Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU33413

A Transfer Learning-Based Approach for Four-Class Lung Cancer Classification from CT Images

Fangyao Liu

School of Mathematics, University of Edinburgh, Edinburgh, United Kingdom
s2629379@ed.ac.uk

Abstract. Lung cancer classification from CT images is clinically important but remains challenging due to the similarity between different pathological subtypes. Using a Kaggle CT image dataset, this study examines a convolutional neural network based on transfer learning that is implemented on the Edge Impulse platform for four-class lung cancer classification. The dataset includes Adenocarcinoma, Large cell carcinoma, Squamous cell carcinoma, and Normal cases. A baseline model was first established using a standard transfer learning workflow, followed by a systematic set of controlled experiments designed to evaluate the impact of different modelling and training factors. In particular, consistent experimental settings were used to examine the impacts of learning rate adjustment, model architecture scaling, input picture resolution, data augmentation, training cycles, and optimizer configuration. The baseline model achieved an accuracy of 52.7%, while the best-performing configuration reached 67.2% after increasing model capacity and input resolution and disabling augmentation. The results demonstrate that model architecture and input resolution play a more critical role in improving classification performance than training hyperparameter adjustments. Overall, while the proposed approach is effective, distinguishing visually similar lung cancer subtypes remains a challenging task.

Keywords: Lung cancer classification, CT imaging, Transfer learning, Convolutional neural networks, Edge Impulse

1. Introduction

The existence of many disease symptoms can lead to impacts on human basic health, daily life, and even the life cycle. To be more specific, lung cancer continues to be a major global health burden and one of the main causes of cancer-related deaths globally [1]. In the 2020 database, among the 2.2 million new global cancer cases, 1.8 million were caused by lung cancer-related causes, leading to a very high mortality rate [1]. According to the research, patients with early-stage lung cancer have a 5-year survival rate of over 75%, but patients with advanced-stage lung cancer have a 5-year survival rate of less than 5% [2]. Because screening with low-dose computed tomography can greatly improve survival outcomes for patients detected at earlier stages, early diagnosis of lung cancer is crucial [2]. Therefore, medical imaging is important in lung cancer investigation and image-based analysis has strong clinical value. Deep learning has been extensively investigated for Computed Tomography (CT) image interpretation, and CT is a frequently used imaging technique in

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU33413

lung cancer screening and diagnosis [3]. A significant amount of research has previously been done on classification, segmentation, and predictive modeling, however there are still challenges [4].

However, image interpretation can be difficult and complex, also those manual analysis strongly depends on radiologists, which is also necessary to take into account the high labor costs and time costs. To improve it, automated deep learning systems could facilitate consistent and effective diagnosis, resulting in quick and precise picture analysis with automatic prediction and categorization [5]. In fact, Convolutional Neural Network (CNN)-based methods are common for CT analysis [4], existing studies have explored subtype and histology classification, also compared multiple pretrained models [5], there are clearly the settings of adenocarcinoma, squamous cell carcinoma, and the other category [6].

Although previous studies have reported promising results, model generalizability and external validation remain important concerns in this field. In particular, the performance of deep learning models may vary across different datasets, task settings, and experimental conditions, indicating that their robustness still requires further evaluation and analysis [7]. Apart from that, although CT-based deep learning approaches have been widely explored in lung cancer research, image classification and detection tasks still present considerable challenges. In particular, limitations in dataset size, task complexity, and the difficulty of distinguishing between different cancer types may affect model performance. Also, previous review studies indicate that CT-based lung cancer classification systems frequently rely on careful hyperparameter tuning, and their efficiency may not consistently be ideal under varying situations [8].

Stimulated by these difficult problems, this research project studies the classification of lung cancer CT images by making use of deep learning methods. Specially, a volume convolution nerve network model is constructed and tested on a many-class CT image data collection. This research carries out further analysis on the ways in which dataset features and model design could have influence on classification effect, meanwhile it uses an edge-based machine learning platform (Edge Impulse) to let efficient model development and deployment become possible.

2. Method

2.1. Dataset preparation

The dataset used in this study was obtained from Kaggle [9], consisting of lung CT scan images for multi-class classification. Adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal cases are the four classes of the categorization task. The dataset contains a total of 903 images, with 326 Adenocarcinoma images, 163 Large cell carcinoma images, 255 Squamous cell carcinoma images, and 159 Normal images. Also, those CT images are provided in Grey format, and standardization is required prior to model training since the initial image resolutions differ between samples.

For guaranteeing consistency, by making use of Edge Impulse's built-in preprocessing flow, all photographs were scaled to 96 × 96 pixels in order to ensure uniformity, hence the dataset was automatically split into training and testing sets through the use of an 80/20 ratio. Furthermore, in the step of feature generation making, the input photograph pictures were normalized through processing to scale the pixel numerical values into a unified uniform range. This procedure guarantees all input characteristics make proportional contribution in the training period, thus preventing characteristics that have bigger digital values from taking control of the studying process. In addition, the normalization can help promote the training stability and accelerate the convergence of the neural network. Beside this, data expansion methods, which include horizontal turn-over and

DOI: 10.54254/2755-2721/2026.GU33413

small rotation, were utilized to promote generalization ability and partly avoid the problem of class imbalance. Figure 1 gives out representative CT pictures coming from the four classes. Visual distinctions are comparatively obvious for several ordinary situations, while the three cancer subtypes present great overlapping in visual manifestation, hence pointing out the hardness of fine-granularity classification.

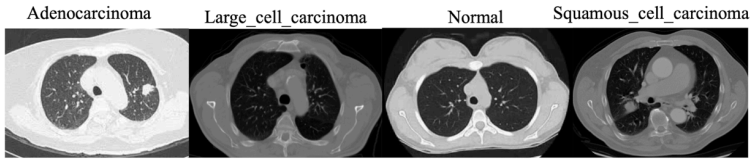


Figure 1. Representative CT images from the four lung cancer classes [9]

2.2. Edge impulse-based CNN model

Convolutional Neural Networks (CNNs) are extensively utilized for image classification since they have the ability to automatically learn hierarchical feature representations directly from original image data. Through stacked convolutional layers, nonlinear activation functions and pooling procedures, CNNs step by step extract low-level vision patterns such as edges and textures, and combine these into higher-level semantic structures. This level-by-level feature study lets CNNs especially fit for medical image analysis, in which tiny space and structure changes may hold meaning for disease judgment [10,11].

Rather than training a CNN entirely from scratch, this study adopts a transfer learning strategy. Transfer learning reuses visual features learned from a source task and adapts them to a new target problem, which is particularly useful when labelled medical image datasets are limited in size. Early layers of a pre-trained network usually capture general low-level features such as edges and textures, while deeper layers encode more task-specific representations. Transfer learning can increase convergence stability and lessen overfitting in medical picture classification by optimizing these representations on the target dataset [12]. This strategy is also relevant to lung cancer CT analysis. Previous studies have applied transfer learning to lung cancer CT image classification and shown that pre-trained convolutional features can support pathological subtype recognition under limited-data conditions [13].

Next, this paper gives consideration to Edge Impulse, it is a cloud-based MLOps platform which is used for making and putting into practice edge and embedded machine learning systems. In addition, the main objective is to decrease brokenness in edge AI development through providing an integrated working flow which connects data processing, model making, learning, optimization, and putting into use within one alone environment. Because this platform can let the whole flow of end-to-end image classification work together with model-side configuration and evaluation which faces deployment, hence it was used by this research as the realization environment for the baseline classifier [14].

The baseline model of this study was realized as a four-category CNN grader which makes use of a transfer learning-based image classification work flow. On the model layer, the system is composed by an image processing module, which is followed by a convolutional neural network that is based on transfer learning. This network carries out classification into four diagnosis types. Instead of designing a convolutional architecture from the beginning zero, this method carries out fine-tuning on a pre-trained model to let it fit the target task. Key hyperparameters were set for the regulating of the learning process when model training is conducted inside the given pipeline. These

DOI: 10.54254/2755-2721/2026.GU33413

here include the count of training circles, study speed, and optimization disposition settings, which together decide the convergence behavior and generalization capability of the model. The baseline model has its root in a MobileNetV2 architecture, which gives a light weight and effective design that fit image classification works. The detailed collocation arrangement of the baseline model is concluded in Table 1, therefore the entire model working flow is shown by Figure 2. This basic line establishment acts as the reference arrangement for later comparison experiments.

Table 1. Baseline model configuration of the edge impulse-based CNN classifier

Component	Setting
Backbone architecture	MobileNetV2 (width multiplier 0.35)
Input feature size	27,648 features
Number of training cycles (epochs)	20
Learning rate	0.0005
Batch size	(default, Edge Impulse)
Optimizer	Default (learned optimizer disabled)
Training processor	CPU
Data augmentation (training stage)	Enabled
Final layer configuration	16 neurons, dropout = 0.1

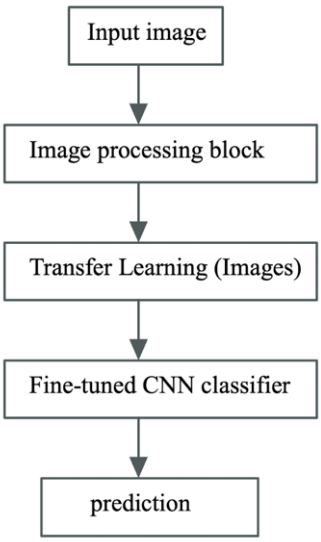


Figure 2. Workflow of the baseline CNN classification model (Picture credit: Original)

3. Results and discussion

3.1. Baseline model performance

The performance of the basic CNN model upon the validation collection is concluded in Table 2, with detailed per-class results which are displayed in Table 3. The model has obtained a total accuracy of 52.7% with a verification loss of 1.60, therefore it shows middle-level performance. Just like what Table 3 displays, this model carries out its work in an uneven way among different classes. Very high accuracy can be seen for Normal (88.9%) and Adenocarcinoma (82.5%), thus Large cell

DOI: 10.54254/2755-2721/2026.GU33413

carcinoma (19.2%) and Squamous cell carcinoma (23.1%) have obviously lower effect. The confused matrix shows that big wrong classification exists among cancer sub-types. In the concrete situation, Large cell carcinoma is frequently given prediction as Adenocarcinoma (76.9%), therefore, Squamous cell carcinoma is often mixed up together with both Adenocarcinoma (51.3%) and Large cell carcinoma (20.5%). This shows that the model is not able to learn enough features which have discrimination ability, for the separation of classes that are similar on vision.

This limitation can be attributed to three main factors. First, CT images of different lung cancer subtypes share similar textures and structural patterns, making feature separation inherently difficult. Second, the baseline model may not have sufficient capacity to capture fine-grained differences required for subtype classification. Third, class imbalance, particularly the smaller number of Large cell carcinoma samples, may bias the model towards dominant classes such as Adenocarcinoma. Overall, while the model distinguishes normal tissue effectively, it shows clear limitations in differentiating between similar cancer subtypes, highlighting the need for further model improvements.

Table 2. Overall performance of the baseline model on the validation set

Metric	Value
Accuracy	52.7%
Loss	1.60
Weighted F1 score	0.48
AUC	0.75

Table 3. Confusion matrix of the baseline CNN model on the validation set

	Adenocarcinoma	Large cell carcinoma	Squamous cell carcinoma	Normal
Adenocarcinoma	82.5%	7.0%	8.8%	1.8%
Large cell carcinoma	76.9%	19.2%	3.8%	0%
Squamous cell carcinoma	51.3%	20.5%	23.1%	5.1%
Normal	11.1%	0%	0%	88.9%
F1 score	0.65	0.23	0.33	0.80

3.2. Effect of learning rate on model performance

To investigate the effect of learning rate, the baseline value of 0.0005 was reduced to 0.0001 while keeping all other settings unchanged. The performance of the modified model is compared against the baseline. The results show a significant decrease in performance. The model achieved an accuracy of 29.0%, compared to 52.7% in the baseline, and the weighted F1 score dropped from 0.48 to 0.24. This indicates that the smaller learning rate negatively affected the training process. From the confusion matrix, the model fails to correctly classify most samples across all classes. For example, Adenocarcinoma is correctly classified in only 5.3% of cases, while a large proportion is misclassified as Large cell carcinoma. Similarly, predictions for other classes are highly scattered, indicating unstable classification behavior.

The occurrence of this performance decrease can be explained by the learning rate that is excessively small. An extremely small learning rate makes the speed of parameter updating become slower, therefore it prevents the model from achieving effective convergence in the number of

DOI: 10.54254/2755-2721/2026.GU33413

training epochs which is limited. As the consequence, this model is not able to study meaningful feature expressions, and therefore it still stays in underfitting condition. On the whole, this experiment has proven that an excessively tiny learning rate can greatly obstruct the performance of the model, hence emphasizing that it is very important for people to select a suitable optimization scale.

3.3. Effect of model architecture and input resolution

For the investigation of the influence brought by model architecture and input resolution, the baseline configuration is modified through replacing MobileNetV2 (96×96, width 0.35) with MobileNetV2 (160×160, width 0.75), thus other relevant settings are kept unchanged. The improved model that we have made exhibits an obvious promotion, with correct rate rising from 52.7% to 65.6% and the weighted F1 value from 0.48 to 0.66. According to what the confusion matrix shows, the classification effect has promotion on all kinds of cancer subtypes. As a special case, the accuracy of Large cell carcinoma is raised from 19.2% to 50%, therefore the accuracy of Squamous cell carcinoma is from 23.1% to 66.7%. The Normal category achieves 100% accuracy, which therefore shows a more forceful separation between normal and pathological modes.

This promotion can be given to two main reasons. Firstly, higher input resolution can retain fine-grained structure information inside CT images, which are necessary for distinguishing the tiny differences among cancer subtypes. Second, the bigger model structure brings bigger representational ability, hence making the network can learn more distinguishing features. Put together, all these elements can promote the model's capability to divide the classes that are similar on vision. However, this promotion comes with the price of the increment of calculation resources. The time of inference ascends from 1165 ms to 2930 ms, and the usage of both RAM and flash has the obvious growth, therefore this points out a trade-off that lies between classification effect and deployment efficiency. On the whole, raising model holding ability and input resolving rate greatly enhances grading effect, especially for hard cancer sub-types, hence this is at the cost of bigger calculation expense.

3.4. Effect of data augmentation

For the purpose of researching what influence data augmentation brings, the function of augmentation was turned off, at the same time, the model structure and training configuration were maintained without change. The model we obtained reached a verification accurate rate of 67.2% and a weighted F1 score of 0.64, it displays a tiny promotion when compared with the past configuration that uses augmentation. Starting from the confusion matrix (which is displayed in Table 4), the classification effect for Adenocarcinoma has promoted to 89.5%, meanwhile Squamous cell carcinoma still keeps a relatively high value which is 56.4%. The category of Normal continuously reaches a one hundred percent perfect classification result (100%). However, Large cell carcinoma still is a difficult problem, it has relatively low accuracy (23.1%) and often gets mixed up with other cancer types by mistake.

This outcome gives indication that, in this data collection, data expanding does not have obvious promotion to model universalization and may even bring changeableness which lets learning become more hard. Because the data collection already includes many different kinds of CT image patterns, extra man-made transformations cannot give useful new information, but on the contrary they play the role of noise. As the consequence, the model which was trained not using augmentation is able to learn more stable and discriminative feature. On the whole, the influence

DOI: 10.54254/2755-2721/2026.GU33413

that data augmentation brings to this task is limited. Although it usually brings benefits to small data sets, its effectualness is dependent on the inherent properties of the data and the categories of transformations that have been put into use.

Table 4. Confusion matrix of the best-performing model on the validation set

	Adenocarcinoma	Large cell carcinoma	Squamous cell carcinoma	Normal
Adenocarcinoma	89.5%	0%	8.8%	1.8%
Large cell carcinoma	46.2%	23.1%	30.8%	0%
Squamous cell carcinoma	38.5%	0%	56.4%	5.1%
Normal	0%	0%	0%	100%
F1 score	0.76	0.38	0.59	0.86

4. Conclusion

This research has proven that the enhancement of feature expression capability has higher importance than the adjustment of training hyperparameters in the classification of lung cancer CT images. The increment of model capacity and input resolution has brought more obvious performance promotion, while the modifications of such items as learning rate, training cycles and optimizer parameter settings produced relatively small influence.

But, these results also point out the inside difficulty of telling apart alike cancer types, which still is a main hard problem in this work. This restriction is therefore caused by the not big enough sample size and the overlapping of visual characteristics between the groups. Therefore, hence, for the purpose of better catching small pathological changes, future research should therefore put focus on promoting data quality and feature extraction methods.

References

[1] Li, C., Lei, S., Ding, L., Xu, Y., Wu, X., Wang, H., Zhang, Z., Gao, T., Zhang, Y., & Li, L. (2023). Global burden and trends of lung cancer incidence and mortality. Chinese Medical Journal, 136(13), 1583–1590. https: //doi.org/10.1097/CM9.0000000000002529

[2] Lancaster, H. L., Heuvelmans, M. A., & Oudkerk, M. (2022). Low-dose computed tomography lung cancer screening: Clinical evidence and implementation research. Journal of Internal Medicine, 292(1), 68–80. https: //doi.org/10.1111/joim.13480

[3] Thanoon, M. A., Zulkifley, M. A., Mohd Zainuri, M. A. A., & Abdani, S. R. (2023). A review of deep learning techniques for lung cancer screening and diagnosis based on CT images. Diagnostics, 13(16), 2617. https: //doi.org/10.3390/diagnostics13162617

[4] Atmakuru, A., Chakraborty, S., Faust, O., Salvi, M., Barua, P. D., Molinari, F., Acharya, U. R., & Homaira, N. (2024). Deep learning in radiology for lung cancer diagnostics: A systematic review of classification, segmentation, and predictive modeling techniques. Expert Systems with Applications, 255, 124665. https: //doi.org/10.1016/j.eswa.2024.124665

[5] Shatnawi, M. Q., Abuein, Q., & Al-Quraan, R. (2025). Deep learning-based approach to diagnose lung cancer using CT-scan images. Intelligence-Based Medicine, 11, 100188. https: //doi.org/10.1016/j.ibmed.2024.100188

[6] Chaunzwa, T. L., Hosny, A., Xu, Y., & Aerts, H. J. W. L. (2021). Deep learning classification of lung cancer histology using CT images. Scientific Reports, 11(1), 5471. https: //doi.org/10.1038/s41598-021-84630-x

[7] Wulaningsih, W., Villamaria, C., Akram, A., Benemile, J., Croce, F., & Watkins, J. (2024). Deep learning models for predicting malignancy risk in CT-detected pulmonary nodules: A systematic review and meta-analysis. Lung, 202(5), 625–636. https: //doi.org/10.1007/s00408-024-00706-1

[8] Javed, R., Abbas, T., Khan, A. H., et al. (2024). Deep learning for lung cancer detection: A review. Artificial Intelligence Review, 57, 197. https: //doi.org/10.1007/s10462-024-10807-1

DOI: 10.54254/2755-2721/2026.GU33413

[9] Kabil007. (2024). LungCancer4Types-ImageDataset [Data set]. Kaggle. https: //www.kaggle.com/datasets/kabil007/lungcancer4types-imagedataset
[10] Chen, X., Wang, X., Zhang, K., Fung, K. M., Thai, T. C., Moore, K., Mannel, R. S., Liu, H., Zheng, B., & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. Medical image analysis, 79, 102444. https: //doi.org/10.1016/j.media.2022.102444
[11] Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E. J., & Dean, J. (2021). Deep learning–enabled medical computer vision. Annual Review of Biomedical Engineering, 23, 221–248. https: //doi.org/10.1146/annurev-bioeng-110220-012203
[12] Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: A literature review. BMC Medical Imaging, 22(1), 69. https: //doi.org/10.1186/s12880-022-00793-7
[13] Wang, S., Dong, L., Wang, X., & Wang, X. (2020). Classification of pathological types of lung cancer from CT images by deep residual neural networks with transfer learning strategy. Open Medicine, 15, 190–197. https: //doi.org/10.1515/med-2020-002
[14] Hymel, S., Banbury, C., Situnayake, D., Elium, A., Ward, C., Kelcey, M., Baaijens, M., Majchrzycki, M., Plunkett, J., Tischler, D., Grande, A., Moreau, L., Maslov, D., Beavis, A., Jongboom, J., & Reddi, V. J. (2022). Edge Impulse: An MLOps platform for tiny machine learning. arXiv. https: //doi.org/10.48550/arXiv.2212.03332