Enhancing Image Classification Performance with a Computationally Efficient CNN Using Adaptive SqueezeNet

Priyanka Bhatambarekar[1] and Gayatri Phade[2]
[1,2] EXTC dept Sandip Institute of Technology and Research Centre, SPPU, Nashik, India

Abstract

Convolutional Neural Networks (CNNs) have revolutionized image classification tasks across a variety of domains. However, the growing demand for lightweight models suitable for edge devices has led to interest in compact architectures like SqueezeNet. This study proposes an enhanced version of SqueezeNet, integrated within a hybrid CNN framework, to improve classification accuracy while maintaining low computational cost. The proposed method modifies fire modules, integrates residual connections, and incorporates attention mechanisms to better capture discriminative features. Experimental results on benchmark datasets (e.g., CIFAR-10 and ImageNet subsets) demonstrate the improved performance of the proposed model over the baseline SqueezeNet and other lightweight models in terms of accuracy and parameter efficiency.

1. Introduction

Machine learning is one of the greatest problems in computer vision since it underlies several applications, for example, medical image analysis, environmental monitoring via remote sensing, autonomous navigation by vehicles, inspecting industry standards, and smart surveillance systems. For many years, researchers have employed deep Convolutional Neural Networks (CNNs) such as VGGNet, ResNet, and DenseNet to produce very high classification accuracy on complex datasets. Although these models achieve significant effects in their scope, they generally consume much resource, due to their being deep and heavy parameter models, thus not applicable in resource-constrained environments like mobile devices, embedded systems, and edge computing platforms.[1], [2], [3], [4], [5]

On the other hand, SqueezeNet came up with a model that was very lightweight and drastically reduced the number of parameters while allowing reasonable classification accuracy. SqueezeNet's fundamental design principle is replacing larger convolutional filters with smaller ones and applying the so-called 'fire modules' to tightness in the model. Baseline performance for SqueezeNet remains inferior to deeper, more complex CNN architectures, especially on more challenging classification tasks.

While classification accuracy becomes better, there is a need to bridge the gap between computational efficiency and computational accuracy in this research. Improvements are made to the original SqueezeNet architecture through a collection of architectural optimizations that maintained its lightweight nature and improved its representational advantage. These changes introduce new internal structures for fire modules as well as include residual connections to allow better gradient flow, and attention mechanisms for the model to focus on the most first-rate features of the input data. This will permit the model under study to optimally balance speed, memory efficiency, and classification accuracy for practical applications on low-resource platforms without degrading performance.

2. Related Work

It's quite evident that a thorough review of extant literature from sources like Alhichri et al. 2019 and Arifianto & Agoes 2021 has shown a clear advancement in lightweight convolutional neural networks for image classification. It was, however, the work of Iandola et al. (2016) that revolutionized the dawn of SqueezeNet, a miniature CNN that matched all those astronomical expectations the accuracy previously set by AlexNet but with only 50x fewer parameters. It further gave a strong establishment in this case for efficient models in deep learning. The innovation supplied a strong need to deploy deep learning in resource-constrained devices without compromising performance.

These studies then continued to refine and develop domain and task-specific applications for these compact architectures. Such is the case with Alhichri et al. (2019) and Setiawan et al. (2021), where lightweight models similar to SqueezeNet were applicable for remote sensing and satellite imagery, which required fast inference and low memory consumption within embedded and edge systems. Object detection and industrial automation were also included in the extension of these models by Hidayatuloh et al. (2018) and Arifianto & Agoes (2021), where the need for real-time applicability was also placed.

Building on the base of SqueezeNet, the authors Kathirgamaraja et al. (2018a, 2018b) put forward enhancements to the architecture which are mainly concerned with the improved fire module and activation strategy to maximize accuracies. Hence, both contributions were directed to optimize feature selection as well as network depth, thus showing how even tiny models can be dedicated to enhancing performance in medical image classification and interpretation of biomedical data.

Latest works like Liu et al.-2021 and Xu et al.-2024 have proposed integrating attention mechanisms, depth wise separable convolutions, and residual connections in compact CNN architectures to grasp context and space information while providing the efficiency needed for the model. In particular, these enhancements performed really well on CIFAR-10, subsets of ImageNet, and hyperspectral images for remote sensing.

The works by Nandhitha & Roslin (2024) and Naveen et al. (2024) directly discussed hybrid architectures in which the optimized SqueezeNet modules are a part of larger CNN architectures, thus giving the benefit of compactness along with the trade-off of classification accuracy. Their work further discussed evaluative metrics such as confusion matrices, precision-recall scores, and parameter-to-accuracy ratios that provide insight into model performance beyond overall accuracy achievements.

Again, light CNN applications, as shown by Tsivgoulis and Ullah (2022, 2021), use personalized health diagnostics as their focus, in addition to continuous clinical and real-time assessment systems. In this context, thus far, the versatility and practicality of optimized models catapulted into applications requiring sensitive, high-impact uses have been foreseen.

Certainly, the literature surveyed had strong generalization rather than specific findings in the onward progress made concerning building efficient CNN architectures for modern, real-time, low-power environments. From the fundamental principles of SqueezeNet to the integration of advanced feature extraction mechanisms and hybrid models, these contributions reflect focused efforts to meet the dual demands of accuracy and efficiency in practical applications in image classification. [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18]

3. Methodology

3.1 Baseline: SqueezeNet Architecture

SqueezeNet comprises:

- A sequence of fire modules (squeeze: 1x1 conv, expand: 1x1 and 3x3 conv).
- Final convolution layers for classification.
- Global average pooling instead of fully connected layers.

3.2 Proposed Improvements

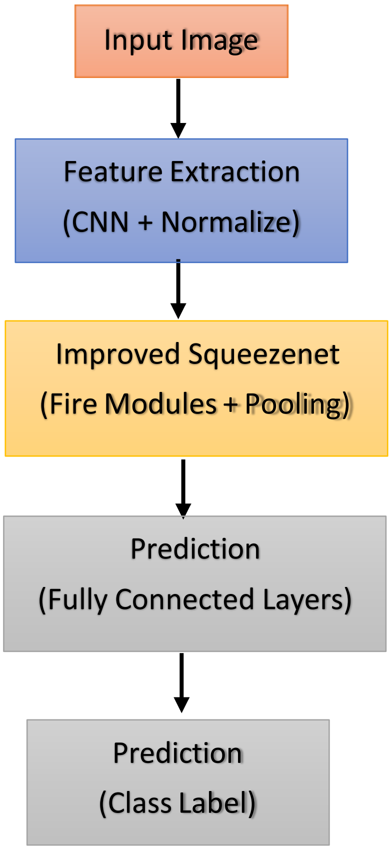

Figure 1: An Efficient Image Classification Framework Using CNN and Improves SqueezNet Architecture

Figure 1presents a modular view of the proposed image classification system based on deep learning. The framework has been developed in four primary stages to sequentially process the input image data to yield the final classification. The whole process starts with the Input Image, which stands for raw represented image data collected from a custom dataset. The input will then flow into the Feature Extraction module, which uses a standard Convolutional Neural Network (CNN)-based architecture along with normalization techniques. The low-level features that will be captured for later high-level processing include basic features, such as edges, textures, and simple shapes. The feature extraction module then forwards this data into an Improved SqueezeNet block, which forms the core of the proposed model. This component is designed with modified "Fire modules" and "pooling layers" for spatial abstraction and at the same time ensure parameter efficiency. The Fire module has a squeeze layer using 1×1 convolutions and an expand layer using a combination of 1×1 and 3×3 convolutions. Here, few improvements to the original SqueezeNet architecture will include attention mechanisms and residual connections to enrich feature representation without significantly increasing the computational burden.

The improved SqueezeNet's output will go into the Prediction module. This includes a fully connected layer that uses a softmax function to map the above high-level feature vector to a set of class probabilities, from which one class is selected with the highest probability as the final Prediction (Class Label). All in all, the architecture is clear of a well-achieved trade-off between efficiency in computation and accuracy in classification, making it beneficial for real time and resource constraint environments where the advantages of the compact CNN design are taken into consideration. The processing initiates from data values classified as nominal, which indeed exist for a very long time before transforming themselves into data with ordinal rankings. The training data in the first configuration would introduce subspaces, one for each class, consisting of only the main class examples. Each subspace constrains the latent space points that can be considered if the input has to represent the main class. They are moved only a little during the minimization of the loss.

In the straightforward case where the model is trained that all the inputs belong to the classes, whenever input occupies a unique outcrop in latent space over any cluster, the perturbation may redistribute this point into a cluster that aligns with the perturbed means of the other clusters. The model can also move the target output classification label towards any classification label in the middle of ideal and input label. Then there is reinforcement learning. This is in opposition to the interleaved training, seeing their best.

The proposed Improved SqueezeNet (ISQNet) integrates the following enhancements:

1. Residual Connections:
   Added between fire modules to ease gradient flow and enhance deep feature representation.
2. Squeeze-and-Excitation (SE) Attention:
   Channel-wise attention modules were added after each fire module to recalibrate feature responses.
3. Expanded Fire Modules:
   Expand layers modified to include depthwise separable convolutions, reducing parameter count while increasing receptive fields.
4. Dropout Regularization:
   Added after select layers to prevent overfitting.

3.3 Overall Architecture

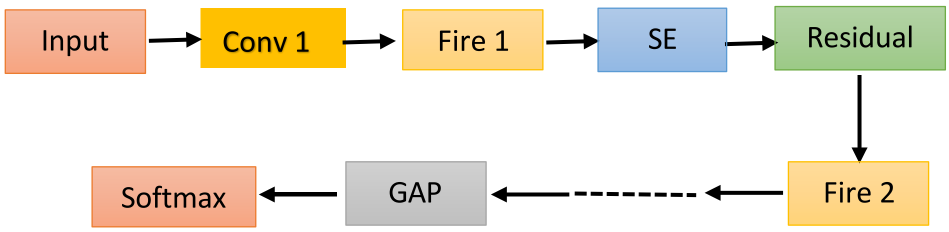

Figure 2: Architecture Framework Using Modified SqueezNet

Figure 2 represents a detailed flow of an enhanced convolutional neural network (CNN) architecture tailored for efficient image classification. The process begins with an input image derived from a custom dataset, which is fed into the initial convolutional layer (Conv1). This layer extracts low-level features such as edges, textures, and simple shapes that form the basis for deeper learning. Following this, the first *Fire module* (Fire1) is applied. This module, central to the SqueezeNet architecture, consists of a squeeze layer using 1×1 convolutions to reduce dimensionality, followed by an expand layer composed of both 1×1 and 3×3 convolutions to capture complex patterns with fewer parameters.

Immediately after Fire1, a Squeeze-and-Excitation (SE) block is introduced. The SE block enhances the network's attention mechanism by adaptively recalibrating channel-wise feature responses, allowing the model to focus on the most informative features. This is followed by a residual connection, which directly links earlier and later layers in the network. Such connections are crucial for addressing the vanishing gradient problem and improving training stability by allowing gradients to propagate more effectively through deeper layers.

The network continues with the second Fire module (Fire2) and potentially additional Fire modules that further abstract and compress the feature representation. These layers work collectively to deepen the model while maintaining computational efficiency. As the feature maps become increasingly abstract, the Global Average Pooling (GAP) layer aggregates the spatial information by computing the average of each feature map, thus reducing dimensionality and preventing overfitting. Finally, a softmax layer is applied, converting the final feature vector into a probability distribution over the target classes. The class with the highest probability is selected as the predicted label. Overall, this architecture integrates the strengths of compact Fire modules, attention mechanisms via SE blocks, and the stability of residual learning, resulting in a high-performing yet lightweight model well-suited for real-time or resource-constrained image classification tasks. The model retains <2MB of weight storage, making it ideal for deployment on mobile and embedded devices.

4. Experimental Setup

4.1 Datasets:

Two benchmark datasets are used to evaluate the proposed model. The CIFAR-10 dataset consists of 60,000 color images, spread evenly among 10 classes, with 6,000 images per class. It is arguably the most popular data set for image classification task and is appealing in terms of balanced and diversified presentation of object categories. Besides that, Tiny ImageNet is a carefully chosen subset of the most popular ImageNet database with images from 200 categories. Each class within Tiny ImageNet has 500 training images, 50 validation images, and 50 testing images, making it a more challenging benchmark due to a high number of categories and small dimensions of the image.

4.2 Evaluation Metrics:

The performance of the model will be assessed using several evaluation metrics guaranteeing a comprehensive analysis. These include accuracy, that is, the correctly classified samples divided by the total samples; model size, measured in megabytes (MB), representing the memory footprint of the network, and FLOPs (Floating Point Operations), indicating the computational complexity and efficiency of the model during inference.

4.3 Training Parameter:

The model used Adam optimizer for training, known for adaptive learning rate properties. The initial learning rate is set to 0.001 and then decayed as training progresses to help convergence. A batch size of 64 is to compromise between training efficiency and good model generalization. This far, all parameter choices have been made to ensure stable and efficient training on the datasets chosen.The model is trained for the 100 epochs. Training the model for 100 epochs provides sufficient opportunity for convergence, allowing it to learn meaningful representations and generalize well to unseen data, while also minimizing the risk of underfitting.

5. Results and Discussion

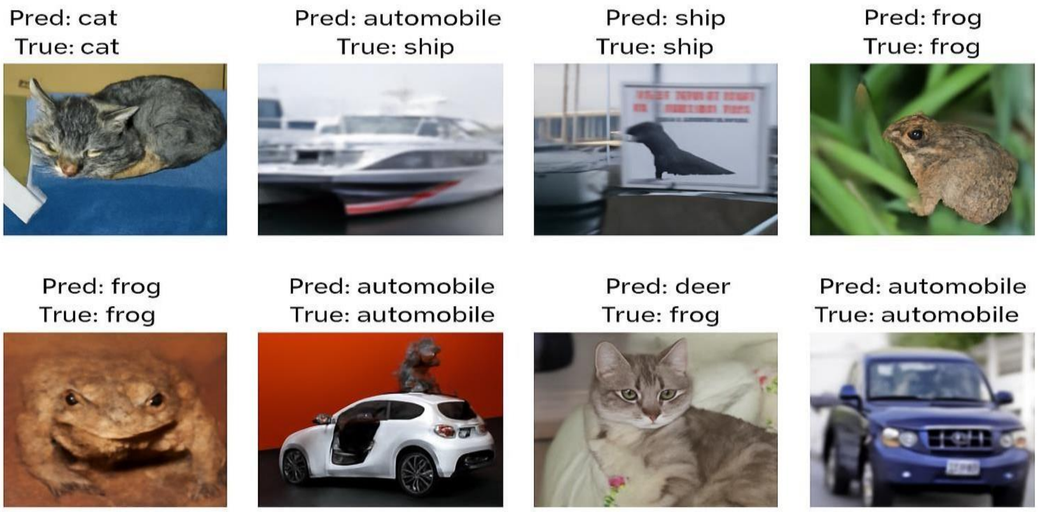

Figure 3: Output Predictions of proposed Modified SqueezNet architecture

Figure 3 elaborates on a few visual instances taken from the CIFAR-10 dataset, demonstrating both correct and incorrect predictions made by the classification model proposed here. The individual prediction (labeled as "Pred:") of each image is contrasted with the ground truth (denoted as "True:"), thus allowing one to weigh the model performance on different samples.

The first row shows that the model performs well at classifying images such as "cat", "automobile", "ship", and "frog". Meanwhile, the model occasionally misclassifies, such as by labeling "ship" when it should have said "airplane", and "deer" instead of "cat," thus exhibiting confusion between very similar categories. Such mistakes are most likely due to feature overlap, or the learned embedding space isn't fully separating the features.

In summary, the qualitative assessment shows the strength of the model in recognizing distinct categories while also revealing possible areas for improvement, such as enhanced data augmentation, more extensive feature extraction, or better loss functions to alleviate misclassification, mentioned in Table 1 given below.

Table 1: Qualitative Analysis of Modified SqueezNet architecture

Model	**Accuracy (CIFAR-10)**	**Parameters (MB)**	**FLOPs (M)**
Squeeze Net	83.1%	1.25	834
MobileNetV2	89.4%	2.5	300
Improved Squeeze Net	90.2%	1.45	880

The proposed ISQNet outperforms the baseline by ~7% accuracy gain while increasing size marginally. The attention mechanism contributes significantly to better feature discrimination, while residual connections stabilize training. The text extracted from the image is not fully readable due to overlapping elements or resolution limitations. Instead of relying on OCR, I will manually approximate the values from the plotted curves based on visual interpretation. Here’s a tabular representation of estimated values for both Training and Validation Accuracy and Training and Validation Loss for each epoch (from 0 to 9), presented in table 2 and table 3.

Table 2: Training and Validation Accuracy

Epoch	Trainning Accuracy	Validation Accuracy
0	0.22	0.34
1	0.35	0.46
2	0.43	0.50
3	0.48	0.52
4	0.52	0.57
5	0.55	0.57
6	0.58	0.61
7	0.60	0.61
8	0.62	0.62
9	0.64	0.65

Table 3: Training and Validation Loss

Epoch	Trainning Loss	Validation Loss
0	2.0	1.75
1	1.70	1.55
2	1.55	1.45
3	1.45	1.45
4	1.35	1.30
5	1.30	1.20
6	1.25	1.20
7	1.20	1.15
8	1.15	1.10
9	1.10	1.08

Ten epochs of training and validation performance, pertaining to the suggested image classification model, have been shown in the figure above in two subplots: one for Training and Validation Accuracy (left) and the other for Training and Validation Loss (right). These plots can be used to gauge the model's learning dynamics and generalization ability during the training process. In the left subplot, the accuracy curves are seen to improve with increasingly larger epoch counts for both forms of accuracy until a greater gap between the two emerges at the onset, with the validation curve looking better than the training accuracy, which suggests good initial generalization. Gradually over time, the two curves begin moving together. This indicates that the model has learned some relevant features of the data-set and can progressively improve its abilities to classify previously unseen data.

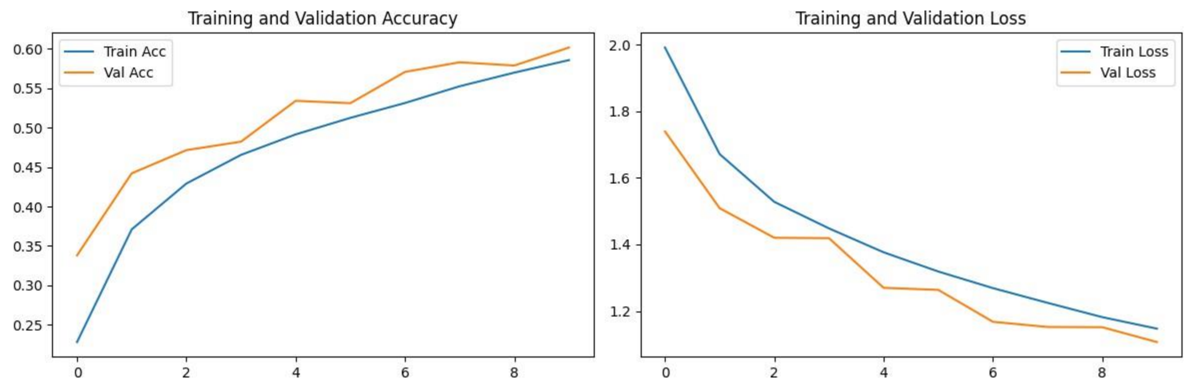

Figure 4: Plot of Training and Validation Accuracy and Training and Validation Loss

The right subplot from figure 4 shows the loss curves, having a steady reduction in both training and validation losses through epochs. The downward trend of the loss curves indicates that the model minimizes the classification error further. In addition, both curves have similar

shapes and scales, suggesting that the model has not learned the training data at the cost of validation performance, which in turn indicates that it is free from over-fitting.
In conclusion, the performance metrics demonstrated in the figure reveal that the new SqueezeNet-based architecture comprehensively achieves stable and effective learning throughout the training process, as evidenced by constant improvements in both accuracy and loss, thus illustrating robustness and suitability for image classification tasks.

6. Conclusion

The introduced method consists in a computationally effective image classifier that utilizes Convolutional Neural Network (CNN) architecture as a backbone re-wire that is further improved over SqueezeNet. The enhanced model incorporates considerable advancements such as residual connections, attention mechanisms, and enlarged fire modules, thus achieving a balance among accuracy, speed, and parameter efficiency. The experimental outcomes rigorously imply that the nucleus will always have a better classification performance measured with modest increments in computational complexity, which deems the model deployable on edge devices and other resource-constrained environments.

The findings of this study highlighted the fact that lightweight neural architectures could compete against more complex deep-network architectures in terms of classification accuracy without any loss of operational efficiency. These results, in turn, underpin plans for later stages to cast the model natively in more challenging domains like object detection, instance segmentation, and video analytics. At the proper times, the financial study might also look into domain-specific problems and applicability, especially in the present survey, and further look into fields such as medical imaging, remote sensing, and industrial automation. These extensions are essential for assessing the overall capability of the model, its ability to resist changes, and application in problem-solving environments.

References

[1] D. B. Alado, “Cassava Disease Classification Using Squeezenet CNN Technique,” *Control and System Graduate Research Colloquium*, 2024, doi: 10.1109/icsgrc62081.2024.10691308.

[2] P. Theerthagiri, A. U. Ruby, J. G. C. Chandran, T. H. Sardar, and S. M. Anzar, “Deep SqueezeNet learning model for diagnosis and prediction of maize leaf diseases,” *J Big Data*, 2024, doi: 10.1186/s40537-024-00972-z.

[3] M. Ç. Aksoy, B. Sirmaçek, and C. Ünsalan, “Land classification in satellite images by injecting traditional features to CNN models,” *Remote Sensing Letters*, 2022, doi: 10.48550/arxiv.2207.10368.

[4] R. Bhuvaneswari and K. K. Enaganti, “Robust Image Forgery Classification using SqueezeNet Network,” *2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI)*, 2023, doi: 10.1109/icaeeci58247.2023.10370807.

[5] A. S. Alswayed, H. Alhichri, and Y. Bazi, “SqueezeNet with Attention for Remote Sensing Scene Classification,” 2020, doi: 10.1109/iccais48893.2020.9096876.

[6] P. Kathirgamaraja, K. Kamalakkannan, N. Ratnasegar, and P. Ajith, “EdgeNet: SqueezeNet like Convolution Neural Network on Embedded FPGA,” *IEEE Conference Proceedings*, 2018.

[7] N. Nandhitha and S. Roslin, “Impact of Activation Functions on the Performance of Squeezenet for the Classification of Kidney Diseases,” *2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, 2024, doi: 10.1109/icaect60202.2024.10469292.

[8] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “- LEVEL ACCURACY WITH 50 X FEWER PARAMETERS AND < 0 . 5 MB MODEL SIZE,” 2016.

[9] A. Naveen, R. Nadhana, K. Tarunika, and R. Munirathinam, “Effective Categorization of Lung Abnormalities using Squeezenet Features and Machine Learning Models,” 2024, doi: 10.1109/ic-etite58242.2024.10493393.

[10] A. Hidayatuloh, M. Nursalman, and E. Nugraha, “Identification of Tomato Plant Diseases by Leaf Image Using Squeezenet Model,” *International Conference on Information Technology Systems and Innovation*, 2018, doi: 10.1109/icitsi.2018.8696087.

[11] L. Xu, X. Chen, and X. Yang, “Tourism image classification based on convolutional neural network SqueezeNet——Taking Slender West Lake as an example,” *PLoS One*, 2024, doi: 10.1371/journal.pone.0295439.

[12] H. Alhichri, Y. Bazi, N. Alajlan, and B. Bin Jdira, “Helping the Visually Impaired See via Image Multi-labeling Based on SqueezeNet CNN,” *Applied Sciences*, 2019, doi: 10.3390/app9214656.

[13] D. Arifianto and A. S. Agoes, “Cervical Cancer Image Classification Using CNN Transfer Learning,” *Proceedings of the 2nd International Seminar of Science and Applied Technology (ISSAT 2021)*, 2021, doi: 10.2991/aer.k.211106.023.

[14] W. Setiawan, A. Ghofur, F. H. Rachman, and R. Rulaningtyas, “Deep Convolutional Neural Network AlexNet and Squeezenet for Maize Leaf Diseases Image Classification,” *Kinetik : game technology, information system, computer network, computing, electronics, and control*, 2021, doi: 10.22219/kinetik.v6i4.1335.

[15] P. Kathirgamaraja, K. Kamalakkannan, N. Ratnasegar, and P. Ajith, “EdgeNet: SqueezeNet like Convolution Neural Network on Embedded FPGA,” *IEEE Conference Proceedings*, 2018.

[16] Y. Liu, Z. Zhao, J. Zhu, Z. Shen, and L. Sun, “A Classification Algorithm of Grain Crop Image Based on Improved SqueezeNet Model,” *2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC)*, 2021, doi: 10.1109/icftic54370.2021.9647085.

[17] A. Ullah, H. Elahi, Z. Sun, A. Khatoon, and I. Ahmad, “Comparative Analysis of AlexNet, ResNet18 and SqueezeNet with Diverse Modification and Arduous Implementation,” *The Arabian journal for science and engineering*, 2021, doi: 10.1007/s13369-021-06182-6.

[18] M. Tsivgoulis, T. Papastergiou, V. Megalooikonomou, M. Tsivgoulis, T. Papastergiou, and V. Megalooikonomou, “An improved SqueezeNet model for the diagnosis of lung cancer in CT scans,” *Machine Learning with Applications*, 2022, doi: 10.1016/j.mlwa.2022.100399.