INTERNATIONAL JOURNAL OF
BUSINESS INNOVATION AND SMART TECHNOLOGIES
IJOBIST

International Journal of
Business Innovation and Smart Technologies
Volume xx : xx-xx, Month Year
E-ISSN: xxxx-xxxx | https://doi.org/xx/ijobist.xx

Literature Review of Accuracy Analysis in Digital Image Classification Using Deep Learning

Ni Luh Widi Rahayu[1],
Institut Teknologi dan Bisnis Muhammadiyah Bali[1]
Corresponding author: luhwidirahayu@gmail.com; Tel.: 08983161086

ABSTRACT

The development of technology, particularly the advancement of Artificial Intelligence (AI), including machine learning, presents significant challenges in the field of research, especially in digital image classification using machine learning techniques. Digital image classification is a process based on the detection and identification of objects within images across various object categories. Several machine learning methods commonly used for digital image classification include K-Nearest Neighbor (KNN), Gray Level Co-occurrence Matrix (GLCM), Support Vector Machine (SVM), and Convolutional Neural Network (CNN). This study focuses on analyzing the accuracy level of one of the most widely used digital image classification methods, namely the Convolutional Neural Network (CNN). The analysis of CNN accuracy in digital image classification across various application domains is conducted based on a literature review of 40 journal articles related to digital image classification. The accuracy achieved in digital image classification is strongly influenced by the classification method, dataset characteristics, supporting computational tools, and architectural modifications applied to specific methods. In addition, the quality of image data used for both training and testing significantly affects the resulting accuracy. Therefore, performing data preprocessing or image enhancement on training and testing datasets can improve classification accuracy during evaluation. The results indicate that the average accuracy achieved using CNN methods outperforms Support Vector Machine (SVM), suggesting that the application of Convolutional Neural Networks in digital image classification across various objects demonstrates a relatively high level of accuracy.

ARTICLE INFO

Article history:
Received: Month 1st, Year
Revised: Month 1st, Year
Accepted: Month 1st, Year
Published: Month 1st, Year

Keywords: Digital Image Classification; Convolutional Neural Network(CNN); Food Image Recognition; Deep Learning Architecture; Accuracy Evaluation

1. INTRODUCTION

The utilization of technology, particularly advancements in Artificial Intelligence (AI) such as machine learning (Nurhikmat. T & Purwaningsih. T, 2018), provides a highly effective medium for enhancing knowledge in the field of education, especially in digital image classification. With the development of machine learning, which learns from sets of relatively simple data, a deeper subfield has emerged, namely deep learning. Deep learning is an algorithm inspired by and analogous to the structure and function of the human brain, commonly referred to as artificial neural networks (neurons) (Brownlee, 2016). By leveraging technological media, particularly Artificial Intelligence (AI), it is possible to develop systems capable of classifying digital images across various domains into meaningful information related to the desired output. The objective of image classification is to replicate human capabilities in understanding digital image information, enabling computers to classify objects in images in a manner similar to human perception.

Several methods have been employed in digital image classification research, including K-Nearest Neighbor (KNN), Gray Level Co-occurrence Matrix (GLCM), Support Vector Machine (SVM), and Convolutional Neural Network (CNN). This study focuses on analyzing the accuracy level of one of the most widely used digital image classification methods, namely the Convolutional Neural Network (CNN). The motivation for analyzing CNN accuracy stems from findings in previous studies indicating that CNN achieves superior accuracy compared to other digital image classification methods. This superiority was notably demonstrated by Alex Krizhevsky's victory in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 using a CNN-based approach. Furthermore, CNN has been identified as a highly suitable algorithm for image and audio specialization (LeCun Y, 1990), as achieving high classification and prediction accuracy requires hierarchical, large-scale processing and strong connectivity.

Previous studies on machine learning-based classification across various domains have predominantly employed CNN methods with high classification accuracy, including applications in transportation (Arfian, 2018), arts (Salsabila, 2018), botany (Kusumaningrum, 2018), (Rachardi, 2020), and food recognition (Darma Udayana et al., 2020). These findings demonstrate that the Convolutional Neural Network algorithm is capable of effectively and accurately solving classification problems with a high level of accuracy. This study conducts an analysis of the accuracy level of Convolutional Neural Network (CNN) implementation in digital image classification across multiple domains. The research is based on a literature review of 40 journal articles related to digital image classification. The analysis covers several aspects, including: (1) classification implementations on digital images across various object categories; (2) commonly used classification algorithms and their corresponding architectures; (3) average accuracy levels achieved by each algorithm across different objects; (4) comparative accuracy analysis between CNN and SVM methods; (5) analysis based on publication year; (6) analysis based on research location (domestic and international); (7) analysis based on data or image sources, including existing datasets and newly generated datasets; (8) analysis based on computational time (short, long, or unspecified); (9) analysis based on computational specifications, including the hardware used; and finally, (10) conclusions drawn from the overall analysis of CNN accuracy in digital image classification.

2. METHODS

Of the 40 journal articles reviewed, 27 articles discuss digital image classification and recognition across various object categories, 5 articles focus on the use of Convolutional Neural Network (CNN) methods and their architectures, 2 articles examine the accuracy of CNN methods, and 6 articles address comparisons of Convolutional Neural Network (CNN) methods. The data sources were obtained from academic search engines, including Google Scholar, ScienceDirect, and IEEE Xplore. In conducting this study, a schematic diagram was employed as a reference to represent the steps undertaken in analyzing the accuracy level of the Convolutional Neural Network (CNN) method. The schematic diagram presents four main steps that must be completed prior to data processing and analysis. These steps were designed to identify the key aspects to be analyzed in this study. The detailed schematic diagram of the research process is shown in Figure 4.

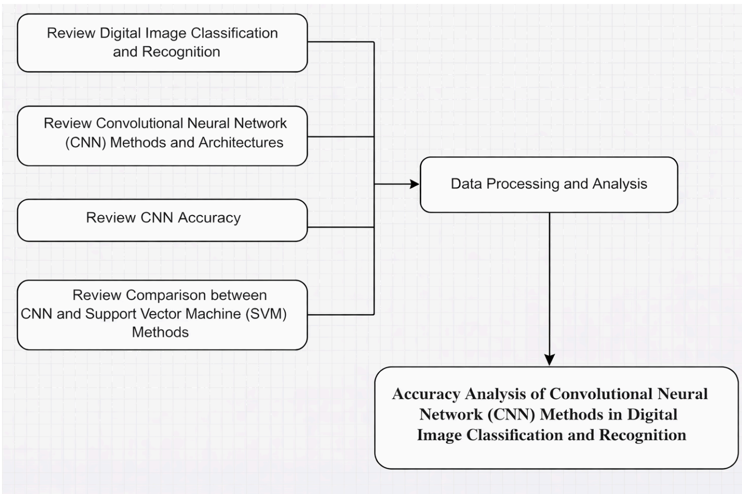

Figure 4. Research Schematic Diagram

Figure 4 illustrates that this study involves a literature review of digital image classification and recognition, which can be performed using various Deep Learning methods, one of which is the Convolutional Neural Network (CNN). The review of CNN methods is conducted to identify the parameters and architectures employed in digital image classification and recognition tasks. To evaluate the effectiveness of the applied methods, an analysis of the accuracy levels of the methods and architectures is required.

Furthermore, a comparison between Convolutional Neural Network (CNN) and other methods is necessary to assess classification accuracy and to analyze the strengths and advantages of the applied approach. Therefore, this study includes a comparative review between the Convolutional Neural Network (CNN) method and another widely used digital image processing method, namely the Support Vector Machine (SVM).

3. RESULTS AND DISCUSSION

Based on a review of 40 journal articles related to the implementation of the Convolutional Neural Network (CNN) method, the reviewed studies were obtained from several sources, including Google Scholar, IEEE Xplore, Gaussian Journal, ALGOR Journal, and other academic journals. Of the 40 journals reviewed, 31 studies were conducted internationally, while 9 studies were conducted in Indonesia. This research aims to analyze the accuracy level of CNN implementation in digital image classification and recognition, beginning with classification and recognition tasks applied to digital images across various object categories.

3.1 Digital Image Recognition and Classification Implementation

An image is a representation, resemblance, or imitation of an object. Mathematically, an image can be expressed as a continuous function of light intensity over a two-dimensional plane. A visible image is the result of light reflected from an object. Images are classified into two types: continuous images, obtained from optical systems that receive analog signals (such as the human eye and analog cameras), and discrete (digital) images, which are produced through the digitization of continuous images. The digitization process consists of sampling and quantization processes (Peryanto et al., 2020).

Digital image classification has been extensively studied by researchers, with implementations applied to various objects such as food and beverage images, plant images, mask usage images, spice and seasoning images, meat type images, land surface images, fruit images, and facial expression images. One study on food image classification was conducted by Agus Eka et al., focusing on food image prediction to estimate calorie intake for patients with Diabetes Mellitus. The study used 10,000 food images as training data and tested system accuracy on 20 food classes, each with 50 new images (1,000 test images). The results showed that the developed system achieved a maximum accuracy of 98% and a minimum accuracy of 66%, indicating variability in classification results (Nugraha & Cipta, 2020).

Subsequent research by Subhi et al. examined the improvement of Malaysian food classification. The study used 5,800 Malaysian food images and 3,300 local food images distributed across 11 categories, with 300 images per category. A CNN method was applied using image dimensions of 220 × 220 × 3, incorporating a 24-layer architecture consisting of 21 convolutional layers and 3 fully connected layers. The results demonstrated that increasing the number of layers and utilizing larger datasets significantly improved classification accuracy (Subhi & Ali, 2019).

Further research by Rajesh et al. focused on automatic fruit image classification using 200 fruit images, consisting of apples, mangoes, oranges, and grapes. The methods applied included median filtering, K-Means clustering, and classification using CNN, Backpropagation Neural Network (BPNN), and SVM. The highest accuracy of 90% was achieved using CNN, with individual accuracies of 88% for apples, 91% for mangoes, 91% for oranges, and 90% for grapes (Yamparala & Challa, 2020).

Additional studies include Yadav et al., who applied SqueezeNet and VGG-16 for food image classification using the Food-101 dataset, achieving a training accuracy of 93.47% and validation accuracy of 77.20%. Simon et al. implemented the NutriNet model on 225,953 food and beverage images, achieving an accuracy of 86.72% (Mezgec & Seljak, 2019). Abdul et al. utilized CNN models on Food-101, UEC-256, and Yummly-28K datasets, extracting nutritional composition features using convolutional layers (Salam & Abdul, 2019). Dhiksa applied CNN with three convolution layers, achieving a peak validation accuracy of 78.9% (Solanki, 2020).

Further studies by Kanjanapan et al. combined YOLOv3 and MobileNets for Thai fast-food classification, achieving an average accuracy of 92.83%, with a maximum of 96% (Sukvichai et al., 2019). Kagaya et al. achieved an accuracy of 93.8% for calorie recognition using CNN. Deepak et al. reported an average accuracy of 85% using optimized CNN models with TensorFlow 2.0 (Sukvichai et al., 2019). Rajayogi et al. applied transfer learning with InceptionV3, VGG16, VGG19, and ResNet, achieving the highest accuracy of 87.9% using InceptionV3 (Sukvichai et al., 2019).

Additional implementations include Narit et al., who achieved 100% accuracy on certain Thai fast-food categories using CNN and InceptionV3 (Hnoohom & Yuenyong, 2018). Ahmad et al. classified mustard leaf varieties using CNN and achieved 84% accuracy (Kurniadi & Fal Sadikin, 2020). Arham et al. classified mask usage using CNN with accuracy exceeding 90% across multiple scenarios (Rahim et al., 2020). Nugroho et al. classified human facial expressions using CNN, achieving 0.9032 training accuracy and 0.66 validation accuracy (Nugroho et al., 2020). Felix et al. identified plant species from leaf images using CNN, achieving 76% accuracy (Felix et al., 2020). Duan-Yu et al. implemented a two-stage CNN classifier for food localization and recognition, achieving 92.63% precision in 0.25 seconds (Chen & Wang, 2018). Chin-Li et al. classified food categories and predicted nutritional quantities using CNN, achieving an average accuracy of 70% (Chin et al., 2017) .

The implementation of digital image classification across various object categories is illustrated in Figure 1.

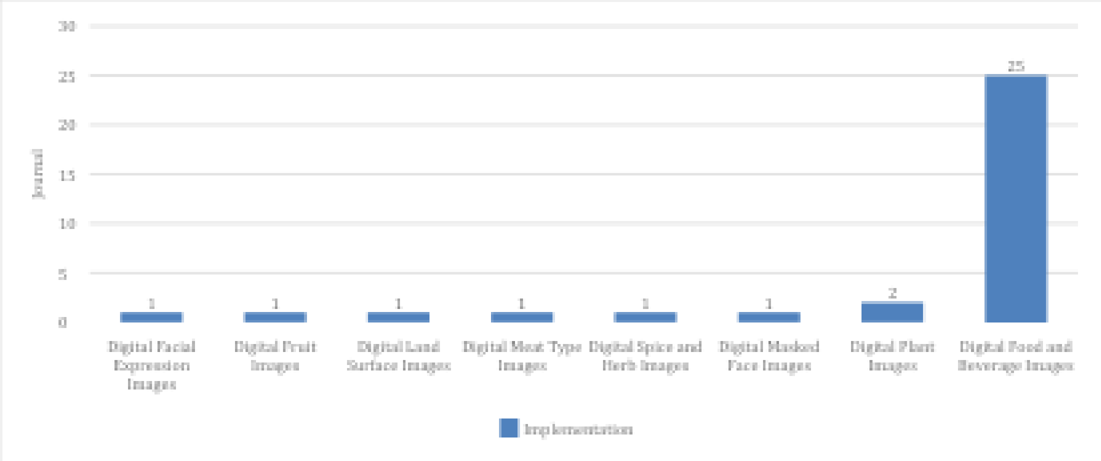

Figure 1. Graph of Digital Image Classification Implementation Across Various Objects

The graph above presents data based on digital image recognition and classification across various object categories. The number of implementations for facial expression images is 1, fruit images is 1, land surface images is 1, meat type images is 1, spice and seasoning images is 1, mask user images is 1, plant images is 2, and the highest number of implementations is found in food and beverage images, totaling 25 studies. From these data, it can be observed that the implementation of digital image classification for food and beverage objects dominates the topic addressed in this literature review. Accordingly, this review focuses on analyzing the accuracy levels achieved in food and beverage image classification using CNN methods. Studies whose implementations are not related to food and beverages are used as comparative references for evaluating accuracy levels.

3.2 Methods and Algorithms for Digital Image Classification

In digital image classification, various methods can be applied, including Convolutional Neural Networks (CNN) and their architectures such as AlexNet, ResNet, GoogLeNet, VGG-16, and others. In addition, the Support Vector Machine (SVM) method has also been employed in several studies.

A study conducted by Isna et al. focused on spice and seasoning image classification using the CNN algorithm. The dataset consisted of three spice categories, namely ginger, galangal, and ginseng, with a total of 300 images, 100 images per category. The dataset was divided into training and testing sets with an 80%:20% ratio. The CNN model with hyperparameter tuning achieved a training accuracy of 98.75% and a training loss of 0.0769. The testing accuracy reached 85%, while classification accuracy on new data was 88.89%. These results indicate that a two-layer CNN model demonstrated good performance, supported by high training accuracy and acceptable testing results (Wulandari et al., 2020).

Subsequent research by Yaman et al. examined a comparison between CaffeNet and AlexNet architectures in CNN-based food image classification. The dataset comprised 16,950 training images across 11 categories. The proposed deep CNN architecture consisted of five convolutional and pooling layers, two normalization layers, and four fully connected layers. The comparison results showed classification accuracies of 80.51% for CaffeNet and 82.07% for AlexNet, while the proposed structure achieved 70.12% accuracy (Yaman et al., 2017).

Another study by Neneng et al. investigated meat type image classification based on texture using GLCM feature extraction and SVM classification. The dataset included images of goat, buffalo, horse, and beef, captured at distances of 20 cm, 30 cm, and 40 cm. Each meat type consisted of 150 images, resulting in a total of 1,800 images, with 120 images used for testing. The best classification accuracy of 87.5% was achieved at a pixel distance of d = 2 and a GLCM direction of 135 degrees, indicating that images captured at a 20 cm distance are recommended for meat image classification (Neneng et al., 2016).

Research by Gianluigi et al. explored food recognition across multiple countries using CNN architectures combined with SVM as an alternative end-to-end classifier. A newly constructed dataset comprising 20 food categories and 11,943 images was used. Seven CNN architectures, including AlexNet, Caffe-Reference, GoogLeNet, VGGNet-16, VGGNet-19, InceptionV3, and ResNet-50, were evaluated. The overall average accuracy achieved was 92.03%, while the adapted architecture reached 92.13%. The best performance was obtained by Inception-v3 with 95.39%, followed by GoogLeNet at 94.49%, ResNet-50 at 92.90%, and MobileNet-v2 at 92.88% (Ciocca et al., 2020).

Further research by Shuqiang et al. proposed a multi-scale CNN activation fusion approach to generate more discriminative aggregate features for food recognition. The datasets used were ETH-101, containing 101,000 images across 101 categories, and VireoFood-172, comprising 110,241 food images across 172 categories. The study combined three CNN-based methods, namely AlexNet, CaffeNet, and MSMVFA (DenseNet-161). The results demonstrated a Top-1 accuracy of 90.61% and a Top-5 accuracy of 98.1% on the VireoFood-172 dataset (Jiang et al., 2020).

Another study by Lili Pan et al. developed the DeepFood framework for automatic food ingredient classification using deep learning. The dataset consisted of 41 food ingredient classes, with 100 images per class. The proposed framework integrated ResNet deep features, Information Gain feature selection, and SMO classification. The DeepFood framework achieved an average multi-class classification accuracy of 87.78%, with ResNet and Information Gain producing the highest performance. The average performance of AlexNet was comparable to that of CaffeNet (Pan et al., 2017).

Additional studies employing various digital image classification methods and algorithms are presented in Figure 2.

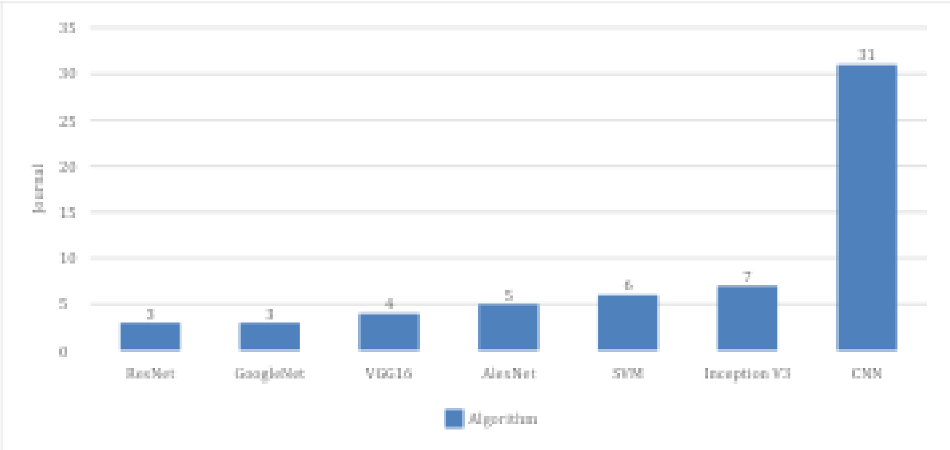

Figure 2. Graph of the Number of Journals Based on the Use of Digital Image Classification Methods

The graph above presents data on the number of journal articles based on the methods and architectures applied in the reviewed studies. The methods used include CNN ResNet architecture with a total of 3 studies, CNN GoogLeNet architecture with 3 studies, CNN AlexNet architecture with 4 studies, the SVM method with 6 studies, the Inception V3 method with 7 studies, and the most widely used method is CNN, which appears in 31 studies.

The use of CNN methods in 31 studies is primarily based on the high accuracy levels achieved by CNN. In addition, one related study states that CNN is a method with superior accuracy compared to other digital image classification methods. This claim is supported by the victory of Alex Krizhevsky in the ImageNet Large Scale Visual Recognition Challenge 2012 using the CNN method.

3.3 Average Accuracy Level of Each Method

The application of various image classification methods across different object categories is influenced by multiple factors, one of which is the dataset used. The accuracy results obtained from food and beverage image classification using CNN differ from those obtained from image classification of mask users, even when the same method is applied.

A study conducted by Halimatu et al. discussed the development of precision agriculture for plant image recognition and classification. The dataset used in this study consisted of images extracted from corn plantations involving several corn varieties. The method applied was Convolutional Neural Network (CNN), which was considered the most effective approach for optimizing corn variety classification. The experimental results of the developed CNN model achieved an average accuracy of 99.58% (Abdullahi et al., 2017).

Another study by Punnarumol et al. examined Thai food image recognition in smartphone applications and analyzed its accuracy using a relatively small dataset. The dataset consisted of 1,987 Thai food images across 40 food categories, with 50–60 images per category and a resolution of 300 × 300 pixels. Additionally, a distorted dataset was used, comprising 5% of the total data, involving random brightness, scaling, and cropping. The method applied was CNN-based food image classification using a transfer learning model with Inception V3, which was then integrated into a smartphone application. The results showed that the training data achieved accuracy levels ranging from 68.1% to 73.4%, while the distorted image dataset achieved accuracy levels ranging from approximately 73.4% to 75.2%. Calorie information was displayed according to the provided dataset (Temdee & Uttama, 2017).

Subsequent research by Chakkrit et al. focused on improving image accuracy for Thai food recognition. The dataset used for testing was THFOOD-50, consisting of 15,770 images from 50 well-known Thai food categories. The methods applied included GoogleNet, BN-Inception, NU-InNet 1.0, and NU-InNet 1.1 models. The results indicated that for NU-InNet 1.0, top-1 accuracy increased from 70.14% to 79.68%, 78.62%, and 75.44% for depths of 4, 8, and 12, respectively. The highest top-1 accuracy was achieved by NU-InNet 1.0 with a depth of 4, resulting in an improvement of 9.54%. Similarly, for NU-InNet 1.1, top-1 accuracy increased from 74.24% to 80.34%, 79.75%, and 75.69% for depths of 4, 8, and 12, respectively. The highest top-1 accuracy was obtained by NU-InNet 1.1 with a depth of 4, achieving an improvement of 6.10% (Termritthikun & Kanprachar, 2017).

Additional studies related to the average accuracy levels achieved by each method and their implementations across various object categories are presented in Figure 3.

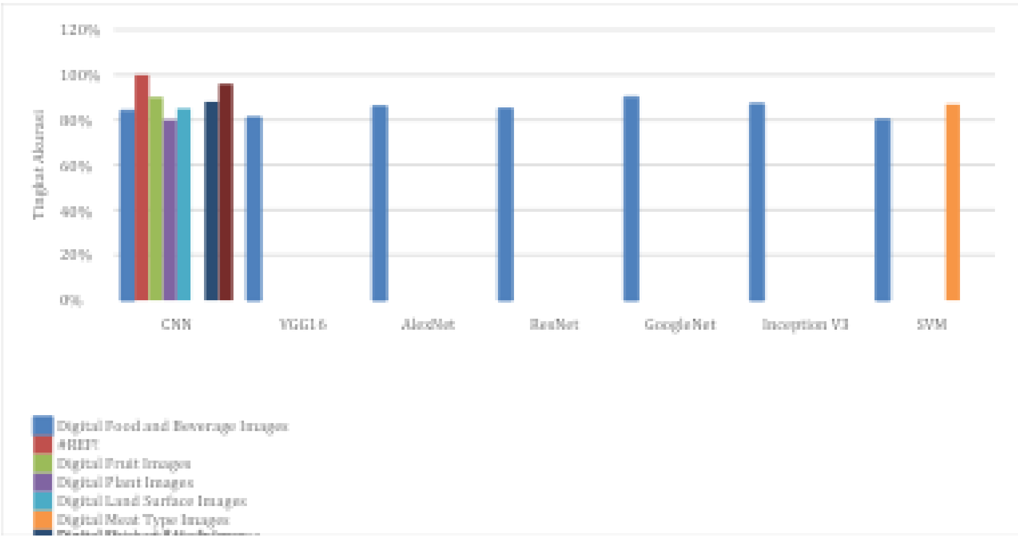

Figure 3. Graph of Average Accuracy Levels

The graph above presents the average accuracy levels obtained from the applied methods and architectures. Image classification using the CNN method achieved the highest average accuracy in the implementation on mask user digital images, with an average accuracy of 96%. Meanwhile, the implementation on food and beverage images using CNN achieved an average accuracy of 84%, VGG16 achieved 81%, AlexNet achieved 86%, ResNet achieved 85%, GoogLeNet achieved 90%, Inception V3 achieved 87%, and SVM achieved 80%.

The highest average accuracy obtained using the CNN method in mask user image classification is attributed to the use of highly complex and detailed training data, which resulted in high accuracy levels. In this study, the training dataset consisted of 4,326 images obtained from Kaggle.com. During testing, data preprocessing techniques such as image wrapping and cropping were applied to generate labeled mask and non-mask datasets, thereby facilitating the classification process.

In addition, the high average accuracy achieved using CNN methods is influenced by their widespread adoption and the fact that most studies focus on food and beverage image classification. These studies commonly utilize existing training datasets and apply architectural modifications and enhancements to CNN models in the proposed testing scenarios, resulting in higher accuracy levels compared to other methods applied to non-food and non-beverage digital image objects.

3.4 Comparison of Accuracy Levels between Convolutional Neural Network (CNN) and Support Vector Machine (SVM)

The accuracy levels obtained from various digital image classification methods need to be compared in order to determine the most effective method. A study conducted by Rajesh et al. discussed automatic fruit image classification using a dataset of 200 fruit images, consisting of 50 apple images, 50 mango images, 50 orange images, and 50 grape images. The methods applied included nonlinear processing techniques such as median filtering, segmentation using K-Means clustering, and classification using CNN, BPNN, and SVM. The results indicated that the highest accuracy of 90% was achieved using CNN. The classification accuracy for apples was 88%, mangoes 91%, oranges 91%, and grapes 90%, resulting in an average accuracy of 90%.

Another study by David J. et al. focused on real-time calorie monitoring for dietary management. The dataset used was Food-101, consisting of 101,000 food images across 101 categories. The applied method was a neural network configuration using the Google Inception V3 model, with calorie estimation measured per 100 g food portion. The results showed that using the Food-101 dataset, the method achieved a high accuracy of 86.97%, which was the highest accuracy reported by the authors compared to SVM, which achieved only 50.76% accuracy on the same dataset (Attokaren et al., 2017).

Further research by Hongsheng He et al. examined automatic food classification using DietCam, a system that detects food ingredients and classifies food types. The dataset consisted of 15,262 food images used for training and testing, divided into 55 food categories with an average of 277 images per category. The method applied was Multi-Kernel SVM. The results showed that DietCam achieved approximately 90% recognition accuracy for general food categories and 85% accuracy for DC-5 food categories. In contrast, recognition accuracy using SIFT, which is commonly applied in fast food classification, reached only 60% for DC-1 food categories and dropped to 30% for DC-4 to DC-6 categories due to the complexity of food composition prediction (He et al., 2016).

Subsequent research by Yoshiyuki et al. discussed real-time food recognition using smartphones. The dataset consisted of 50 food categories for training and 6,781 food images for testing, with 100 images per category. The applied method used GrabCut for food segmentation, SURF-based features for color histogram extraction, and linear SVM for classification. The results showed a classification accuracy of 81.55% for the top five categories in the testing data. The cumulative classification rates with estimation shifts of less than ±20% and ±40% errors were 31.81% and 50.34% for a 15% shift, and 34.54% and 54.16% for a 25% shift (Yanai & Kawano, 2014).

Another study by Taichi et al. examined an automatic food image recognition system for recording dietary habits using Multiple Kernel Learning. The dataset consisted of 50 food categories with 100 images per category. The applied method integrated MKL-based features with SVM, extracting various image features such as bag-of-features, color histograms, and Gabor texture features from training images. The results showed that the best classification rate using a single feature was 38.18% based on color histograms, while MKL-based feature fusion achieved a classification rate of 61.34% for 50 food image categories. The prototype system achieved a classification accuracy of 37.55% for 166 food images uploaded by users via mobile devices (Joutou & Yanai, 2019).

Comparative results of average accuracy levels obtained using CNN methods across various object categories and SVM methods across different object categories are presented in Figure 4.

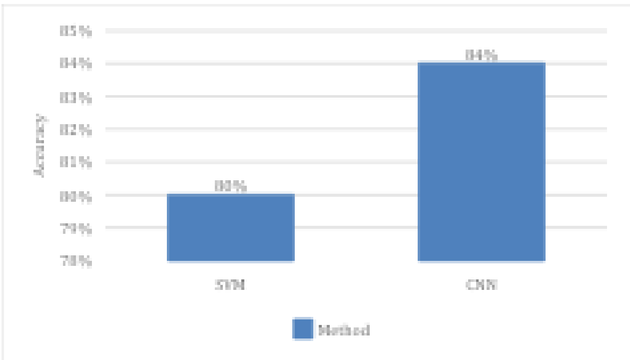

Figure 4. Comparison Graph of Accuracy Levels between Convolutional Neural Network (CNN) and Support Vector Machine (SVM)

The graph above presents a comparison of the average accuracy values obtained from digital image classification using CNN and SVM methods. The results show that the average accuracy achieved using the

CNN method is 84%, calculated from 34 journal articles that applied CNN methods. Meanwhile, the average accuracy obtained using the SVM method is 80%, calculated from 6 journal articles that employed SVM methods in their studies.

The higher average accuracy achieved by the CNN method is attributed to its advantages, including the availability of various architectures and the ability to modify these architectures to improve classification performance. In addition, one related study highlights the superiority of CNN methods, as evidenced by Alex Krizhevsky's victory in the ImageNet Large Scale Visual Recognition Challenge 2012 using a CNN-based approach.

3.5 Development of Image Classification

Research on image classification conducted by various researchers has continued to increase over the years. This trend is driven by the widespread adoption of deep learning across multiple domains, with digital image classification representing a significant challenge and opportunity for researchers and practitioners in the field of deep learning. The distribution of journal articles based on their publication year is illustrated in Figure 5.

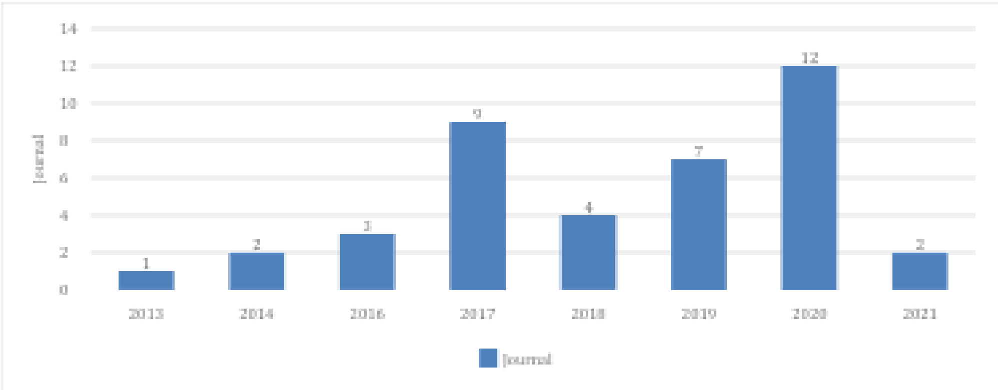

Figure 5. Graph of Image Classification Development

The graph above presents data on the number of journal articles based on their publication year. Within the period from 2013 to 2021, one study was conducted in 2013, two studies in 2014, three studies in 2016, nine studies in 2017, four studies in 2018, seven studies in 2019, twelve studies in 2020, and two studies in 2021.

3.6 Dataset Usage in Digital Image Classification

Several studies have reported that the datasets used for evaluating the proposed methods vary, ranging from newly constructed datasets used for testing to existing datasets utilized for training purposes. A study conducted by Md Tohidul et al. discussed efficient food image classification using the Food-11 dataset. The dataset consisted of 16,643 food images grouped into 11 food categories and divided into three subsets: 9,866 training images, 3,430 validation images, and 3,347 evaluation images. The method applied in this study was a CNN model based on the Inception V3 architecture (Islam et al., 2018).

Subsequent research by Anusrhee et al. examined a comparison between NDVI features and CNN for aerial land cover image classification. The dataset used was SAT-6, consisting of a total of 405,000 image patches with dimensions of 28 × 28 pixels, covering six land cover classes: barren land, trees, grassland, roads, buildings, and water bodies. The study compared NDVI features and CNN with a multi-layer perceptron architecture incorporating batch normalization (Ramanath et al., 2019).

Another study by Suartika et al. investigated image classification using CNN on the Caltech-101 database. The dataset consisted of 390 images from Caltech-101, including 150 poultry images and 80 images each for crocodile, cougar, and face categories. The CNN method employed a two-stage process using feedforward and backpropagation. The results showed a classification success rate of 20% for poultry categories and 50% for cougar, crocodile, and face categories. Overall, the classification accuracy ranged between 20% and 50% (Suartika E. P, I Wayan, Wijaya Arya Yudhi, 2016).

Further research by Gianluigi et al. addressed food recognition using a newly constructed dataset aimed at monitoring dietary intake. The dataset consisted of 3,616 food images across 73 food categories. The methods applied included benchmarking algorithms and CNN-based classification. The highest food recognition accuracy was achieved using a patch-based approach on the UNIMIB2015 dataset, measured using SA and MAA metrics, yielding accuracies of 99.05% and 99.03%, respectively (Ciocca et al., 2017).

Another study by Paritosh et al. focused on food recognition using a pipeline-based CNN architecture. The datasets used included ETH Food-101 and 50 categories of Indian food images, with each category containing 100 images. The methods applied were the EnsembleNet architecture and fine-tuned CNN AlexNet models. The results showed that for Indian food images, the accuracy achieved was 73.50% for Top-1, 94.40% for Top-5, and 97.60% for Top-10. For the ETH Food-101 dataset, the results were 72.12% for Top-1, 91.61% for Top-5, and 95.95% for Top-10 (Pandey et al., 2017).

Subsequent research by Keiji et al. discussed food image recognition using Deep Convolutional Neural Networks with pre-training and fine-tuning architectures. The datasets used included Japanese food image datasets UEC-FOOD100 and UEC-FOOD256, representing 100 and 256 food classes, respectively. In addition, a supplementary dataset of 1,730,441 food-related tweet images containing one of the target food names was incorporated. The applied method utilized a Deep CNN with pre-training and fine-tuning. The experimental results achieved the best classification accuracies of 78.77% and 67.57% for the UEC-FOOD100 and UEC-FOOD256 datasets, respectively (Yanai & Kawano, 2014).

Another study by K. Kogias et al. examined a food classification system designed for individuals with Diabetes Mellitus. The datasets used for system evaluation were Food-101 and UEC-Food256, comprising a total of 3,248 food images. The method employed a two-stage CNN approach combining AlexNet and VGGNet architectures for classification. The developed algorithm for first-level classification on the NTUA-Food 2017 dataset improved classification accuracy on the benchmark Food Image Dataset (FID) to 97.08%. The refined AlexNet-based classification outperformed other approaches, achieving an accuracy of 97.08%. Overall system accuracy reached 84.18%, while DPM calculations revealed an average CE error of 1.97 g per food serving (Kogias et al., 2018).

The distribution of dataset usage across previous studies, including Caltech-101, NutriNet, Food-101, UEC-Food256, and THFOOD-50 as training and testing data, is illustrated in Figure 5.

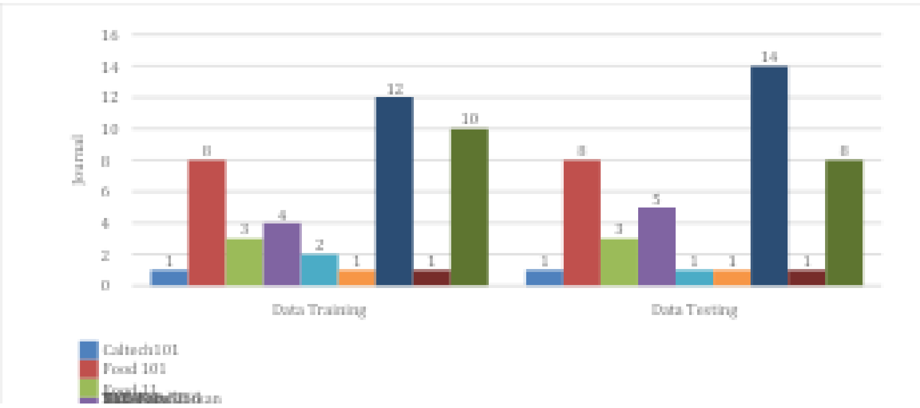

Figure 5. Graph of Dataset Usage in Digital Image Classification

The graph above presents data on the number of journal articles based on the data or images used in the studies, where the datasets are divided into two categories: training data and testing data. Based on the graph, the use of newly constructed datasets as testing data was reported in 14 studies, while the use of newly constructed datasets as training data was reported in 12 studies. Previously established datasets such as Caltech101, NutriNet, Food-101, Food-11, SAT-6, UEC-Food256, and THFOOD-50 were used as both training and testing data.

The Caltech101 dataset was used as training and testing data in one journal article. The Food-101 dataset was used as training and testing data in eight journal articles. The Food-11 dataset was used as training and testing data in three journal articles. The UEC-Food256 dataset was used as training data in four journal articles and as testing data in five journal articles. The THFOOD-50 dataset was used as training data in two journal articles and as testing data in thirteen journal articles. The NutriNet dataset was used as training and testing data in one journal article, while the SAT-6 dataset was used as training and testing data in one journal article.

3.7 Computational Time in Digital Image Classification

In digital image classification, the classification process using training and testing data requires varying computational time depending on the dataset size and the classification method applied. The grouping of journal articles based on computational time reported in previous studies is presented in Figure 6.

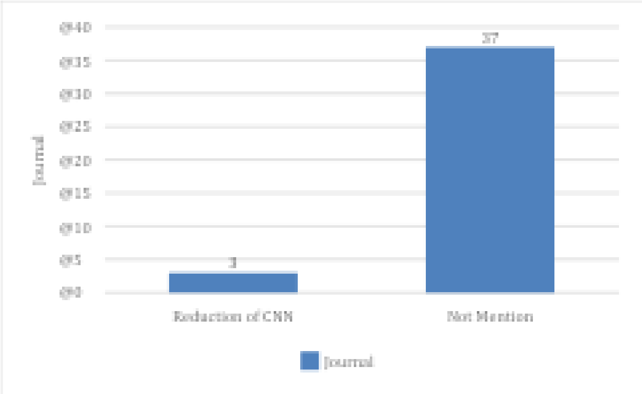

Figure 6. Graph of Computational Time in Digital Image Classification

The graph above presents data on the number of journal articles based on the computational time reported during the research process. Computational time refers to the duration required by a CNN method to perform recognition or classification of digital images during both training and testing phases. It is explained that CNN models modified through reduction techniques tend to achieve shorter computational times, whereas CNN models applied without reduction require longer computational times. The graph indicates that three journal articles reported the use of modified CNN models, while the remaining 37 articles did not specify computational time.

The 37 journal articles that did not report computational time are most likely associated with the use of CNN methods that were not modified or reduced, resulting in longer computational durations that were not explicitly stated. Additionally, some studies did not report computational time because their primary focus was not on computational efficiency, but rather on classification correctness and accuracy levels achieved in the conducted experiments.

3.8 Specifications of Computational Tools Used

In previous studies, some researchers reported the specifications of the computational tools used for digital image classification, while others did not disclose such information. The grouping of journal articles based on the hardware used is presented in Figure 7.

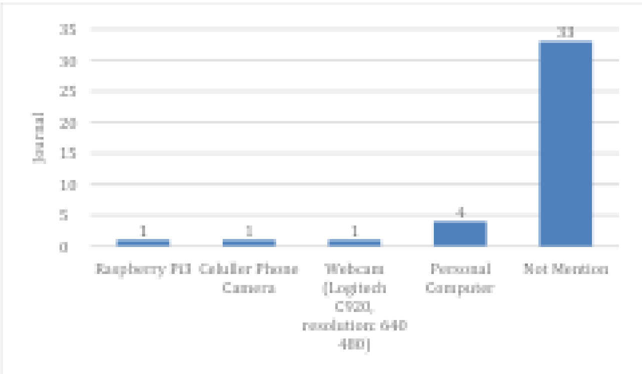

Figure 7. Graph of Computational Hardware Specifications Used

The graph above presents data on the number of journal articles based on the specifications of computational tools used in the studies. One journal reported that the research was conducted using supporting hardware in the form of a Raspberry Pi 3. Two other journals mentioned the use of supporting hardware such as smartphone cameras and webcams. Four journals stated that the research was conducted using personal computers for testing purposes, while the remaining 33 journals did not specify the computational tools used in their studies.

The 33 journals that did not report computational hardware specifications are most likely studies that did not focus on discussing hardware details, but instead emphasized classification correctness and the accuracy levels obtained from the conducted experiments.

4. CONCLUSION

Based on the analysis of the accuracy levels of Convolutional Neural Network (CNN) methods through a review of 40 journal articles, it can be concluded that the accuracy achieved in digital image classification is strongly influenced by the classification method used, the dataset, supporting computational tools, and architectural modifications applied to certain methods. In addition, the quality of image data used for both training and testing significantly affects the resulting accuracy. Therefore, applying preprocessing or image enhancement techniques to training and testing datasets can improve accuracy performance during evaluation.

A comparison between two widely used classification methods, Convolutional Neural Network (CNN) and Support Vector Machine (SVM), shows that CNN achieved a higher average accuracy of 84%, compared to 80% for SVM. This indicates that the application of Convolutional Neural Network (CNN) methods in digital image classification across various object domains provides relatively high accuracy. Furthermore, the ability to reduce and combine CNN architectures can further enhance the accuracy of digital image classification results.

REFERENCES

Abdullahi, H. S., Sheriff, R. E., & Mahieddine, F. (2017). *Convolution neural network in precision agriculture for plant image recognition and classification*. *Intech*, 1–3. https://doi.org/10.1109/intech.2017.8102436

Arfian. (2018). Implementasi Convolutional Neural Network Terhadap Transportasi Tradisional Menggunakan Keras. *Yogyakarta: Universitas Islam Indonesia.*

Attokaren, D. J., Fernandes, I. G., Sriram, A., Murthy, Y. V. S., & Koolagudi, S. G. (2017). Food classification from images using convolutional neural networks. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, *2017-Decem*, 2801–2806. https://doi.org/10.1109/TENCON.2017.8228338

Brownlee, J. (2016). Master Machine Learning Algorithms. Discover How They Work and Implement Them From Scratch. *In Machine Learning Mastery With Python*.

Chen, D. Y., & Wang, H. S. (2018). Handheld food localization and food recognition using convolutional neural network. *ACM International Conference Proceeding Series*, 61–64. https://doi.org/10.1145/3299852.3299862

Chin, C. L., Huang, C. C., Lin, B. J., Wu, G. R., Weng, T. C., & Chen, H. F. (2017). Smartphone-based food category and nutrition quantity recognition in food image with deep learning algorithm. *2016 International Conference on Fuzzy Theory and Its Applications, IFuzzy 2016*, 54387913. https://doi.org/10.1109/iFUZZY.2016.8004962

Ciocca, G., Micali, G., & Napoletano, P. (2020). State Recognition of Food Images Using Deep Features. *IEEE Access*, *8*, 32003–32017. https://doi.org/10.1109/ACCESS.2020.2973704

Ciocca, G., Napoletano, P., & Schettini, R. (2017). Food Recognition: A New Dataset, Experiments, and Results. *IEEE Journal of Biomedical and Health Informatics*, *21*(3), 588–598. https://doi.org/10.1109/JBHI.2016.2636441

Darma Udayana, I. P. A. E., Sudarma, M., & Surya Cipta Nugraha, P. G. (2020). Implementation of Convolutional Neural Networks to Recognize Images of Common Indonesian Food. *IOP Conference Series: Materials Science and Engineering*, *846*(1). https://doi.org/10.1088/1757-899X/846/1/012023

Felix, Wijaya, J., Sutra, S. P., Kosasih, P. W., & Sirait, P. (2020). Implementasi Convolutional Neural Network Untuk Identifikasi Jenis Tanaman Melalui Daun. *Jurnal SIFO Mikroskil*, *21*(1), 1–10.

He, H., Kong, F., & Tan, J. (2016). DietCam: Multiview food recognition using a multikernel SVM. *IEEE Journal of Biomedical and Health Informatics*, *20*(3), 848–855. https://doi.org/10.1109/JBHI.2015.2419251

Hnoohom, N., & Yuenyong, S. (2018). Thai Fast Food Image Classification Using Deep Learning. *1st International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering, ECTI-NCON 2018*, 116–119.

Islam, M. T., Karim Siddique, B. M. N., Rahman, S., & Jabid, T. (2018). Image Recognition with Deep Learning. *2018 International Conference on Intelligent Informatics and Biomedical Sciences, ICIIBMS 2018*, *3*, 106–110. https://doi.org/10.1109/ICIIBMS.2018.8550021

Jiang, S., Min, W., Liu, L., & Luo, Z. (2020). Multi-Scale Multi-View Deep Feature Aggregation for Food Recognition. *IEEE Transactions on Image Processing*, *29*(XX), 265–276. https://doi.org/10.1109/TIP.2019.2929447

Joutou, T., & Yanai, K. (2019). A Food Image Recognition System with Multiple Kernel Learning. *Icip 2019*, 285–288.

Kogias, K., Andreadis, I., Dalakleidi, K., & Nikita, K. S. (2018). A Two-Level Food Classification System for People with Diabetes Mellitus Using Convolutional Neural Networks. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2018-July*, 2603–2606. https://doi.org/10.1109/EMBC.2018.8512839

Kurniadi, A., & Fal Sadikin, M. (2020). Implementasi Convolutional Neural Network Untuk Klasifikasi Varietas Pada Citra Daun Sawi Menggunakan Keras Implementation of Neural Network Convolutionals For Classification of Variety on Image of Collards Meat Leaves Using The Keras. *Journal of Computer and Information Technology*, *4*(1), 25–33. http://e-journal.unipma.ac.id/index.php/doubleclick

Kusumaningrum, T. F. (2018). Implementasi Convolutional Neural Network (CNN) Untuk Klasifikasi Jamur Konsumsi Di Indonesia Menggunakan Keras. *Yogyakarta: Universitas Islam Indonesia.*

LeCun Y. (1990). *Handwritten Digital Recognition with a Back Propagation Network*.

Mezgec, S., & Seljak, B. K. (2019). Using Deep Learning for Food and Beverage Image Recognition. *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, 5149–5151. https://doi.org/10.1109/BigData47090.2019.9006181

Neneng, N., Adi, K., & Isnanto, R. (2016). Support Vector Machine Untuk Klasifikasi Citra Jenis Daging Berdasarkan Tekstur Menggunakan Ekstraksi Ciri Gray Level Co-Occurrence Matrices (GLCM). *Jurnal Sistem Informasi Bisnis*, *6*(1), 1. https://doi.org/10.21456/vol6iss1pp1-10

Nugraha, P. G. S. C., & Cipta, S. (2020). Prediksi Citra Makanan Menggunakan Convolutional Neural Network Untuk. *Jurnal Teknologi Informasi Dan Komputer*, 30–38.

Nugroho, P. A., Fenriana, I., & Arijanto, R. (2020). Implementasi Deep Learning Menggunakan Convolutional Neural Network (Cnn) Pada Ekspresi Manusia. *Algor*, *2*(1), 12–21.

Nurhikmat. T, & Purwaningsih. T. (2018). Implementasi Deep Learning Untuk Image Classification Menggunakan Algoritma Convolutional Neural Network (CNN) Pada Citra Wayang Golek. *ResearchGate*, 1–113. https://doi.org/10.13140/RG.2.2.10880.53768

Pan, L., Pouyanfar, S., Chen, H., Qin, J., & Chen, S. C. (2017). DeepFood: Automatic Multi-Class Classification of Food Ingredients Using Deep Learning. *Proceedings - 2017 IEEE 3rd International Conference on Collaboration and Internet Computing, CIC 2017*, *2017-Janua*, 181–189. https://doi.org/10.1109/CIC.2017.00033

Pandey, P., Deepthi, A., Mandal, B., & Puhan, N. B. (2017). FoodNet: Recognizing Foods Using Ensemble of Deep Networks. *IEEE Signal Processing Letters*, *24*(12), 1758–1762. https://doi.org/10.1109/LSP.2017.2758862

Peryanto, A., Yudhana, A., & Umar, R. (2020). Klasifikasi Citra Menggunakan Convolutional Neural Network dan K Fold Cross Validation. *Journal of Applied Informatics and Computing*, *4*(1), 45–51. https://doi.org/10.30871/jaic.v4i1.2017

Rachardi, F. (2020). *Deteksi Gambar Gestur Kosakata Bahasa Isyarat Indonesia dengan Convolutional Neural Network*. 192.

Rahim, A., Kusrini, K., & Luthfi, E. T. (2020). Convolutional Neural Network untuk Kalasifikasi Penggunaan Masker. *Inspiration: Jurnal Teknologi Informasi Dan Komunikasi*, *10*(2), 109. https://doi.org/10.35585/inspir.v10i2.2569

Ramanath, A., Muthusrinivasan, S., Xie, Y., Shekhar, S., & Ramachandra, B. (2019). Ndvi versus cnn features in deep learning for land cover classification of aerial images. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2019-July*, 6483–6486. https://doi.org/10.1109/IGARSS.2019.8900165

Salam, A., & Abdul, B. (2019). Classification of Food Nutrients Composition using Deep Learning. *Malaysian Journal of Information and Communication Technology*, *4*(2), 66–83.

Salsabila. (2018). Penerapan Deep Learning Menggunakan Convolutional Neural Network Untuk Klasifikasi Citra Wayang Punakawan. *Yogyakarta: Universitas Islam Indonesia.*

Solanki, D. (2020). *Detection and Classification of Food Consumption Using Convolutional Neural Networks*. *9*(12), 802–805. https://doi.org/10.21275/SR201130135219

Suartika E. P, I Wayan, Wijaya Arya Yudhi, S. R. (2016). Klasifikasi Citra Menggunakan Convolutional Neural Network (Cnn) Pada Caltech 101. *Jurnal Teknik ITS*, *5*(1), 76. http://repository.its.ac.id/48842/

Subhi, M. A., & Ali, S. M. (2019). A deep convolutional neural network for food detection and recognition. *2018 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2018 - Proceedings*, 284–287. https://doi.org/10.1109/IECBES.2018.08626720

Sukvichai, K., Maolanon, P., Sawanyawat, K., & Muknumporn, W. (2019). Food categories classification and Ingredients estimation using CNNs on Raspberry Pi 3. *10th International Conference on Information and Communication Technology for Embedded Systems, IC-ICTES 2019 - Proceedings*. https://doi.org/10.1109/ICTEmSys.2019.8695967

Temdee, P., & Uttama, S. (2017). *Food Recognition on Smartphone using Transfer Learning of Convolution Neural Network*. IEEE.

Termritthikun, C., & Kanprachar, S. (2017). Accuracy improvement of Thai food image recognition using deep convolutional neural networks. *2017 International Electrical Engineering Congress, IEECON 2017*, 2–5. https://doi.org/10.1109/IEECON.2017.8075874

Wulandari, I., Yasin, H., & Widiharih, T. (2020). Klasifikasi Citra Digital Bumbu Dan Rempah Dengan Algoritma Convolutional Neural Network (Cnn). *Jurnal Gaussian*, *9*(3), 273–282. https://doi.org/10.14710/j.gauss.v9i3.27416

Yaman, D., Eyiokur, F. I., & Ekenel, H. K. (2017). *Comparison of convolutional neural network models for document image classification*. 1–4. https://doi.org/10.1109/siu.2017.7960562

Yamparala, R., & Challa, R. (2020). *Computerized Classification of Fruits using Convolution Neural Network*. 10–13.

Yanai, K., & Kawano, Y. (2014). FOOD IMAGE RECOGNITION USING DEEP CONVOLUTIONAL NETWORK WITH PRE-TRAINING AND FINE-TUNING Keiji Yanai Yoshiyuki Kawano Department of Informatics , The University of Electro-Communications , Tokyo , Japan. *International Conference on Multimedia & Expo Workshops (ICMEW) . IEEE*, 1–6.