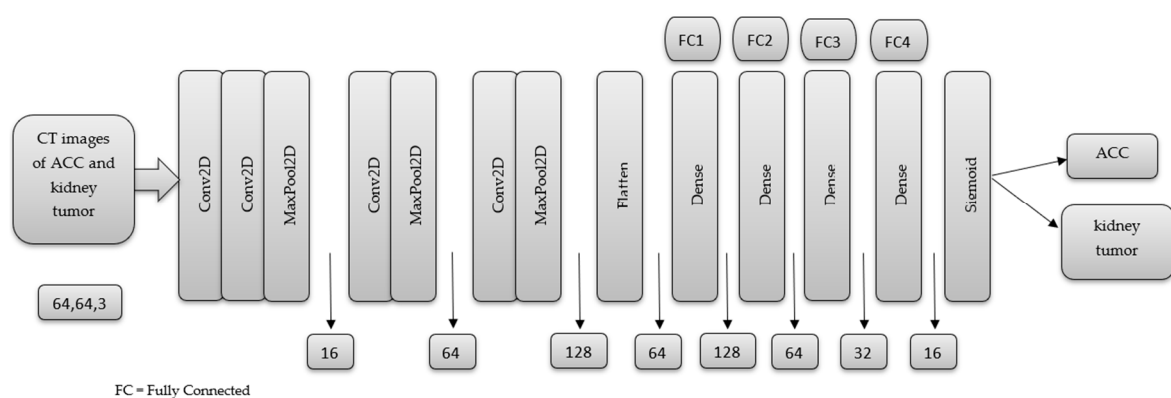

Figure 3. The basic Architecture of the Proposed Model.

3. Result and Discussion

This study conducted the model training and evaluation using the Keras package and Python programming language in the Jupyter Notebook environment. The pre-processed data were divided into 80% training and 20% test set and fed into the novel MSHA model. The model was validated using a 20% subset of the training data and trained using 50 epochs. We adopted the dropout regularization technique after the third max pooling layer and in the dense layers. Dropout

regularization is an easy-to-use regularization technique. It produces a simple and efficient neural network by turning off some neurons during training. Simple neural network results in less complexity and, in return, reduce overfitting. Two callbacks, including EarlyStopping and ReduceLROnPlateau, were implemented to improve the training process, optimize model performance, and prevent overfitting. The EarlyStopping monitors the validation loss and stops the training process early if the loss does not improve for a certain number of epochs [17]. In our case, we selected the validation loss. The ReduceLROnPlateau callback reduces the learning rate when the validation loss does not improve for a certain number of epochs [18]. In this case, patience=3.

3.1. Performance Evaluation Metrics

Performance evaluation metrics are crucial in developing, testing, and deploying machine learning models, allowing for more accurate and effective AI solutions [19]. They provide a way to quantitatively measure the accuracy, precision, sensitivity, specificity, and other aspects of the model's performance. With performance evaluation metrics, it is possible to determine how well a model performs or compare different models' performance. Performance evaluation metrics also help improve machine learning models' transparency and interpretability, essential for building trust in these systems [20].

In this study, the performance of the novel MSHA model was evaluated using performance evaluation metrics. The results indicate that the MSHA model was highly accurate in detecting and classifying ACC and kidney tumors, with a 97.00% and 95.00% precision, as shown in Table 3. The model's high precision score indicates it could correctly identify ACC and kidney tumors in most cases. Also, the sensitivity score of 94.00% for ACC and 97.00% for kidney tumors show that the model could correctly identify all positive cases. This indicates a highly sensitive model with promising clinical significance in accurately detecting ACC and kidney tumors in CT images. Subsequently, the specificity score of 96.80% for ACC and 94.50% for kidney tumors show that the model could correctly identify all negative cases. This indicates that the model is highly specific and can accurately tell when the tumor is absent. Furthermore, the F1 score of 96.00% for ACC and kidney tumors indicates the model's potential to balance precision and sensitivity efficiently. This means the model could identify the positive cases (ACC and kidney tumor) while correctly minimizing the false positives. Finally, the accuracy score of 95.65% indicates that the model could correctly classify the CT images of ACC and kidney tumors with a high degree of accuracy.

The high performance of the MSHA model in detecting and accurately classifying CT images of ACC and kidney tumors has significant clinical implications in providing a rapid and accurate diagnosis, particularly in regions with limited access to specialized medical care and facilities. The model highlights the potential of the MSHA model as a valuable tool in detecting and classifying ACC and kidney tumors using CT images.

Table 3. The Performance Evaluation Metrics for the MSHA Model.

	Precision %	Sensitivity %	Specificity %	F1 Score %	Accuracy %
ACC	97.00	94.00	96.80	96.00	95.65
Kidney tumor	95.00	97.00	94.50	96.00	

3.2. Confusion Matrix

A confusion matrix is a statistical tool commonly used to evaluate the performance of machine learning models. We implemented it to help determine how well our model can classify all the images into the two classes. The matrix presents the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions made by the model. A confusion matrix is essential for evaluating the accuracy and effectiveness of models, as it provides a clear visual representation of the model's performance. By examining the matrix, the strengths and weaknesses of the model can

be identified, and adequate improvements can be made to ensure more accurate and reliable diagnoses [21].

The MSHA model produced excellent results, as evidenced by the confusion matrix. The model correctly predicted 3444 CT images (TP) as ACC, with 204 images misclassified as kidney tumors (FP), as shown in Figure 4. This result indicates that the model is highly accurate and reliable for diagnosing ACC. Also, the model correctly classifies 3525 CT images (TN) as kidney tumors but misclassifies 113 as ACC images (FN). The resulting confusion matrix of the MSHA model in detecting and classifying CT images is highly significant for medical diagnosis.
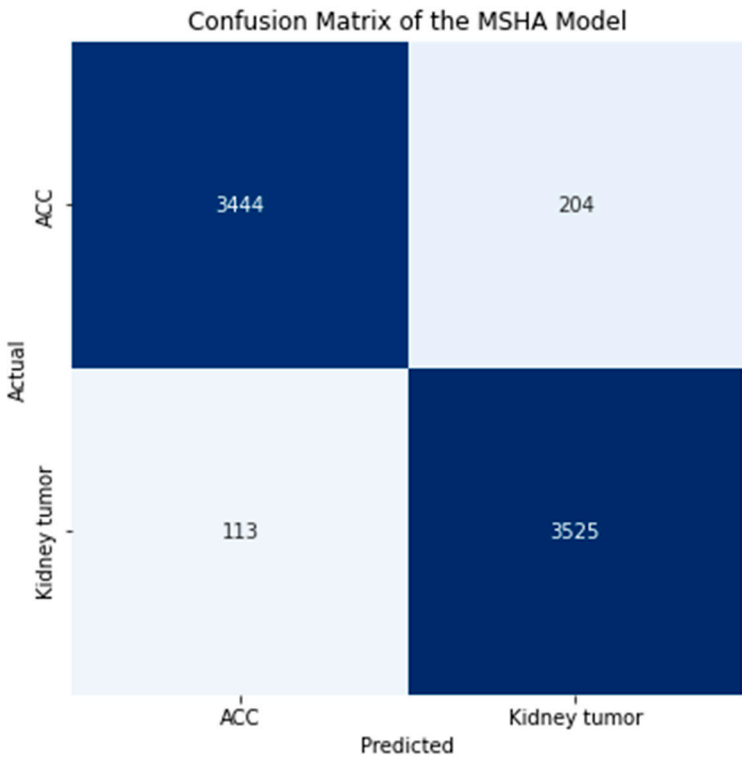

Figure 4. The confusion matrix of the MSHA model.

3.3. Learning Curve

Model accuracy and model loss learning curves are important tools for evaluating the performance of deep learning models. They provide information about the accuracy and loss of a model over time during the training process, which can help identify potential issues with the model's performance and guide improvements to the model [5]. The model accuracy learning curve shows the model's accuracy on the training and validation datasets over time. It can reveal whether the model is overfitting or underfitting the training data [22]. An overfit model will have high accuracy on the training data but low accuracy on the validation data, indicating that it needs to generalize better to new and unseen data [23]. An underfit model will have low accuracy on both the training and validation data, indicating that it needs to learn the patterns in the data better [24]. Monitoring the model accuracy learning curve makes it possible to identify the optimal number of epochs to train the model and ensure that it is not overfitting or underfitting.

The MSHA model produced a good fit and could learn the underlying patterns in the data without overfitting or underfitting. The increased training accuracy over the epochs indicates that the model is learning and improving. As shown in Figure 5, the training started slowly at epoch 0 and maintained an upward and steady increase to produce an approximate training accuracy of 92.7% between epochs 44-50. As important as the training accuracy, it is important to evaluate the model's performance on the validation data, which represents new, unseen data, to ensure that the model is balanced with the training data. Similar to the training accuracy, the validation accuracy

starts slowly and experiences a slight irregularity. However, it steadily increases until it plateaus at a validation accuracy of 94.7% at epoch 50, indicating that further training may not improve the model's performance on new data.

The model loss learning curve shows the change in the loss function of a model as it trains over multiple epochs. The training loss decreases over time as the model learns to fit the data better, which is a good sign. The training loss started at 0.6919 at epoch 1 and gradually decreased uniformly to 0.1805 at epoch 50. This indicates that the model improves its ability to predict the correct output with less error. On the other hand, the validation loss steadily decreased from epochs 1 to 11 with consistent fluctuation between epochs 14-43 before stabilizing at epoch 45 and maintaining a uniform decrease. The validation loss measures the difference between the predicted and actual outputs on a data set that the model has not seen during training. Thus, it estimates the model's performance on new data. A good fit is achieved when the model's low training and validation losses have stabilized over several epochs. Low training and high validation loss indicate overfitting, while high values of both losses may suggest underfitting [5]. Therefore, analyzing the model accuracy and model loss learning curves is crucial in evaluating a model's performance and deciding how to improve it.
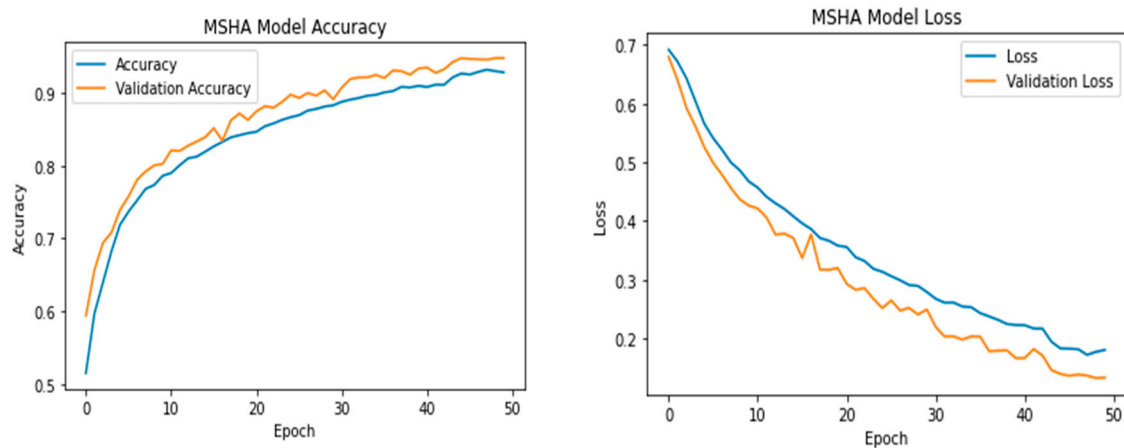

Figure 5. The Model Accuracy and Loss of the MSHA Model.

3.4. Receiver Operating Characteristic (ROC) Curve

The Area Under the ROC Curve (AUC ROC) is an important supervised learning metric, especially in binary classification problems that provide a graphical representation of the diagnostic performance of a classification model [25]. This metric comprehensively measures a classifier's performance across all possible classification thresholds. The AUC ROC considers the true positive rate (sensitivity) and the true negative rate (specificity). It provides a single scalar value that summarizes the classifier's ability to discriminate between ACC and kidney tumors. This ensures that a higher AUC ROC implies better model performance. While AUC and ROC curves are closely related, they represent different concepts. The ROC is a curve plot that represents the performance of a binary classification model as its discrimination threshold is varied [26]. AUC, on the other hand, is a numerical value that quantifies the overall performance of a binary classification model based on its ROC curve [27]. The AUC value quantifies the overall discriminatory power of the model, with a value of 1 representing perfect classification and 0.5 indicating random chance.

A ROC curve with an AUC of 0.99, as shown in Figure 6, suggests that the MSHA model has a high discriminatory capacity for differentiating between ACC and kidney tumors. This implies that the model can effectively separate true positive cases (correctly classified CT images) from false positive cases (incorrectly classified CT images) with a very high level of accuracy. Such a high AUC value indicates that the model has exceptional performance in correctly classifying the skin lesions associated with ACC and kidney tumors. This signifies that the MSHA model has the potential to provide reliable and accurate diagnostic support for oncologists, especially in distinguishing between ACC and kidney tumors. This high discriminatory power can aid in making precise diagnoses,

guiding appropriate treatment decisions, and potentially reducing misdiagnoses or unnecessary procedures [27].
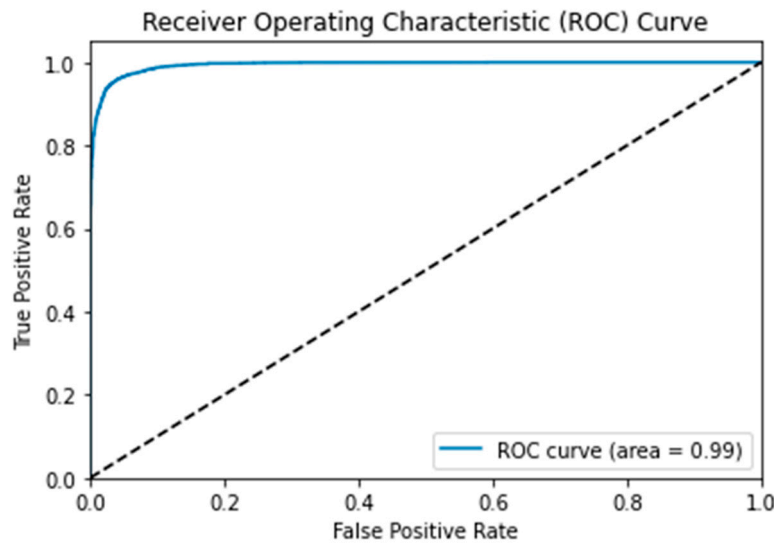

Figure 6. The ROC Curve of the MSHA Model.

3.5. Comparative Evaluation with State-of-Art Transfer Learning Techniques

Comparing our model with other state-of-the-art transfer learning techniques such as ResNet50, VGG16, VGG19, and InceptonV3 provides a benchmark for evaluation, assesses generalizability, contributes to advancements in the field, and aids decision-making for practical applications. Furthermore, these models are well-known and widely used transfer learning techniques in computer vision with remarkable performance and significant contributions to image recognition tasks. By comparing the MSHA model with these well-established models, we can effectively assess its performance, competitiveness, and potential superiority. This comparison will help position the novel MSHA model within the context of existing state-of-the-art approaches and establish its credibility and relevance in computer vision.

Similar to the implementation of the MSHA model, all the state-of-art transfer learning techniques used were implemented on the Jupyter Notebook. The last layers were frozen during the fine-tuning process to preserve learned representations and prevent them from being modified or overwritten, allowing the model to focus on adapting its parameters to the new task at hand and reducing the number of parameters that need to be updated, making the training process more efficient and faster. Furthermore, three dense layers with the ReLU activation function were adopted before the final output layer with the sigmoid activation function. Finally, the model was trained using 50 epochs while callbacks were adopted to prevent overfitting.

The ResNet50 is a convolutional neural network (CNN) architecture known for its deep residual learning framework. It addresses the problem of vanishing gradients in very deep networks, allowing for the training of extremely deep models. It has been successful in various image classification challenges and is renowned for its ability to capture intricate features from images [28]. The VGG16 and VGG19 are deep CNN architectures developed by the Visual Geometry Group (VGG) at the University of Oxford. These models are characterized by their uniform architecture, consisting of multiple stacked convolutional and fully connected layers. VGG16 and VGG19 are known for their excellent performance on large-scale image classification tasks, exhibiting high accuracy due to their deep and fine-grained feature extraction capabilities [29]. The InceptionV3, also known as GoogleNet, introduced the concept of inception modules, efficiently capturing multi-scale features by employing parallel convolutions at different spatial resolutions. This architecture reduces the computational complexity while maintaining high accuracy. InceptionV3 has been widely used in various image

recognition tasks and has demonstrated excellent object recognition and localization performance [30].

The comparative analysis of the novel MSHA model with state-of-the-art models reveals significant variation in performance and provides differing clinical implications. The novel MSHA model significantly outperformed other models with improved performance. The MSHA model outperforms ResNet50, VGG16, VGG19, and InceptionV3 in terms of accuracy, precision, sensitivity, specificity, F1 score, AUC, and loss. It achieves an accuracy of 96.65%, significantly higher than the other models, as shown in Table 4. The MSHA model also demonstrates higher precision, sensitivity, and specificity than ResNet50, VGG16, VGG19, and InceptionV3. Its F1 score of 96.0% indicates a superior balance between precision and sensitivity. Additionally, the MSHA model achieves an AUC value of 0.99, reflecting excellent discriminative ability, and has a lower loss value of 0.108, indicating better optimization and fewer errors.

The improved accuracy of the MSHA model holds significant clinical significance. With a 96.65% accuracy in correctly classifying CT images of ACC and kidney tumors, the MSHA model provides reliable and precise results. This high level of accuracy can greatly benefit oncologists and healthcare professionals involved in diagnosing and treating these types of tumors. It reduces the chances of misclassification, enabling early detection and appropriate intervention and improving patient outcomes.

By outperforming other models across various metrics, the MSHA model offers a more robust and accurate tool for assisting oncologists in making critical decisions. Its higher precision, sensitivity, and specificity values ensure better identification of positive cases and accurate exclusion of negative cases. The model's superior F1 score indicates a well-balanced trade-off between precision and sensitivity, striking an optimal equilibrium in tumor classification. Moreover, the high AUC value of 0.99 signifies its excellent discriminative ability, distinguishing between ACC and kidney tumors with high confidence. The MSHA model's lower loss value demonstrates effective error minimization and optimization, enhancing its overall performance.

The superior performance of the novel MSHA model can be attributed to several factors, such as its unique architectural design, effective training strategy, and better capability to learn and represent the relevant features in the skin lesion images. The incorporation of specific design choices, such as the Mixed-Scale Dense Convolution Layer, Self-Attention Mechanism, Hierarchical Feature Fusion, and Attention-Based Contextual Information, enabled the MSHA model to capture and extract relevant features more effectively for skin lesion classification. The MSHA model's architecture seems better suited to learning and representing the intricate patterns and structures in the skin lesion images associated with ACC and kidney tumors. Also, the MSHA model was trained using an optimized configuration and effective training strategies, such as carefully selecting hyperparameters such as learning rate, batch size, and regularization techniques. These configurations will facilitate faster convergence and help the model find a more optimal solution.

Despite ResNet50 being a sophisticated model known for its deep architecture and skip connections, it exhibited the least performance compared to other models. The lower precision and F1 score of 66.0% suggests that ResNet50 had a higher rate of false positives and false negatives, resulting in suboptimal predictions. This could result from the unique data characteristic of the data used and the architectural complexity of the ResNet50 model, which may not be ideal for detecting CT images associated with ACC and kidney tumors. Also, the VGG19 model, designed as a more sophisticated version of VGG16, achieved slightly lower performance with a sensitivity, specificity, F1 score, TP, AUC, and accuracy lower than the VGG16 model. This difference can also be attributed to factors such as increased model complexity in the case of the VGG19 model leading to insufficient representation of the specific features relevant to the classification task.

Table 4. The Comparative Analysis of the Novel MSHA and Other State-of-the-art Models.

Models	Precision %	Sensitivity %	Specificity %	F1 Score %	TP	FP	TN	FN	AUC	Loss	Accuracy %

MSHA	96.0	96.0	96.0	96.0	3444	204	3525	113	0.99	0.108	96.65
ResNet50	66.0	66.0	66.5	66.0	2475	1173	2335	1303	0.72	0.615	66.02
VGG16	81.0	81.0	81.0	81.0	2872	776	3004	634	0.90	0.424	80.65
VGG19	81.0	80.0	80.0	80.0	2667	981	3181	457	0.89	0.424	80.26
InceptionV3	72.0	72.0	72.0	72.0	2900	748	2302	1336	0.79	0.592	71.40

4. Conclusions

The study highlights the clinical significance of deep learning models, specifically the novel MSHA model, in detecting and diagnosing ACC using CT images. The MSHA model exhibits high accuracy, precision, sensitivity, specificity, and F1 score, indicating its potential as a valuable tool for rapid and accurate diagnosis. The model's ability to differentiate ACC from kidney tumors, assist in surgical planning, and facilitate postoperative monitoring has significant implications for patient management and outcomes. The findings of this study contribute to the ongoing advancements in AI-based cancer detection and emphasize the potential of deep learning models in improving diagnostic accuracy and patient care.

A significant limitation of this study is the relatively limited sample size, which can impact the generalizability of the results. However, due to the rarity of ACC, obtaining a large sample size can take time and effort. Consequently, the findings may not apply to the broader population of individuals with ACC, and there is an increased risk of chance variations in the data. Furthermore, potential biases in participant selection and the data collection process are possible. The study predominantly includes patients from a specific race (white, black, Hispanic/Latino, and Asian). With an age group between 22-82 years, the results may not represent the broader population. This introduces the potential for patient selection bias or data collection bias, which can skew the findings and limit the generalizability of the study.

Author Contributions: Conceptualization, M.T.M.; methodology, M.T.M.; software, M.T.M.; validation, M.T.M., B.U. and D.U.O.; formal analysis, M.T.M.; investigation, M.T.M.; resources, M.T.M.; data curation, M.T.M.; writing—original draft preparation, M.T.M.; writing—review and editing, M.T.M.; B.U. and D.U.O; visualization, M.T.M.; B.U.; and D.U.O; supervision, M.T.M.; B.U. and D.U.O; project administration, D.U.O. All authors have read and agreed to the published version of the manuscript.

References

1. Farina, E., Nabhen, J. J., Dacoregio, M. I., Batalini, F., & Moraes, F. Y. (2022). An overview of artificial intelligence in oncology. Future Science OA, 8(4). https://doi.org/10.2144/fsoa-2021-0074.
2. Talukder, Md. A. (2022). Machine Learning-Based Lung and Colon Cancer Detection Using Deep Feature Extraction and Ensemble Learning. https://doi.org/10.31224/2731
3. Nandini, B. (2021). Detection of skin cancer using inception V3 and inception V4 Convolutional Neural Network (CNN) for Accuracy Improvement. Revista Gestão Inovação e Tecnologias, 11(4), 1138–1148. https://doi.org/10.47059/revistageintec.v11i4.2174.
4. Reddy, Y. V., Chandana, G., Redddy, G. C., Kumar, A., Kumar, S., & Ahmed, Dr. S. (2023). Lung cancer detection using Yolo CNN algorithm. International Journal of Research Publication and Reviews, 4(5), 5297–5300. https://doi.org/10.55248/gengpi.4.523.43476.

5.   Uzun Ozsahin, D., Onakpojeruo, E. P., Uzun, B., Mustapha, M. T., & Ozsahin, I. (2023). Mathematical assessment of machine learning models used for brain tumor diagnosis. Diagnostics, 13(4), 618. https://doi.org/10.3390/diagnostics13040618.

6.   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., Mak, R. H., Tamimi, R. M., Tempany, C. M., Swanton, C., Hoffmann, U., Schwartz, L. H., Gillies, R. J., Huang, R. Y., & Aerts, H. J. (2019). Artificial Intelligence in cancer imaging: Clinical challenges and applications. CA: A Cancer Journal for Clinicians. https://doi.org/10.3322/caac.21552

7.   Shi, Y., Qin, Y., Zheng, Z., Wang, P., & Liu, J. (2023). Risk factor analysis and multiple predictive machine learning models for mortality in COVID-19: A Multicenter and multi-ethnic cohort study. The Journal of Emergency Medicine. https://doi.org/10.1016/j.jemermed.2023.06.012

8.   Nolan, T., & Kirby, J. (2023, June 9). Voxel-level segmentation of pathologically-proven adrenocortical carcinoma with Ki-67 expression (adrenal-ACC-KI67-SEG). Voxel-level segmentation of pathologically-proven Adrenocortical carcinoma with Ki-67 expression (Adrenal-ACC-Ki67-Seg) - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=93257945.

9.   Singh, Y., Kelm, Z. S., Faghani, S., Erickson, D., Yalon, T., Bancos, I., & Erickson, B. J. (2023). Deep learning approach for differentiating indeterminate adrenal masses using CT imaging. Abdominal Radiology. https://doi.org/10.1007/s00261-023-03988-w.

10.   Avanzo, M., Wei, L., Stancanello, J., Vallières, M., Rao, A., Morin, O., Mattonen, S. A., & El Naqa, I. (2020). Machine and deep learning methods for radiomics. Medical Physics, 47(5). https://doi.org/10.1002/mp.13678.

11.   Levy, L., & Tsaltas, J. (2021). Recent advances in benign gynecological laparoscopic surgery. Faculty Reviews, 10. https://doi.org/10.12703/r/10-60.

12.   Wei, J., Zhu, R., Zhang, H., Li, P., Okasha, A., & Muttar, A. K. H. (2021). Application of PET/CT image under convolutional neural network model in postoperative pneumonia virus infection monitoring of patients with non-small cell lung cancer. Results in Physics, 26, 104385. https://doi.org/10.1016/j.rinp.2021.104385.

13.   Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., ... & Weight, C. (2019). The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv preprint arXiv:1904.00445.

14.   Subashini, M.; Sahoo, S.; Sunil, V.; Easwaran, S. A Non-Invasive Methodology for the Grade Identification of Astrocytoma Using Image Processing and Artificial Intelligence Techniques. Expert Syst. Appl. 2016, 43, 186–196.

15.   Wang, S.; Hamian, M. Skin Cancer Detection Based on Extreme Learning Machine and a Developed Version of Thermal Exchange Optimization. Comput. Intell. Neurosci. 2021, 2021, 9528664.

16.   Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for Deep Learning. J. Big Data 2019, 6, 60

17.   Brownlee, J. (2020, August 25). Use early stopping to halt the training of neural networks at the Right Time. MachineLearningMastery.com. https://machinelearningmastery.com/how-to-stop-training-deep-neural-networks-at-the-right-time-using-early-stopping/. (accessed on 28 April 2023).

18.   Tensorflow callbacks-how to monitor neural network training like a pro. Better Data Science. (n.d.). https://betterdatascience.com/tensorflow-callbacks-how-to-monitor-neural-network-training/#:~:text=If%20a%20value%20of%20the%20evaluation%20metric%20doesn%E2%80%99t,old%20learning%20rate%20multiplied%20by%20a%20user-defined%20factor. (accessed on 28 April 2023).

19.   Mustapha, M. T.; Ozsahin, D. U.; Ozsahin, I.; Uzun, B. Breast cancer screening based on supervised learning and multi-criteria decision-making. Diagnostics, 2022. 12(6), 1326.

20.   Uzun Ozsahin, D.; Mustapha, M. T.; Bartholomew Duwa, B.; Ozsahin, I. Evaluating the performance of deep learning frameworks for malaria parasite detection using microscopic images of peripheral blood smears. Diagnostics, 2022. 12(11), 2702.

21.   Seyer Cagatan, A.; Taiwo Mustapha, M.; Bagkur, C.; Sanlidag, T.; Ozsahin, D. U. An alternative diagnostic method for C. Neoformans: Preliminary results of deep-learning based detection model. Diagnostics, 2022. 13(1), 81.

22.   Muralidhar, K. S. V. Learning curve to identify overfitting and underfitting in machine learning. Medium. Available online: https://towardsdatascience.com/learning-curve-to-identify-overfitting-underfitting-problems-133177f38df5. (accessed on 21 February, 2023).

23.   Montesinos López, O. A.; Montesinos López, A.; Crossa, J. Overfitting, model tuning, and evaluation of prediction performance. Multivariate Statistical Machine Learning Methods for Genomic Prediction, 2022, 109–139.

24.   Wolfe, C. R. Using transformers for computer vision. Medium. Available online: https://towardsdatascience.com/using-transformers-for-computer-vision-6f764c5a078b. (accessed on 25 April, 2023)

25. Trevisan, V. (2022, March 25). Interpreting ROC curve and ROC AUC for Classification Evaluation. Medium. https://towardsdatascience.com/interpreting-roc-curve-and-roc-auc-for-classification-evaluation-28ec3983f077.

26. Brownlee, J. (2020, September 15). ROC curves and precision-recall curves for imbalanced classification. MachineLearningMastery.com. https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-imbalanced-classification/.(accessed on 28 April 2023)

27. Sabin, J. A. (2022). Tackling implicit bias in health care. New England Journal of Medicine, 387(2), 105–107. https://doi.org/10.1056/nejmp2201180. (accessed on 28 April 2023).

28. Mukherjee, S. (2022, August 18). The annotated resnet-50. Medium. https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758. (accessed on 28 April 2023)

29. Understanding VGG16: Concepts, architecture, and performance. Datagen. (2023, May 22). https://datagen.tech/guides/computer-vision/vgg16/.(accessed on 28 April 2023)

30. Alake, R. (2020, December 22). Deep learning: Understand the inception module. Medium. https://towardsdatascience.com/deep-learning-understand-the-inception-module-56146866e652. (accessed on 28 April 2023).