International Journal of Chemical and Biochemical Sciences
(ISSN 2226-9614)

Journal Home page: www.iscientific.org/Journal.html

An Accurate Deep Learning Based Framework for Detection of Thyroid Cancer Using Ultrasound Images

Samaa A. Nasr[1*], Hanan M. Abdel-Fattah[2], Mohamed M. Abdelsalam[3], Hossam El-Din Moustafa[2]

[1]Biomedical Engineering Program, Faculty of Engineering, Mansoura University, Egypt

[2]Electronics and Communications Engineering Department, Faculty of Engineering, Mansoura University, Egypt

[3]Computers Engineering and Control System Department, Faculty of Engineering, Mansoura University, Egypt

Abstract

Thyroid cancer tends to have a relatively high survival rate and is often treatable if detected early. However, its prevalence and impact can vary by country due to factors like healthcare access, screening practices, and environmental factors. The ultrasound images can detect abnormal changes in the structure or size of the thyroid gland. This includes the presence of nodules or growths within the thyroid tissue. The present study presents a framework to classify thyroid cancer cells using ultrasound images into two main specified cell types using deep learning algorithms. The DDTI (The digital database of Thyroid Ultrasound Images) is used in this investigation. Depending on their internal composition, echogenicity, margins, calcifications, and TI-RADs, experts categorize the cells into two main distinct types which are benign and Malignant cells. Introducing a pipeline to improve diagnostic accuracy by experimenting keras application models and choosing the appropriate model result is the main contribution. The study reached the conclusion that deep learning could improve ultrasound screening classification results. The most suitable algorithms for this application appear to be VGG16, MobileNet, and InceptionResNetV2. The VGG16, MobileNet, and InceptionResNetV2 have the highest accuracy of 99.9%. These results support the proposed framework as a reliable diagnostic tool for thyroid cancer.

Keywords: Thyroid disease diagnosis, Ultrasound images, Deep Learning, Fine tuning, CNN, Binary classification.

Full length article **Corresponding Author*, e-mail: samaanasr96@gmail.com

1. Introduction

Cancer is the body's abnormal cells growing uncontrollably. Uncontrolled growth and division of cancer cells allow them to invade healthy tissues and organs and subsequently spread throughout the body [1]. The thyroid gland, resembling a butterfly in shape, plays a vital role in regulating key physiological functions such as body temperature, heart rate, and blood pressure in humans [2]. In recent decades, there has been a growing focus on this crucial organ due to the increasing detection of malignant thyroid nodules [3]. This trend has garnered global attention, with over 300 million individuals diagnosed with thyroid-related conditions in 2018, and the numbers continue to rise today [4]. Notably, thyroid cancer, which is the most prevalent cancer in men aged 30 to 39 years, affects women three times more frequently and has a threefold higher diagnosis rate [5]. The thyroid gland consists of two primary types of cells: follicular cells and parafollicular cells, with follicular cells dominating the tissue. These cells are responsible for producing thyroid hormones that contain iodine. The Thyroid Imaging Reporting and Data System (TI-RADS) is a contemporary classification system designed to offer a comprehensive assessment, primarily focusing on malignancy categorization. This system includes the following classifications: TIRAD 1 (indicating normal findings, occasionally benign with no need for Fine Needle Aspiration or FNA), TIRAD 2 (suggesting a benign cluster), TIRAD 3 (likely benign), TIRAD 4 (indicating a suspicious abnormality), TIRAD 5 (signifying a high likelihood of malignancy), and TIRAD 6 (confirming proven malignancy).

Despite the continual increase in thyroid disease cases, diagnostic procedures in clinical settings have seen limited advancements since the 20th century [6]. The early detection and diagnosis of thyroid nodules can significantly impact lives by preventing cancer, ultimately leading to reduced rates of illness and mortality. Thyroid nodules are a common clinical finding, with a prevalence of approximately 19-68% in the general population, as reported in various studies. While the majority of thyroid nodules are benign, the ability to accurately distinguish between benign and malignant nodules is crucial for appropriate patient management and to reduce unnecessary surgical interventions [7-8]. Ultrasound imaging is a widely used modality for the evaluation of thyroid nodules due to its safety, accessibility, and ability to provide real-time, high-resolution images. Thyroid ultrasound, an imaging method employing sound waves, is utilized to evaluate various aspects of thyroid health, including the size, location, and characteristics of the primary nodule. Additionally, it helps identify the presence of other thyroid nodules, such as those in the contralateral lobe, and assesses suspicious-looking lymph nodes. It's important to note that the size of a nodule alone does not reliably indicate whether it is malignant or benign. However, the subjective interpretation of ultrasound images by radiologists can lead to variations in diagnostic accuracy [9]. Therefore, radiologists rely on ultrasound findings, coupled with a physical examination of the neck. Further diagnostic tests, such as thyroid CT, thyroid MRI, and tissue sampling, may be necessary to confirm a diagnosis. To address the challenge and improve the accuracy of thyroid nodule classification, Computer-Aided Diagnosis (CAD) systems have emerged as a promising solution. CAD systems are computer-based tools designed to assist radiologists and clinicians in the interpretation of medical images [10-11]. In the context of thyroid nodule assessment, CAD systems can provide an objective and standardized approach to nodule classification, aiding in the early detection of malignant nodules and reducing the rate of unnecessary biopsies or surgeries. For instance, the presence of thyroid nodules ranging from 1.0 to 1.9 cm in diameter on an ultrasound is often indicative of an elevated cancer risk. Deep learning techniques, specifically within the realm of Computer-Aided Diagnosis (CAD), have gained widespread application in clinical practice, spanning disease prediction, diagnosis, and treatment. In diagnosing conditions like hypothyroidism and hyperthyroidism, researchers have explored various statistical methods and machine learning algorithms, including k-nearest neighbor, linear discriminant analysis, and decision trees [12-15]. Notably, deep learning algorithms have been prominently employed in the classification of thyroid nodules from medical images. Researchers have dedicated substantial efforts to enhancing the accuracy of thyroid disease diagnosis through the analysis of ultrasound images. Results have shown that the accuracy of detecting malignant thyroid nodules using ultrasound images falls within the range of 72% to 92%, often surpassing the capabilities of less-experienced radiologists [16-20].

1.1. Type A: Papillary carcinoma

Papillary carcinoma is the most common and typically less aggressive form of thyroid cancer, originating from follicular cells and often forming finger-like projections as shown in Figure 1 (a) [21].

1.2. Type B: Follicular carcinoma

Follicular carcinoma is a thyroid cancer type originating from follicular cells and is more aggressive than papillary carcinoma, with a greater tendency to spread as shown in Figure 1 (b) [22].

1.3. Type C: medullary carcinoma

Medullary carcinoma is a less common and aggressive type of thyroid cancer originating from C cells. It can occur sporadically or hereditarily and is not linked to thyroid hormone production as shown in Figure 1 (c) [23].

1.4. Type D: anaplastic carcinoma

Anaplastic carcinoma is an uncommon and highly aggressive thyroid cancer that arises from follicular or papillary thyroid cells, rapidly transforming into an undifferentiated and fast-growing form. It tends to spread quickly and is challenging to treat as shown in Figure 1 (d) [24].

2. Related work

The study entailed a comprehensive exploration of modern approaches for early thyroid cancer diagnosis. It encompassed a thorough comparative analysis of these approaches, considering factors like the datasets, classification algorithms, performance metrics, and publication details, all presented in Table 1. This extensive comparison revealed a significant finding: none of the models employed in prior studies achieved a comparable level of accuracy to the proposed models. The models stand out not only for its ease of implementation but also for its exceptional accuracy, distinguishing it from existing approaches.

3. Materials and methods

In the investigation, Thyroid ultrasound images were sourced from the Digital Database Thyroid Image (DDTI), an openly accessible repository of such images, accessed via reference [46]. The dataset comprised a total of 260 cases and 350 thyroid ultrasound images in JPG format, with 61 categorized as benign and 289 as malignant, spanning a wide age range from 21 to 92 years. Each image presented in 8-bit grayscale [47]. Figure 2 visually illustrates both a benign and a malignant image from the dataset for reference. These thyroid ultrasound images underwent a meticulous evaluation and categorization process carried out by experienced radiologists using the Thyroid Imaging Reporting and Data System (TIRADS). The TIRADS system classifies these images into seven distinct categories, each contingent on specific ultrasound features: (3) absence of suspicious features, (4a) presence of one suspicious feature, (4b) presence of two suspicious features, (4c) presence of three or four suspicious features, and (5) presence of five or more suspicious features [46]. In the study, the focus was on selecting images categorized as benign, specifically those classified as TIRADS 2 and 3. Conversely, Images were classified as malignant when falling into categories 4a, 4b, 4c, or 5, based on a categorization strategy established in prior research findings [16,48].

Radiologists assessed sonographic characteristics and measurements, such as composition, echogenicity, calcification, margin, and shape, using the TI-RADS guidelines proposed by Kwak et al., (2011) [50] as a reference. Nodule compositions were categorized into three primary groups: solid, cystic, or spongiform. A spongiform composition indicated that the nodule primarily consisted of small cystic spaces. Echogenicity was classified as hyperechoic, isoechoic, hypoechoic, or markedly hypoechoic by comparing nodule echoes to those from normal thyroid tissues. Marked hypo-echogenicity was identified when the nodule displayed lower echogenicity than the surrounding strap muscle. Microcalcification referred to tiny, punctate, or "dot-like" foci without posterior acoustic artifacts, each measuring less than 1 mm in diameter. On the other hand, calcifications were grouped into macrocalcification or triangular reverberation artifacts, with a diminishing width known as comet-tail artifacts. Margins were categorized based on their smoothness, with options including irregular, lobulated, ill-defined, halo, or displaying extrathyroidal extension (ETE). Lastly, Nodule shapes were classified as either wider or taller in comparison to their natural proportions [49]. Fine-tuning is a specific form of transfer learning that involves taking a pre-trained model and further training it on a task-specific dataset. Instead of training from scratch, fine-tuning modifies the top layers of the pre-trained model to adapt it to the target task while retaining the valuable features learned from the original data. This approach is particularly useful when working with limited data or for customizing pre-trained models for specific applications. In this study, A total of twenty-four fine-tuned models were employed with the primary goal of identifying the optimal model that could achieve the peak classification effectiveness. To accomplish this, a comprehensive comparison among diverse tested algorithms was conducted. These methodologies encompassed a range of well-known models, including ResNet (50, 50V2, 101, 101V2, 152, and 152V2), MobileNet (V1 and V2), Inception (V3 and ResNetV2), Xception, VGG (16), DenseNet (121, 169, and 201), NasNet (Mobile and Large), and EfficientNet (B0, B1, B2, B3, B4, B5, and B6).

3.1. ResNet

ResNet, short for Residual Networks, is a pivotal advancement in deep learning and convolutional neural networks (CNNs). It was introduced to overcome the problem of vanishing gradients in deep networks by introducing residual connections, allowing information to flow more effectively [51]. I extensively explored multiple ResNet models, including ResNet50, ResNet50V2, ResNet101, ResNet101V2, ResNet152, and ResNet152V2, among which ResNet101V2 exhibited the most promising outcomes. Enclosed is a Figure illustrating the architecture of this particularly successful model in Figure 3.

3.2. MobileNet

MobileNet is a family of neural network architectures tailored for mobile and resource-limited devices. Their key feature is efficiency, achieved through depth-wise separable convolutions that reduce computational demands while maintaining accuracy. It striked a balance between speed and accuracy, thus serving as a valuable tool for deploying deep learning models on devices with limited computing power [53]. I delved into various MobileNet models—MobileNet, and MobileNetV2. Among these, the performance of MobileNet stood out as the most impressive, showcasing superior results. Figure 4 included a visual representation detailing the architecture of this impactful model for reference. (A) The complete structure of MobileNet and (B) a detailed elucidation of the DS layer.

3.3. Inception

Inception, also known as GoogleNet, is a pioneering deep neural network architecture. Inception modules employ parallel convolutional layers with different filter sizes to capture multi-scale features efficiently. Inception was crafted to tackle the challenges of deep neural networks, offering a balance between model depth and computational efficiency. This architecture's ability to reduce parameters while maintaining high accuracy has made it a crucial component in image classification [55]. I thoroughly examined several Inception models, including InceptionV3, and InceptionResNetV2. Among these variations, InceptionResNetV2 emerged as the standout performer, exhibiting exceptional results. Additionally, Figure 5 incorporated a visual representation elucidating the architecture of this highly effective model for review.

3.4. Xception

Xception is a neural network architecture that extends the Inception model by utilizing depth-wise separable convolutions. This innovative technique isolates spatial and channel-wise convolutions, resulting in a significant reduction in model parameters without compromising performance. Xception's strength lies in its efficient feature extraction, making it particularly well-suited for computer vision tasks like image classification [57]. Figure 6 included a graphical illustration that clarifies the structure of this exceptionally efficient model for examination.

3.5. VGG

VGG, or Visual Geometry Group, is a widely recognized deep neural network architecture known for its straightforward and effective design. It distinguishes itself by employing a deep stack of 3x3 convolutional filters, resulting in a uniform and easy-to-understand structure. Although VGG models are relatively deep and parameter-heavy compared to some alternatives, their simplicity has made them a pivotal benchmark in the realm of deep learning. VGG networks have demonstrated remarkable performance in image classification [59]. I employed the VGG16 model for my analysis, and Figure 7 serves as a graphical depiction elucidating its architectural layout.

3.6. DenseNet

DenseNet, or Densely Connected Convolutional Networks, is an influential deep learning architecture distinguished by its dense interlayer connections. Unlike conventional convolutional neural networks (CNNs), DenseNet employs dense connections where each layer is directly connected to every subsequent layer. This dense connectivity promotes enhanced feature reuse, smoother gradient flow, and more compact model designs [61]. I extensively explored multiple DenseNet models—DenseNet121, DenseNet169, and DenseNet201.

Due to its superior performance compared to the other two models, DenseNet201 stands out notably, especially when trained with a batch size of 24. Figure 8 included architectural representations that depict this model's configuration.

3.7. NasNet

NASNet, or Neural Architecture Search Network, is an automated deep learning architecture that employs reinforcement learning to autonomously design neural network structures. Unlike traditional manual design methods, NASNet uses a search algorithm to discover optimal architectures for specific tasks, reducing the need for human expertise in architecture crafting [63]. I delved into different NasNet models, encompassing NasNetMobile and NasNetLarge. NasNetLarge distinguishes itself significantly due to its better performance in comparison to the other model. Figure 9 provided architectural representations illustrating the configuration of this standout model.

3.8. EfficientNet

EfficientNet represents a family of convolutional neural network architectures celebrated for their exceptional performance-efficiency balance. Through a unique scaling technique that adjusts network width, depth, and resolution uniformly, EfficientNet optimizes model size and computational demands [65]. I thoroughly examined various EfficientNet models—EfficientNetB0, EfficientNetB1, EfficientNetB2, EfficientNetB3, EfficientNetB4, EfficientNetB5, EfficientNetB6, and EfficientNetB7. Given that all EfficientNet models yield identical results, Figure 10 opted to utilize EfficientNetB0 as a representative example for showcasing the EfficientNet architecture (Figure 10).

4. Evaluation metrics

Key performance metrics of significant importance include precision, recall, sensitivity, specificity, and accuracy. To compute these evaluation metrics effectively, the analysis relies on four essential variables. Metrics such as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) play a critical role in accurately evaluating the model's performance [67].

4.1. Accuracy

Accuracy denotes the proportion of correctly identified cases among the total instances.

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

4.2. Precision

Precision represents the fraction of accurately predicted positive results among all outcomes identified as positive.

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

4.3. Recall

Recall signifies the ratio of accurately predicted events to all the events that were anticipated.

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

4.4. Sensitivity

Sensitivity measures the average fraction of true positives that are accurately recognized among all actual positives.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (4)$$

4.5. Specificity

Specificity quantifies the proportion of correctly identified negative values among all actual negatives.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (5)$$

5. Results and discussion

The dataset was divided into two parts, with 80% allocated for training and 20% for testing purposes. The training and testing phases were executed using Google Colab as the computing environment. The experiments were carried out specifically on the DDTI dataset, and their outcomes are elaborated upon in the subsequent discussions. In this section, the outcomes of the experiments are outlined. The efficiency of the network is significantly influenced by the learning rate, batch size, and dimensions of the input image. These factors play substantial roles in determining the network's performance.

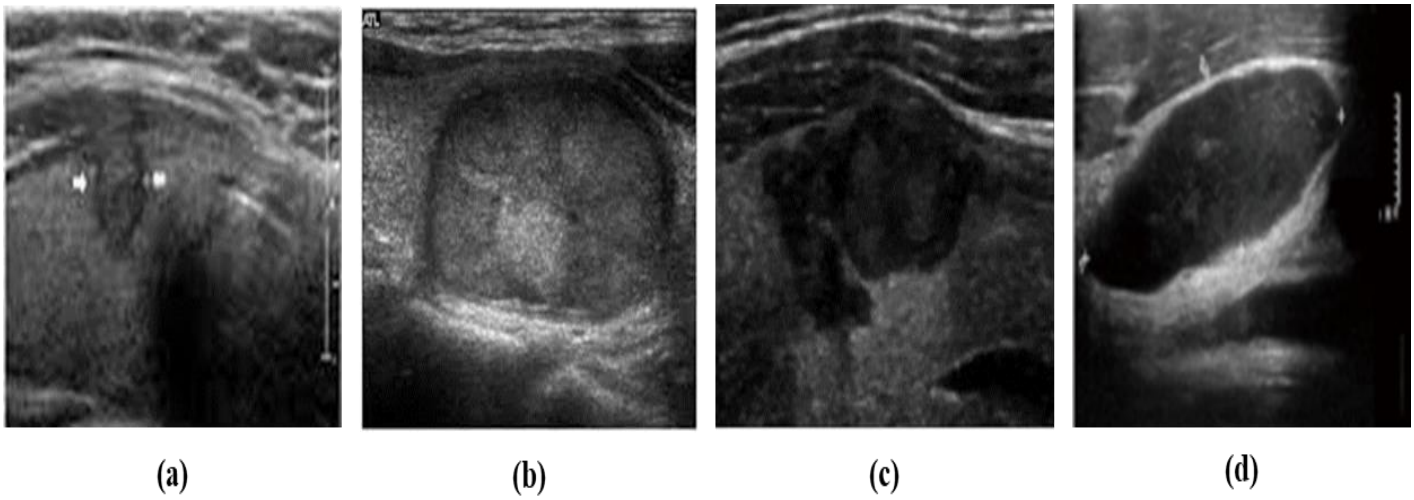

Figure 1: Ultrasound images for different types of thyroid carcinoma: (a) Papillary carcinoma, (b) Follicular carcinoma, (c) Medullary carcinoma, (d) Anaplastic carcinoma.

Table 1: A comparison of the relevant research from the perspectives of image modality, classification methods, outcomes, and publishing information

Author, Year	Image Modalities	Method	Sensitivity	Specificity	Accuracy
L.C. Long 2023 [25]	Ultrasound images	CascadeMaskR-CNN	93%	95%	94%
X.Zhang 2022 [26]	Ultrasound imaging and computed tomography (CT)	Xception neural network	94%	—	98%
Y.J.Lin 2021 [27]	Whole slide imaging	VGG16	94%	—	99%
A.Naglah 2021 [28]	MRI	Multi-input CNN	69%	97%	87%
A.Naglah 2021 [29]	MRI	Multi-input CNN	82%	-	88%
Y.Liu 2021 [30]	Ultrasound images	ThyNet	94%	81%	—
S.Peng 2021 [31]	Ultrasound images	ThyNet	94%	81%	89%
W.K.Chan 2021 [32]	Ultrasound images	ResNet101	72.5%	81.4%	77.6%
J.H.Lee 2020 [33]	CT images	Xception	80.2%	83%	82.5%
M.S.Kavetha 2020 [34]	Post-ablation whole-body planar scans (RxWBSs)	MFDN	—	85%	93%
V.Kumor 2020 [35]	Sonographic images	Multiprong CNN (MPCNN)	88%	73%	—
C.Sun 2020 [36]	Ultrasound images	SVM + CNN	96.4%	83.1%	92.5%
Y.Wang 2020 [37]	Ultrasound images	VGG16	63%	80%	74%
S.W.Kwon 2020 [38]	Ultrasound images	VGG16	70%	92%	—
F.Abdolali 2020 [39]	Ultrasound images	Mask R-CNN	79%	—	—
Y. Lu 2020 [40]	Ultrasound images	CascadeMaskR-CNN	93%	95%	94%
J.H.Lee 2019 [41]	CT images	VGG19	89%	94%	92%
X.Li 2019 [42]	Sonographic images	Deep convolutional neural network (DCCN)	93%	86%	89%
Q.Guan 2019 [43]	Ultrasound images	Inception v3	93.3%	87.4%	~95%
P. Tsou 2019 [44]	Histopathology images	Google inception v3	—	—	95%
H.Li 2018 [45]	Ultrasound images	R-CNN	81%	—	—

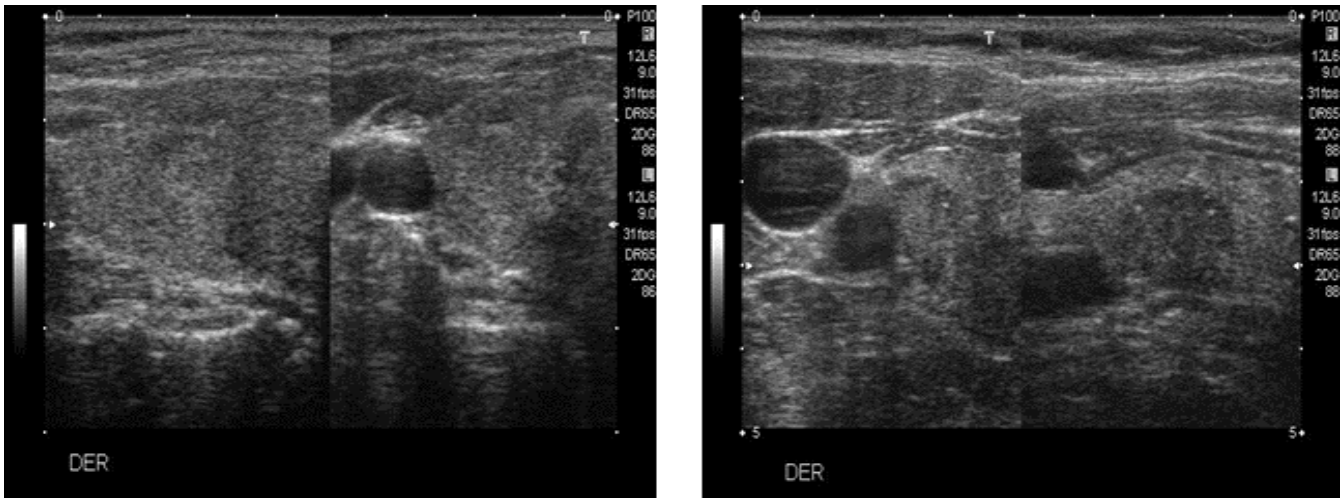

Figure 2: TIRADS 2 has ill-defined margins and micro-calcification on the left; TIRADS 5 on the right has well-defined margins and micro-calcification and is labelled "malignant."

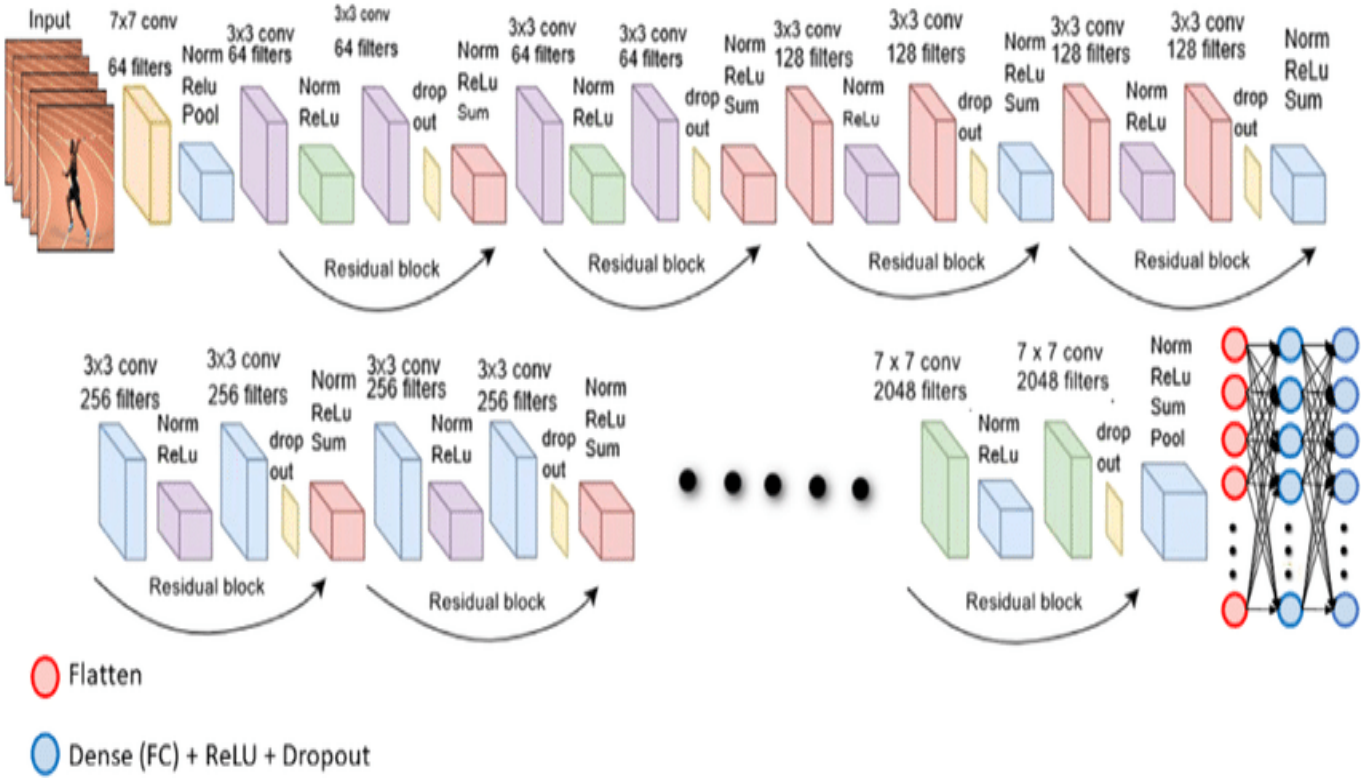

Figure 3: ResNet101V2 architecture diagram [52].

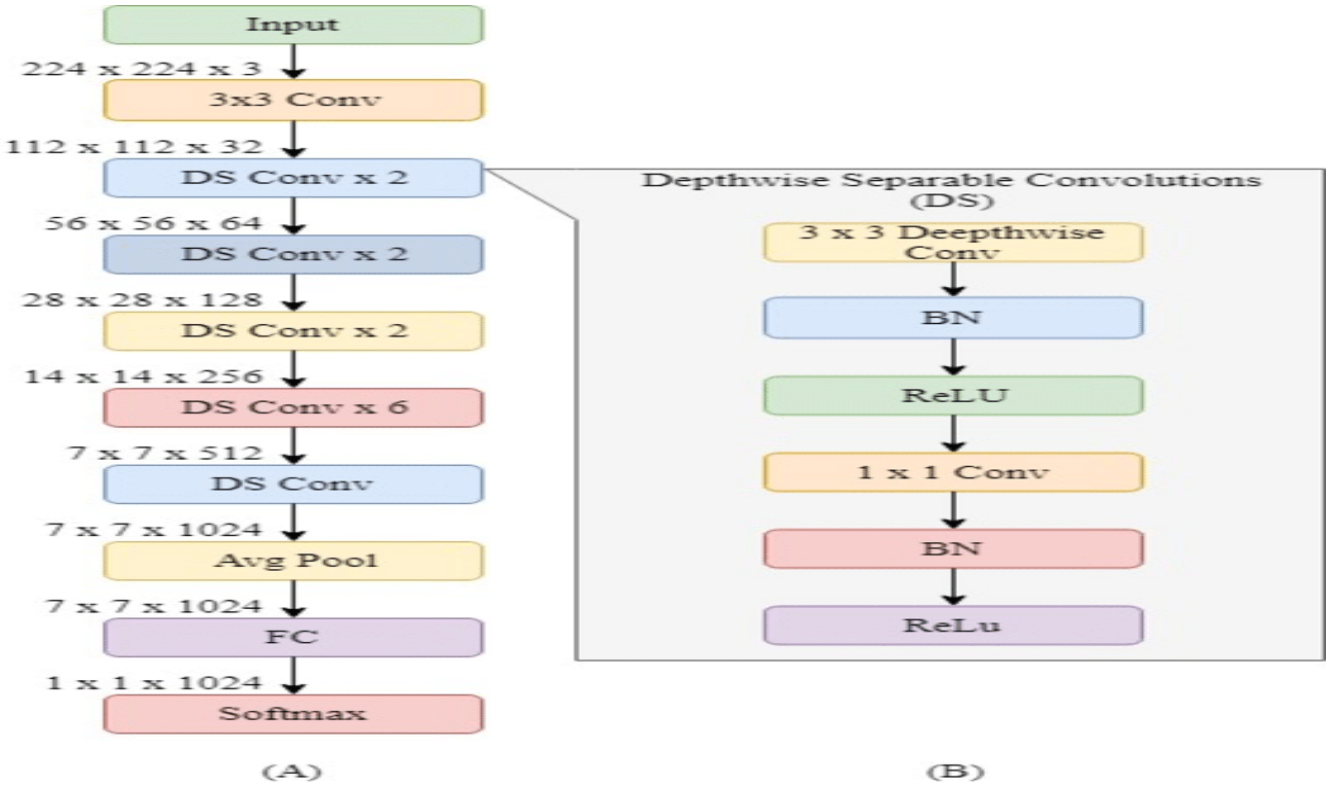

Figure 4: MobileNet architecture diagram [54].

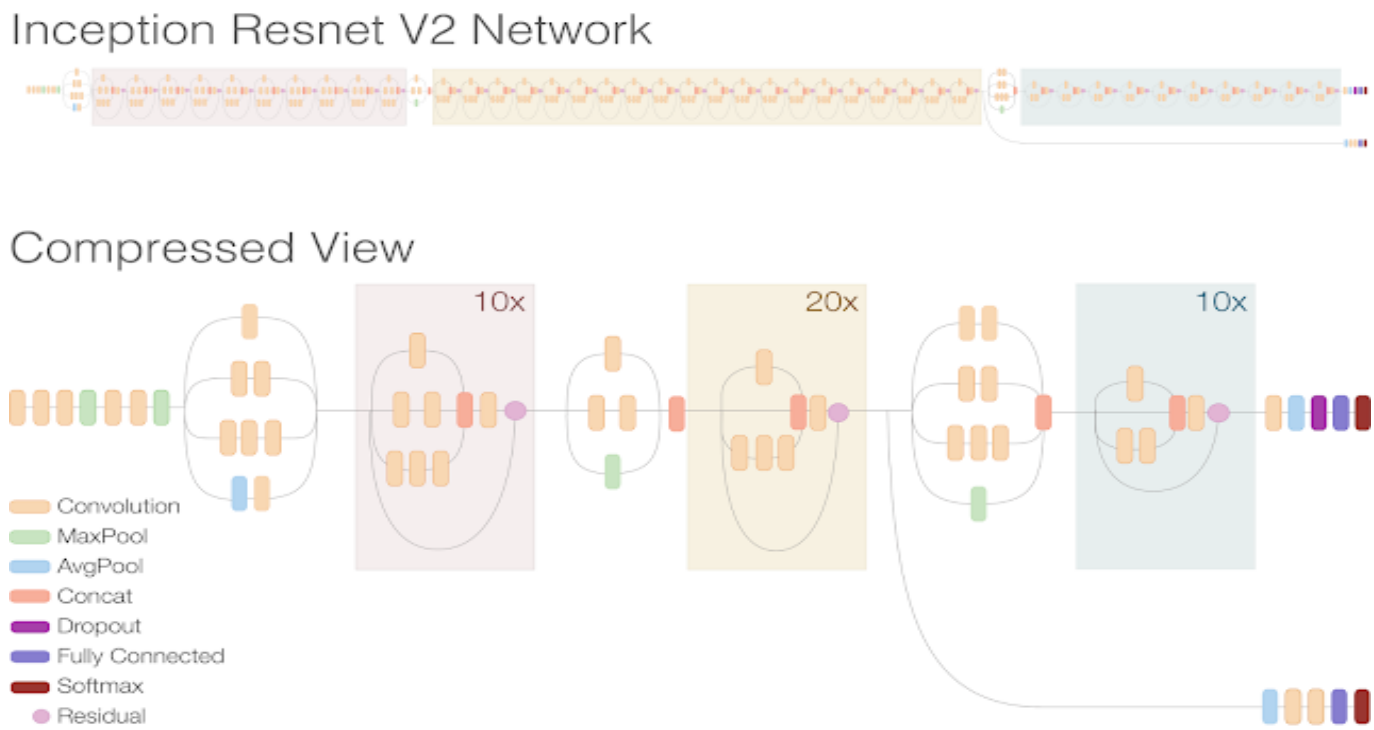

Figure 5: InceptionResNetV2 architecture schematic diagram [56].

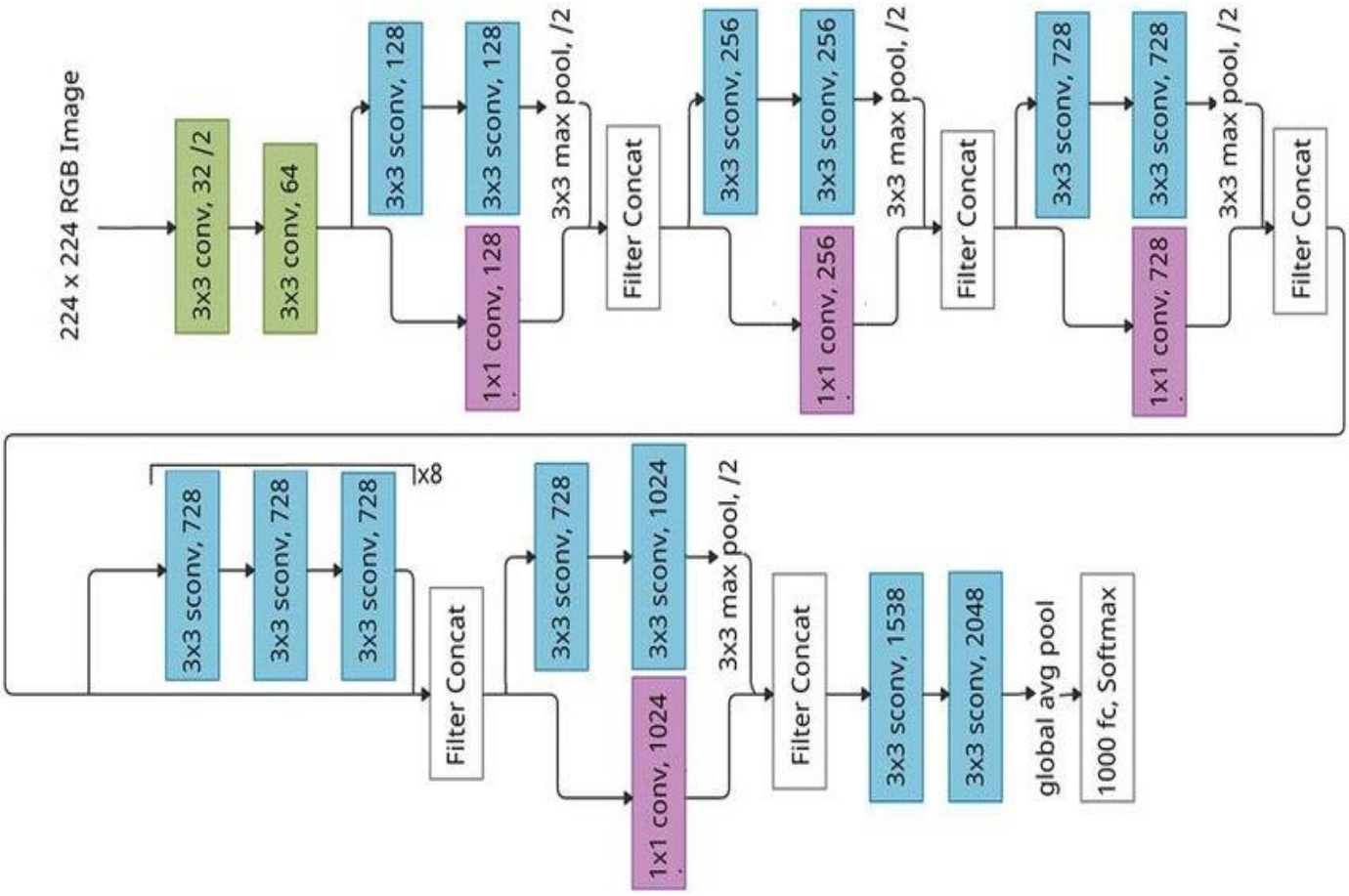

Figure 6: Xception deep CNN architecture diagram [58].

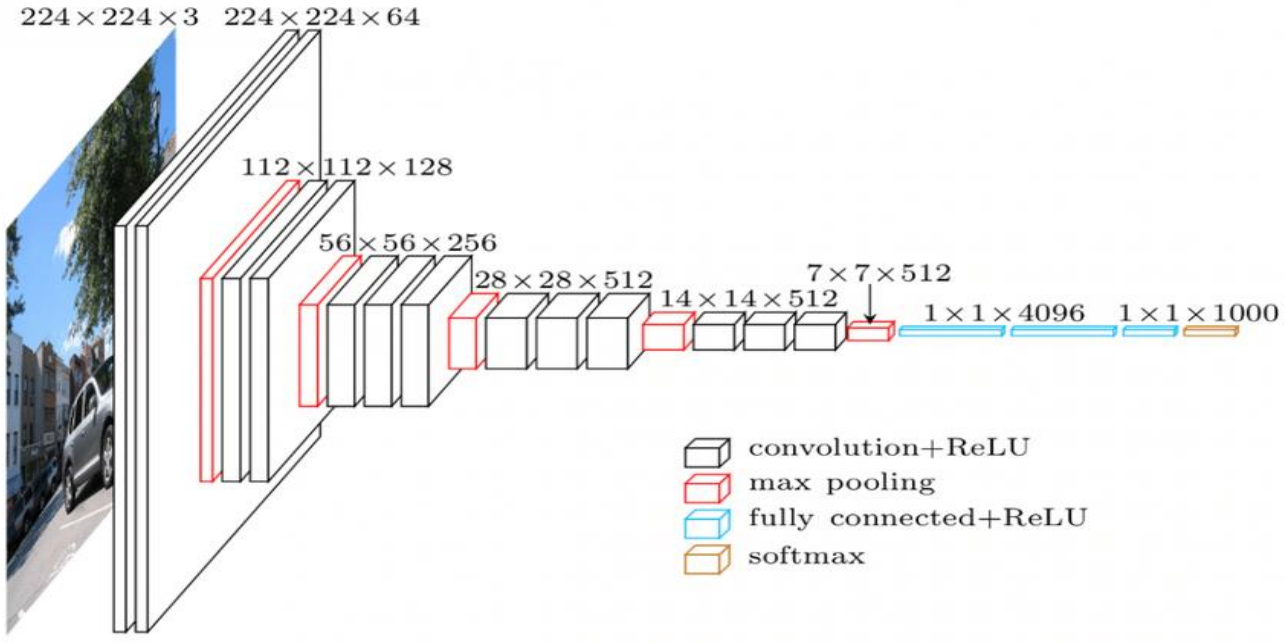

Figure 7: Vgg16 neural network architecture diagram [60].

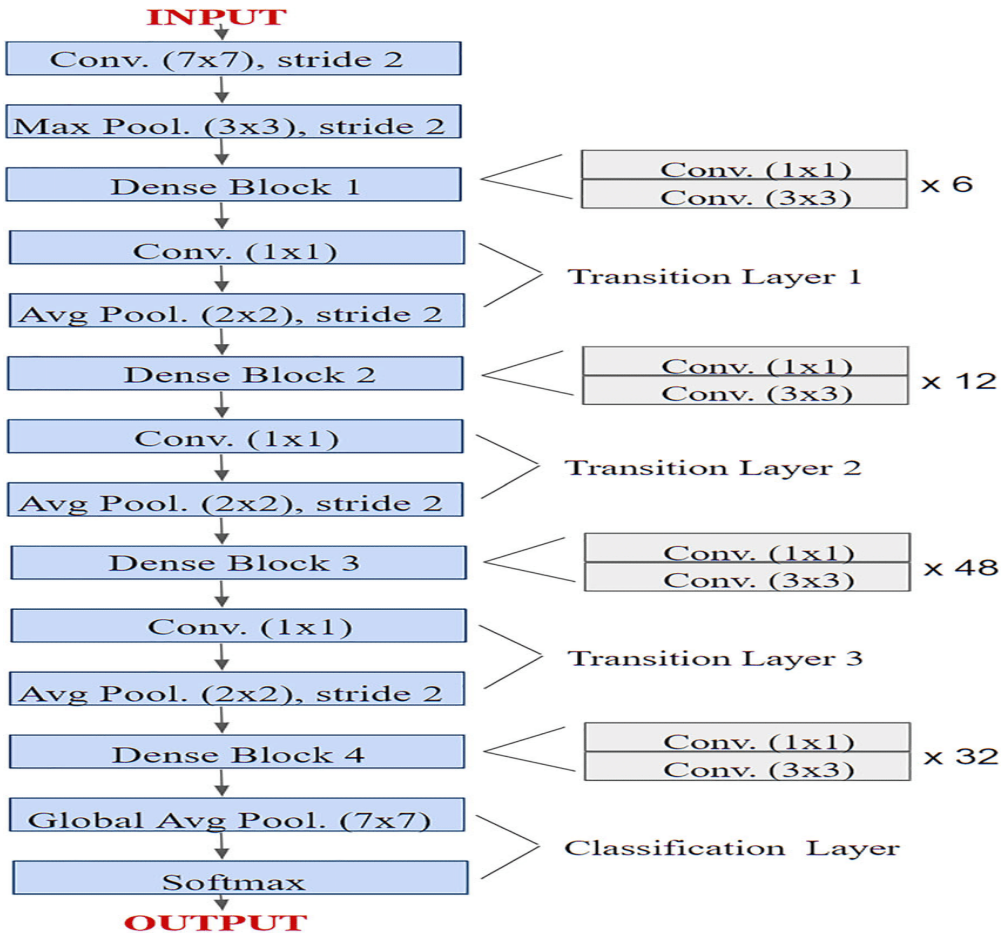

Figure 8: DenseNet201 layered architecture diagram [62].

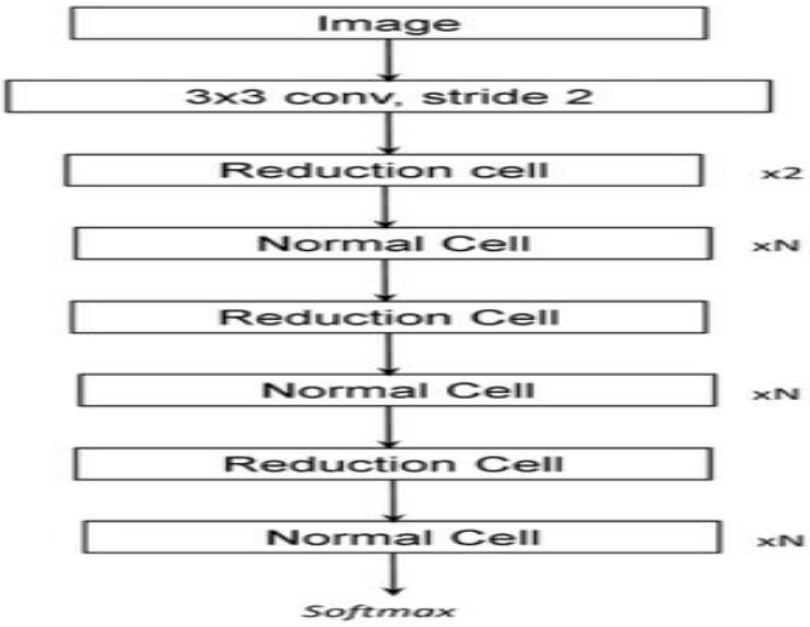

Figure 9: NasNetLarge architecture diagram [64].

Table 2: Comparison of various Fine-tuned models in binary classification task using batch size of 32.

Model Name	Accuracy	precision	recall	sensitivity	Specificity
ResNet50	82.3%	82.6%	82.6%	95.1%	95.1%
ResNet50V2	**94.1%**	**94.7%**	**94.7%**	**99.9%**	**99.9%**
ResNet101	82.3%	82.5%	82.5%	93.4%	93.4%
ResNet101V2	**97%**	**94%**	**94%**	**99.7%**	**99.7%**
ResNet152	82.3%	82.5%	82.5%	93.4%	93.4%
ResNet152V2	91.1%	91.1%	91.1%	99.9%	99.9%
MobileNet	**99.9%**	**99.7%**	**99.7%**	**99.7%**	**99.7%**
MobileNetV2	**94.1%**	**94.1%**	**94.1%**	**99.9%**	**99.9%**
InceptionV3	82.3%	82.5%	82.5%	96.1%	96.1%
InceptionResNetV2	**99.9%**	**99.9%**	**99.9%**	**99.9%**	**99.9%**
Xception	91.1%	92.5%	92.5%	99.2%	99.2%
Vgg16	**99.9%**	**99.9%**	**99.9%**	**99.9%**	**99.9%**
DenseNet121	82.3%	82.5%	82.5%	97.9%	97.9%
DenseNet169	85.2%	85.6%	85.6%	96.6%	96.6%
DenseNet201	85.2%	84.1%	84.1%	97.3%	97.3%
NasNetMobile	82.3%	82.5%	82.5%	95.3%	95.3%
NasNetLarge	85.2%	84.3%	84.3%	96.3%	96.3%
EfficientNetB0	82.3%	82.5%	82.5%	82.5%	82.5%
EfficientNetB1	82.3%	82.5%	82.5%	82.5%	82.5%
EfficientNetB2	82.3%	82.5%	82.5%	91.1%	91.1%
EfficientNetB3	82.3%	82.5%	82.5%	89.5%	89.5%
EfficientNetB4	82.3%	82.5%	82.5%	85.8%	85.8%
EfficientNetB5	82.3%	82.5%	82.5%	90.3%	90.3%
EfficientNetB6	82.3%	82.5%	82.5%	92.7%	92.7%

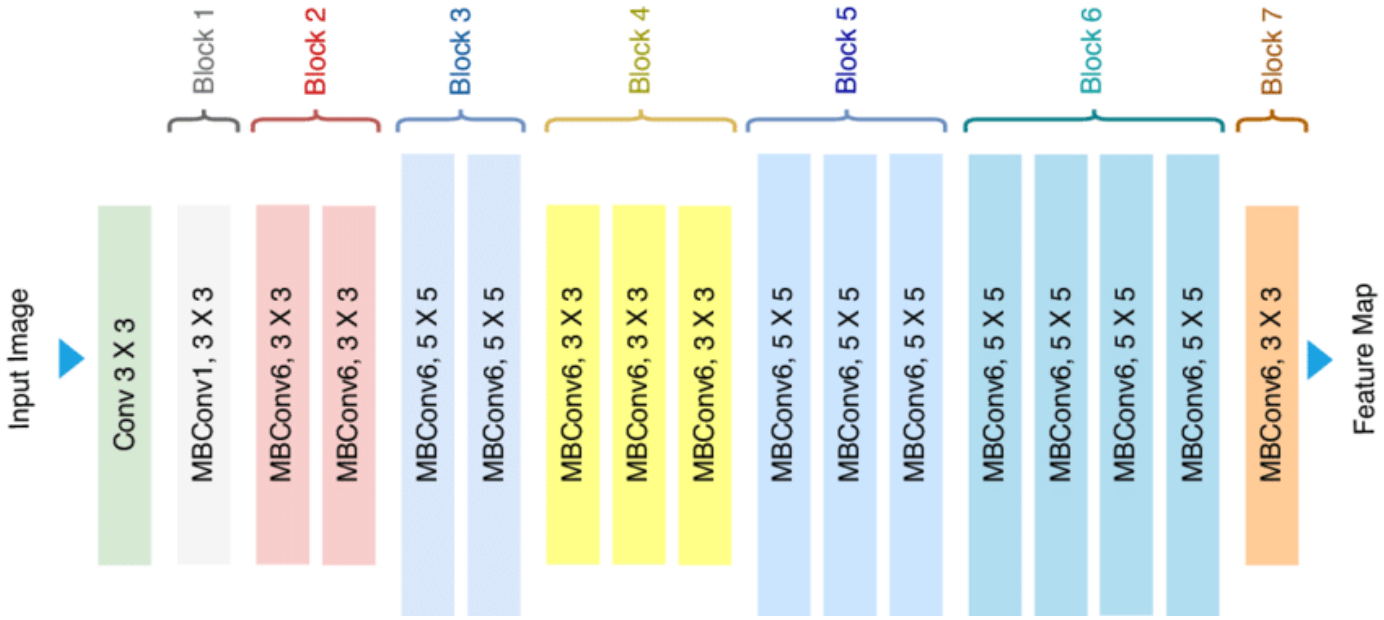

Figure 10: EfficientNetB0 architecture diagram [66].

Table 3: Comparing different Fine-tuned models in binary classification task employing batch size of 24.

Model Name	Accuracy	Precision	Recall	Sensitivity	Specificity
ResNet50	82.3%	82.5%	82.5%	96.1%	96.1%
ResNet50V2	**94.1%**	**94.7%**	**94.7%**	**99.7%**	**99.7%**
ResNet101	82.3%	82.5%	82.5%	93.4%	93.4%
ResNet101V2	91.1%	91.9%	91.9%	98.9%	98.9%
ResNet152	82.3%	82.5%	82.5%	95%	95%
ResNet152V2	85.2%	89.3%	89.3%	99.2%	99.2%
MobileNet	**99.7%**	**99.7%**	**99.7%**	**99.7%**	**99.7%**
MobileNetV2	91.1%	92.7%	92.7%	99.9%	99.9%
InceptionV3	82.3%	82.7%	82.7%	95.2%	95.2%
InceptionResNetV2	**94.1%**	**96%**	**96%**	**99.7%**	**99.7%**
Xception	**94.1%**	**94.2%**	**94.2%**	**99.5%**	**99.5%**
Vgg16	**99.9%**	**99.7%**	**99.7%**	**99.9%**	**99.9%**
DenseNet121	82.3%	82.5%	82.5%	97.9%	97.9%
DenseNet169	82.3%	85.6%	85.6%	96.6%	96.6%
DenseNet201	85.2%	84.1%	84.1%	98.1%	98.1%
NasNetMobile	85.2%	82.8%	82.8%	95.5%	95.5%
NasNetLarge	85.2%	84.6%	84.6%	96.8%	96.8%
EfficientNetB0	82.3%	82.5%	82.5%	89%	89%
EfficientNetB1	82.3%	82.5%	82.5%	82.5%	82.5%
EfficientNetB2	82.3%	82.5%	82.5%	85.6%	85.6%
EfficientNetB3	82.3%	82.5%	82.5%	90.1%	90.1%
EfficientNetB4	82.3%	82.5%	82.5%	85.2%	85.2%
EfficientNetB5	82.3%	82.5%	82.5%	88.2%	88.2%
EfficientNetB6	82.3%	82.5%	82.5%	90.8%	90.8%

Differences in classification sensitivity, specificity, precision, recall, and accuracy were observed across various batch sizes while maintaining a constant split ratio, image size, number of epochs, learning rate, and optimizer (20%–80%, 224 × 224, 400, 0.0001, and Adam, respectively) for the DDTI dataset. Notably, the most optimal classification performance was achieved specifically when utilizing a batch size of 32. Results are summarized in Table 2 and Table 3. A comprehensive comparison of twenty-four alternative models revealed that Vgg16, MobileNet, and inceptionresnetv2 models consistently delivered the most favorable outcomes across various metrics using batch size of 32. These advancements in deep learning have spurred the creation and evaluation of sophisticated models specifically for identifying malignant thyroid nodules. In response, a versatile framework was devised, leveraging CNN architecture and the aforementioned twenty-four models. This study aimed to address the challenge of early thyroid cancer detection using thyroid ultrasound images. For this purpose, a Convolutional Neural Network model employing transfer learning and fine-tuning with pretrained models was employed to distinguish between malignant and benign images. The experiment's effectiveness was gauged through accuracy, precision, recall, sensitivity, and specificity metrics. Notably, Vgg16, MobileNet, and inceptionresnetv2 models, employing Transfer Learning for accelerated learning, achieved an outstanding recognition accuracy of 0.999. The results, delineated in Table 2 for a batch size of 32 and Table 3 for a batch size of 24, were obtained utilizing the adam optimizer across 400 epochs.

6. Conclusions

Diagnosing the severity of thyroid cancer with exceptional accuracy represents a formidable task in current biomedical research. This study addresses this challenge by introducing a deep learning-based approach designed to effectively detect thyroid diseases using a dataset comprising 350 samples categorized into six TIRAD (Thyroid Imaging Reporting and Data System) classes but with binary target classification. To achieve the objective, Multiple deep learning algorithms, including ResNet, MobileNet, Inception, Xception, DenseNet, NasNet, and EfficientNet, were extensively explored and evaluated. This was followed by a comprehensive performance comparison of these algorithms. In the evaluation of models, VGG16, MobileNet, and InceptionResNetV2 demonstrated exceptional performance, achieving accuracy rates of 99.9%, respectively, along with precision scores of 99.9% but MobileNet has precision score 99.7%. Following closely, ResNet101V2 secured the second position with an accuracy of 97% and a precision of 94%. These evaluations encompassed both validation and testing phases. Future research endeavors will involve the implementation of more sophisticated machine learning (ML) and deep learning (DL) algorithms. These algorithms will be applied to diverse datasets, including both statistical data and medical images. The primary aim is to enhance the efficiency and effectiveness of thyroid cancer disease prediction and diagnosis, particularly in the early stages of detection.

References

[1] K. P. Win, Y. Kitjaidure, K. Hamamoto, T. Myo Aung. (2020). Computer-assisted screening for cervical cancer using digital image processing of pap smear images. Applied Sciences. 10 (5): e1800.

[2] M. D. Bethesda. (2018). SEER Cancer Stat Facts Thyroid Cancer. National Cancer Institute (accessed on 10 May 2021).

[3] J. Kim, J. E. Gosnell, S. A. Roman. (2020). Geographic influences in the global rise of thyroid cancer. Nature Reviews Endocrinology. 16 (1): e17-e29.

[4] Y. Liu, Y. Xu, X. Meng, X. Wang, T. Bai. (2020). A study on the auxiliary diagnosis of thyroid disease images based on multiple dimensional deep learning algorithms. Current Medical Imaging. 16 (3): e199-e205.

[5] R. M. Feng, Y. N. Zong, S. M. Cao, R. H. Xu. (2019). Current cancer situation in China: good or bad news from the 2018 Global Cancer Statistics? Cancer communications. 39 (1): e1-e12.

[6] D. A. Watters, A. T. Ahuja, R. M. Evans, W. Chick, W. W. King, C. Metreweli, A. K. Li. (1992). Role of ultrasound in the management of thyroid nodules. The American journal of surgery. 164 (6): e654-e657.

[7] Y. Yao, X. Chen, S. Wu, L. Guo, H. Zhang, Q. Zhu, J. Tang, F. Luan, Y. Zhao, F. Lv, Y. He. (2018). Thyroid nodules in centenarians: prevalence and relationship to lifestyle characteristics and dietary habits. Clinical interventions in aging. 1 (1): e515-e522.

[8] H. J. Tae, D. J. Lim, K. H. Baek, W. C. Park, Y. S. Lee, J. E. Choi, J. M. Lee, M. I. Kang, B. Y. Cha, H. Y. Son, K. W. Lee, S. K. Kang. (2007). Diagnostic value of ultrasonography to distinguish between benign and malignant lesions in the management of thyroid nodules. Thyroid. 17 (5): e461-e466.

[9] A. S. McQueen, K. S. Bhatia. (2015). Thyroid nodule ultrasound: technical advances and future horizons. Insights into imaging. 6 (2): e173-e188.

[10] A. Prochazka, S. Gulati, S. Holinka, D. Smutek. (2019). Classification of thyroid nodules in ultrasound images using direction-independent features extracted by two-threshold binary decomposition. Technology in cancer research & treatment. 18 (1): e1533033819830748.

[11] Y. Mokli, J. Pfaff, D. P. Dos Santos, C. Herweh, S. Nagel. (2019). Computer-aided imaging analysis in acute ischemic stroke–background and clinical applications. Neurological Research and Practice. 1 (1): e1-e13.

[12] N. Chambara, S. Y. Liu, X. Lo, M. Ying. (2021). Diagnostic performance evaluation of different TI-RADS using ultrasound computer-aided diagnosis of thyroid nodules: An experience with adjusted settings. PloS one. 16 (1): e0245617.

[13] W. Ahmad, A. Ahmad, C. Lu, B. A. Khoso, L. Huang. (2018). A novel hybrid decision support system for thyroid disease forecasting. Soft Computing. 22 (1): e5377-e5383.

[14] G. R. Banu. (2016). Predicting thyroid disease using linear discriminant analysis (LDA) data mining technique. Communications on Applied Electronics (CAE). 4 (1): e4-e6.

[15] G. R. Banu. (2016). A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease. International Journal of Computer Sciences and Engineering. 4 (11): e64-e70.

[16] D. T. Nguyen, J. K. Kang, T. D. Pham, G. Batchuluun, K. R. Park. (2020). Ultrasound image-based diagnosis of malignant thyroid nodule using artificial intelligence. Sensors. 20 (7): e1822.

[17] Y. Chang, A. K. Paul, N. Kim, J. H. Baek, Y. J. Choi, E. J. Ha, K. D. Lee, H. S. Lee, D. Shin, N. Kim. (2016). Computer-aided diagnosis for classifying benign versus malignant thyroid nodules based on ultrasound images: a comparison with radiologist-based assessments. Medical physics. 43 (1): e554-e567.

[18] J. Ma, F. Wu, J. Zhu, D. Xu, D. Kong. (2017). A pre-trained convolutional neural network-based method for thyroid nodule diagnosis. Ultrasonics. 73 (1): e221-e230.

[19] S. Y. Ko, J. H. Lee, J. H. Yoon, H. Na, E. Hong, K. Han, I. Jung, E. K. Kim, H. J. Moon, V. Y. Park, E. Lee, J. Y. Kwak. (2019). Deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. Head & neck. 41 (4): e885-e891.

[20] J. Xia, H. Chen, Q. Li, M. Zhou, L. Chen, Z. Cai, H. Zhou. (2017). Ultrasound-based differentiation of malignant and benign thyroid Nodules: An extreme learning machine approach. Computer methods and programs in biomedicine. 147 (1): e37-e49.

[21] R. D. Chernock, I. S. Hagemann. (2015). Molecular pathology of hereditary and sporadic medullary thyroid carcinomas. American journal of clinical pathology. 143 (6): e768-e777.

[22] C. E. Silver, R. P. Owen, J. P. Rodrigo, A. Rinaldo, K. O. Devaney, A. Ferlito. (2011). Aggressive variants of papillary thyroid carcinoma. Head & neck. 33 (7): e1052-e1059.

[23] H. Chen, R. S. Sippel, M. S. O'Dorisio, A. I. Vinik, R. V. Lloyd, K. Pacak. (2010). The North American Neuroendocrine Tumor Society consensus guideline for the diagnosis and management of neuroendocrine tumors: pheochromocytoma, paraganglioma, and medullary thyroid cancer. Pancreas. 39 (6): e775-e783.

[24] S. Yamashita, V. Saenko. (2007). Mechanisms of disease: molecular genetics of childhood thyroid cancers. Nature clinical practice Endocrinology & metabolism. 3 (5): e422-e429.

[25] L. C. Long, Y. Bui Hoang, N. L. Trung, B. T. Dung, T. T. Ha, L. V. Nguyen. (2023, October). A Review in Deep Learning-Based Thyroid Cancer Detection Techniques Using Ultrasound Images. In International Conference on Intelligence of Things (pp. 15-25). Cham: Springer Nature Switzerland.

[26] X. Zhang, V. C. Lee, J. Rong, F. Liu, H. Kong. (2022). Multi-channel convolutional neural network architectures for thyroid cancer detection. PLoS One. 17 (1): e0262128.

[27] Y. J. Lin, T. K. Chao, M. A. Khalil, Y. C. Lee, D. Z. Hong, J. J. Wu, C. W. Wang. (2021). Deep learning fast screening approach on cytological whole slides for thyroid cancer diagnosis. Cancers. 13 (15): e3891.

[28] A. Naglah, F. Khalifa, R. Khaled, A. A. K. Abdel Razek, M. Ghazal, G. Giridharan, A. El-Baz. (2021). Novel MRI-based CAD system for early detection of thyroid cancer using multi-input CNN. Sensors. 21 (11): e3878.

[29] A. Naglah, F. Khalifa, R. Khaled, A. El-Baz. (2021). Thyroid cancer computer-aided diagnosis system using MRI-based multi-input CNN model. In 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) (pp. 1691-1694). IEEE.

[30] Y. Liu, J. Liang, S. Peng, W. Wang, H. Xiao. (2021). A deep-learning model to assist thyroid nodule diagnosis and management–Authors' reply. The Lancet Digital Health. 3 (7): e411-e412.

[31] S. Peng, Y. Liu, W. Lv, L. Liu, Q. Zhou, H. Yang, J. Ren, G. Liu, X. Wang, X. Zhang,Q. Du, H. Xiao. (2021). Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multi-center diagnostic study. The Lancet Digital Health. 3 (4): e250-e259.

[32] W. K. Chan, J. H. Sun, M. J. Liou, Y. R. Li, W. Y. Chou, F. H. Liu, S. T. Chen, S. J. Peng. (2021). Using deep convolutional neural networks for enhanced ultrasonographic image diagnosis of differentiated thyroid cancer. Biomedicines. 9 (12): e1771.

[33] J. H. Lee, E. J. Ha, J. H. Kim. (2019). Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT. European radiology. 29 (1): e5452-e5457.

[34] M. Kavitha, C. H. Lee, K. Shibudas, T. Kurita, B. C. Ahn. (2020). Deep learning enables automated localization of the metastatic lymph node for thyroid cancer on 131I post-ablation whole-body planar scans. Scientific reports. 10 (1): e7738.

[35] V. Kumar, J. Webb, A. Gregory, D. D. Meixner, J. M. Knudsen, M. Callstrom, M. Fatemi, A. Alizad. (2020). Automated segmentation of thyroid nodule, gland, and cystic components from ultrasound images using deep learning. Ieee Access. 8 (1): e63482-e63496.

[36] C. Sun, Y. Zhang, Q. Chang, T. Liu, S. Zhang, X. Wang, L. Niu. (2020). Evaluation of a deep learning-based computer-aided diagnosis system for distinguishing benign from malignant thyroid nodules in ultrasound images. Medical Physics. 47 (9): e3952-e3960.

[37] Y. Wang, W. Yue, X. Li, S. Liu, L. Guo, H. Xu, H. Zhang, G. Yang. (2020). Comparison study of radiomics and deep learning-based methods for thyroid nodules classification using ultrasound images. Ieee Access. 8 (1): e52010-e52017.

[38] S. W. Kwon, I. J. Choi, J. Y. Kang, W. I. Jang, G. H. Lee, M. C. Lee. (2020). Ultrasonographic thyroid nodule classification using a deep convolutional neural network with surgical pathology. Journal of digital imaging. 33 (1): e1202-e1208.

[39] F. Abdolali, J. Kapur, J. L. Jaremko, M. Noga, A. R. Hareendranathan, K. Punithakumar. (2020). Automated thyroid nodule detection from ultrasound imaging using deep convolutional neural networks. Computers in Biology and Medicine. 122 (1): e103871.

[40] F. Abdolali, J. Kapur, J. L. Jaremko, M. Noga, A. R. Hareendranathan, K. Punithakumar. (2020). Automated thyroid nodule detection from ultrasound imaging using deep convolutional neural networks. Computers in Biology and Medicine. 122 (1): e103871.

[41] J. H. Lee, E. J. Ha, J. H. Kim. (2019). Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT. European radiology. 29 (1): e5452-e5457.

[42] X. Li, S. Zhang, Q. Zhang, X. Wei, Y. Pan, J. Zhao, X. Xin, C. Qin, X. Wang, J. Li, K. Chen. (2019). Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. The Lancet Oncology. 20 (2): e193-e201.

[43] Q. Guan, Y. Wang, J. Du, Y. Qin, H. Lu, J. Xiang, F. Wang. (2019). Deep learning-based classification of ultrasound images for thyroid nodules: a large scale of pilot study. Annals of translational medicine. 7 (7).

[44] P. Tsou, C. J. Wu. (2019). Mapping driver mutations to histopathological subtypes in papillary thyroid carcinoma: applying a deep convolutional neural network. Journal of clinical medicine. 8 (10): e1675.

[45] H. Li, J. Weng, Y. Shi, W. Gu, Y. Mao, Y. Wang, W. Liu, J. Zhang. (2018). An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images. Scientific reports. 8 (1): e6600.

[46] L. Pedraza, C. Vargas, F. Narváez, O. Durán, E. Muñoz, E. Romero. (2015, January). An open access thyroid ultrasound image database. In 10th International symposium on medical information processing and analysis. SPIE. 9287 (1): e188-e193).

[47] https://www.kaggle.com/datasets/dasmehdixtr/ddti-thyroid-ultrasound-images/data

[48] K. V. Sai Sundar, K. T. Rajamani, S. Siva Sankara Sai. (2019). Exploring image classification of thyroid ultrasound images using deep learning. In Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB) (pp. 1635-1641). Springer International Publishing.

[49] D. W. Kim, E. J. Lee, H. S. In, S. J. Kim. (2010). Sonographic differentiation of partially cystic thyroid nodules: a prospective study. American journal of neuroradiology. 31 (10): e1961-e1966.

[50] J. Y. Kwak, K. H. Han, J. H. Yoon, H. J. Moon, E. J. Son, S. H. Park, H. K. Jung, J. S. Choi, B. M. Kim, E. K. Kim. (2011). Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. Radiology. 260 (3): e892-e899.

[51] K. He, X. Zhang, S. Ren, J. Sun. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[52] M. Bilal, M. Maqsood, S. Yasmin, N. U. Hasan, S. Rho. (2022). A transfer learning-based efficient spatiotemporal human action recognition framework for long and overlapping action classes. The Journal of Supercomputing. 78 (2): e2873-e2908.

[53] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, H. Adam. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv. 1704 (4): e4861.

[54] S. Phiphiphatphaisit, O. Surinta. (2020). Food image classification with improved MobileNet architecture and data augmentation. In Proceedings of the 3rd International Conference on Information Science and Systems. 1 (1): e51-e56.

[55] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, A. Rabinovich. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition. 1 (1): e1-e9.

[56] https://blog.research.google/2016/08/improving-inception-and-image.html?m=1

[57] F. Chollet. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition. 1 (1): e1251-e1258.

[58] K. Srinivasan, L. Garg, D. Datta, A. A. Alaboudi, N. Z. Jhanjhi, R. Agarwal, A. G. Thomas. (2021). Performance comparison of deep cnn models for detecting driver's distraction. CMC-Computers, Materials & Continua. 68 (3): e4109-e4124.

[59] K. Simonyan, A. Zisserman. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14 (09): e1556.

[60] M. Loukadakis, J. Cano, M. O'Boyle. (2018). Accelerating deep neural networks on low power heterogeneous architectures.

[61] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 3 (1): e4700-e4708.

[62] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, M. Kaur. (2021). Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. Journal of Biomolecular Structure and Dynamics. 39 (15): e5682-e5689.

[63] B. Zoph, V. Vasudevan, J. Shlens, Q. V. Le. (2018). Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 1 (3): e8697-e8710.

[64] S. Albahli, W. Albattah. (2020). Detection of coronavirus disease from X-ray images using deep learning and transfer learning algorithms. Journal of X-ray Science and Technology. 28 (5): e841-e850.

[65] M. Tan, Q. Le. (2019). Efficient net: Rethinking model scaling for convolutional neural networks. In International conference on machine learning. PMLR. 1 (1): e6105-e6114).

[66] T. Ahmed, N. H. N. Sabab. (2022). Classification and understanding of cloud structures via satellite images with Efficient UNet. SN Computer Science. 3 (1): e1-e11.

[67] M. M. El Sherbiny, E. Abdelhalim, H. E. D. Mostafa, M. M. El-Seddik. (2023). Classification of chronic kidney disease based on machine learning techniques. Indonesian Journal of Electrical Engineering and Computer Science. 32 (2): e945-e955.