International Journal of Advanced Technology and Engineering Exploration, Vol 8(74)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2020.S1762126

Research Article

Comparison of deep learning convolutional neural network (CNN) architectures for CT lung cancer classification

Sarah Mohd Ashhar[1], Siti Salasiah Mokri[1*], Ashrani Aizzuddin Abd Rahni[1], Aqilah Baseri Huddin[1], Noraishikin Zulkarnain[1], Nor Aniza Azmi[2] and Thanuja Mahaletchumy[3]
Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia, 43600, UKM Bangi, Selangor, Malaysia[1]
Center of Diagnostic, Therapeutic and Investigation Studies (CODTIS), Faculty of Health Science, Universiti Kebangsaan Malaysia, Jalan Raja Muda Abdul Aziz, 50300, Kuala Lumpur, Malaysia[2]
Center of Nuclear Imaging and Nuclear Medicine, UKM Medical Center, Jalan Ya'acob Latif, Bandar Tun Razak, Cheras, 56000 Kuala Lumpur, Malaysia[3]

Received: 9-October-2020; Revised: 14-January-2021; Accepted: 18-January-2021

Abstract
Lung cancer has become one of the most common deaths amongst the cancer patients. World Health Organisation states that lung cancer is the second most fatal cancer all over the world in 2014. Alarmingly, most of the lung cancer patients are diagnosed at the later stages where the cancer has spreads. Thus, early screening via Computed Tomography scan particularly among active smokers is encouraged. Manual diagnosis of the cancer is made feasible through the integration of Computer Aided Diagnosis system. For the past few years, deep learning method leads most of the artificial based intelligence applications including CAD systems. This paper aims to investigate the performance of five newly established Convolutional Neural Network architectures; GoogleNet, SqueezeNet, DenseNet, ShuffleNet and MobileNetV2 to classify lung tumours into malignant and benign categories using LIDC-IDRI datasets. Their performances are measured in terms of accuracy, sensitivity, specificity and area under the curve of the receiver operating characteristic curve. Experimental results show that GoogleNet is the best CNN architecture for CT lung tumour classification wih an accuracy of 94.53%, specificity 99.06%, sensitivity of 65.67% and AUC 86.84%.

Keywords
Computed tomography, Convolution neural network, Deep learning, Lung cancer.

1.Introduction
Lung cancer is the second most fatal cancer worldwide and ranks at the eighth place in the overall reported deaths in Malaysia [1]. Almost 90% of lung cancer patients are both active and passive smokers. As early symptoms of lung cancer are commonly unnoticeable, most of the diagnosis of these patients are found out when they are at stage 3 or 4. Therefore, early screening is highly anticipated. Lung cancer screening can be carried out through computed tomography (CT) scan, sputum cytology and biopsy [2]. Of these three, the non-invasive CT scan is the most viable method to detect lung lesions. However, it is difficult to manually diagnosis small lesions in CT images with high accuracy especially when the surroundings are unclear and noisy.

To assist the physicians on cancer diagnosis, computer aided detection and diagnostic (CAD) system is used. Among the early applications of CADs are on the diagnosis and management of breasts and thoracic cancers [3]. Of recent, the CAD systems are mostly founded based on machine learning methods led by the deep learning method. The main downside of the conventional machine learning methods is the requirement for feature extraction process. The process is omitted when using deep learning. Although deep learning requires heavy computing power and large datasets for

*Author for correspondence
This work was supported in part by the Center of Research and Instrumentation Management (CRIM), Universiti Kebangsaan Malaysia, Selangor, Malaysia and the Ministry of Education, Malaysia (MOE) under Grant FRGS/1/2019/TK04/UKM/02/5

training, the advancement of computational power, top notch graphic processing unit (GPU) and large open online database are constantly and readily available in order to facilitate deep learning implementation. Therefore, improved deep learning neural networks can be created to stablish the best CAD system [4−6] to achieve an accurate cancer diagnosis, for instance in lung lesion diagnosis application via CT images.

Up till now, there are several deep learning architectures have been proposed for lung lesion classification. In lieu of this, this paper compares the performance of five different deep learning architectures; GoogleNet, SqueezeNet, DenseNet, ShuffleNet and MobileNetV2 in the application of lung lesion classification into either malignant or benign specifically in CT images.

2.Literature review

There are several studies in the past literature that use deep learning to classify the lung lesions in CT images. *Table 1* shows four recent studies for such application. Has compared three deep learning architectures to classify the lung lesions into benign and malignant [4]. They are convolutional neural network (CNN), deep neural network (DNN) and sparse auto encoder (SAE). Their study shows that CNN outperforms the rest of the architectures with the highest accuracy of 84.15%, specificity of 83.96% and sensitivity of 84.32%.

Proposed a two deep 3D customized mixed link network (CMixNet) architecture for detecting and classifying lung nodule [7]. Faster R-CNN is used to learn features from CMixNet and together with U-Net like encoder decoder architecture, nodule detection is achieved. Classification of the lung lesions takes place by learning the features through gradient boosting machine (BGM) from the designed 3D CMixNet structure. Their proposed method achieves an accuracy of 88.79%, specificity of 89.83% and sensitivity of 81.35%.

On the other hand, [8] proposed a hierarchical semantic convolutional neural network (HSCNN) in predicting the malignancy of CT scan pulmonary nodule. The architecture is based on two level output; one is low-level semantic features and another one is high level prediction of malignancy of nodule. The architecture gives an accuracy of 88.9%, specificity of 83.96% and 81.35%.

Compared three CNN architectures; namely Alexnet, VGG16 and their proposed architecture known as SilNet [9]. AlexNet is a very popular CNN architecture for its simplicity whereby it has 5 convolution layers and 3 fully connected layers. AlexNet has three convolution group which are the kernel size of 11x11, 5x5 and 3x3. VGG16 is almost like AlexNet but with the addition of 3x3 filter that replaces large kernel size filter on the AlexNet. It has better accuracy than AlexNet. SilNet consists of a group of layers that makes up a group of convolutions which are convolution layer, batch normalization layer, ReLu and pooling layer. This group of convolutions is used five times in the architecture.

Table 1 Several studies on lung lesions classification using deep learning

Ref	Architecture	Accuracy	Specificity	Sensitivity
[4]	CNN	84.15	83.96	84.32
	DNN	82.37	80.66	83.90
	SAE	82.37	83.96	81.35
[7]	CMixNet	88.79	89.83	93.97
[8]	HSCNN	84.20	88.90	70.50
[9]	AlexNet	79.50	76.50	83.04
	VGG16	97.50	97.60	98.26
	SilNet	98.75	99.40	99.56

In addition, [10] proposed 15 layers CNN based architecture to classify long nodules into nodules and nonnodules types, known as LdcNet. It has both feature extraction layer of 3 convolutional blocks and classifier layer. They achieved a classification accuracy of 97.2%. Also applied their novel Gated-Dilated (GD) network that is founded on CNN for benign and malignant lung lesion classification [11]. It achieved an accuracy of 92.75%. Multiple dilated convolutions are used instead of max-poolings while there is a specific network to capture the input features and to determine the suitable dilated convolution.

In this paper, we aim to compare five CNN based deep learning architectures which are GoogleNet, SqueezeNet, ShuffleNet, DenseNet and MobileNetV2. GoogleNet [12] also known as Inception-v1 is created by Google. This architecture introduces a module called inception module that contains 1x1 convolution used to shrink the dimension to lessen the computational bottleneck. The module also assists in the expansion of the depth and the width of network.

SqueezeNet that is introduced by [13] has the advantage of having 50 times less parameters as compared to AlexNet. This architecture uses 1x1 filter instead of 3x3 filter which makes it 9 times smaller in parameters. It also has a module called fire module that contains a squeeze convolution layer with 1x1 filter. This layer is inputted to the expand layer with both 1x1 and 3x3 convolution filters. To obtain high accuracy in classification, the downsampling is done later in the network so that the activation map is large.

ShuffleNet [14] is created by Megvii Inc. (Face++) that promotes the CNN to achieve not only high accuracy but also high computational efficiency. It uses two components; pointwise group convolution and channel shuffle to realize the computational efficiency. The shuffle channel is where the feature map from the previous group is shuffled and feed to the different group of the next group convolution layer. They also introduce ShuffleNet Unit where pointwise group convolution is implemented before the shuffle operation. To recover the dimension, pointwise group convolution is added twice and 3x3 average pooling is added to increase the dimension of the channel with an adequate computation cost.

DenseNet [15] forms a network that achieves high accuracy with less parameters. This architecture introduces a method in which every layer receives input of feature map from all the previous layers. Hence, this enables the future layer to have bits of information from each layer before it. DenseNet creates a compact and slimmer network as the channel number can be reduced. Other than that, it utilizes computational cost and memory systematically. The last architecture is MobileNetV2 [16] that is created by Google as an improvement of the last MobileNetV1. This architecture introduces a module called inverted residual structure. This module creates a shortcut which is put at the bottleneck layer. It is used as input of low dimension that has been compressed and filtered with simple depthwise convolution. The feature is being casted back to a rendition in low dimension with a linear convolution. This architecture is said to be memory efficient.

3.Methodology

The compared architectures are tested to classify lung lesions in CT images in which the datasets are taken from the Lung Image Database Consortium image collection (LIDC-IDRI) [17]. As many as 1646 datasets are collected from this database that comprise of malignant and benign lesions. For each dataset, 4 annotations done by the radiologists are recorded and the lesion is labelled into 5 categories:

1. likely to be benign
2. fairly likely to be benign
3. intermediately likely to be benign and malignant
4. fairly likely to be malignant
5. likely to be malignant.

We extracted this information from the attached extensible markup language (XML) file of the datasets and classify the data into malignant if the average label value is more than 2.5 and vice versa. From the collected data, 1423 malignant cases and 223 benign cases are obtained. These data then are resized to 244x244 and are saved in PNG file. These data are segregated into 70% training and 30% testing. *Figure 1* shows the example of two datasets after preprocessing.

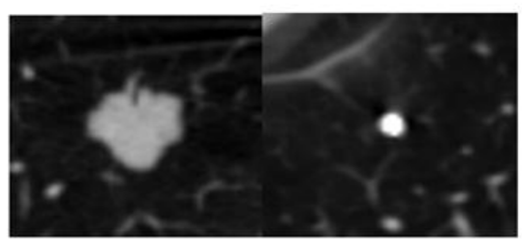

Figure 1 Example of datasets of the CT lung lesions after pre-processing

MATLAB R2020a is used to train and test the deep learning architectures in classifying the lung lesions. Five architectures are tested and compared. They are GoogLeNet, SqueezeNet, ShuffleNet, DenseNet and MobileNetV2. These networks are trained using NVIDIA GTX-1060 6 GB GPU. *Table 2* shows the parameters used in the training process for each architecture. The parameters are common for all architectures for comparable comparison purpose.

Table 2 Training parameters

Parameters	Value
Initial learn rate	0.01
Validation frequency	50
Max Epochs	20
Mini batch size	64
Learn rate drop factor	0.2
Learn rate drop period	5
Momentum	0.9

We evaluate the performance of the five networks based on classification metrics such as accuracy, specificity, sensitivity and area under the curve of the receiver operating characteristic curve (AUC). Accuracy is the measurement of the correct data predicted by the network divided by the whole data. This can be calculated as below:

(TP: True Positive, TN: True Negative; FP: False positive; FN: False Negative

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

Specificity is the number of data correctly labelled as negative over the whole negative data. The formula is as stated:

$$Specificity = \frac{TN}{FP+TN} \quad (2)$$

Sensitivity is the number of data correctly labelled as positive over the whole positive data. It is calculated as:

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

AUC is also being analyzed through the Receiver Operating Characteristics (ROC). This parameter is used to analyze the performance of the network to classify the data into their respective classes. AUC of 1 represents a perfect classification while 0 indicates a definite classification failure. *Figure 2* shows the overall block diagram of the implemented classification using the CNN based networks.

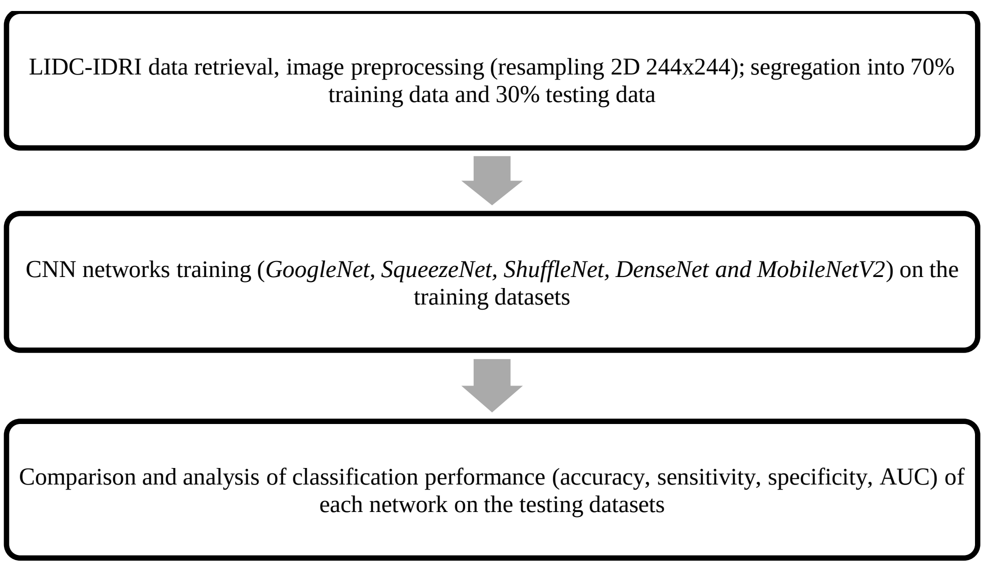

Figure 2 Block diagram of the classification method using deep learning networks

4.Result

Table 3 tabulates the performance of the networks in terms of accuracy, specificity, sensitivity and AUC while *Figure 3* shows the graphical comparison of the performance for each network. GoogLeNet architecture outperforms all the other architectures with an accuracy of 94.53%, specificity of 99.06% and sensitivity of 65.67%. GoogLeNet able to perform the best in accordance to the way the architecture is designed that is to limit the computational bottleneck. Therefore, there is an ease to run the architecture smoothly. The second best performed architecture is SqueezeNet with an accuracy of 94.13%, specificity of 99.06% and sensitivity of 62.69%. Both architectures implement a

1x1 filter in their novel modules. It shows that by integrating 1x1 filter, it can improve the architecture performance as it lessens the burden of facing computational bottleneck.

The next performed classifier is DenseNet with accuracy of 93.52%, specificity of 89.83% and sensitivity of 59.7%. The propagation of input layer to the subsequent layers proves to have beneficial attributes as the front layers able to learn more features and increasing its accuracy. However, the 121 layers that it consists may not be suitable to be used on a simple single GPU computer as it can slow down the running time. ShuffleNet performance is not well received compared to the rest as it tackles the accuracy of 92.91%, specificity of 98.83% and sensitivity of 55.22%. The last out of all the architectures is MobileNetV2 with an accuracy of 92.91%, specificity of 97.65% and sensitivity of 62.64%. However, it still manages to achieve the accuracy of higher than 90%. This show that all the architectures perform greatly with their novelty modules that are designed for high accuracy while maintaining a good number of parameters.

Figure 4 shows the ROC curve for each architecture. Based on the ROC curves, GoogleNet has the highest area under the ROC curve (AUC) of 86.84%. DenseNet achieves AUC of 86.12%, followed by SqueezeNet (85.04%), ShuffleNet (82.34%) and lastly MobileNetV2 (82.11%).

Figure 5 to Figure 9 show the confusion matrix for each architecture in the classification task. The confusion matrix is used to recognise the correctly labelled data to their respected classes and the incorrectly labelled data to their respected classes. This subsequently be calculated in the form of specificity and sensitivity.

The next simulation result is an example of prediction by each architecture on a specific dataset. In this case, the dataset is malignant. *Figure 10* to *Figure 14* show the prediction made by each architecture together with the accuracy values. It is seen that GoogLeNet, SqueezeNet and DenseNet predicted the data correctly while ShuffleNet and MobileNet missed the prediction. GoogleNet gives the highest prediction which is 0.99 while MobileNetV2 wrongly classifies the data into benign with an accuracy of 0.72.

Table 3 Accuracy, specificity, sensitivity, AUC performance

Architecture	Accuracy (%)	Specificity (%)	Sensitivity (%)	AUC (%)
GoogleNet	94.53	99.06	65.67	86.84
SqueezeNet	94.13	99.06	62.69	85.04
ShuffleNet	92.91	98.83	55.22	82.34
DenseNet	93.52	98.83	59.7	86.12
MobileNetV2	92.91	97.65	62.64	82.11

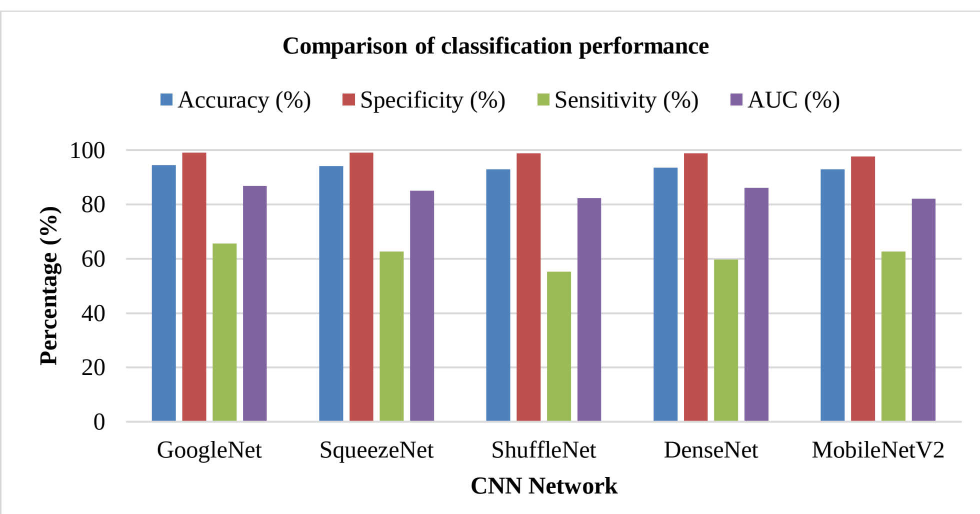

Figure 3 Comparison of accuracy, specificity, sensitivity and AUC of each network

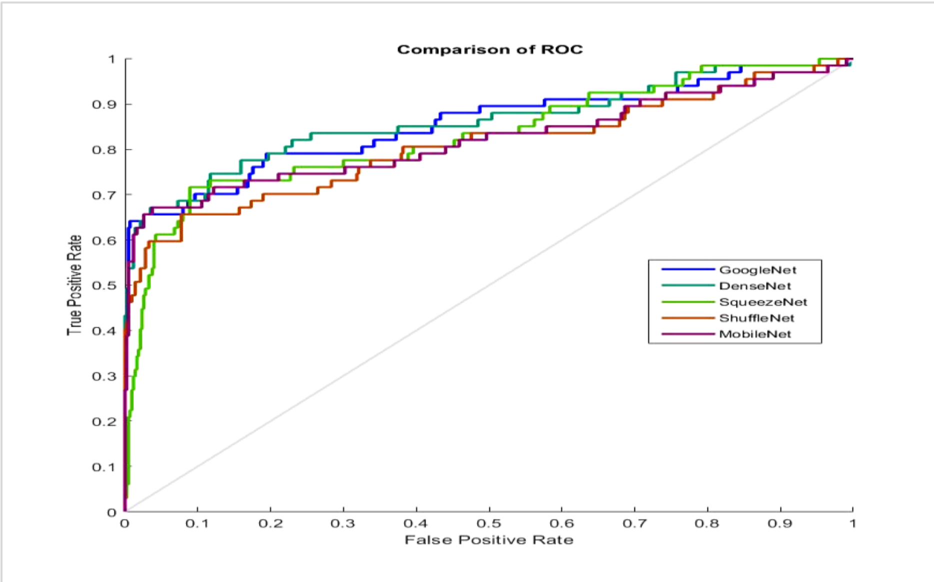

Figure 4 Comparison of ROC curves between GoogleNet, DenseNet, SqueezeNet, ShuffleNet and MobileNet

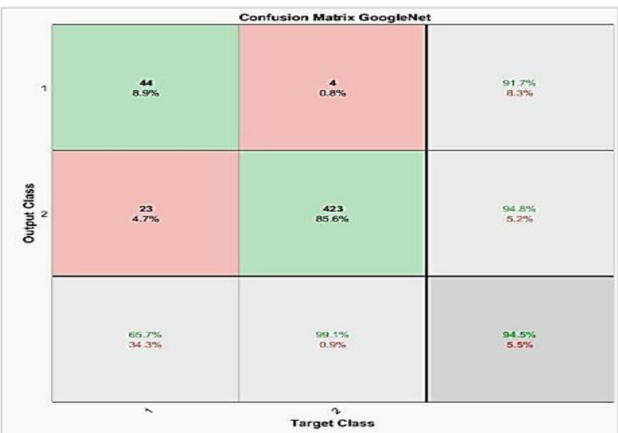

Figure 5 GoogLeNet confusion matrix

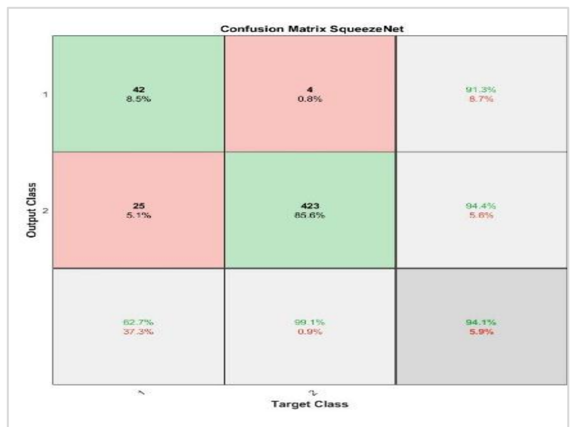

Figure 6 SqueezeNet confusion matrix

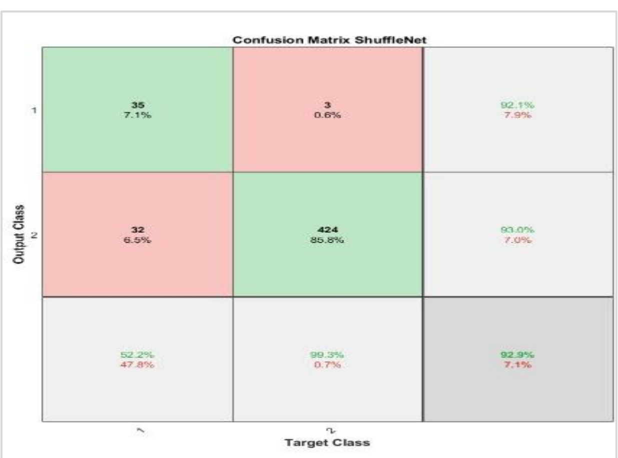

Figure 7 ShuffleNet confusion matrix

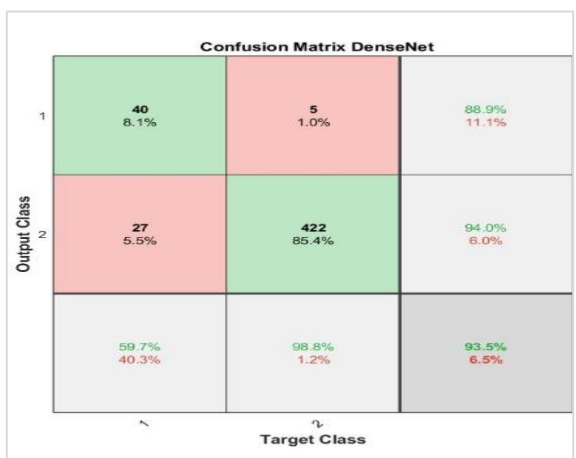

Figure 8 DenseNet confusion matrix

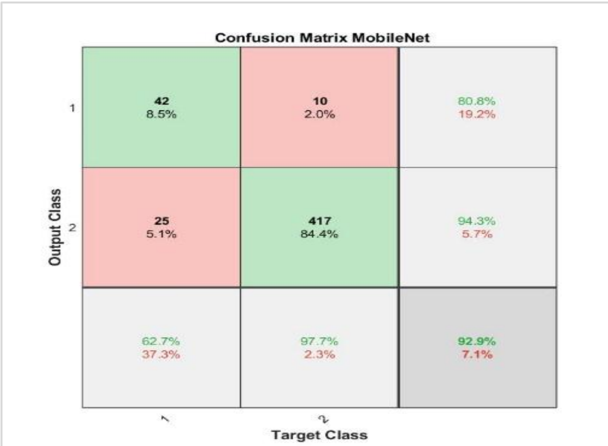

Figure 9 MobileNetV2 confusion matrix

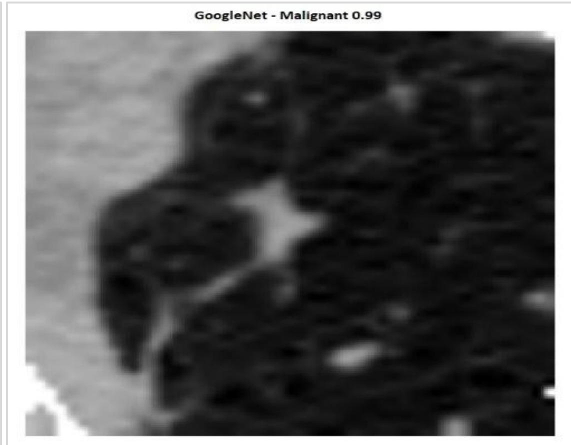

Figure 10 GoogLeNet prediction – malignant 0.99

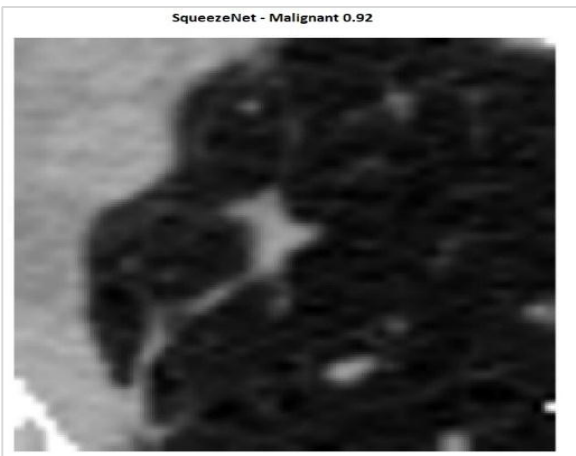

Figure 11 SqueezeNet prediction -malignant 0.92

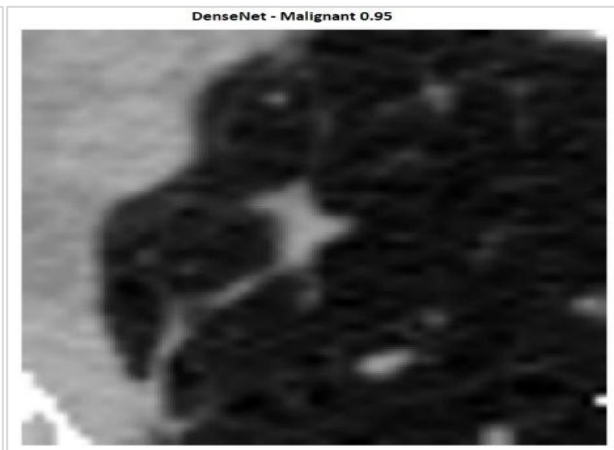

Figure 12 DenseNet prediction – malignant 0.95

Figure 13 ShuffleNet prediction – benign 0.65

Figure 14 MobileNet prediction – benign 0.72

In overall, it is observed that GoogLeNet architecture is the best deep learning classifier to classify the lung lesions in CT images into benign or malignant lesions followed by SqueezeNet, DenseNet, ShuffleNet and lastly, MobileNetV2. Here, our results are only based on LIDC-IDRI database. However, the results are reliable as LIDC-IDRI is the biggest online lung CT dataset available of all.

5.Conclusion

In this paper, the performance of five CNN architectures, namely GoogleNet, SqueezeNet, DenseNet, ShuffleNet and MobileNetV2 to classify the lung lesions in CT images into benign and malignant lesions have been compared and analyzed. The experiment results show that the GoogleNet architecture is the best deep learning classifier to classify the lung lesions in CT images into benign or malignant with an accuracy of 94.53%. The least performed architecture of all is MobileNetV2 with an accuracy of 92.91%. Therefore, further study on the GoogleNet network is required in order to improve the classification accuracy of lung lesions in CT images.

Acknowledgment

The authors would like to acknowledge the Center of Research and Instrumentation Management (CRIM), Universiti Kebangsaan Malaysia, Selangor, Malaysia and the Ministry of Education, Malaysia (MOE) for the Fundamental Research Grant Scheme (FRGS) with project code: FRGS/1/2019/TK04/UKM/02/5 to support this project.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] Kan Chan Siang MD, John CK. A review of lung cancer research in Malaysia. Medical Journal of Malaysia. 2016; 71(1):70-8.
[2] Rossi F, Mokri SS, Rahni AA. Development of a semi-automated combined PET and CT lung lesion segmentation framework. In medical imaging biomedical applications in molecular, structural, and functional imaging 2017 (p. 101370B). International Society for Optics and Photonics.
[3] Shukor MZ, Mokri SS. Comparison of segmentation techniques for nonsmall cell lung cancer (NSCLC) in PET images. Solid State Technology. 2020; 63(1s):958-66.
[4] Song Q, Zhao L, Luo X, Dou X. Using deep learning for classification of lung nodules on computed tomography images. Journal of Healthcare Engineering. 2017:1-8.
[5] Chan HP, Hadjiiski LM, Samala RK. Computer-aided diagnosis in the era of deep learning. Medical Physics. 2020; 47(5): e218-27.
[6] Lee JG, Jun S, Cho YW, Lee H, Kim GB, Seo JB, et al. Deep learning in medical imaging: general overview. Korean Journal of Radiology. 2017; 18(4):570-84.
[7] Nasrullah N, Sang J, Alam MS, Mateen M, Cai B, Hu H. Automated lung nodule detection and classification using deep learning combined with multiple strategies. Sensors. 2019; 19(17):1-19.
[8] Shen S, Han SX, Aberle DR, Bui AA, Hsu W. An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification. Expert Systems with Applications. 2019; 128:84-95.
[9] Abraham GK, Bhaskaran P, Jayanthi VS. Lung nodule classification in CT images using convolutional neural network. In 9th international conference on advances in computing and communication 2019 (pp. 199-203). IEEE.
[10] Tran GS, Nghiem TP, Nguyen VT, Luong CM, Burie JC. Improving accuracy of lung nodule classification using deep learning with focal loss. Journal of Healthcare Engineering. 2019: 1-9.
[11] Al-Shabi M, Lee HK, Tan M. Gated-dilated networks for lung nodule classification in CT scans. IEEE Access. 2019; 7:178827-38.
[12] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In proceedings of the IEEE conference on computer vision and pattern recognition 2015 (pp. 1-9).
[13] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360. 2016.
[14] Zhang X, Zhou X, Lin M, Sun J. Shufflenet: an extremely efficient convolutional neural network for mobile devices. In proceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 6848-56).
[15] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In proceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 2261-9).
[16] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: inverted residuals and linear bottlenecks. In proceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 4510-20).
[17] Armato III SG, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Medical Physics. 2011; 38(2):915-31.

Sarah Mohd Ashhar is an Electronic Engineering student at the Department of Electrical, Electronics and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia majoring in Computer and Communication Engineering from 2016 to 2020.
Email: sarahashhar@gmail.com

Siti Salasiah Mokri is a Senior Lecturer at the Department of Electrical, Electronics and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia. She obtained Bachelor Degree in Electrical Engineering from University of Malaya in 2000, MSc in Mechatronics from International Islamic University Malaysia in 2008 and Phd in Biomedical Engineeing from University Putra Malaysia in 2016. Her research interests are Image Processing, Biomedical Images, Artificial Intelligence, Control System and Mechatronics.
Email: siti1950@ukm.edu.my

Ashrani Aizuddin Abd. Rahni received B.A and M.Eng degrees from the University of Cambridge and Ph.D. from the University of Surrey, United Kingdom. He is a Senior Lecturer at Universiti Kebangsaan Malaysia (UKM) and a graduate member of the Board of Engineers Malaysia and the IEEE. His research interest are Volumetric Medical Image Computing and Image Guided Computer Aided Medical Diagnosis and Intervention.
Email: ashrani@ukm.edu.my

Aqilah Baseri Huddin received her BEng (Hons) in Electrical and Electronics Engineering and her PhD in Electrical and Electronics Engineering from the University of Adelaide, Australia in 2007 and 2015, respectively. She is currently a Senior Lecturer in the Department of Electrical, Electronics and System Engineering in Universiti Kebangsaan Malaysia, UKM. Her research interests are mainly in the field of Image Processing and Artificial Intelligence.
Email: aqilah@ukm.edu.my

Noraishikin Zulkarnain is a Senior Lecturer in the Department of Electrical, Electronic and Systems Engineering, UKM Bangi, Malaysia. She is having 5 years experience in the teaching field. She received Bachelor of Engineering with honour (Electrical and Medical Electronics) degree from Universiti Teknologi Malaysia. Then, she got the opportunity to further her study in PhD with a fast track programme. She received a PhD degree in Control Systems from Universiti Teknologi Malaysia in 2016. Her research area of interest is Control System, Image Processing and Medical Electronics Engineering.
Email: shikinzulkarnain@ukm.edu.my

Nor Aniza Azmi is a Senior Lecturer at the Faculty of Health Sciences, Universiti Kebangsaan Malaysia. She received Bachelor of Diagnostic Imaging & Radiotherapy from Universiti Kebangsaan Malaysia in 2005, Master of Research in Health Sciences, University of Liverpool in 2008 and PhD in Health Sciences from Universiti Teknologi Mara (UiTM) in 2018.
Email: noraniza.azmi@ukm.edu.my

Dr Thanuja Mahaletchumy is a Nuclear Medicine Specialist and currently heads the Department of Molecular Imaging and Nuclear Medicine at Universiti Kebangsaan Malaysia Medical Center. She obtained her medical degree from University Malaysia Sarawak (UNIMAS) and Master of Medicine (Nuclear Medicine) from Universiti Sains Malaysia (USM). She is also a fellow of the Asian Nuclear Medicine Board (ANMB). Her practice mainly focuses on oncologic imaging (PET-CT and SPECT-CT).
Email: mthanu@msn.com