Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.5, October 2025, pp. 2009~2018
e-ISSN: 2548-6861

Indonesian Food Classification Using Deep Feature Extraction and Ensemble Learning for Dietary Assessment

Muhammad Yusuf Kardawi [1]*, Frederic Morado Saragih [2]*, Laksmita Rahadianti [3]*, Aniati Murni Arymurthy[4]*
* Faculty of Computer Science, Universitas Indonesia
my.kardawi@gmail.com [1], frederic.morado@ui.ac.id [2], laksmita@cs.ui.ac.id [3], aniati@cs.ui.ac.id [4]

Article Info

Article history:

Received 2025-08-04
Revised 2025-08-29
Accepted 2025-09-03

Keyword:

Deep Learning,
Dietary Assessment,
Histogram Equalization,
Machine Learning,
Padang Cuisine.

ABSTRACT

Food is a cornerstone of culture, shaping traditions and reflecting regional identities. However, understanding the nutritional content of diverse cuisines can be challenging due to the vast array of ingredients and the similarities in appearance across different dishes. While food provides essential nutrients for the body, excessive and unbalanced consumption can harm health. Overeating, particularly high-calorie and fatty foods, can lead to an accumulation of excess calories and fat, increasing the risk of obesity and related health issues such as diabetes and heart disease. This paper introduces a novel ensemble learning approach with a dictionary that contains food nutrition content for addressing this challenge, specifically on Padang cuisine, a rich culinary tradition from West Sumatera, Indonesia. By leveraging a dataset of nine Padang dishes, the system employs image enhancement techniques and combines deep feature extraction and machine learning algorithms to classify food items accurately. Then, depending on the classification results, the system evaluates the nutritional content and creates a dietary evaluation report that includes the amount of protein, fat, calories, and carbs. The model is evaluated using different evaluation metrics and achieving a state-of-the-art accuracy of 85.56%, significantly outperforming standard baseline models. Based on the findings, the suggested approach can efficiently classify different Padang dishes and produce dietary assessments, enabling personalised nutritional recommendations to provide clear information on a balanced diet to enhance physical and overall wellness.

I. INTRODUCTION

Padang cuisine is not only well-known to be delicious, but also a reflection of Indonesian culture, tradition, and history. It is an essential component of the national identity and plays a significant role in the country's economy and tourism. Exploring the rich flavours of Padang cuisine means diving into the fascinating diversity of Indonesian culture and traditions. While Padang cuisine offers irresistible delights, it is important to remember that overconsumption can bring negative health impacts. An irregular diet can have a high risk of triggering obesity. Padang cuisine is generally high in calories, fat and salt. Rendang, one of Padang's most popular dishes, contains about 193 calories and 7.9 grams of fat per 100 grams. Although Padang culinary richness should be preserved, excessive and unbalanced consumption can be harmful to health. Excessive consumption can lead to the accumulation of calories and fat, leading to obesity.

Obesity is defined as abnormal or excessive fat cumulation that poses a health risk. According to the global burden of illness, this issue has escalated into an epidemic, with over 4 million deaths annually attributed to obesity in 2017. Over 1 billion individuals globally suffer from obesity, comprising 650 million adults, 340 million teenagers, and 39 million children [1]. In Indonesia, easy access to unhealthy foods high in fat, sugar and salt is a significant cause of malnutrition. The 2018 Basic Health Research in Indonesia, according to Health Research of the Republic of Indonesia, one in three adults, one in five children aged five to twelve, and one in seven teenagers aged thirteen to eighteen are overweight or obese [2].

Food computing primarily applies computer science techniques to studies about food. It entails collecting and analyzing food data from many modalities, including food images, food logs, recipes, tastes, and smells [3]. It has become the center of public attention as an essential area in solving food-related issues in human life, such as food selection [4], healthy diet [5], nutrition analysis [6], and food recognition [7]. As one primary task in food computing, food image classification plays an essential role in diet management for personal health management. Nowadays, more chances to assist people in comprehending their everyday eating habits, investigating nutritional patterns, and maintaining a balanced diet are made possible by diet evaluation and nutrition analysis tools. Food image classification can help in determining the nutritional needs of the body [8] and this plays an essential role in its use in health fields such as understanding nutrients in food [9], [10] and healthy diet management [11].

However, automatic food classification based on computer vision is difficult because they are very similar visually. For example, some soupy food classes may resemble other soupy food classes. Additionally, fried foods are also very difficult to distinguish. Finally, Asian food, including Indonesian food, generally has a more diverse composition of recipes than Western food [12]. Also, it is challenging to ascertain the composition of food due to its changeable shape, which depends on the cooking technique [13]. As a result, it is difficult to determine the type of food through image classification because it generally has similarities between its classes and visually has similarities with food in other classes, as well as differences from food in the same class.

Previous research works have attempted to use a variety of methods based on both classic machine learning and deep learning to tackle the challenging job of food classification [3]. Nonetheless, there has been a limited number of studies that utilize Indonesian food datasets. Also, the studies only focus on using and developing one of the traditional machine learning or deep learning models [5]. Thus, in this paper, we propose a deep model that combines deep learning-based methods with traditional machine learning classification models on the Indonesian food dataset. The following is a summary of our contribution:

1. This paper proposes incorporating deep convolutional neural network models for feature extraction. First, we perform image enhancement on the input image. Next, we combine state-of-the-art deep convolutional neural networks (VGG16 and ResNet-50) followed by a concatenate layer to extract features from each input image.
2. We explore traditional machine learning on bagging and boosting using an ensemble model approach for food classification and diet assessment. Specifically, we designed and developed an ensemble system for food image analysis that can generate food type classification and detailed nutrient content of each food item.
3. We developed a nutritional reference database for Padang cuisine, drawing upon the existing "Indonesian Food Image Dataset" from Kaggle website.
4. We evaluated our nutrition analysis system and food classification model through a series of comprehensive experiments. We tried various combinations in the development process and compared the effectiveness of various models using the average accuracy, precision, recall and F1-Score metrics.

II. LITERATURE STUDY

In addressing nutrition issues in food consumption, it is first essential to identify the type of food consumed and what nutrients are present. Images of the same type of food vary in content, shape, size, texture, and color depending on the geography, cultural customs, or ingredients that are available. Sophisticated deep learning approaches that can recognize the distinctive qualities of several dish varieties are needed to solve this challenge [14].

Previous research has created a novel and scalable platform that can record Indonesian traditional food knowledge and classify it using the principles of depth-based CNN and multipath-based CNN [15]. Another study extensively reviews algorithms and methodologies in image-based dietary assessment, Comparing cutting-edge methods for volume estimate and autonomous food identification. It divides food identification techniques into two categories: deep learning-based end-to-end picture recognition and traditional methods using features that are manually created. In a similar vein, other volume estimation techniques are investigated, such as deep learning, perspective transformation, stereo, model-based, and depth camera-based techniques [16]. Another one addresses the challenge of continually learning to classify food images online. It introduces a new method for selecting representative examples from each class by clustering data based on visual similarity and selecting examples based on cluster means using Power Iteration Clustering [17].

Traditional machine learning methods such as using color histogram and Gabor texture methods for feature extraction [18], combining individual dietary patterns with image analysis results [19], using support vector machine (SVM) based image segmentation [20], utilizing Nu Support Vector Regression [21], and using food images with specific domains [22]. Other methods, such as Bayesian network for incremental learning in food detection and food balance estimation [23]. Random Forest models were also used to cluster superpixels from training data [24], and other techniques used include Improved Fisher Vectors (IFV) [25], Bag-of-Words Histogram (BOW) [26], Bag-of-Features (BoF) [27], Randomized Clustering Forests (RCF) [28], Nearest neighbour classifier [29] generating descriptors based on Difference of Gaussian (DoG) and Scale Invariant Feature Transform (SIFT) [30] and Mid-Level Discriminative Superpixels (MLDS) [31].

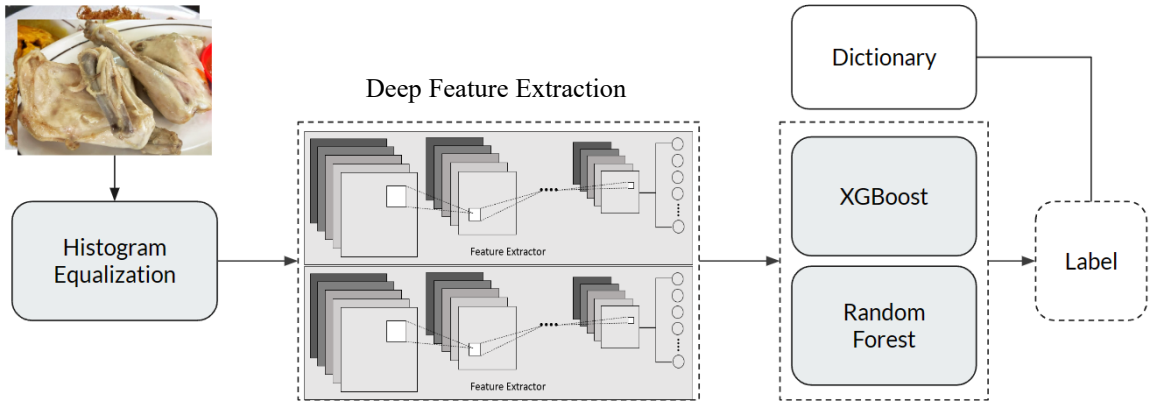

Figure 1. The framework of the proposed method

Deep learning models are commonly used to classify food based on images due to their ability to extract various image features, such as changing the traditional convolution in the Convolutional Neural Network (CNN) algorithm with jumping convolution as the basic unit aimed at food image feature extraction [32], model complexity reduction and data augmentation [33], extracting regions of interests (ROIs) by applying Region Proposal Network [34], and exploiting the rich relationships through bipartite-graph labels (BGL) [35].

The researchers also built a deep CNN model based on AlexNet for food recognition on the Food-101 dataset. They achieved an average accuracy of 50.76% with the help of the Random Forests algorithm in finding food-distinguishing features [36]. Then, another study improved the accuracy to 74% for 10 types of dishes with different CNN configurations [37]. Another study was to extract picture patches for every food item on a grid and feed those patches into a deep CNN [38]. However, these approaches are time-consuming and require complex configuration of model parameters. Transfer learning offers a faster and more effective solution using pre-trained models with minimal configuration for food recognition rather than building deep CNN models from scratch.

III. Materials and Method

This section details the proposed methodology for a food image classification system specifically designed for dietary assessment applications. The system aims to accurately identify Indonesian Padang food items from images and provide corresponding nutritional information. As illustrated in Figure 1, the proposed methodology comprises a four-stage pipeline: image enhancement, feature extraction, classification, and nutritional analysis. Each stage plays a critical role in the overall system performance and will be described in detail in the following subsections. The ultimate goal is to create a system that can assist users in making informed dietary choices by providing readily accessible nutritional data based on visual food input.

A. System Design and Development

We developed several stages in the food image classification system for dietary assessment as shown in Figure 2. Starting from input images that vary significantly in size and different light intensities, some images tend to be dark and images that are very bright. Because of this diversity, we apply the Histogram Equalization (HE) technique to increase the contrast of the input images and stretch the intensity range of the images. Meanwhile, we perform image resizing to 3 x 256 x 256 to overcome the diversity in image size. Then, we apply deep feature extraction to the uniform image using the transfer learning method. We apply the combined architecture of VGG16 [39] and ResNet50 [40] as the feature extractor and add a concatenate layer to combine the features from the two different models.

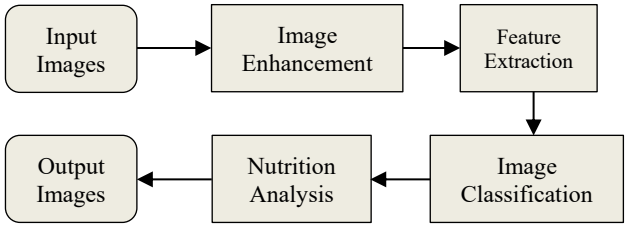

Figure 2. Flowchart of proposed diagram

We selected VGG16 and ResNet50 as feature extractors due to their complementary strengths and proven performance in image classification tasks. VGG16 offers a relatively simple and uniform architecture characterized by its consistent use of small 3x3 convolutional filters, making it easy to implement and understand while providing strong feature representation capabilities. ResNet50, on the other hand, addresses the vanishing gradient problem often encountered in deep networks through its innovative use of residual connections (skip connections). These connections allow gradients to propagate more effectively during training, enabling the network to learn more complex and nuanced

TABLE I
DATASET DESCRIPTION

Class Name	Ayam Goreng	Ayam Pop	Daging Rendang	Dendeng Batokok	Gulai Ikan	Gulai Tambusu	Gulai Tunjang	Telur Balado	Telur Dadar
Train	84	90	81	86	88	78	93	79	92
Test	20	22	20	21	21	19	23	19	22
Total	958 Images								

features from the input images, ultimately resulting in improved performance, especially with limited training data, as is typical in transfer learning. Combining these two models allows us to leverage simplicity and depth, capturing a wider range of features for a more robust food image classification.

In the classification stage, we apply an ensemble algorithm based on bagging and boosting by combining Random Forest [24] and XGBoost [41] classification algorithms. We opted for Random Forest and XGBoost as our ensemble classifiers because they represent two powerful and distinct approaches to ensemble learning, offering a balance of robustness and accuracy. Random Forest, a bagging method, builds multiple decision trees on random subsets of the data and features, reducing variance and overfitting, which is crucial when dealing with high-dimensional feature spaces extracted from deep learning models. XGBoost, a gradient boosting method, sequentially builds trees, with each subsequent tree correcting the errors of its predecessors, leading to high predictive accuracy and the ability to capture complex non-linear relationships within the data. By combining these two algorithms through soft voting, we leverage the strengths of both the stability and generalization ability of Random Forest and the high precision and fine-tuning capabilities of XGBoost.

Following successful food image classification, the system accesses a dedicated nutritional dictionary to retrieve the corresponding dietary information. This dictionary, explicitly built for this project, contains calorie, fat, carbohydrate, and protein values for each of the nine Padang dishes, based on a standard 100g serving size. The output presented to the user is an image of the food, clearly labeled with its identified class and the associated nutritional breakdown. While the seamless integration of classification and nutritional information is a key feature of the system, the critical point of our system's design and development, and the primary focus of our evaluation, is the synergistic combination of the feature extraction and ensemble classification stages.

We hypothesize that this combination leads to superior performance compared to using individual components. We present results for the combined system and conduct a detailed comparative analysis to validate this. This analysis involves systematically evaluating the performance of each feature extractor (VGG16 and ResNet50) and each classifier (Random Forest and XGBoost) in isolation, as well as in all possible pairwise combinations, before finally evaluating the complete ensemble.

B. *Dataset Description*

Our experiments were conducted using the publicly available Padang Cuisine dataset, the Indonesian Food Image Dataset, obtained from the Kaggle platform [42]. This dataset focuses on nine distinct classes of Padang cuisine, a popular and regionally diverse culinary tradition from West Sumatra, Indonesia. Table I provides a detailed breakdown of the dataset's composition, including the class names and the initial number of images per class. The dataset, as obtained initially, comprised 992 images across these nine classes. This dataset provides a valuable resource for developing and evaluating food classification models, particularly within the context of Indonesian cuisine, which often presents challenges due to visual similarities between dishes and variations in preparation and presentation.

To ensure the quality and reliability of our results, we performed a crucial data-cleaning step before model training. This involved manually screening each image within the dataset to identify and remove any images that did not meet our predefined quality criteria. Specifically, we removed images that exhibited significant blurriness, low resolution, or other visual artifacts that could negatively impact the performance of the feature extraction and classification stages. This rigorous cleaning removed 34 images, leaving a final dataset of 958 high-quality images.

Figure 3. Sample images from the dataset

Although the dataset employed in this study contains a relatively limited number of images, it was selected due to its wide recognition and frequent use as a benchmark on the Kaggle platform, which underscores its reliability and relevance within the research community. Furthermore, the

availability of alternative publicly accessible datasets of comparable quality and suitability for the intended tasks is highly constrained, thereby reinforcing the justification for its adoption in this research.

This curated dataset was partitioned into training and testing sets, following a standard 80:20 split. The training set, comprising 80% of the images (766 images), was used to train the various models, while the remaining 20% (192 images) served as a held-out test set for evaluating model performance. Importantly, as shown in Table I, the class distribution within the dataset is relatively balanced, mitigating the need for data balancing techniques such as oversampling or undersampling. Figure 3 provides representative sample images from each of the nine Padang food classes, illustrating the visual characteristics and diversity within the dataset.

C. Image Enhancement

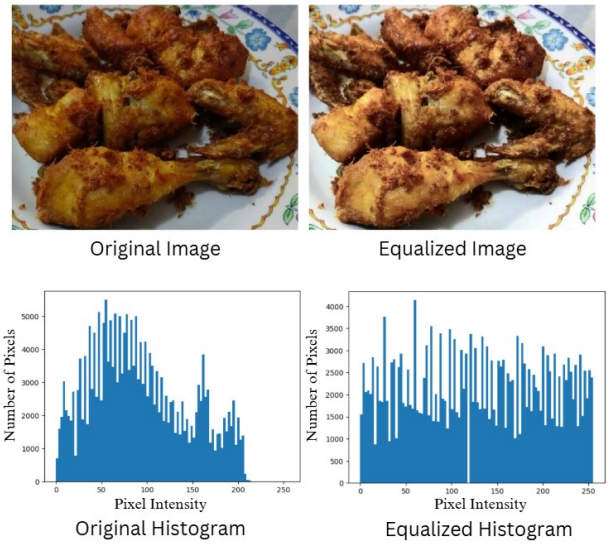

Figure 4. Sample of Implementation for HE

Image enhancement is the process of improving image quality that aims to make the image easier to interpret, and this stage is an initial process in image processing. We apply the histogram equalization (HE) algorithm at this stage, an image processing method used to enhance image contrast by adjusting the histogram. A histogram is a graph that illustrates how frequently each intensity value appears in a picture. Low-contrast images have uneven histograms, with some intensity values appearing very frequently and others appearing only a few times. It causes the image to look blurry and unclear. HE creates a more uniform histogram image by reassigning the histogram pixel values, thereby improving the image contrast. However, the drawback of using HE is that it only considers the probability distribution of pixel values. It does not consider the spatial distribution of pixels in the image, thus allowing for brightness distortion issues to occur [43]. Figure 4 shows the comparison of histograms before and after equalization.

D. Feature Extraction

Feature extraction forms the foundation of our food image classification system, serving as a critical step in transforming raw pixel data into meaningful representations that machine learning algorithms can effectively utilize. In essence, feature extraction aims to capture the food images' most salient and discriminative visual characteristics, enabling the model to distinguish between different food classes. To achieve this, we leverage the power of CNN, a class of deep learning models designed explicitly for processing image data. We employ two distinct and well-established CNN architectures: VGG16 and ResNet50. VGG16, known for its simplicity and uniform structure with stacked 3x3 convolutional layers, provides a robust baseline for feature extraction. ResNet50, with its more profound architecture and innovative use of residual connections (skip connections), allows for the extraction of more complex and hierarchical features, mitigating the vanishing gradient problem that can hinder the training of intense networks. The selection of these two architectures allows combining the strengths of both models.

We employ a transfer learning approach to maximize the effectiveness of these CNN architectures and expedite the training process. Instead of training the networks from scratch, which would require a massive dataset and significant computational resources, we utilize pre-trained versions of VGG16 and ResNet50. These models have been pre-trained on the large-scale ImageNet dataset, containing millions of images across thousands of object categories. This pre-training process allows the networks to learn general-purpose image features transferable to other tasks, including our food classification problem. Specifically, we use VGG16 and ResNet 50 as fixed feature extractors. The final fully connected layer and softmax layer (the "top layer") of each pre-trained network, which is specific to the ImageNet classification task, are removed. The remaining layers encode the learned visual features and are then used to process our Padang food images. The output of VGG16 is 4096 dimensions, and Resnet50 is 2048. The resulting feature vectors from both networks are then concatenated along the channel dimension using a dedicated concatenation layer, resulting in a combined feature vector of 2560 dimensions. This rich, fused feature representation captures a broader range of visual information than either network could provide, forming the input for the subsequent classification stage.

E. Ensemble Classification

The classification stage assigns a class label (one of the nine Padang food types) to each input image based on its extracted features. This is achieved through an ensemble of two distinct classification algorithms: Random Forest and XGBoost. The input to this stage is the 2560-dimensional

TABLE III
REFERENCE FOR NUTRITIONAL CONTENTS

Food (100 g)	Calories (kkal)	Fat (gram)	Carbohydrate (gram)	Protein (gram)
Ayam Goreng	595	0	0	30.5
Ayam Pop	256	0	0	30.5
Daging Rendang	193	7.9	7.8	22.6
Daging Batokok	433	9	7.8	55
Gulai Ikan	106	3.3	2.5	16.5
Gulai Tambusu	130	7.2	1.5	10.9
Gulai Tunjang	122	5.7	3.9	13.8
Telur Balado	162	11.5	0.7	12.8
Telur Dadar	251	19.4	1.4	16.3

feature vector produced by the feature extraction stage, representing a concatenation of the feature maps from the pre-trained VGG16 and ResNet50 models. Random Forest and XGBoost are trained independently on these feature vectors to learn how to map the features and the food classes. We utilize a soft voting mechanism to combine the predictions of these two classifiers. Soft voting calculates the average predicted probability for each class across both classifiers. The class with the highest average probability is then chosen as the final predicted class for the input image. Combining the outputs of diverse classifiers, this ensemble approach enhances the overall robustness and accuracy of the food classification system.

TABLE II
TRAINING TIME

Models	Time (second)
VGG + Random Forest	20.53
VGG + XGBoost	345.56
ResNet + Random Forest	48.08
ResNet + XGBoost	1716.87
VGG + ResNet + Random Forest	37.28
VGG + ResNet + XGBoost	3125.59
VGG + ResNet + Random Forest + XGBoost	1318.53

Table II shows the various combinations of experiments conducted with different configurations of ensemble model types. The last model is the most complete attempt, combining two different classifiers: Random Forest and XGBoost. Details of the time used in the training process can also be seen there. In the context of ensemble models, training time includes the entire process involving training the XGBoost model, training the Random Forest model, and merging the results of both. It includes data processing, decision tree building, parameter tuning, and other steps that may be involved in model training. As for models that only use one of the two types of classifiers, training time is the time the model takes to build the decision tree.

F. Nutrition Analysis

After the food classification has been completed, the last step is for the system to generate a dietary assessment that analyzes the food's nutrients. In this paper, we analyze the number of calories, fat, carbohydrates, and protein of each food type. The nutritional analysis results for dietary assessment will be displayed in the output image. The nutritional references we use are references that we make based on various sources from the internet. Details of the nutritional references used in the dietary assessment can be seen in Table III. We take the weight of each item of food to be 100 grams, the average amount for each meal. Each row represents the food's calories, fat, carbohydrate, and protein nutrients.

For this initial study, we adopted a simplified approach using a standard 100g portion size for each food item to establish a baseline for nutritional analysis. We recognize that this is a simplification and that actual portion sizes vary significantly. This approach allows for a preliminary assessment of the nutritional content based solely on food type identification, but it does not account for portion size variations.

IV. RESULTS AND DISCUSSION

The developed system aims to classify food images and provide nutritional analysis of food images to assist in dietary assessment. This section will discuss and evaluate the proposed methodology, including implementation details, evaluation metrics, and a comparison with existing approaches. The results demonstrate the effectiveness of our combined deep learning and ensemble learning approach.

A. Implementation Details

The proposed method has been developed using the TensorFlow framework. We convert the Blue-Green-Red (BGR) image to Luminance-representation between green and red-representation between blue and yellow (LAB) color space in the HE stages. LAB separates the image information

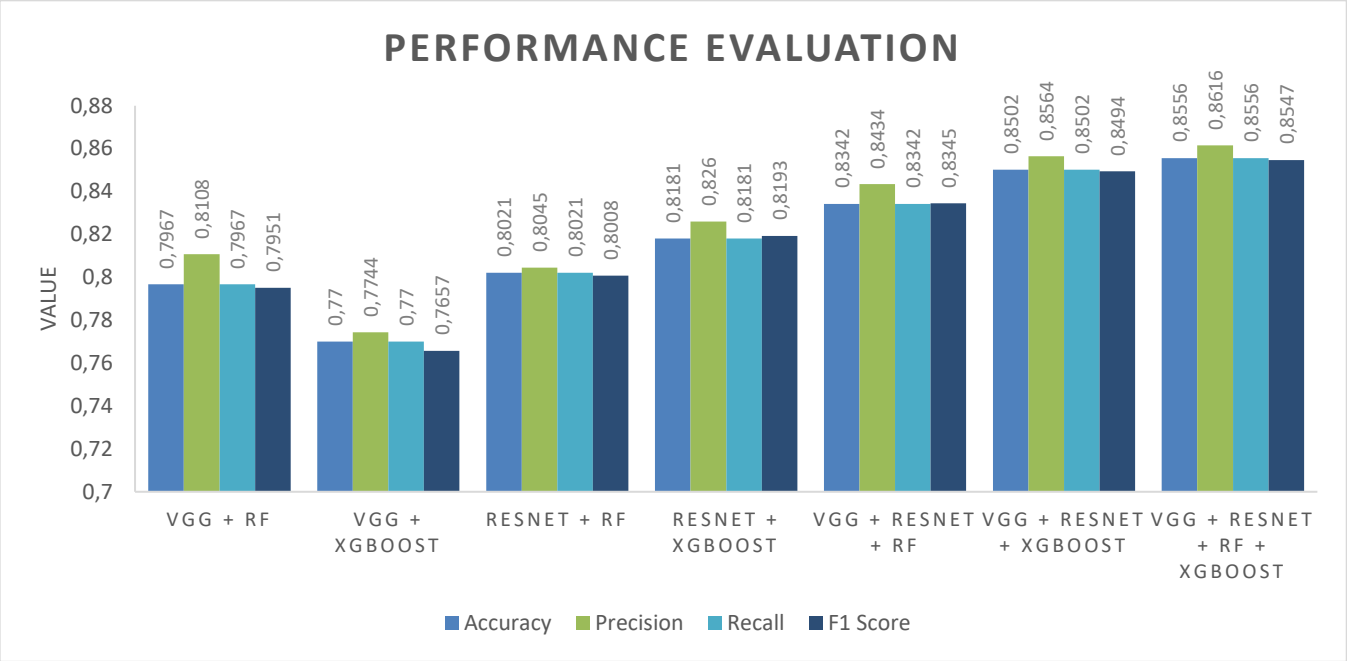

Figure 5. Models performance evaluation

into luminance (L) and two channels for color (A and B). It is helpful because HE works best on the luminance alone. Then, we apply HE to the L channel. This process redistributes the intensity values in the L channel to create a more uniform distribution, effectively enhancing the image contrast. To evaluate the proposed method, we performed various combinations between the respective uses of feature extraction and classifier before finally combining feature extraction and utilizing ensemble learning for classification.

TABLE IV
PARAMETER USED FOR CLASSIFICATION

XGBoost	Value	Random Forest	Value
n_estimators	200	n_estimators	200
max_depth	3	max_depth	15
learning_rate	0.3	min_samples_split	2
subsample	0.8	min_samples_leaf	1
colsample_bytree	0.8	max_features	'sqrt'
reg_lambda	1		
reg_alpha	0		

Table IV explains each parameter in the classification stage using Random Forest and XGBoost. Although both techniques are based on decision trees, there are differences in their approach. Therefore, the hyperparameter tuning process also differs in the parameters used. In this study, we set the number of n_estimators as 200. The n_estimator parameter determines the number of decision trees to be built. The larger the n_estimator value, the more complex the model will be. However, a more complex model will also be more prone to overfitting. Also, the training time depends on the parameter configuration in Table IV.

B. Evaluation Measure

The most used measures are evaluation matrices such as accuracy, precision, recall, and F1 Score. The total average accuracy is calculated by dividing the total number of properly categorized data by the total amount of data. A popular metric for evaluating a classification model's performance is average accuracy. Average accuracy gives a general idea of how well the model can predict the data. Precision measures how accurately the model predicts the positive class. Then, recall measures how completely the model detects the positive class. Meanwhile, the F1 Score is a combination of precision and recall. A high F1 Score denotes a good recall and precision of the model.

C. Food Image Classification

At this stage, we compare all the results obtained from the proposed methods. By trying deep feature extraction and combining it and trying two traditional machine learning classification techniques, we finally perform ensemble learning with a soft voting method to obtain the class prediction probability value before finally customizing each food's nutrition dictionary.

Figure 5 describes the results obtained from each experiment conducted. In this study, we utilized the transfer learning technique on the VGG16 and ResNet50 models, so based on the evaluation results obtained, the ResNet model provides better classification values than the VGG16 model in the case of performing feature extraction with an accuracy value of 80.21% on the Random Forest classifier and 81.81%

Figure 6. Sample of final output with nutrition assessment

on the XGBoost classifier. The transfer learning technique makes the ResNet50 architecture superior as it has more parameters and layers than the VGG16 architecture, which tends to be leaner. In the final stage of the experiment, we obtained the highest accuracy of 85.56% by combining feature extraction and using the ensemble classifier.

ResNet50 generally outperforms VGG16 in this food classification task primarily due to its more profound architecture and the incorporation of residual connections. While VGG16 has a more straightforward, sequential structure, ResNet50's depth allows it to learn more complex and hierarchical features, crucial for distinguishing subtle variations between visually similar Padang dishes. Crucially, ResNet50's residual connections mitigate the vanishing gradient problem that often plagues deep networks, enabling practical training and allowing information to flow more quickly through the network, thus preserving and utilizing more nuanced features. The superior performance of the ensemble (Random Forest + XGBoost with soft voting) stems from combining the strengths of two diverse and robust classifiers. Random Forest reduces variance and overfitting through its bagging approach, while XGBoost boosts accuracy by sequentially correcting errors. Soft voting then leverages the probability distributions from both, effectively averaging out individual classifier biases and uncertainties, leading to a more robust and accurate final prediction that capitalizes on the complementary strengths of each model.

TABLE V
MODEL COMPARISON WITH OTHERS

Model	Accuracy	Time (second)
ResNet18 [40]	0.7593	2762.24
VGG16 [39]	0.7807	2095.86
Proposed	0.8556	1318.53

Table V presents a direct performance comparison between our proposed ensemble method and the baseline models, ResNet18 and VGG16, for food image classification. To ensure a fair evaluation, all models were trained under the same experimental setup, using identical hyperparameters, including a learning rate of 0.0001, batch size of 16, weight decay of 1×10^{-4}, and 100 training epochs with the Adam optimizer, which outperformed alternatives like SGD. Our proposed method achieved superior accuracy while requiring less training time than ResNet18 and VGG16, which attained accuracy scores of 75.93% and 78.07%, with training times of 2762 and 2095 seconds, respectively. This controlled comparison isolates model architecture's impact on performance and training efficiency.

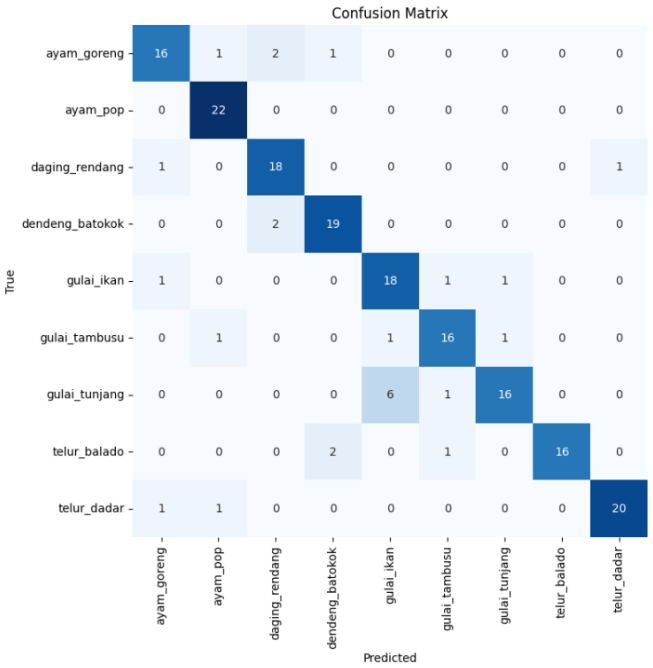

Figure 7. Confusion matrix of selected model

An interesting observation is that there is a challenging class to predict, namely the Gulai Tunjang food type. It is because the appearance and colour between the Gulai Tunjang and Gulai Ikan classes are very similar, so it would be challenging to classify, especially since both food types have similar types of soup. For example, in the case of fish curry,

if there is a lot of curry/gravy, it will make the fish that should be in the food invisible and confuse the model. However, the result of the confusion matrix as shown in figure 7 is outstanding overall. There is only one class that is not optimal due to the characteristics of the food dataset. The proposed system's final step in campaigning for a healthy diet is to analyze each food image's nutritional content. As explained in the previous section, our nutritional analysis assumes a base weight of 100 grams per food. The example of the final output can be seen in Figure 6. The food image classification and nutrition analysis system help users maintain a healthy diet. The system identifies food types, calculates their nutritional information, and displays it on the output image. The benefit is that users can track their nutritional intake, choose foods that suit their needs, and maintain a balanced diet.

To provide a comprehensive evaluation of our proposed ensemble deep learning approach, our analysis extends beyond simply comparing individual components of our system (VGG16 and ResNet50 as feature extractors, Random Forest, and XGBoost as classifiers). We also benchmark our results against two widely used, standalone deep learning models for image classification: VGG16 and ResNet18. These models serve as crucial baselines, allowing us to assess the performance gains achieved by combining feature extraction and ensemble learning and comparing them to complete, end-to-end deep learning solutions commonly used in the field. This broader comparison strengthens our findings validity and demonstrates our proposed method's advantages within a broader context of image classification techniques.

V. CONCLUSION

This paper explores food image classification by combining deep learning and traditional machine learning techniques. We utilize deep learning techniques to perform feature extraction and machine learning techniques to perform classification. We combine two types of deep learning architectures and two types of classifier algorithms from traditional machine learning. We obtained the best model combination in the model with VGG16-ResNet50 as a deep feature extractor and Random Forest-XGBoost as a classifier with the application of soft voting. We obtained the best evaluation score compared to other model combinations, which are 85.56% accuracy, 86.16% precision, 85.56% recall, and F1 score of 85.47%. These evaluation values are much better than fine-tuned baseline architectures such as ResNet18 and VGG16. Our model also reduces the training time to only 1318.53 seconds, thanks to the utilization of transfer learning in the deep feature extraction stage. The fact that images in the same cuisine category have traits related to color or pattern, combining multiple deep feature extractors, and utilizing ensemble learning can increase classification accuracy in the same food class.

While our system demonstrates strong performance in classifying Padang cuisine and providing nutritional information, it is essential to acknowledge its limitations. The current model is primarily trained and validated on a dataset of nine Padang dishes, meaning its applicability to other cuisines or broader food categories is not guaranteed. It would require further training and data collection. Furthermore, the nutritional analysis is based on a standardized 100-gram serving size for each food item. While this provides a valuable baseline, it does not account for variations in portion sizes consumed in real-world scenarios, nor does it capture individual-specific dietary needs or customizations in food preparation that might affect nutritional content.

We use the classification results to analyze the nutrition of the classified food and display a report of the number of calories, fat, carbohydrates, and protein in the output image. We conducted extensive experiments to evaluate the efficiency and effectiveness of our system. The results show that our proposed solution achieves excellent performance and has the nutrition of the meal that was identified is examined using the classification results, and a report on the amount of calories, fat, carbs, and protein in the final image is displayed. To assess the usefulness and efficiency of our system, we carried out several comprehensive evaluations. Conclusions demonstrate the outstanding performance of our suggested approach and its enormous potential for promoting a healthy diet to avoid overweight and obesity.

ACKNOWLEDGMENT

The author(s) disclosed receipt of the following financial support for the publication of this article. This research was conducted during studies funded by Lembaga Pengelola Dana Pendidikan (LPDP) under Master program scholarship.

DAFTAR PUSTAKA

[1] World Health Organization, “World Obesity Day 2022 – Accelerating action to stop obesity.” Accessed: Dec. 20, 2023. [Online]. Available: https://www.who.int/news/item/04-03-2022-world-obesity-day-2022-accelerating-action-to-stop-obesity

[2] UNICEF Indonesia, “Indonesia: Angka orang yang kelebihan berat badan dan obesitas naik di semua kelompok usia dan pendapatan.” Accessed: Dec. 27, 2023. [Online]. Available: https://www.unicef.org/indonesia/id/siaran-pers/indonesia-angka-orang-yang-kelebihan-berat-badan-dan-obesitas-naik-di-semua-kelompok

[3] W. Min, S. Jiang, L. Liu, Y. Rui, and R. Jain, “A survey on food computing,” *ACM Comput. Surv.*, vol. 52, no. 5, Sep. 2019, doi: 10.1145/3329168.

[4] P. J. Brady *et al.*, “A Qualitative Study of Factors Influencing Food Choices and Food Sources Among Adults Aged 50 Years and Older During the Coronavirus Disease 2019 Pandemic,” *J. Acad. Nutr. Diet.*, vol. 123, no. 4, pp. 602-613.e5, Apr. 2023, doi: 10.1016/j.jand.2022.08.131.

[5] C. Liu *et al.*, “A New Deep Learning-Based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure,” *IEEE Trans. Serv. Comput.*, vol. 11, no. 2, pp. 249–261, Mar. 2018, doi: 10.1109/TSC.2017.2662008.

[6] F. P. W. Lo, Y. Guo, Y. Sun, J. Qiu, and B. Lo, “An Intelligent Vision-Based Nutritional Assessment Method for Handheld Food Items,” *IEEE Trans. Multimed.*, vol. 25, pp. 5840–5851, 2023, doi: 10.1109/TMM.2022.3199911.

[7] W. Min *et al.*, “Large Scale Visual Food Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, 2023, doi: 10.1109/TPAMI.

[8] G. Waltner *et al.*, “Personalized Dietary Self-Management Using Mobile Vision-Based Assistance,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2017, pp. 385–393. doi: 10.1007/978-3-319-70742-6_36.

[9] A. Myers *et al.*, “Im2Calories: Towards an Automated Mobile Vision

Food Diary," in *2015 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Dec. 2015, pp. 1233–1241. doi: 10.1109/ICCV.2015.146.

[10] Q. Thames *et al.*, "Nutrition5k: Towards Automatic Nutritional Understanding of Generic Food," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, 2021, pp. 8899–8907. doi: 10.1109/CVPR46437.2021.00879.

[11] Y. Lu, T. Stathopoulou, M. F. Vasiloglou, S. Christodoulidis, Z. Stanga, and S. Mougiakakou, "An Artificial Intelligence-Based System to Assess Nutrient Intake for Hospitalised Patients," *IEEE Trans. Multimed.*, vol. 23, pp. 1136–1147, 2021, doi: 10.1109/TMM.2020.2993948.

[12] X. J. Zhang, Y. F. Lu, and S. H. Zhang, "Multi-Task Learning for Food Identification and Analysis with Deep Convolutional Neural Networks," *J. Comput. Sci. Technol.*, vol. 31, no. 3, pp. 489–500, May 2016, doi: 10.1007/s11390-016-1642-6.

[13] S. Mezgec and B. K. Seljak, "Nutrinet: A deep learning food and drink image recognition system for dietary assessment," *Nutrients*, vol. 9, no. 7, Jul. 2017, doi: 10.3390/nu9070657.

[14] G. Suddul and J. F. L. Seguin, "A comparative study of deep learning methods for food classification with images," *Food Humanit.*, vol. 1, no. July, pp. 800–808, 2023, doi: 10.1016/j.foohum.2023.07.018.

[15] A. Wibisono, H. A. Wisesa, Z. P. Rahmadhani, P. K. Fahira, P. Mursanto, and W. Jatmiko, "Traditional food knowledge of Indonesia: a new high-quality food dataset and automatic recognition system," *J. Big Data*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40537-020-00342-5.

[16] F. P. W. Lo, Y. Sun, J. Qiu, and B. Lo, "Image-Based Food Classification and Volume Estimation for Dietary Assessment: A Review," *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 7, pp. 1926–1939, 2020, doi: 10.1109/JBHI.2020.2987943.

[17] J. He and F. Zhu, "Online Continual Learning for Visual Food Classification," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2021-Octob, no. 1, pp. 2337–2346, 2021, doi: 10.1109/ICCVW54120.2021.00265.

[18] H. Hoashi, T. Joutou, and K. Yanai, "Image recognition of 85 food categories by feature fusion," in *IEEE International Symposium on Multimedia, ISM 2010*, 2010, pp. 296–301. doi: 10.1109/ISM.2010.51.

[19] and S. B. Giovanni Maria Farinella, Dario Allegra, Filippo Stanco, *On the Exploitation of One Class Classification to Distinguish Food Vs Non-Food Images*, vol. 9281. in Lecture Notes in Computer Science, vol. 9281. Springer International Publishing, 2015. doi: 10.1007/978-3-319-23222-5.

[20] S. Inunganbi, A. Seal, and P. Khanna, "Classification of Food Images through Interactive Image Segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2018, pp. 519–528. doi: 10.1007/978-3-319-75420-8_49.

[21] N. D. Martinez-Lara, C. L. Garzon-Castro, and A. Filomena-Ambrosio, "Nu-Support Vector Classification Training for Feature Identification in 'Arepas': A Colombian Traditional Food," in *13th International Symposium on Advanced Topics in Electrical Engineering, ATEE 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ATEE58038.2023.10108229.

[22] H. Yang, D. Zhang, D. J. Lee, and M. Huang, "A sparse representation based classification algorithm for Chinese food recognition," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2016, pp. 3–10. doi: 10.1007/978-3-319-50832-0_1.

[23] K. Aizawa, Y. Maruyama, H. Li, C. Morikawa, and G. C. De Silva, "Food balance estimation by using personal dietary tendencies in a multimedia food log," *IEEE Trans. Multimed.*, vol. 15, no. 8, pp. 2176–2185, 2013, doi: 10.1109/TMM.2013.2271474.

[24] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.

[25] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: Theory and practice," *Int. J. Comput. Vis.*, vol. 105, no. 3, pp. 222–245, 2013, doi: 10.1007/s11263-013-0636-x.

[26] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, pp. 2169–2178, 2006, doi: 10.1109/CVPR.2006.68.

[27] M. M. Anthimopoulos, L. Gianola, L. Scarnato, P. Diem, and S. G. Mougiakakou, "A food recognition system for diabetic patients based on an optimized bag-of-features model," *IEEE J. Biomed. Heal. Informatics*, vol. 18, no. 4, pp. 1261–1271, 2014, doi: 10.1109/JBHI.2014.2308928.

[28] F. Moosmann, E. Nowak, and F. Jurie, "Randomized clustering forests for image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1632–1646, 2008, doi: 10.1109/TPAMI.2007.70822.

[29] F. Kong and J. Tan, "DietCam: Automatic dietary assessment with mobile camera phones," *Pervasive Mob. Comput.*, vol. 8, no. 1, pp. 147–163, 2012, doi: 10.1016/j.pmcj.2011.07.003.

[30] F. Kong and J. Tan, "DietCam: Regular shape food recognition with a camera phone," *Proc. - 2011 Int. Conf. Body Sens. Networks, BSN 2011*, pp. 127–132, 2011, doi: 10.1109/BSN.2011.19.

[31] S. Singh, A. Gupta, and A. A. Efros, "Unsupervised discovery of mid-level discriminative patches," *Comput. Vision-European Conf. Comput. Vis.*, vol. 7573 LNCS, no. PART 2, pp. 73–86, 2012, doi: 10.1007/978-3-642-33709-3_6.

[32] L. Xiao, T. Lan, D. Xu, W. Gao, and C. Li, "A Simplified CNNs Visual Perception Learning Network Algorithm for Foods Recognition," *Comput. Electr. Eng.*, vol. 92, Jun. 2021, doi: 10.1016/j.compeleceng.2021.107152.

[33] M. Chun, H. Jeong, H. Lee, T. Yoo, and H. Jung, "Development of Korean Food Image Classification Model Using Public Food Image Dataset and Deep Learning Methods," *IEEE Access*, vol. 10, pp. 128732–128741, 2022, doi: 10.1109/ACCESS.2022.3227796.

[34] L. Jiang, B. Qiu, X. Liu, C. Huang, and K. Lin, "DeepFood: Food Image Analysis and Dietary Assessment via Deep Model," *IEEE Access*, vol. 8, pp. 47477–47489, 2020, doi: 10.1109/ACCESS.2020.2973625.

[35] F. Zhou and Y. Lin, "Fine-grained image classification by exploring bipartite-graph labels," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2016, pp. 1124–1133. doi: 10.1109/CVPR.2016.127.

[36] L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101 - Mining discriminative components with random forests," *Proceeding Eur. Conf. Comput. Vis.*, vol. 8694 LNCS, no. PART 6, pp. 446–461, 2014, doi: 10.1007/978-3-319-10599-4_29.

[37] H. Kagaya, K. Aizawa, and M. Ogawa, "Food detection and recognition using convolutional neural network," *Proc. 2014 ACM Conf. Multimed.*, no. 3, pp. 1085–1088, 2014, doi: 10.1145/2647868.2654970.

[38] S. Christodoulidis, M. Anthimopoulos, and S. Mougiakakou, "Food recognition for dietary assessment using deep convolutional neural networks," *Int. Conf. Image Anal. Process.*, vol. 9281, pp. 458–465, 2015, doi: 10.1007/978-3-319-23222-5_56.

[39] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *Int. Conf. Learn. Represent.*, Sep. 2015, [Online]. Available: http://arxiv.org/abs/1409.1556

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[41] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[42] Faldo Fajri Afrinanto, "Padang Cuisine (Indonesian Food Image Dataset)," Kaggle. [Online]. Available: https://www.kaggle.com/dsv/4053613

[43] D. Xiang, H. Wang, D. He, and C. Zhai, "Research on Histogram Equalization Algorithm Based on Optimized Adaptive Quadruple Segmentation and Cropping of Underwater Image (AQSCHE)," *IEEE Access*, vol. 11, pp. 69356–69365, 2023, doi: 10.1109/ACCESS.2023.3290201.