Automatic Grading System of Incoming Raw Unclean Edible Bird Nest using Deep Learning Model

KHOR KHYE JIM

UNIVERSITI SAINS MALAYSIA

2021

Automatic Grading System of Incoming Raw Unclean Edible Bird Nest using Deep Learning Model

by

KHOR KHYE JIM

**Thesis submitted in fulfilment of the requirements
for the degree of
Bachelor of Mechanical Engineering**

June 2021

Declaration

I hereby declare that the work presented in this thesis is the result of my own work. Acknowledgement is credited to materials taken from various sources, wither published or unpublished by giving explicit references.

Name of Student: Khor Khye Jim

Matric Number: 138285

Signature: Jim

Date: 12-07-2021

ACKNOWLEDGEMENT

Firstly, I would like to express my gratitude to my supervisor, Ir. Dr Yen Kin Sam, for his endless support and motivation in my research work. His guidance and advice have supported me in selecting the suitable technique and equipment for the preparation work. Without his effort, I would not achieve the relevant information and result for meeting the objectives of the research work.

A great thanks go to the Chan Sek Ying and Khor Wen Hui from the quality department of Tian Ma Sdn. Bhd for providing a conducive workspace for capturing the EBN samples to my research work. Thanks to their guidance and teaching in the categorizing method of EBN grades according to the relevant features. Their efforts have helped me solve various problems and barriers in analyzing and preparing the EBN samples collected.

Last but not least, I would also thank USM Library for developing the USM Thesis Formatting program with the team members; Mdm. Jamilah Hassan Basri, Mdm. Lizawati Muhammadan, Mdm. Noor Adilah Azmi, Mdm. Rosnani Ahmad and Ms. Rabiahtul Adauwiyah Abu Hanifah. The information and knowledge from them are vital and helping me to write the thesis appropriately. Thank you for the tremendous positive feedbacks given, along with the continuous support received.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AMD	Advanced micro devices
ANFIS	Adaptive neuro-fuzzy inference system
ANN	Artificial neural network
AUC	Area of curve
AUROC	Area under the receiver operating characteristics
BA	Bat algorithm
CAM	Class activation mapping
CBAM	Convolutional Block Attention Module
CNN	Convolutional neural network
CPU	Central processing unit
DNCNN	Deep Normalized convolutional neural network
EBN	Edible bird nest
Grad-CAM	Gradient-weighted Class Activation Mapping
GTX	Giga Texel Shader eXtreme
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
KM	K-means clustering
KMBA	New hybrid bat algorithm clustering based on k-means
kNN	k-nearest neighbour
ReLU	Rectifier Linear Unit
VIPS	Valid images per second
VOPS	Valid FLOPs per second

LIST OF APPENDICES

Sistem Penggredan Automatik Sarang Burung Yang Mentah Dengan Menggunakan Model Deep Learning

ABSTRAK

Sistem penggredan untuk sarang burung mentah yang boleh dimakan memainkan peranan penting dalam menentukan harga pasaran antara industri EBN dan peternakan burung walit. Sistem ini juga bertindak sebagai proses utama untuk memantau kualiti EBN dalam barisan pengeluaran. Walau bagaimanapun, sistem penggredan berdasarkan pemerhatian manusia adalah subjektif dan bergantung pada pengalaman pekerja. Hal ini telah menghalang prestasi tinggi industri EBN dalam sistem penggredan. Pelaksanaan pengelasan pembelajaran mesin seperti ANFIS dan KMBA lebih standard dan tepat, tetapi mereka memerlukan pengalaman pekerja yang matang dengan spesifik teknik operasi. Oleh itu, model pembelajaran mendalam dengan kemampuan belajar sendiri pada proses pengekstrakan ciri dan penglibatan manusia yang rendah dikajikan untuk menyelesaikan kelemahan sistem penggredan berdasarkan pemerhatian manusia dan pengelasan konvensional. Pendekatan pembelajaran pemindahan dapat menjimatkan lebih banyak kuasa komputasi melalui model pra-latihan yang penanda aras berbanding dengan membina model dari awal. Hal ini juga mengurangkan masalah intensif tenaga kerja dan penggunaan masa yang panjang dalam mengumpulkan dataset untuk melatih model. Hasilnya, model terbaik yang diselaraskan ialah ResNet50 dengan ketepatan tertinggi 92.51% antara lima model pra-latihan yang dipilih dalam menentukan sebanyak 13 gred EBN. Prestasi model yang diselaraskan mengaungguli pengelasan konvensional ANFIS (88.24%) dan KMBA (85.60%) dalam sistem penggredan EBN. Pengaktifan Neuron dan analisis Grad-CAM diusulkan untuk memvisualisasikan ramalan model pada nilai EBN. Sampel EBN juga memasukkan gambar impian mendalam untuk meningkatkan ciri yanf telah dikesan oleh model untuk menunjukkan nilai EBN masing-masing. Analisis ini bertujuan untuk berbuktikan bahawa model yang diselaraskan telah mempelajari ciri khas dan relevan untuk meramalkan gred EBN. Cara-cara ini dapat memberikan pemahaman yang lebih jelas dan telus kepada manusia dalam ramalan model untuk meningkatkan kebolehpercayaan model dalam sistem penggredan EBN secara automatik.

Automatic Grading System of Incoming Raw Unclean Edible Bird Nest using Deep Learning Model

ABSTRACT

The grading system for raw unclean EBN plays a vital role in determining the market price between the EBN industry and swiftlet farming. The system also acts as a primary process to monitor the quality of EBN in the production line. However, the human visual system is subjective and based on the workers' experience, hindering a high performance in the grading system. Although the machine learning classifiers such as ANFIS and KMBA were more standardized and accurate, they required experience workers with the specific operation technique for the application. Therefore, a deep learning model with the self-learning ability on the feature extraction process and low human intervention was developed to solve the drawbacks of the human visual system and conventional algorithms. The transfer learning approach could save more computational power via a pre-trained model than build the model from scratch. It also reduces the labour-intensive and time-consuming issues in collecting the vast dataset to train the model. As a result, the best-fine-tuned model was ResNet50, with the highest accuracy of 92.51% among the five pre-trained models selected in identifying 13 of the EBN grades. The performance of the fine-tuned model outperformed the conventional classifiers of ANFIS (88.24%) and KMBA (85.60%) in the EBN grading system. Neuron activation and Grad-CAM analyses were proposed for visualizing the model's prediction on the EBN grades. The investigations aim to provide strong evidence that the fine-tuned model had learned the distinctive and relevant features for predicting the EBN grades. The EBN samples also fed into the deep dream images to enhance the features had detected by the model to indicate the respective EBN grades. The methods provide a better understanding to humans in the model's prediction for increasing the trustability of the model in the automatic EBN grading system.

Chapter 1 INTRODUCTION

1.1 Introduction

Edible Bird Nest (EBN) is the traditional functional food, and therapeutic herbal medicine people consume. It has high nutritional and medicinal value for combating malnutrition, anti-ageing, and enhancing people's metabolism and immune body system [1]. After collecting EBN from swiftlet premises, they go through the primary process of sorting into their respective grades without any cleaning process known as raw unclean EBN [2]. Afterwards, they undergo a sequence of operations show in figure 1.1 to produce the finished products known as raw, clean EBN before selling to the market.

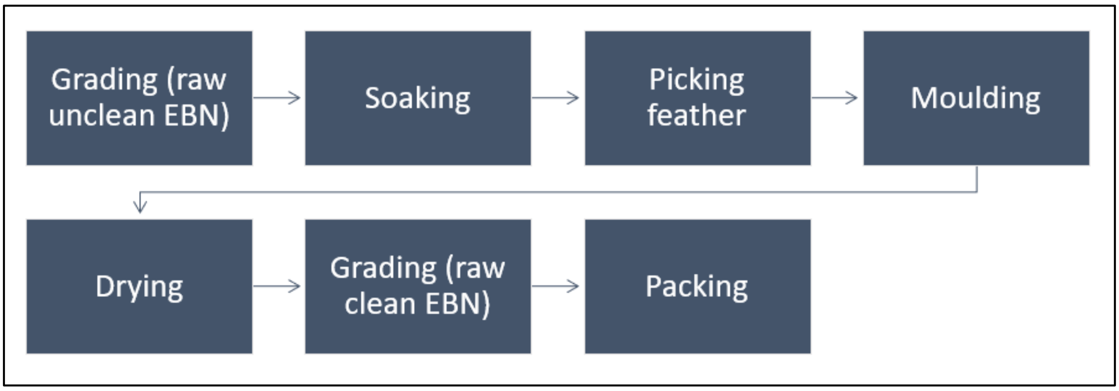


Figure 1.1: The flow chart of the traditional cleaning process of EBN [3]

The external morphological features of EBN, such as size, shape, colour, and impurities, are vital for the grading process [4]. The grading system acts as a guideline for the quality control of EBN and determines the market price of EBN [5]. Therefore, a human visual inspection system relies on professional panels to distinguishing the EBN grades according to the Standards of Malaysia of Edible Bird Nest [6]. However, the practice is prone to human fatigue issues which caused inconsistent results. The problems, such as misjudgment, low concentration of labours over long working time, and behaviour of panels have influenced the quality of grading and sorting system. As a result, the judgment always different between the panels as the skills and experiences are different. It increased the probability of misclassifying on the EBN grades [7]. Thus, productivity decreased in terms of efficiency and utilization performance because of the human fatigue issues.

The introduction of a machine learning algorithm has solved the limitations of the human visual system. It is also challenging in real-life applications. The real-life applications have various constraints such as lightning condition, characteristics of the object, and type of data required different processing techniques for building a flexible model. The process is labour intensive, especially in the workspace setup. Furthermore, high dimensional data increased the difficulty of choosing suitable algorithms [8]. The algorithm required a large amount of data as the training example to avoid overfitting issues. The correct label data also limited for large datasets as they may not obtain complete information well from a single observation, which affected the model accuracy [8]. Besides, the visual data is more complex. The problem, also known as concept drift, is developing a learning algorithm capable of adjusting quickly with the changes [8]. The ability of self-learning of deep learning can solve all these limitations.

Recently, deep learning played a vital role in solving various problems such as computer vision, text recognition, sound, and natural language processing. A convolutional neural network (CNN) is a type of deep learning model problem-solving by visual understanding. CNN learns and solves the problems through the input images by the multiple hidden layers [9]. Besides, the deep learning model can achieve a precise and consistent result without a manual feature extraction by a domain expert. For instance, AlexNet has achieved the champion in the 2012 ImageNet competition with only 15.3% of the top-5 error rate compared with the second place of 26.2% [10]. The potential of deep learning was studied in the EBN grades classification using the transfer learning approach.

The research aims to develop an automatic EBN grading system with a deep learning model. The transfer learning method uses the pre-trained model benchmark to improve the computational cost and accuracy performance in the EBN classification grades. Therefore, the pre-trained models such as AlexNet, SqueezeNet, ResNet18, MobileNetV2, and ResNet50 were compared based on their memory consumption, strength, and accuracy performance for choosing the best model in EBN grades classification. Further investigation also applied to the model selected for discovering the model's reliability in the EBN grading system through the visualization method. The purpose was to observe and monitor the internal working of the model for the prediction of EBN grades.

1.2 Problem Statement

A machine learning algorithm helped build a system with high standardization and automation. The performance also outperformed the human visual system in various vision tasks with high speed and consistency. However, it is limited in handling the variety of complex vision data in the manufacturing system. It also requires manual feature extraction by a domain expert to guide the system in learning specific features for the final classification. The high-dimensional data raised problems in fitting many parameters for generalizing the system. Thus, it is time-consuming and labour-intensive to build a flexible machine learning algorithm in the unpredictable condition of the production line. Hence, the potential of deep learning was studied to implement in the EBN grading system due to its self-learning capability and fruitful achievements in a variety of vision tasks.

1.3 Objective

1. To compare the pre-trained models in developing an automatic grading system for incoming raw unclean of EBN.
2. To assess the accuracy of the deep learning model in the EBN grading system.
3. To investigate the relationship between the predicted outcome and the extracted features in the deep learning model.

1.4 Thesis outline

The thesis was summarised as follows. Chapter 2 described the reading of published resources used as the guidelines for completing the thesis. It involved the current automation system in the EBN grading system, deep learning in the agriculture fields, the performance metrics, and the visualization method for explaining the interpretability and internal working of the model. While Chapter 3 listed the equipment used for developing the deep learning model. The discussion on pre-processing techniques to prepare the EBN dataset for model training. The vital features for determining the EBN grades according to the industrial standard also included. Besides, the steps and parameters to train and tune a deep learning model. The performance metrics used to evaluate the performance of the model. Therefore, the visualization method explained the decision-making of the model for the final output prediction. Last but not least, chapter 4 discussed the performance of the model in predicting the EBN grades. The table form displayed the results and information obtained for explanation purposes. The highest activation channels in the specific layers were extracted for visualizing the distinctive features learned by the model toward the final prediction. Moreover, the formation of the heatmap displayed the vital region on the input images used by the model to predict the EBN grades. Finally, chapter 5 discussed the conclusion drawn from the research work. It also discussed the thoughts and ideas of the research work synthesized to develop an automatic EBN grading system. The chapter ended with the recommendation and suggestion for future work.

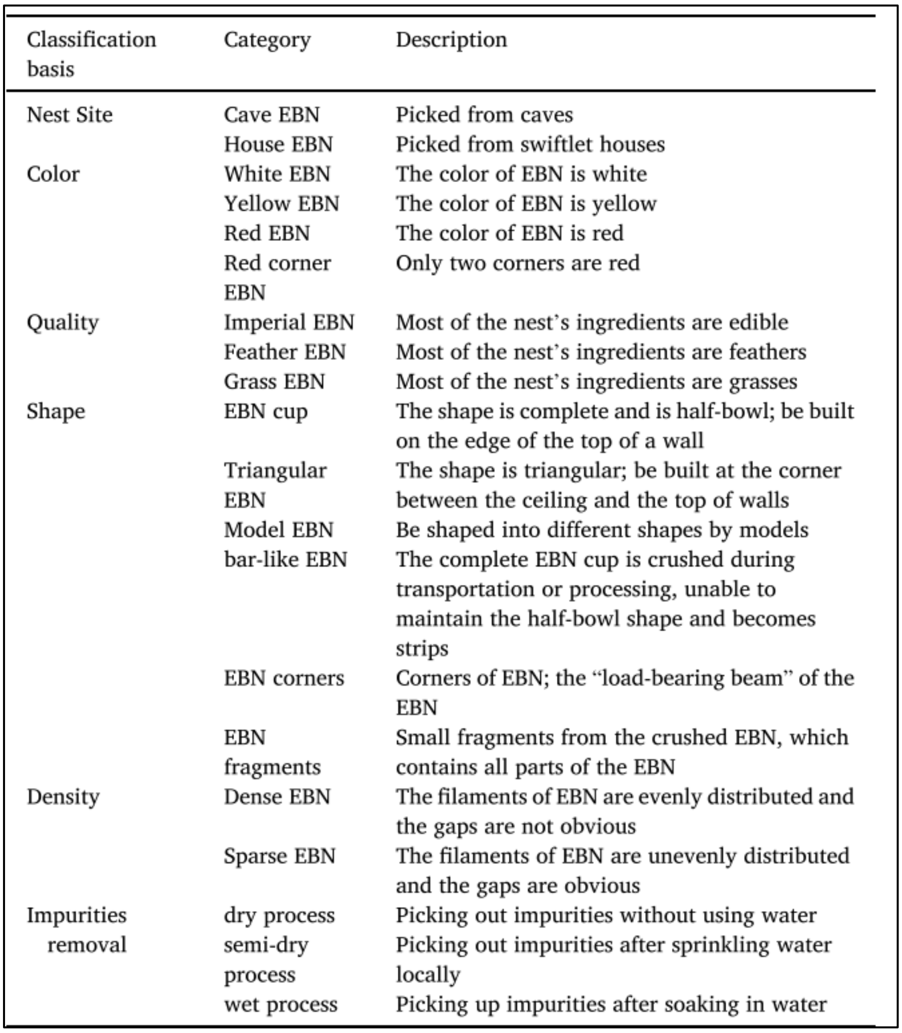
Chapter 2 LITERATURE REVIEW

2.1 Introduction

The research works and achievements related to the automatic grading system will be reviewed and discussed in this chapter. The method to identify the EBN grades will be studied according to industrial and Malaysian standards. The challenges and limitations for the current practices, such as the human visual system and conventional classifiers, will be reviewed. The benefits of the deep learning approach for various vision tasks will be studied, especially in agriculture. This chapter also explored the potential of the deep learning model to solve the limitation of the current practices in the grading system.

2.2 The EBN grading system

The harvesting EBN from a cave or house-building will send to the industry for the treatment process before selling to the customers. From Y.Dai et al. [12], the treatment process is divided into three categories: primary, deep and biotechnological processing. The primary method is to categorize the raw unclean EBN into their respective grades based on the impurities, shape, colour, and other influences before proceeding to the following process shown in figure 2.1 [12]. This process is to trace and monitor the raw EBN quality from the different sources. From L.Qi Hao et al. [5], a high-quality standard of raw and processed EBN is vital to lead a stable market price in both local and international trading. Hence, a grading system is required to grade the raw EBN more standardize and accurately. The system also assists in monitoring and controlling the quality standard of EBN from swiftlet farming.



Classification basis	Category	Description
Nest Site	Cave EBN	Picked from caves
	House EBN	Picked from swiftlet houses
Color	White EBN	The color of EBN is white
	Yellow EBN	The color of EBN is yellow
	Red EBN	The color of EBN is red
	Red corner EBN	Only two corners are red
Quality	Imperial EBN	Most of the nest's ingredients are edible
	Feather EBN	Most of the nest's ingredients are feathers
	Grass EBN	Most of the nest's ingredients are grasses
Shape	EBN cup	The shape is complete and is half-bowl; be built on the edge of the top of a wall
	Triangular EBN	The shape is triangular; be built at the corner between the ceiling and the top of walls
	Model EBN	Be shaped into different shapes by models
	bar-like EBN	The complete EBN cup is crushed during transportation or processing, unable to maintain the half-bowl shape and becomes strips
	EBN corners	Corners of EBN; the "load-bearing beam" of the EBN
	EBN fragments	Small fragments from the crushed EBN, which contains all parts of the EBN
Density	Dense EBN	The filaments of EBN are evenly distributed and the gaps are not obvious
	Sparse EBN	The filaments of EBN are unevenly distributed and the gaps are obvious
Impurities removal	dry process	Picking out impurities without using water
	semi-dry process	Picking out impurities after sprinkling water locally
	wet process	Picking up impurities after soaking in water

Figure 2.1: The classification features of EBN [12]

2.3 Features extraction for EBN grading system

The current practice for the grading system is the human visual inspection system by observing the vital features such as impurities, shape, and colour to identify the EBN grades [13][4]. In figure 2.2, A-B represents the length of the EBN, C-D represents the height of the EBN, and D-E represents the depth of the EBN. From M. Y. Koay et al. [4], the conventional way was using the index finger, middle finger, and ring finger to determine the length of the EBN between points A and B. The suggested size for the high-grade EBN was the same size with three stated fingers or bigger [4]. Moreover, the number of impurities and colours were categorized based on the visual of professional panels. However, the conventional method usually produced inconsistent results as it varies with the skills of the panels. The technique was also prone to human fatigue issues which influence the quality control of the EBN. Hence, the machine learning classifiers is introduced to solve the limitations of the human visual system and generate a more standardize EBN grading system.

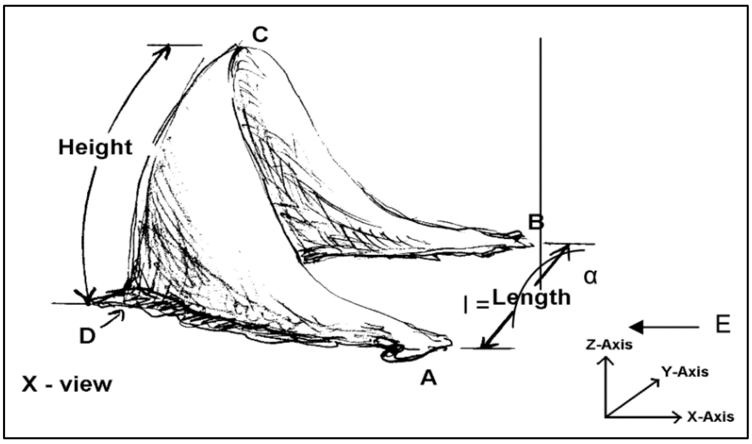


Figure 2.2: The auxiliary view of EBN [6]

M. Koay et al. [4] proposed Adaptive Neuro-Fuzzy Inference System (ANFIS) to identify three grades (AA, A, and B) of EBN. ANFIS is the combination of the Artificial Neural Network (ANN) and Fuzzy Inference System (FIS) by using the learning ability of ANN in optimizing the if-then rules and membership function parameters of FIS [4]. The accuracy performance was then compared with the k-nearest neighbour (kNN) classifier for the performance evaluation. From the result, the k-nearest neighbour (kNN) classifier achieved an accuracy performance of 83.27% with the linearised data, while the Fuzzy C-Means in ANFIS achieved better accuracy up to 88.24% [4].

Next, J. E. Gan [13] proposed a New Hybrid Bat Algorithm Clustering based on K-Means (KMBA) to grade three grades of EBN (AA, A, and B). KMBA is the combination of Bat Algorithm (BA) and K-Means Clustering (KM). This method has covered the limitation of both algorithms. In the working principle, KM is limited only in the local search, and BA assigned the random value for the optimization [13]. KM will provide the initial point for BA in generating the interim solutions with the combination method. As a result, the standard BA achieved 80.29% accuracy with the decimal scaled data, and the KMBA achieved 85.60% accuracy in the prediction [13]. Thus, the machine learning classifiers outperformed human visual inspection in the quality control of EBN.

2.4 Limitation of machine learning classifiers

The machine learning classifiers can produce higher accuracy performance in solving the vision tasks than the human observation. However, the success of the machine learning classifiers highly relied on the excellent and well-defined features extraction process by the domain experts, as shown in figure 2.3 of shallow machine learning [11]. These are limited to the machine learning classifiers for the real-time application as they are sensitive to the camera's view variation, illustration, deformation, and background clutter. In short, many constrained parameters have increased the difficulty to prepare a dataset for the classifiers.

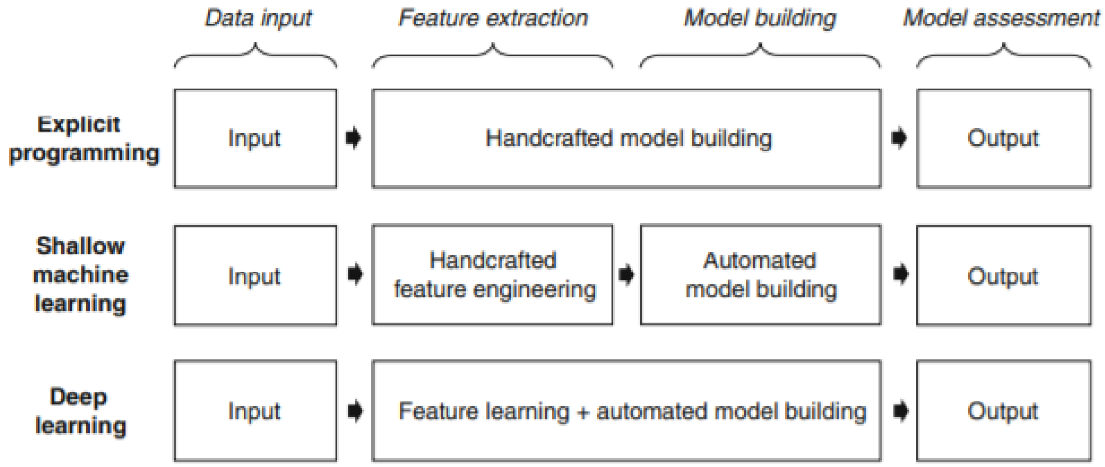


Figure 2.3: The process of the machine learning model buildings [11]

The common challenges of the machine learning classifiers are extracted the relevant features from the available data, which compromise different formats and parameters [14]. The quantity, quality, and labels data are vital in training the algorithms by ignoring the irrelevant and redundant features that will affect the training performance. However, the current machine learning classifiers only can handle the continuous and nominal data [14]. It was challenging to determine the factors and parameters setting of the algorithm to get hold and secure on any type of data.

Besides, the data pre-processing has become another challenge in preparing a dataset. The pre-processing method heavily depends on the selected classifiers and type of available data, which increased the difficulty in making selection [14]. The missing and imbalanced data also affected the model performance during training. The data cleaning tasks are essential to filter the wrong information in a dataset. The variation of classifiers also challenged selected suitable machine learning classifiers as they have their specific parameters tuned [14]. It is labour-intensive and time-consuming to choose the classifiers for developing a model that can handle the abrupt change of available data in the manufacturing process.

2.5 The deep learning approach

The deep learning model can solve the limitations of handcrafted feature engineering in machine learning classifiers with minimal human intervention. The deep learning model mimics human brain behaviour in solving complex tasks with minimal human intervention during making decisions [11]. The automatic feature learning process is hierarchical, which assembled the learned features from the top layer to the bottom layer of the model [11]. Convolutional neural network (CNN) is widely used in the deep learning model to solve various vision tasks. This is because CNN is similar to the biological neural network. The model can be adjusted with the depth and width to solve the specific problems and has strong recognition in the natural images [15]. It proved when the AlexNet won the ILSVRC 2012 competition with only a 16.4% error rate in the large-scale image classification compared with the 28.2% for the traditional machine learning algorithm [10,15]. Hence, the deep learning model can be applied directly to high-dimensional data with its self-learning capability. The model betters in coping with the variation of illustration, background clutter, and unstructured data than machine learning classifiers [11]. Nevertheless, different deep learning architectures have different performances, computational resources, and interpretability with the type of data used [11]. However, there is a lack of the established guidelines in selecting the suitable deep learning approach to solve a specific problem.

2.6 The application of deep learning in the agriculture

Recently, deep learning became a popular approach used by researchers and industries to solve the various visual tasks due to their outperformed performance. Deep learning has been implemented in the smart agriculture fields, mainly for the plant and crop classification in terms of yield predictions, disaster monitoring, and others because of their strong image processing capability [16]. N. Zhu et al. [17] introduced the transfer learning approach to extract the information of cultivated land for land classification and area estimation [17]. As a result, the overall precision is around 90% for the deep learning model with transfer learning [17][18].

Besides, M. M. Raikar et al. [19] has applied the deep learning model in the Okra ladies finger grading system. The dataset consists of 3200 images taken by the mobile camera of 13 megapixels under sunlight [19]. The ladies' fingers had four classes according to their respective length such as small (6-8cm), medium (9-15cm), large (16-21cm), and extra-large (>22cm) [19]. Therefore, the different deep neural network architectures such as AlexNet, GoogleNet, and ResNet50 are compared to select the best architecture for classifying the length of ladies fingers. From the result, ResNet50 has the best accuracy performance of 99.17%, followed by GoogleNet with 68.99%, and the lowest was AlexNet with 63.45% [19].

A. Pande et al. [20] also proposed the transfer learning approach with the Inception V3 model to classify and grade the apples. The apples were categorized as spoilt apples and good apples with the three grades. The images captured by the camera of the Raspberry Pi [20]. The machine-learning algorithm then pre-processed the captured images to remove the background and noise from the images. As a result, the Inception V3 achieved the top 5 accuracy of 90%, while the decision tree achieved only around 72% accuracy [20].

The deep learning model has eliminated the heavily feature extraction process by a domain expert in data preparation. However, different pre-trained models have their advantages in various classification tasks as they depend on knowledge transfer. The type and complexity of the data also have a considerable influence on the model selection.

2.7 Transfer Learning

Transfer learning is a method that repurposes a pre-trained model on the primary task (source domain) to solve another task (target domain) via the leveraging feature representation method [21]. There is a training and testing set categorized as the source and target domain, respectively. The model uses the source domain with a label to train and predict the target domain not involved in the training process. The basic implementation of transfer learning is simply replacing the top layers of the pre-trained model used in classifying 1000 objects according to the original dataset with new layers that suit the new tasks.

There are three general criteria for using knowledge transfer in the image classification's problems: time to transfer, type of knowledge transfer, and transfer strategy [22]. Firstly, the target domain must have a similar high feature with the source domain data. The high similarity and sufficient data will enhance the knowledge transfer and achieve impressive target domain data results. Second, the type of knowledge transfer is inductive transfer learning, instance transfer learning, and parameter transfer learning [22]. Inductive transfer learning uses the source domain instances and labels to transfer the knowledge gain for the classification of EBN.

There are three methods used in the transfer strategy, as shown in figure 2.4 [23]. Strategy one uses the pre-trained model and trains all the layers with the new dataset by updating all the weights and biases. It is used when the data similarity between the source domain and target domain is relatively small with a larger dataset. Still, it consumes a high computational power for the model. Next, strategy two freezes parts of the layers and trains the rest of the layers. The bottom layers of the architecture are learning the general features, while the top layers are learning the high-level detail of the features. The top layers allowed to retrain the weights during backpropagation for learning the new features. Furthermore, strategy three known as feature extraction, which is a basic implementation of transfer learning. This strategy is suitable for the small dataset and high data similarity between the target and source domains as shown in figure 2.5.

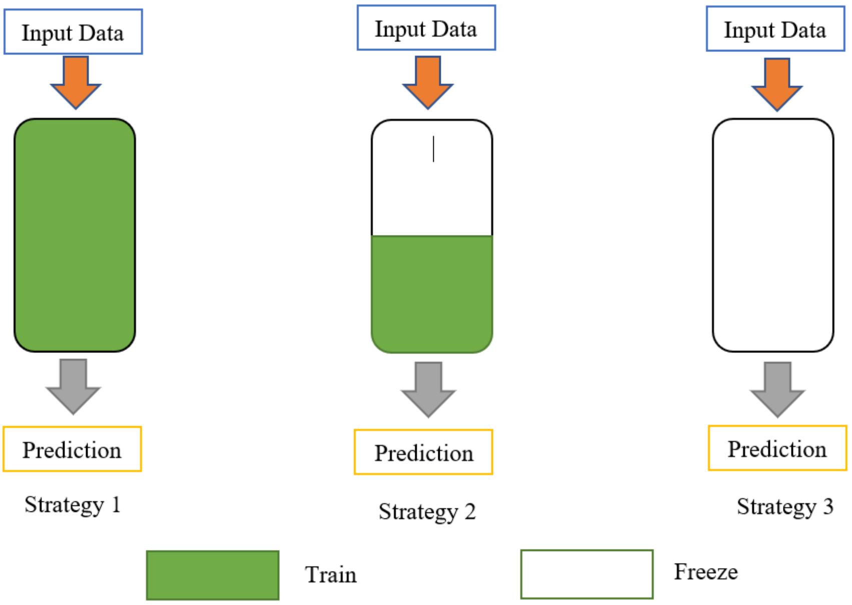


Figure 2.4: The transfer strategy for fine-tuning the model

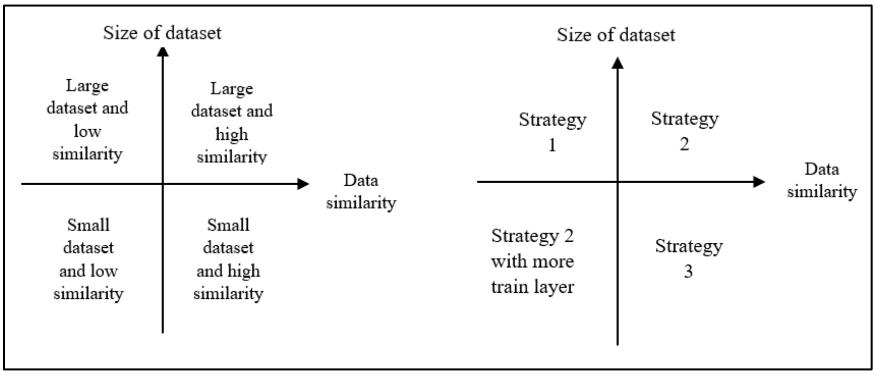

Size of dataset
Large dataset and low similarity
Large dataset and high similarity
Data similarity
Small dataset and low similarity
Small dataset and high similarity
Size of dataset
Strategy 1
Strategy 2
Data similarity
Strategy 2 with more train layer
Strategy 3

Figure 2.5: Data similarity matrix and decision making for fine-tuning the model [23]

2.8 Selection of Pre-Trained Model

2.8.1 AlexNet

Alex Krizhevsky introduced AlexNet, the first convolution neural network that improved model performance via graphics processing unit (GPU). It has contributed to many developments of deep learning in the application of computer vision fields. It has eight learnable layers in the network with five convolutional layers and three fully connected layers [10], as shown in figure 2.6. Rectified Linear Units (ReLUs) can achieve the same accuracy as the hyperbolic tangent activation function (TanH) with a lower training time. The gradient descent also contributes positively to improve the performance of models on the large volume of the dataset. It also uses overlapping pooling by setting the stride value lower than kernel size. This can reduce the model's size, achieve a lower error percentage, and be more resistant to overfitting issues. Dropout had applied by turning off the hidden neuron of convolutional layers temporarily and randomly with predefined probability. This technique forced the model to learn more robust features with different random subsets during forwarding and backward propagation, thus lowering each epoch's training time.

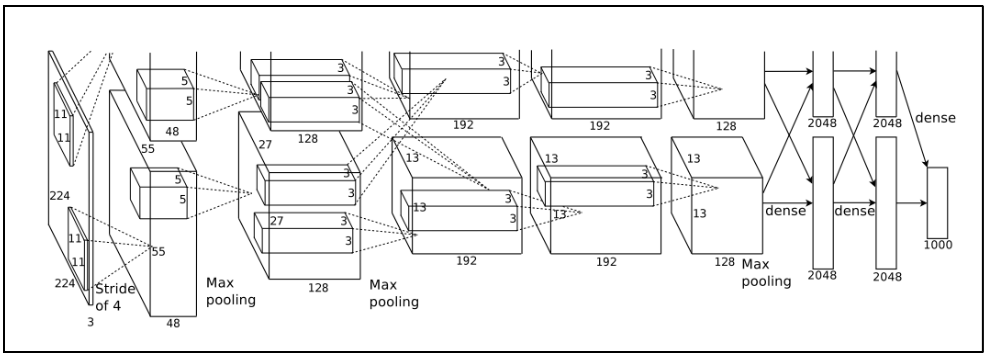


Figure 2.6: The AlexNet architecture [10]

2.8.2 **SqueezeNet**

SqueezeNet released on 22 February 2016 by Forrest Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, Bill Dally, and Kurt Keutzer. It consists of Fire modules building blocks formed by the squeeze layer and expanded layer. The squeeze layer is a convolutional layer with only 1x1 filters size, while the expand layer is a combination of 1x1 and 3x3 convolution filters [24]. The squeeze layer has a lower feature map than the expanded layer, which will form a compression effect on the extracted feature map to reduce the number of parameters in the layers. These Fire modules will be stacked together to build the SqueezeNet architecture. It uses the pointwise filters to replace the 3x3 filters to reduce the number of parameters in the layers, thus decreasing computation effort. It also reduces the number of input channels to 3x3 filters for lowering the model weights [25]. The small filter size downsamples the layers, especially in the early layer. It keeps the high activation maps, which leads to higher accuracy in classification performance. SqueezeNet preserves a high top-5 ImageNet accuracy of 80.3% with a low quantity of parameters.

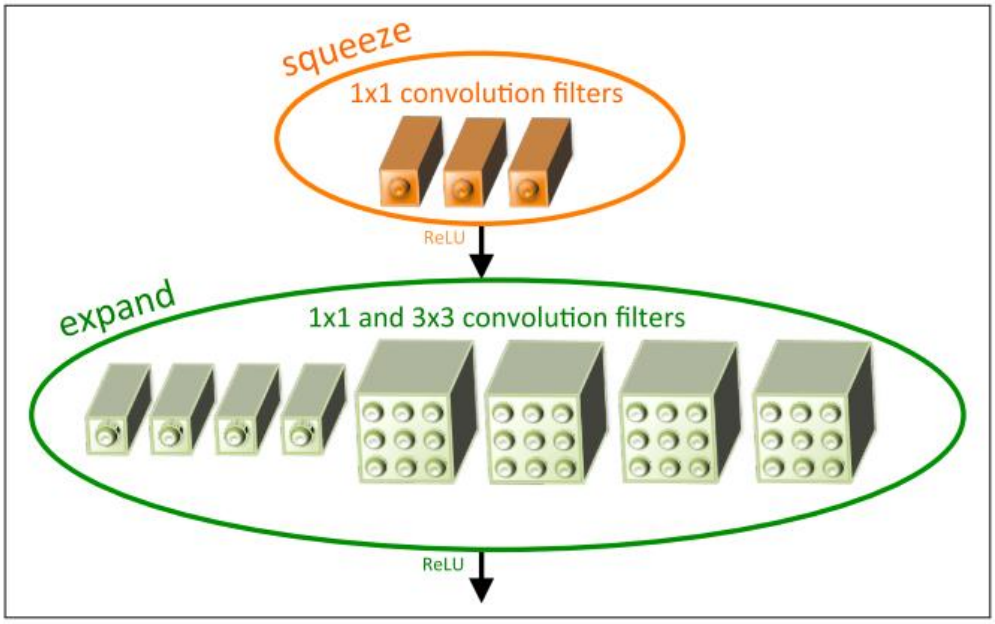


Figure 2.7: The fire module of SqueezeNet model [24]

2.8.3 **MobileNetv2**

MobileNetV2 contains 53 layers in the convolutional neural network. It has a vital building block named depthwise separable convolutions for improving the model's efficiency, as shown in figure 2.8. The building block is divided into two layers which are depthwise (3x3) convolution and pointwise (1x1) convolutions [26]. The depthwise layer filters the input channel via a single convolution filter with a small computational cost. In contrast, the pointwise layer uses linear combinations of the input to build new features. It also uses an inverted residual similar to a residual block by constructing a shortcut path between the bottlenecks, as shown in figure 2.9 [26]. The intuition of inverted residual ensures the gradient propagate across the multi-layers, especially in the deeper layers. The inverted design aims to improve the model efficiency by reducing the usage of memory during training.

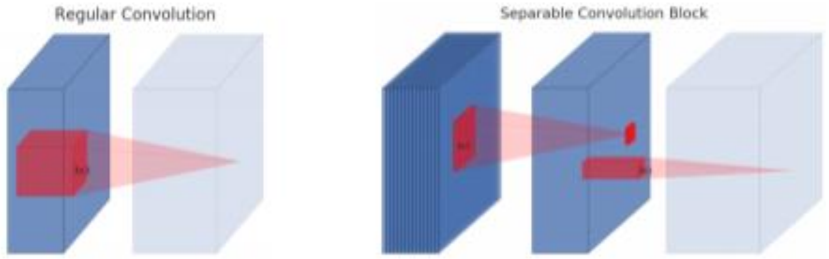


Figure 2.8: The regular convolution vs separable convolution block [26]

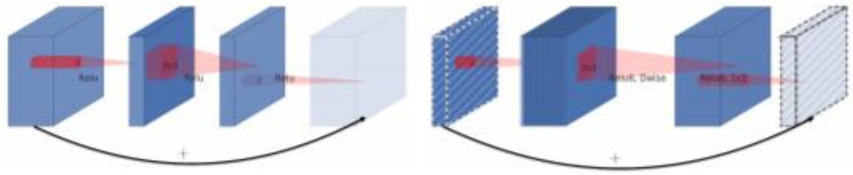


Figure 2.9: The residual block versus inverted residual block [26]

2.8.4 **ResNet 18 & 50**

ResNet, one of the pre-trained models, also known as Residual Network, is a classic network introduced by Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun in their paper named "Deep Residual Learning for Image Recognition" in 2015 [27]. ResNet has outperformed the human level in the various complexities of computer vision tasks. ResNet 50 has 49 convolutional layers in the architecture, while ResNet 18 has 17 layers. The deeper the model, the more complex features can be learned. However, it is easier exposed to the degradation problem with accuracy saturated. Thus, ResNet introduced the concept of skip connections to overcome the problem [27]. The plain network stacked the convolutional layers one after the other. The ResNet is similar to a plain network but with additional input to the convolutional block's output, which has the same image size [27]. Figure 2.10 shows the input is multiplied by the layer's weights, followed by adding bias and adding an "x". This was known as an identity shortcut [27]. This identity shortcut able to help in reducing the degradation problems by allowing the gradient to flow through via the alternate path and ensure the higher layer learn as good as, the lower layer [27]. The reformulation helps the solvers find the perturbation of the multiple nonlinear layers for approximating the identity mapping to solve the degradation problem.

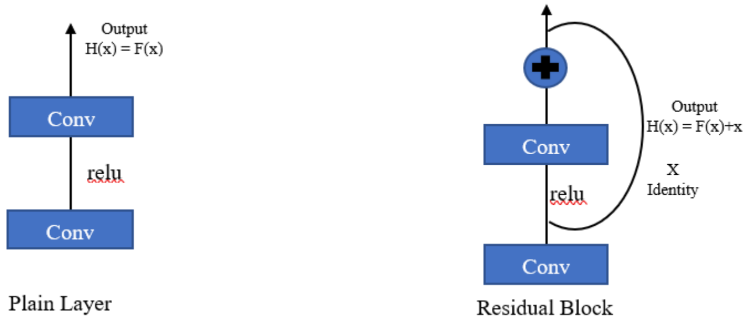


Figure 2.10: The plain layer vs residual block [27]

Since there is no correct and proper guideline to select the pre-trained models in various vision tasks. The models also have different accuracy performances in different tasks. Hence, the models will be studied via experiment to determine the best model in the grading system. C. Luo et al. [21] proposed two metrics which are Valid Images Per Second (VIPS) and Valid FLOPs Per Second (VOPS), which will be used to calculate the memory consumption of models. VIPS calculated how long will be processed by the model for the one image. At the same time, VOPS is the calculation of memory used by the model in processing the images [21]. The memory consumption is related linearly to the computational time and the cost of hardware required to install the model in the devices. These help to rank the ability of the models in various vision tasks.

$$VIPS = \sum_{i=1}^{n} accuracy_i \times \frac{1}{time_i} \tag{2.1}$$

$$VOPS = \sum_{i=1}^{n} accuracy_i \times FLOP_s \times \frac{1}{time_i} \tag{2.2}$$

where:

- $accuracy_i$ is the validation accuracy of the model of i^{th} test
- $time_i$ is the average processing time per image of i^{th} test
- $FLOPs_i$ is the memory of the model for the i^{th} test

2.9 Model optimization

The proper optimization method helps improve the accuracy performance of the model. D. Motta et al. [28] proposed that the hyperparameters such as initial learning rate, mini-batch size, and optimizers have an immense influence on the learning process of the model. The hyperparameters were fine-tuned through the random search method under the specification of the hardware system used. Four famous optimizers are Stochastic gradient descent with momentum, adaptive gradient algorithm (Adagrad), root mean square propagation (RMSprop), and Adaptive Moment Estimation (Adam), which was evaluated in various pre-trained models to search for the best optimizer. From the result, DenseNet201 with the Adam optimizer achieved the highest accuracy, which is 93.5% for the six classes of adult mosquitoes [28]. Thus, the Adam optimizer is a powerful optimizer in reducing the model's loss function during the training.

2.10 Model interpretability and explanation for the output predictions

However, a high accuracy performance did not mean that the model has high interpretability. The deep learning model is the black box operation in which the internal working of the model remains unexplainable [29]. The question is whether the judgement is made based on the relevant features and understandable by human knowledge. Therefore, it is essential to interpret the relevant features of the deep learning model while making a judgement for eliminating the bias issues. Thus, the visualization method is introduced to explain the feature learning representations by the model in making decisions [29]. A high interpretability model helps build a robust and safe system for real-life applications.

2.10.1 **Neurons activation algorithm**

F. Hohman et al. [30] proposed the neuron's activation to explain the judgment made by the model. The strongest activation channel was computed from the convolutional layer by a few input images. The strongest activation channel of the input images was sorted, then recorded by the Pareto principle, starting from the largest number of activated channels until the cumulative sum of all the recorded channels [30]. Later, the channels were recorded into the aggregated activation matrix to display the most activated channels and the final class representation in the model. Moreover, F. Jia et al. [31] extracted the activation strength from the last two convolutional layers of the model to visualize the features learned. This method enabled the visualization of the kernel in the convolutional layer extracted to understand the feature learning of the model. The activation index was used to measure the activation degree of a kernel. This is because the mostly activated channel represented the pattern of the signal which the kernel likely to see. H. Wong et al. [32] proposed that a drastic drop in the total overlapping channel of the last convolution layer of AlexNet show the feature learned by the model toward the final classification between two different classes became distinctive and exclusive to each other.

2.10.2 **Deep Dream**

The deep dream is an algorithm that loosely mimics the actual visual cortex of the human brain to generate an output image. It could visualize the feature learned by the model during training and provided new remix visual concepts [33]. It tends to pick up specific shapes and textures from the guide images to generate the feature learned on the images [33]. It also can visualize any layer with the gradient ascent method to evaluate the correctness of the features learned by the model. However, the technique can identify the model that had learned the relevant features from the object but weak in providing the information to differentiate a high similarity object, especially in the grading system.

2.10.3 **Grad-CAM analysis**

Next, the Gradient-weighted Class Activation Mapping (Grad-CAM) is used in visualizing the vital region on the input images for the prediction. I. Kim et al. [34] proposed that Grad-CAM is better than Class Activation Mapping (CAM). It can generate the essential visual explanation on any of the CNN layers instead of CAM only applicable to the model with global average pooling and one fully connected layer [34][35]. Grad-CAM evaluated the localization of the model by identifying the discriminative region of interest (ROI) on the input images to predict the output class with heatmap formation. From the result, Grad-CAM can determine ROI for representing the discriminate feature learned by the model in classifying six different modality classes. Still, it also led to the noise effect and included some of the background shown in figure 2.11 [34]. S. Woo et al. [36] used Grad-CAM to visualize the final predictions of the pre-trained model with the Convolutional Block Attention Module (CBAM) on the classification of Image Net-1K. The combination model focuses more on the relevant ROI than the conventional model for predicting their valid class in figure 2.12 [36]. However, Grad-CAM analysis weak in providing the trustworthy information, especially in classifying the differences between similar objects. There was a lack of information to predict the Eskimo dog just based on the eyes as the biased result is suspected show in figure 2.12.

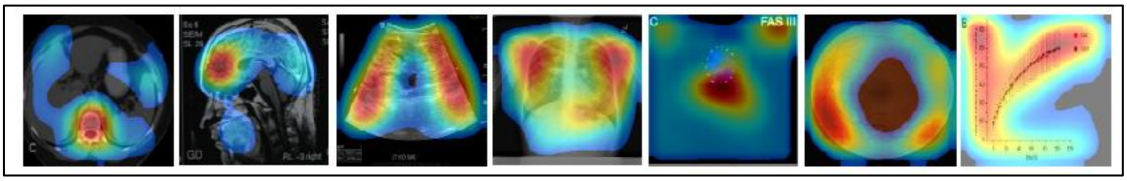


Figure 2.11: The Grad-CAM visualization on the modality class [34]

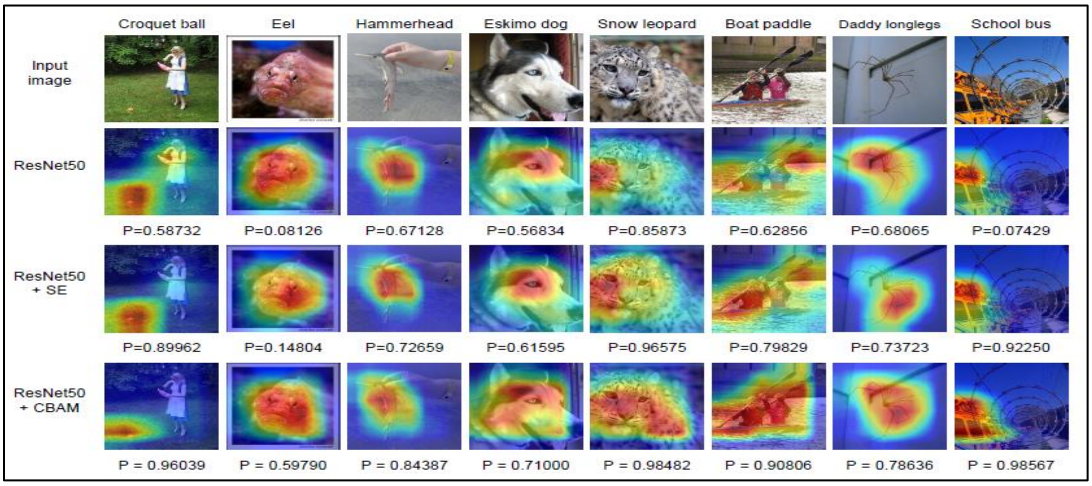


Figure 2.12: The Grad-CAM visualization result on the Image Net-1k [36]

2.11 Summary

This chapter had reviewed the method to determine the EBN grades based on the industrial standard. The grading system is vital to ensure a stable market price. However, the current practice with human vision is subjective and varied with the skills of panels. Machine learning classifiers are labour-intensive and time-consuming as they had various constraints in the feature extraction process and required professional workers with specific operation techniques. Thus, the deep learning approach was proposed as it achieved fruitful achievement in agriculture, such as classification in land area, apple, and okra ladies with minimal human intervention. The transfer learning approach with the knowledge transfer can save more computation cost in developing a model instead of from scratch. However, the accuracy performance depends on the data similarity with the original dataset. The higher the data similarity, the lower the computational cost to achieve a high accuracy performance. Since the deep learning model is a black box operation, a visualization method was applied to interpret the internal working of the model's prediction. Hence, the neuron activation algorithm, deep dream, and Grad-CAM analysis were used to monitor the model for learning the relevant feature and exclusive to each other in the EBN grading system. The method can explain the model had learned the appropriate features for the prediction. However, it still lacks good information explaining the prediction on a similar object with different classes or features.

Chapter 3 METHODOLOGY

3.1 Introduction

The chapter discussed the data preparation for developing a deep learning model. It involved the crucial steps in pre-processing the EBN images used for model training. It also listed the criteria in selecting the pre-trained model to determine the model's characteristics in developing an automatic grading system. The performance metrics used to verify the performance of the model in the EBN grading system. Next, it explained the working principle of the Grad-CAM analysis and neuron activation algorithm in explaining the feature learning by the model from the EBN images.

3.2 List of the equipment

3.2.1 Hardware

The camera system with two megapixels and lenses of f/1.9 was used in the image acquisition process. The camera system was equipped with diffuse light for reducing the shadow during the image capturing process. The camera was installed on the portable stand for easing the image capturing process from the workspace. The computing system used in developing the deep learning model was CPU AMD RYZEN 5 3600, 8GB of RAM, 8GB of memory, and GeForce GTX 1660 graphic card.

3.2.2 Software

The MATLAB version R2021a education version was used in the development of a deep learning model. It involved the pre-processing and visualizing algorithms for processing the EBN images. The pre-trained models were also selected from MATLAB for developing an automatic EBN grading system.

3.3 Data Preparation

The EBN images were captured from the Tian MA Sdn Bhd according to their grading system, as shown in table 3.1. There were 13 EBN grades were collected based on their vital features with the advice of the professional panels of the EBN industry. The EBN grades were named according to the industrial standard. Table 3.2 shows the total EBN samples and images captured.

Table 3.1: The industrial standard of the EBN grading system

Impurities	**Shape**	**Colour**	**Crack**	**Grade**
Few (A)	180° (A)	White (W)	-	AA-W
Few (A)	180° (A)	Yellow (Y)	-	AA-Y
Few (A)	160/135)° (B)	White (W)	-	AB-W
Few (A)	(160/135)° (B)	Yellow (Y)	-	AB-Y
Intermediate (B)	180° (A)	White (W)	-	BA-W
Intermediate (B)	180° (A)	Yellow (Y)	-	BA-Y
Intermediate (B)	(160/135)° (B)	White (W)	-	BB-W
Intermediate (B)	(160/135)° (B)	Yellow (Y)	-	BB-Y
Heavy (C)	-	-	Small crack	C1
Heavy (C)	-	White (W)	-	C-W
Heavy (C)	-	Yellow (Y)	-	CY
-	-	-	Severe crack	H
-	-	Colour difference	-	L

Table 3.2: The EBN samples and images collected according to the grades

No	**EBN Grades**	**Number of samples**	**Number of images**
1	AA-W	143	429
2	AA-Y	357	1071
3	AB-W	110	330
4	AB-Y	285	855
5	BA-W	132	396
6	BA-Y	390	1170
7	BB-W	33	99
8	BB-Y	397	1191
9	C1	146	438
10	C-W	87	261
11	C-Y	211	633
12	H	54	162
13	L	12	36
14	**Total**	**2357**	**7071**

3.4 Data pre-processing

The EBN samples were captured in the three views, which were back (b), top (f), and side (L), as shown in figure 3.1. The three views included all the EBN surfaces to improve the model toward the randomness effect. Furthermore, augmentation techniques such as rotation and translation were applied to generate the additional EBN images to meet the deep learning model requirement. The balanced data helped improve the model's accuracy during training as deep learning was a large-scale image classification. Besides, the EBN images were cropped to remove the unnecessary background from the images. The background of the images was also pre-processed into black by reducing the noise effects via the colour thresholding method. Moreover, the size of the images resized according to the pre-trained models' requirement [37].

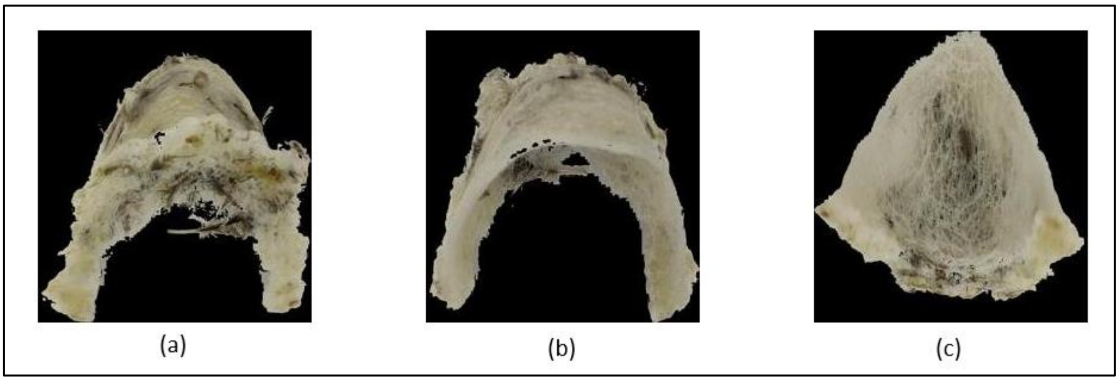


Figure 3.1: The EBN samples collected (a) back view (b) top view (c) side view

3.5 Data Splitting

The percentage split for the total EBN images was 80% of the training and 20 % of the validation set. The purpose of data splitting was to find the model hyperparameter and minimize the model error of the model, which helps in improving the generalization performance of the model selected during training [38].

3.6 Selection of hyperparameter

The hyperparameter tuning was an essential criterion in improving the model performance and reducing the error rate. The hyperparameters such as initial learning rate, mini-batch size, and others were fine-tuned to find suitable values to improve the model performance [28]. The grid search method was applied to find the best value for the model training. The log-10 scale was used in tuning the initial learning rate (0.001 – 0.00001) and the log-2 scale for the mini-batch size (2^6-2^8) [28]. Next, the different optimizers such as Adam, SGDM, and RMSprop were compared on their performance in the error minimization for the cost function of the model [28]. Besides, the dataset was shuffle for every epoch to minimize the training loss and prevent any bias during model training. The L2 regularisation, known as ridge regression, was applied to solve the model overfit issues for improving model performance. Table 3.3 shows the fixed variables used during the model training for every pre-trained model selected.

Table 3.3: The fixed variables during the model training

Parameters	**Details**
Number of epochs	20
Shuffle	Every-epoch
Learn rate schedule	Piecewise
Learn rate drop period	10
Learn rate drop factor	0.1
L2 Regularization	$1e^{-4}$

3.7 Model comparison and training

The pre-trained models selected were AlexNet, SqueezeNet, ResNet18, MobileNetV2, and ResNet50. The characteristics of the models were listed in table 3.4. Five models were compared via the experimental result to determine the best model for the EBN grading system. Three transfer strategies, feature extraction, fine-tuning with specific convolution layers, and fine-tuning with all convolutional layers, were also compared to identify the best strategy in training the model.

Table 3.4: The comparison of the selected pre-trained models

Feature	**AlexNet**	**SqueezeNet**	**ResNet 18**	**MobileNet V2**	**ResNet 50**
Depth	8	18	18	53	50
Size	227MB	4.60MB	44.0MB	13.0MB	96.0MB
Input Images (pixel)	227 x 227	227 x 227	224 x 224	224 x 224	224 x 224
Parameter	61.0M	1.24M	11.7M	3.50M	25.6M

FLOPs	7.25×10^{8}	8.33×10^{8}	1.8×10^{9}	3.0×10^{8}	3.8×10^{9}
Strength	Overlapping pooling	Fire Module	Residual Block	Depthwise Separable Convolution	Residual Block
	ReLUs	1x1 & 3x3 filters size	Skip Connection	Inverted Residual	Skip Connection
	Dropout				
Top 1 accuracy on ImageNet	63.30	57.0	69.57	72.834	75.99
Top 5 accuracy on ImageNet	84.60	80.0	89.24	91.06	92.98

3.8 Evaluation of performance metric

The confusion matrix was plotted after the model training to evaluate the predicted and actual values from the trained model [39]. The number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) will be obtained from the confusion matrix for computing five of the performance metrics to evaluate the overall performance of the trained model [40]. The performance metrics are:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.1}$$

$$Precision = \frac{TP}{TP+FP} \tag{3.2}$$

$$Recall \ = \frac{TP}{TP+FN} \tag{3.3}$$

$$Specificity = \frac{TN}{TN+FP} \tag{3.4}$$

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision+Recall} \tag{3.5}$$

Besides, the Area Under the Receiver Operating Characteristics (AUROC) curve was plotted to evaluate the trained model's classification. The AUROC will provide