

Research Article

SapotaNet: Paving the Way for Efficient Deep Learning with Lightweight Network Architecture

Anita Bhatt ^{1,*} , Maulin Joshi ² 

¹ Electronics and Communication Department, Gujarat Technological University, Ahmedabad, 382424, India

² Department of Electronics and Communication, Sarvajanik College of Engineering and Technology, Surat, 395001, India

*Corresponding Author: Anita Bhatt, E-mail: anbhatt@bvmengineering.ac.in

Article Info	Abstract
Article History	Over the years, agricultural production in India has consistently increased. There remains a substantial gap between per capita demand and supply due to losses, including those incurred during post-harvest processing. Research in cutting-edge technologies, such as computer vision and artificial intelligence, has significantly improved fruit quality evaluation in recent years. Fruits quality evaluation by humans is an expensive and time-consuming task. Sapota fruit is a source of income for people living in tropical regions. However, having a short post-harvest life and a lack of human resources result in high post-harvest losses in the case of Sapota fruit production. Fruit grading using technology can resolve related issues. This work presents an innovative Convolutional Neural Network (CNN) model tailored for the precise grading of Sapota fruit. The proposed CNN model SapotaNet accurately grades Sapota fruit, significantly reducing the need for manual inspection and lowering post-harvest losses. This technology-driven approach offers improved efficiency, consistent grading, and higher-quality output in commercial Sapota production. Since no prevailing database exists for Sapota fruit, a new dedicated database has been created, and a comparative assessment has been undertaken, evaluating knowledge transfer models, as well as our proposed lightweight SapotaNet model. A total of 3,400 Sapota fruit images are assessed for Sapota fruit grading. In our experimental evaluations, our proposed model, SapotaNet, exhibited higher testing accuracy and the least inference time per Sapota fruit among all models assessed.
Received Oct 10, 2024	
Revised Feb 05, 2025	
Accepted Feb 13, 2025	
Keywords	
Computer Vision	
Convolutional Neural Network	
Image Processing	
Deep Learning	
Sapota fruit	



Copyright: © 2025 Anita Bhatt and Maulin Joshi. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license.

1. Introduction

India is an agricultural country, and agriculture-related activities are a vital source of livelihood for millions of people in rural India. The country has huge export potential for fruits and vegetables. However, due to a lack of adequate processing facilities, labor shortages, a cold chain, and inadequate storage facilities, problems persist regarding fruit wastage in exports, transportation, and marketing. In India, Sapota fruit is mainly grown in Gujarat, Maharashtra, Karnataka, West Bengal, Tamil Nadu, Andhra Pradesh, and other parts of the country [1]. In certain regions of India, the Sapota fruit is also known as Chikoo. Sapota

fruit is a nutritious tropical fruit that serves as a food source and income for many people in tropical countries. Sapota fruit is available around the year, but the main season of Sapota fruit is between November and February. Sapota fruit is cultivated in various varieties like Cricket Ball, Kalipatti, Pala, Guthi, Kirti Bairathi, etc. [2]. Kalipatti is a variety that is most widely grown in Gujarat, India.

The purpose of grading fruit according to color, textures, etc. parameters is to improve product quality, export, and yield [3]. According to the first advance estimation report by Agri Exchange Apeda, Gujarat state in India produced the most Sapota fruit in 2021–2022, with 273.83 tonnes of production [1]. The first advance estimation report of Sapota fruit production is shown in Figure 1. Color similarity, texture similarity, and seasonal effect are challenges in grading Sapota fruits. Instead of using only traditional machine learning algorithms for grading, other cutting-edge algorithms were tested to improve the outcome [4].

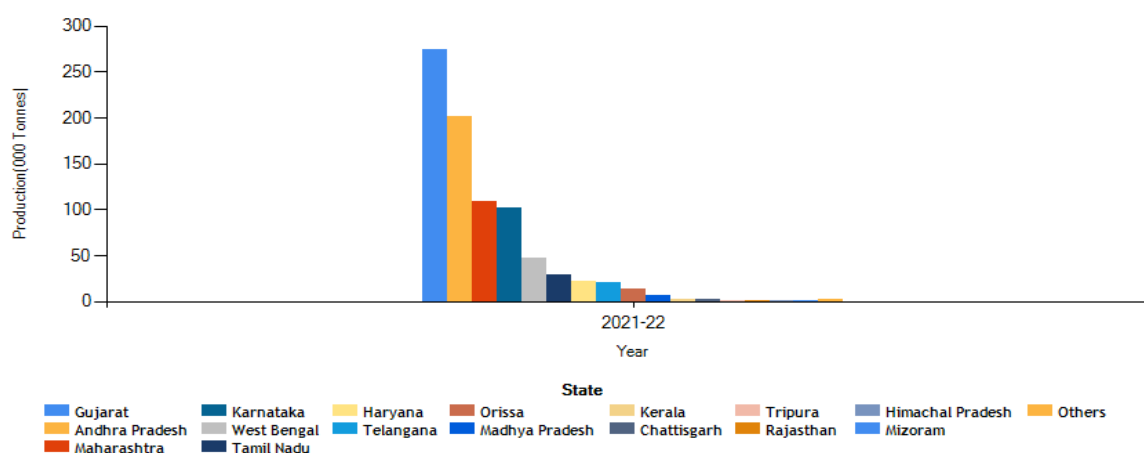


Figure 1. Sapota fruit production in the year 2021-22 in various states of India [1]

The study conducted for this research is on the raw form of Sapota fruit. The main objectives of this research work are (a) The creation of a comprehensive, robust dataset of harvested Sapota fruit and (b) Grade the Sapota fruit into fresh and spoiled categories using various Deep-learning models.

The significant contributions of this paper are as follows:

- This work proposes state-of-the-art Deep learning models for grading Sapota fruit into Fresh and spoiled quality.
- This work is a part of the initiative to develop the first-ever dataset solely dedicated to Sapota fruit images. To the best of our knowledge, no such specific dedicated dataset of Sapota fruit images has been created to date.
- A knowledge transfer approach is applied for Sapota fruit grading, transitioning from shallow to deep model architectures. The process is complemented by fully connected layers, utilizing pre-trained models such as VGG16, VGG19, InceptionV3, DenseNet121, and Xception as feature extractors.

- Following the success of knowledge transfer models, a novel SapotaNet model, specifically tailored for Sapota fruits, offers fewer parameters and a simpler design than existing models while still delivering high accuracy.

The rest of the paper contains related work, materials and methods, results and discussion, conclusion, and future scope.

2. Literature Review

Incorporating artificial intelligence (AI) into the health sector and agriculture domains has fundamentally transformed data processing and operational efficiency [4, 5]. A clean and well-structured dataset is a fundamental prerequisite for developing reliable and accurate AI models in real-time environments [6]. Visual features and shape features are frequently used to categorize fruits and vegetables. In agriculture, the prices of vegetables and fruits vary according to their quality. Classification algorithms are applied to extract knowledge bases that subsequently decide to grade the fruits and vegetables for unknown cases.

In computer vision systems, a wide variety of techniques, including Artificial Neural Networks (ANN), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Convolutional Neural Networks (CNN), have been developed for fruit quality assessment classification [6, 7]. Transfer learning for fruit classification using deep convolutional neural networks achieves good accuracy [8, 9]. Machine learning models were assessed for mango grading and sorting. To improve performance, ensemble techniques such as bagging, boosting, and stacking were implemented, yielding a stacking ensemble model with 98.63% accuracy [10].

For Rice diseases, classification includes individual CNNs, Transfer learning, and ensemble models. Compared to the standard Seresnext101 model, a transfer learning approach using Seresnext101 boosted accuracy by 17%. The ensemble model with DenseNet121, EfficientNetB7, and Xception architectures (DEX) achieved the best performance, reaching an accuracy of 98% [11]. Transfer learning for different varieties of apples, evaluated across all DenseNet201 models, yielded the best classification accuracy at 97.48%. Deep features from DenseNet201 are utilized and trained in several traditional machine learning models, where SVM performed best, achieving an accuracy of 98.28%.

The effect of dimensionality reduction on classification performance, utilizing deep features, principal component analysis (PCA), and machine learning (ML) models, was investigated. MLP achieved an

accuracy of 99.77%, outperforming SVM (99.08%), RFC (99.54%), and KNN (91.63%) [12]. The maturity of papaya fruit is classified using a machine learning and transfer learning approach, where weighted k-NN with the histogram of gradients and VGG19 achieves 100% accuracy [13]. The author developed a lightweight cotton ball segmentation approach for harvesting purposes using CNN Models with an accuracy of more than 90% [14].

A knowledge transfer-based model and vision transformers were used to evaluate the health status of cacti. The proposed model-based Visual image Transformer (ViT) achieved a high accuracy of 88.73% [15]. The YOLOv8–Transformer model can, on average, accurately classify the quality grades of cashew nuts based on their images with a precision of 98.4% [16].

In VGG models, convolution layers mostly use 3×3 filters with a stride of 1 and padding of 1, which aid in preserving spatial information. Max-pooling layers reduce the size of the feature maps, and fully connected layers are responsible for classification based on the extracted features. VGG16 and VGG19 are easier to understand and implement due to their simple and uniform architecture [17]. Inception V3 combines various-sized convolutional filters and inception modules to achieve high accuracy in image classification with fewer parameters, making it both efficient and effective [18].

Xception architecture is crucial in computer vision tasks by offering state-of-the-art performance with a highly efficient deep learning model. Each layer in a dense net has a direct feed-forward connection to every other layer [19]. Compared to conventional CNNs, this connectivity pattern reduces the number of required parameters. DenseNet121 is a network with 121 layers organized into four dense blocks. Each dense block is made up of densely connected convolution layers. The transition layer in DenseNet-121 effectively utilizes techniques such as batch normalization, 1x1 convolution, and average pooling to manage computational load efficiently while preserving key features in the model's design. These modules enable the model to learn intricate patterns and record data at various scales. The dense connectivity principle is used by DenseNet121 [20]. The DenseNet121 model is consequently more parameter efficient.

3. Materials and Methods

3.1. Experimental Setup

This experiment utilizes a Lenovo Legion 5 Pro laptop (32 GB RAM, 1 TB SSD, Windows 11 Home, and Nvidia RTX 3070 8 GB GPU) with a Jupyter Notebook environment.

3.2. Image Acquisition

This proposed work considers the Kalipatti variety of Sapota fruit for grading. Kalipatti variety of Sapota is mainly grown in Gujarat. Between November 2022 and February 2023, already-graded Sapota fruits were collected from expert farmers in the Navsari district of Gujarat, India. In the context provided, 'graded' signifies categorization according to their freshness status, distinguishing between fresh and spoiled specimens. This work creates a new dataset encompassing diverse attributes of Sapota fruit. Sapota fruit images are captured on a test bench comprising a white background during daylight hours. A Realme X mobile phone camera with a 48-megapixel lens was used to capture the images. The camera position was fixed, and shots were captured from various angles by shifting the fruit's position. A summary of the Sapota fruit sample collection is given in Table 1.

Table 1. Sapota fruit sample collection details

Sr. No.	Particular	Description
1	Fruit Name	Sapota
2	Variety	Kalipatti
3	Region and year	Navsari, Gujarat, India 2022-2023
4	Grading	Fresh Fruit and Spoiled Fruit
5	Camera specification	Device Name - Realme X Mobile phone Camera - 48 MP, Dimension - 4000x3000 pixels

A total of 3400 images are used for model examinations. 1700 images of Fresh Sapota fruit and 1700 Spoiled Sapota fruits are considered for grading. Further, 1360 training and 340 testing images are considered in each category. Figure 2 illustrates the conventional method where workers manually grade the Sapota fruits. Table 2 shows the grade distribution of Sapota fruit.



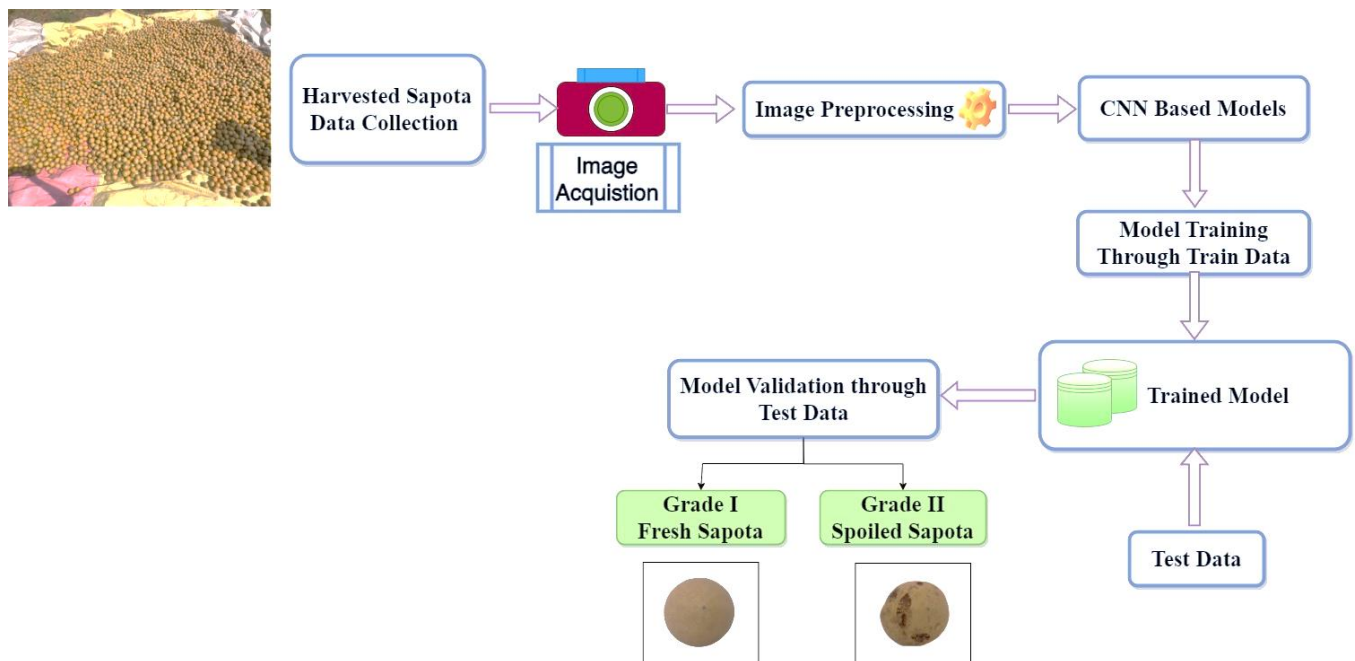
Figure 2. Typical scenario presenting the manual process of grading Sapota fruits by workers

Table 2. Grade distribution of the Sapota fruit based on visual description

Sr. No.	Particular	Visual Description
		Outer Surface and Color
1	Grade-I (Fresh Sapota)	Brown in color or shades of brown and defect-free surface
2	Grade-II (Spoiled Sapota)	Dark brown in color or Shades of brown, and scratches can be seen on the surface. White and Black spots on the surface.

3.3. Workflow for Sapota Fruit Grading Process

The workflow is initiated by collecting harvested Sapota fruits from the field. Subsequently, each graded Sapota by the expert farmer is passed through an image acquisition stage, followed by preprocessing stages, to enhance the image quality. Following preprocessing, the dataset is divided into an 80-20 split, with 80% reserved for training and 20% for testing, ensuring a balanced and representative distribution for robust model evaluation. The developed model is adeptly structured to categorize Sapota fruits. A comprehensive illustration of this workflow is presented in Figure 3, and sample images of the dataset for fresh and spoiled Sapota fruits are shown in Figure 4.

**Figure 3.** A step-by-step illustration process of Sapota fruit categorization

Two distinct methodologies were employed for Sapota fruit grading: Knowledge transfer, utilizing a pre-trained model, and a novel lightweight network, SapotaNet. Both approaches were evaluated to assess their effectiveness in classifying fruit quality. This comparison explored the potential of leveraging existing knowledge versus developing specialized architecture. Performance metrics were used to determine the efficacy of each method.



Figure 4. Sample images of the Sapota fruit dataset: The first row shows Grade-I Fresh Sapota and the second row shows Grade-II Spoiled Sapota

3.4. Knowledge Transfer Approach for Sapota Fruit Grading

In the Knowledge transfer approach, a model already successfully trained for one application is used to train a new model for a similar-looking application [21]. The knowledge transfer approach is the layered freezing strategy for using pre-trained models efficiently. The models used for our experimentation with the knowledge transfer approach are primarily pre-trained on the ImageNet dataset, which comprises over 14 million annotated images [22]. A knowledge transfer model was trained and evaluated using a dataset of 1360 training images and 340 testing images for each category. Data augmentation techniques are used to artificially increase the size and diversity of training datasets. These methods encompass rotations up to 20 degrees, horizontal mirroring, and fill adjustments. The top layers of knowledge transfer models are also referred to as fully connected layers, which categorize input data into distinct categories [23].

We can use a particular model as a feature extractor by removing the top layers and freezing weighted layers. In the first stage of the proposed work, five pre-trained models, namely, VGG16, VGG19, InceptionV3, DenseNet121, and Xception, are used as feature extractors for the Sapota fruit grading task. According to the type of base model, the images are resized and augmented. Accordingly, layers are frozen during the training phase. A pre-trained base model is integrated as the initial layer, followed by a global average pooling layer. This is succeeded by a fully connected dense layer with 128 neurons, a dropout layer with a rate of 0.5, and another dense layer with 64 neurons, all activated by the ReLU function. Fully connected layers are a type of neural network layer characterized by dense connections, where each neuron in the layer connects to every neuron in the preceding and following layers. They play an important role in traditional CNNs for aggregating features and performing classification. The output layer is designed with a dense layer featuring two neurons corresponding to the two categories in the dataset. This output layer

uses a Sigmoid activation function, enabling the categorization task. For optimization, the model employs the Adam optimizer with a learning rate of 0.001 and utilizes binary cross-entropy as the loss function to effectively optimize the training process for the Binary-class classification problem. The grading approach for Sapota fruit, contrasting the knowledge transfer method with the proposed SapotaNet methodology, is illustrated in Figure 5

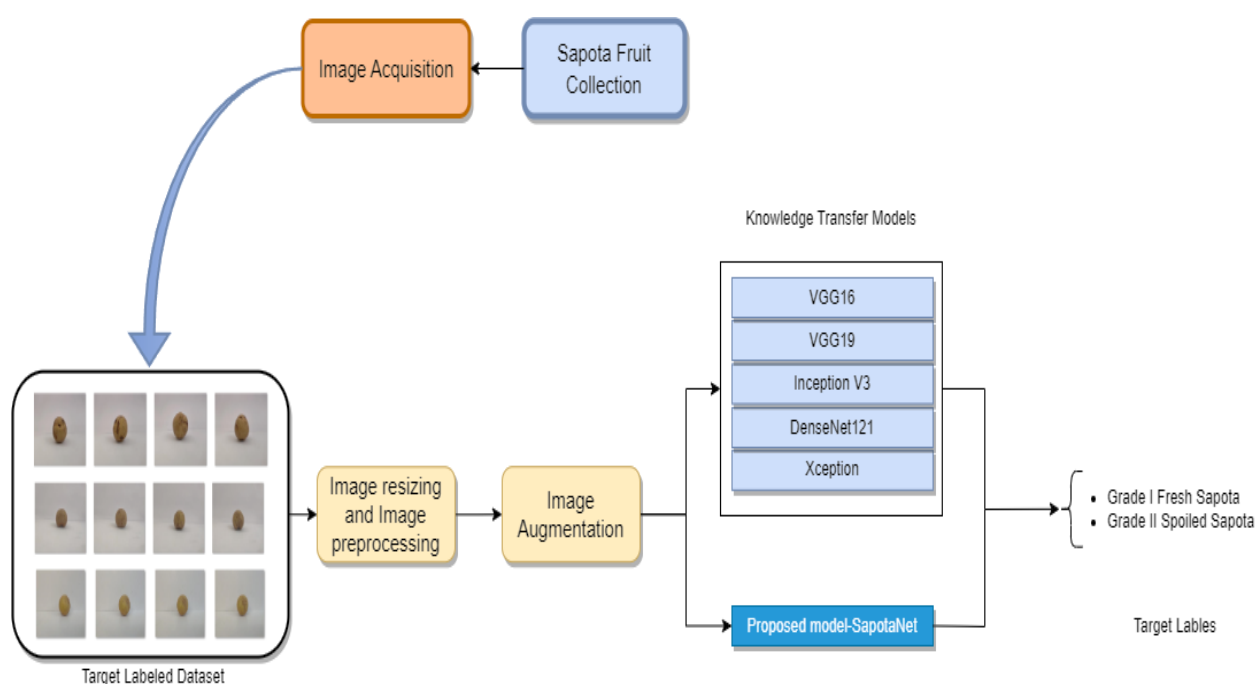


Figure 5. Sequential procedure for the grading of Sapota fruits using knowledge transfer techniques and the innovative SapotaNet model

3.5. Proposed Model SapotaNet for Sapota Fruit Grading

3.5.1. Image Preprocessing and Augmentation

In the presented methodology, the input images were resized to a dimension of $224 \times 224 \times 3$ and subsequently normalized. Resizing standardized image dimensions is a requirement for many machine-learning models. This also offers computational benefits by reducing the amount of data that needs to be processed. Normalization scales pixel intensities to a consistent range, frequently from 0 to 1. This scaling improves model training by promoting faster convergence, enhancing numerical stability, and performing better generalization to unseen data. Normalization prevents features with inherently larger values from disproportionately influencing the learning process. Image augmentation methods enhanced the model's resilience and data diversity. These methods include rotations up to 20 degrees, horizontal mirroring, and fill adjustments.

3.5.2. Identification Model

The model begins with a convolutional layer employing a ReLU activation function, utilizing eight filters of size 3×3 with a stride of 2×2 and the same padding. This is followed by a depth-wise convolutional layer that uses a 3×3 filter and has a stride of 1×1 , maintaining the same spatial dimensions due to the 'same' padding. It is followed by a pointwise convolutional layer with 32 filters of size 1×1 , also with 'same' padding. After that, both convolutional layers are activated using the ReLU function. The block concludes with a max-pooling layer with a pool size of 2×2 and a stride of 2×2 . Subsequently, a third convolutional layer with 64 filters of size 3×3 and stride 1×1 , activated by the ReLU function, is added to the architecture, followed by another max-pooling layer with a pool size of 2×2 with stride 2×2 . Further layers include a flattening layer to transform the 2D feature maps into a 1D vector, followed by a fully connected dense layer comprising 128 neurons activated by the ReLU function. A dropout layer with a rate of 0.2 is introduced to enhance model generalization by randomly setting 20 percent of the input units forced to zero during training. This is succeeded by another dense layer with 64 neurons and a ReLU activation function. The output layer is designed with a dense layer featuring two neurons corresponding to the two categories in the dataset and utilizes a Sigmoid activation. Figure 6 illustrates the architecture of the proposed model-SapotaNet.

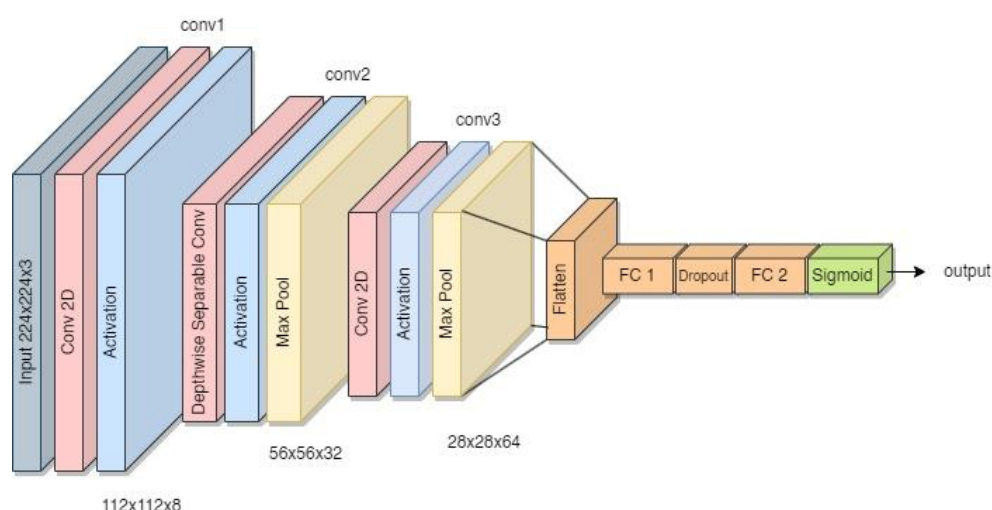


Figure 6. Proposed model-SapotaNet: A lightweight architecture for Sapota fruit grading

The model is trained using the Adam optimizer with a learning rate of 0.001 and adopts binary cross-entropy as the loss function to optimize the training process effectively. Table 3 outlines the training parameters for the proposed SapotaNet model compared to the Knowledge Transfer models. The convolutional layer in neural networks captures features from input data. Activation functions introduce non-linear

elements, enhancing the network's ability to recognize intricate patterns. Max-pooling diminishes computational load and emphasizes essential features, enhancing the network's adaptability to input fluctuations.

Table 3. Training parameters of proposed model-SapotaNet and knowledge transfer models

Parameters	Proposed model-SapotaNet	Knowledge transfer-based models
Optimizer	Adam	Adam
Learning Rate	0.001	0.001
Loss	Binary cross entropy loss	Binary cross entropy loss
Batch size and Epoch	32 and 10	32 and 10
Dropout rate	0.2	0.5
Classifiers	Sigmoid	Sigmoid
Train-Test Split	80-20	80-20

A fully connected layer, commonly referred to as a dense layer, connects every neuron in the preceding layer, processing input data through weighted sums and activation functions to capture intricate patterns that inform final predictions in neural networks. Figure 7 presents a comprehensive summary detailing the proposed SapotaNet approach.

Layer (type)	Output Shape	Param #
conv2d_3 (Conv2D)	(None, 112, 112, 8)	224
activation_3 (Activation)	(None, 112, 112, 8)	0
depthwise_conv2d_1 (DepthwiseConv2D)	(None, 112, 112, 8)	80
conv2d_4 (Conv2D)	(None, 112, 112, 32)	288
activation_4 (Activation)	(None, 112, 112, 32)	0
max_pooling2d_2 (MaxPooling2D)	(None, 56, 56, 32)	0
conv2d_5 (Conv2D)	(None, 56, 56, 64)	18496
activation_5 (Activation)	(None, 56, 56, 64)	0
max_pooling2d_3 (MaxPooling2D)	(None, 28, 28, 64)	0
flatten_1 (Flatten)	(None, 50176)	0
dense_2 (Dense)	(None, 128)	6422656
dropout_1 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 64)	8256
dense_4 (Dense)	(None, 2)	130
Total params: 6,450,130		
Trainable params: 6,450,130		
Non-trainable params: 0		

Figure 7. Summary of model parameters for proposed SapotaNet

4. Results and Discussion

The proposed work aims to grade Sapota fruit into two categories: Grade-I Fresh Sapota fruit and Grade-II spoiled Sapota fruit. In the presented approach to Sapota fruit grading, a transition from simpler to more complex model architectures is implemented, leveraging a knowledge transfer strategy. A range of pre-trained models serves as feature extraction mechanisms, including VGG-16, VGG-19, InceptionV3, DenseNet-121, and Xception. In all these models, fully connected layers were finally appended to each architecture to facilitate Sapota fruit grading. After deriving performance evaluations of knowledge transfer-based models, lightweight model architecture fully dedicated to Sapota fruit is conceptualized. A detailed analysis is conducted to evaluate the performance of knowledge transfer-based models and the proposed SapotaNet model. The performance of each model is assessed by considering the maximum accuracy achieved during training, training time, testing time, Loss, and parameters. Other evaluation parameters considered are the F1 score, Precision, and Recall for model evolution as defined in the equations:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} * 100\% \quad (1)$$

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

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

$$F1 - Score = \frac{(2*Precision*Recall)}{(Precision+Recall)} \quad (4)$$

Here TP-True Positive, TN-True Negative, FP-False Positive and FN-False Negative.

Evaluation results, as shown in Table 4 and Table 5, depict that the proposed SapotaNet model outperforms the knowledge transfer models in terms of all evaluation parameters.

Table 4. Performance evaluation of knowledge transfer and proposed model SapotaNet for Sapota fruit grading

Model Name	Maximum accuracy during training	Training time (seconds)	Testing accuracy	Testing loss	Testing time (seconds)	Parameters (million)
VGG16	96.58	186.24	93.97	0.24	2	14.71
VGG19	96.80	193.45	94.41	0.21	2	23.24
InceptionV3	99.41	200.71	96.18	0.43	2	28.36
DenseNet121	99.37	198.04	96.62	0.27	3	13.46
Xception	99.30	190.60	96.03	0.44	2	33.71
Proposed Model SapotaNet	99.49	178.47	98.82	0.13	1	6.45

Table 5. F1 Score, Precision, Recall of knowledge transfer, and proposed model SapotaNet for Sapota fruit grading

Model Name	F1 score	Precision	Recall
VGG16	93.97	93.98	93.97
VGG19	94.41	94.41	94.41
InceptionV3	96.17	96.19	96.17
DenseNet121	96.61	96.62	96.61
Xception	96.02	96.04	96.02
Proposed Model SapotaNet	98.82	98.85	98.82

In the designed SapotaNet model, rigorous testing resulted in an impressive accuracy of 98.82%. This high performance was achieved while maintaining a relatively lightweight architecture of approximately 6.45 million parameters. Depth-wise separable convolution merges depth-wise and pointwise convolutions, processing spatial and channel details independently [24]. This method offers computational efficiency, resulting in faster training and inference without compromising model accuracy. The same padding preserves the spatial dimensions, allowing the model to maintain spatial information as it progresses through the layers. To mitigate potential overfitting issues, dropout regularization was incorporated within the model architecture. This regularization technique enhances the model's generalization capability by preventing the co-adaptation of feature detectors, thereby promoting robust performance on unseen data. This high level of accuracy of the SapotaNet model underscores the model's effectiveness in accurately categorizing Sapota fruits based on the defined criteria. The proposed SapotaNet model outperforms other knowledge transfer approaches, achieving peak accuracy while reducing computational time and parameters. This efficiency renders it lightweight, making it particularly suitable for deployment in mobile applications and resource-constrained hardware environments [24, 25]. Exposing models to images captured under diverse lighting conditions, including natural light, enhances their robustness and generalization performance by fostering adaptability to real-world environments.

Figures 8 and 9 illustrate the progression of our CNN model's learning using two key metrics: Accuracy, which represents the percentage of correctly predicted samples, and Loss, quantified by the binary cross-entropy loss function, plotted against the number of epochs. The visual representation highlights a consistent improvement in training and validation accuracy, accompanied by a decrease in loss as the epochs progress. SapotaNet is a lightweight Convolutional Neural Network (CNN) designed to efficiently and accurately classify Sapota fruits as fresh or spoiled. This specialized model uses simplified architecture

with fewer parameters than larger, general-purpose CNNs. This makes it computationally efficient, allowing for rapid inference and eliminating the need for manual inspection in commercial Sapota grading. SapotaNet's design addresses the specific challenges of Sapota fruit classification, such as variations in shape, color, and spoilage patterns, while maintaining high accuracy. Several methods can be used to design more lightweight models, including pruning, knowledge distillation, neural architecture search, and combinations of these techniques.

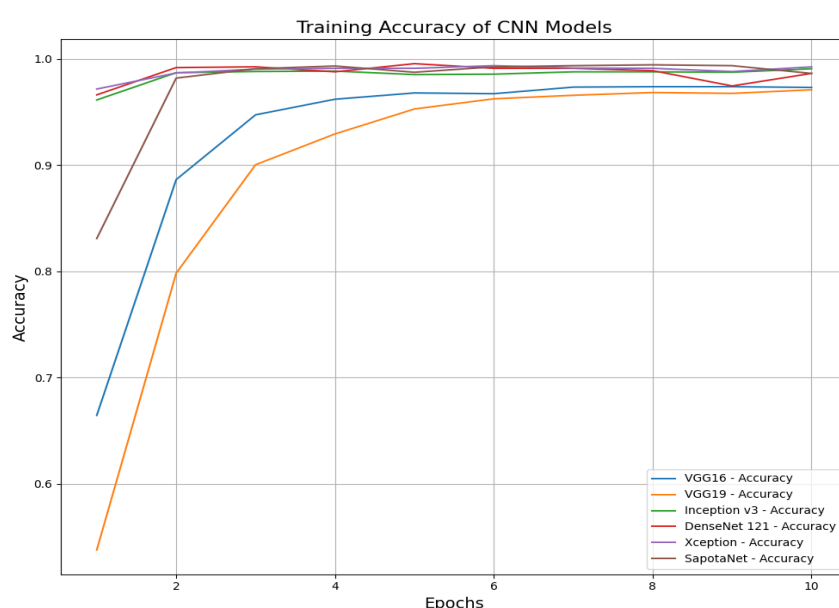


Figure 8. The progress of our model's learning process through training accuracy, representing the percentage of correctly predicted samples

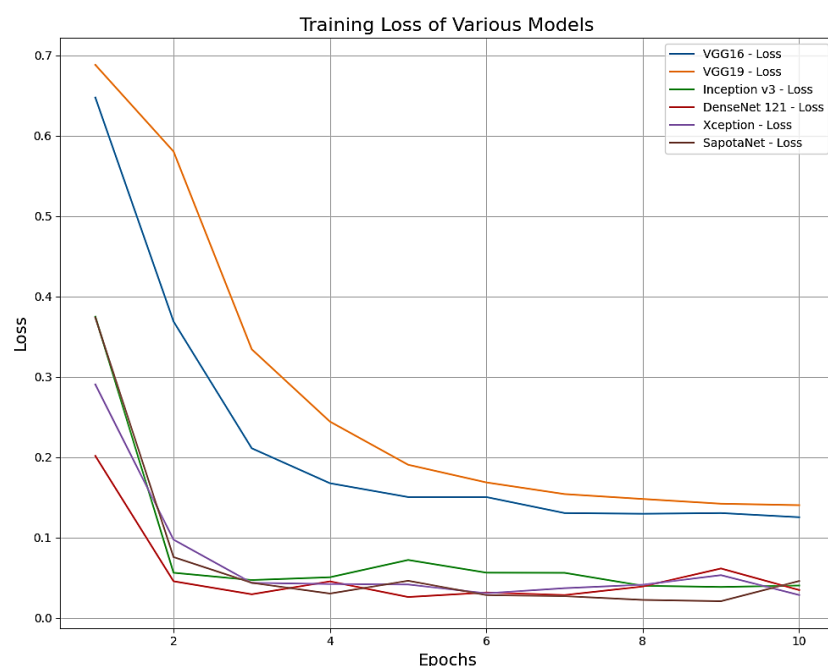


Figure 9. The progress of our model's learning process through training loss

Figure 10 depicts the confusion matrix and corresponding classification report for our proposed SapotaNet model. The confusion matrix visually represents the true and predicted labels. This allows for easy identification of categories where the model performs well and those where it struggles.

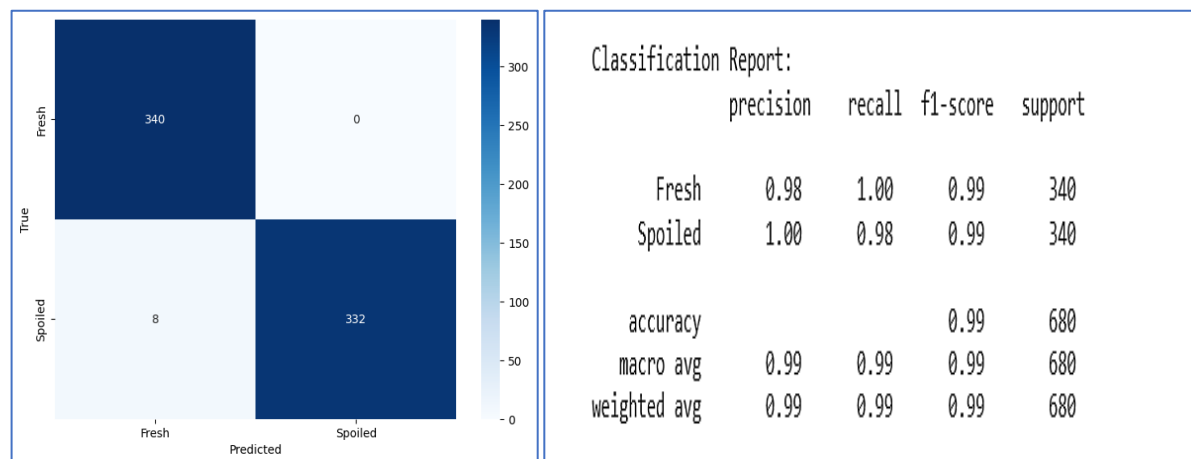


Figure 10. Confusion matrix and classification report of proposed model SapotaNet

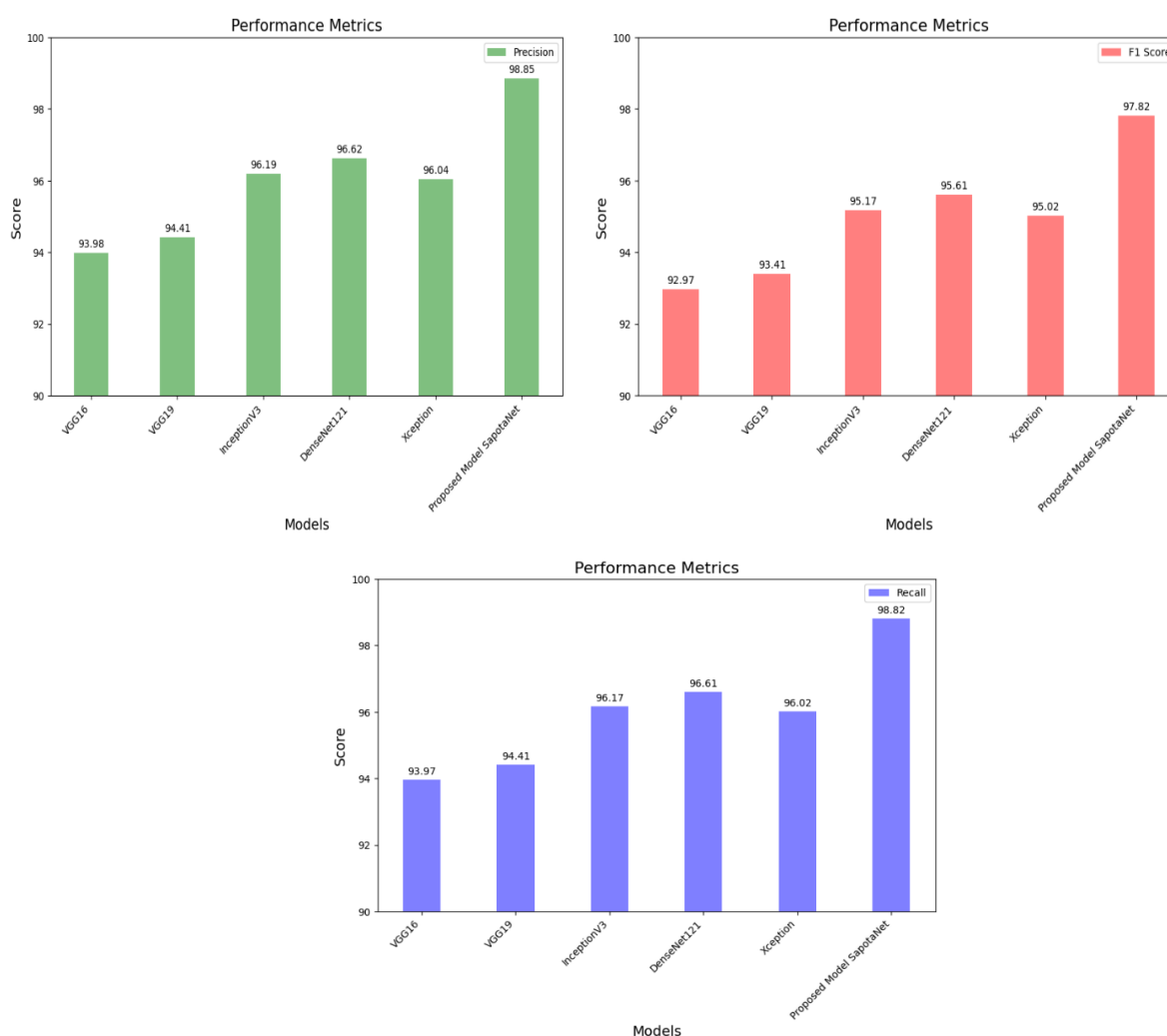


Figure 11. Detailed visualization showcasing the Precision, F1 score, and Recall metrics of knowledge transfer techniques and the proposed SapotaNet Model

Figure 11 compares Precision, F1 score, and Recall across various knowledge-transfer techniques and the proposed SapotaNet model. This visualization underscores SapotaNet's competitive or superior performance in capturing class-specific features compared to baseline transfer-learning strategies. The consistently high metrics for SapotaNet validate its architectural enhancements and fine-tuning strategies, emphasizing the benefits of our customized design.

5. Conclusions

This work makes a significant contribution to the expanding literature on automated fruit grading by creating robust datasets for Sapota fruit. A non-destructive grading of the harvested Sapota fruit dataset performed well in evaluation with various neural network architectures and depths. Through comparative evaluations, shallow and deep networks are assessed on the Sapota fruit dataset to determine optimal performance for real-time grading. Our proposed lightweight SapotaNet model provides low-latency performance with high accuracy, making it suitable for mobile and embedded deployments in Sapota fruit grading. SapotaNet Model's practical implications can be helpful for agricultural sector stakeholders and farmers. On-site grading systems can enhance efficiency and reduce post-harvest losses. This research offers scalable, cost-effective solutions for diverse stakeholders in the fruit supply chain, utilizing machine vision. Future research can also explore the combination of harvested Sapota fruit with on-tree grading. The Sapota fruit can be classified by combining with other fruit varieties.

Acknowledgments: The authors would like to thank Navsari Agriculture University, Gujarat, India, and the farmers of Navsari District for permission to visit farms, take true images of farms, and share their knowledge to create the dataset.

Data availability statement: Data available upon request.

Funding: The authors declare that no funding was received for this research.

Declaration of Competing Interest: The authors declare they have no known competing interests.

Research Involving Human/Animal Participants: This article does not involve any studies conducted by authors on animals or human participants.

References

- [1] AgriXchange. "India's Sapota Production." [Online]. <https://agriexchange.apeda.gov.in/Production/Indiacat/Index> (accessed December 02, 2024).
- [2] K. Kishore and K. K. Mahanti, "Codification and description of phenological growth stages of sapota (Manilkara zapota)

- according to the extended BBCH scale," *Scientia Horticulturae*, vol. 211, pp. 431-439, 2016, doi: <http://dx.doi.org/10.1016/j.scienta.2016.09.034>.
- [3] V. Meshram, K. Patil, V. Meshram, D. Hanchate, and S. D. Ramkteke, "Machine learning in agriculture domain: A state-of-art survey," *Artificial Intelligence in the Life Sciences*, vol. 1, no. 2021, pp. 1-11, 2021, doi: <http://dx.doi.org/10.1016/j.ails.2021.100010>.
 - [4] R. F. Jader, S. Aminifar, and M. H. M. Abd, "Diabetes detection system by mixing supervised and unsupervised algorithms," *Journal of Studies in Science and Engineering*, vol. 2, no. 3, pp. 52-65, 2022, doi: <http://dx.doi.org/10.53898/josse2022234>.
 - [5] M. Wakchaure, B. K. Patle, and A. K. Mahindrakar, "Application of AI techniques and robotics in agriculture: A review," *Artificial Intelligence in the Life Sciences*, vol. 3, no. November 2022, pp. 100057-100057, 2023, doi: <http://dx.doi.org/10.1016/j.ails.2023.100057>.
 - [6] A. Bhargava and A. Bansal, "Fruits and vegetables quality evaluation using computer vision: A review," vol. 33, ed: King Saud bin Abdulaziz University, 2021, pp. 243-257.
 - [7] A. Bhargava and A. Bansal, "Machine learning based quality evaluation of mono-colored apples," *Multimedia Tools and Applications*, vol. 79, no. 31-32, pp. 22989-23006, 2020, doi: <http://dx.doi.org/10.1007/s11042-020-09036-9>.
 - [8] T. B. Shahi, C. Sitaula, A. Neupane, and W. Guo, "Fruit classification using attention-based MobileNetV2 for industrial applications," *PLoS ONE*, vol. 17, no. 2, pp. 1-21, 2022, doi: <http://dx.doi.org/10.1371/journal.pone.0264586>.
 - [9] K. Yan, M. K. C. Shisher, and Y. Sun, "A Transfer Learning-Based Deep Convolutional Neural Network for Detection of Fusarium Wilt in Banana Crops," *AgriEngineering*, vol. 5, no. 4, pp. 2381-2394, 2023, doi: <http://dx.doi.org/10.3390/agriengineering5040146>.
 - [10] N. D. Tai, W. C. Lin, and N. M. Trieu, "Development of a Mango-Grading and -Sorting System Based on External Features , Using Machine Learning Algorithms," *Agronomy*, Vol. 14, no. 4: p. 831, 2024, doi: <http://dx.doi.org/10.3390/agronomy14040831>.
 - [11] M. T. Ahad, Y. Li, B. Song, and T. Bhuiyan, "Comparison of CNN-based deep learning architectures for rice diseases classification," *Artificial Intelligence in Agriculture*, vol. 9, pp. 22-35, 2023, doi: <http://dx.doi.org/10.1016/j.aiia.2023.07.001>.
 - [12] A. Taner, M. T. Mengstu, K. Ça, I. Gür, and N. Ungureanu, "Apple Varieties Classification Using Deep Features and Machine Learning," *Agriculture*, Vol. 14, no. 2, p. 252, 2024, doi: <http://dx.doi.org/10.3390/agriculture14020252>.
 - [13] S. K. Behera, A. K. Rath, and P. K. Sethy, "Maturity status classification of papaya fruits based on machine learning and transfer learning approach," *Information Processing in Agriculture*, vol. 8, no. 2, pp. 244-250, 2021, doi: <http://dx.doi.org/10.1016/j.inpa.2020.05.003>.
 - [14] N. Singh, V. Tewari, P. Biswas, and L. Dhruw, "Lightweight convolutional neural network models for semantic segmentation of in-field cotton bolls," *Artificial Intelligence in Agriculture*, vol. 8, pp. 1-19, 2023, doi: <https://doi.org/10.1016/j.aiia.2023.03.001>.
 - [15] A. Berka, A. Hafiane, Y. Es-Saady, M. El Hajji, R. Canals, and R. Bouharroud, "CactiViT: Image-based smartphone application and transformer network for diagnosis of cactus cochineal," *Artificial Intelligence in Agriculture*, vol. 9, pp. 12-21, 2023, doi: <http://dx.doi.org/10.1016/j.aiia.2023.07.002>.
 - [16] V.-N. Pham, Q.-H. Do Ba, D.-A. Tran Le, Q.-M. Nguyen, D. Do Van, and L. Nguyen, "A Low-Cost Deep-Learning-Based System for Grading Cashew Nuts," *Computers*, vol. 13, no. 3, p. 71, 2024, doi: <http://dx.doi.org/10.3390/computers13030071>.
 - [17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1-14, 2015.
 - [18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer

- Vision," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 2818-2826, 2016, doi: <http://dx.doi.org/10.1109/CVPR.2016.308>.
- [19] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 1800-1807, 2017, doi: <http://dx.doi.org/10.1109/CVPR.2017.195>.
- [20] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2261-2269, 2017, doi: <http://dx.doi.org/10.1109/CVPR.2017.243>.
- [21] Y. Zhou, X. Zhang, Y. Wang, and B. Zhang, "Transfer learning and its application research," *Journal of Physics: Conference Series*, vol. 1920, no. 1, 2021, doi: <http://dx.doi.org/10.1088/1742-6596/1920/1/012058>.
- [22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, 2009: Ieee, pp. 248-255, doi: <http://dx.doi.org/10.1109/CVPR.2009.5206848>.
- [23] W. Zhi, Z. Chen, H. W. F. Yueng, Z. Lu, S. M. Zandavi, and Y. Y. Chung, "Layer removal for transfer learning with deep convolutional neural networks," in *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II 24*, 2017: Springer, pp. 460-469, doi: http://dx.doi.org/10.1007/978-3-319-70096-0_48.
- [24] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017. [Online]. Available: <http://arxiv.org/abs/1704.04861>.
- [25] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 4510-4520, 2018, doi: <http://dx.doi.org/10.1109/CVPR.2018.00474>.