

# Analysis of the Impact of Data Augmentation on the Performance of Deep Learning Models in Multispectral Food Authenticity Identification

Yaru Zhang, Arif Yilmaz, Mirela Popa and Christopher Brewster

Department of Advanced Computing Sciences,

Faculty of Science and Engineering,

Maastricht University, 6200 MD, The Netherlands,

Email: {yaru.zhang, a.yilmaz, mirela.popa, christopher.brewster}@maastrichtuniversity.nl

**Abstract**—Food authenticity is a significant concern in the meat industry, demanding effective detection methods. This study explores the use of multispectral imaging (MSI) and deep learning for meat adulteration detection. We evaluate different deep learning models using transfer learning and preprocessing techniques in a multi-level adulteration classification task. In addition, we propose a novel approach called one-band mixed augmentation for band selection in MSI data, which outperforms traditional reflectance-based feature selection and enhances model robustness. Furthermore, employing the nine-crop approach for dataset augmentation improved the accuracy from 0.63 to 0.74 for DenseNet201 model without transfer learning. This research contributes to advancing food safety assessment practices and provides insights into the application of deep learning for preventing food adulteration. The proposed one-band mixed augmentation approach offers a novel strategy for handling band selection challenges in MSI data analysis.

## I. INTRODUCTION

FOOD safety has become a major issue in recent years, garnering significant attention from regulators and industry stakeholders alike. This issue is particularly critical when it comes to minced meat, which lacks distinctive morphological characteristics, making it more susceptible to intentional adulteration. Such fraudulent practices not only pose serious health risks to consumers but also undermine the integrity of the entire food supply chain, eroding public trust in the food industry. Consequently, it is imperative to adopt proactive measures for detecting and preventing food adulteration, ensuring the delivery of safe, reliable, and high-quality food to consumers.

Traditional methods for detecting meat adulteration typically involve destructive sample analysis, such as PCR analysis [1], are time-consuming and require specialized environments and trained professionals. Consequently, the need for effective and efficient techniques to detect food adulteration has become increasingly urgent. To address this challenge, researchers have investigated the use of non-contact technologies to address meat safety concerns, including the detection of fraud in processed meat using non-destructive spectroscopic methods [2]. Additionally, gas sensors have been employed for monitoring meat quality [3], while electronic noses have been utilized for monitoring meat spoilage [4]. These advanced

technologies provide cost-effective and rapid alternatives to traditional methods, and their integration into food safety regulations reflects their increasing importance in ensuring the integrity of the food supply [5].

Multispectral imaging (MSI) has received considerable attention in recent years as a fast and non-destructive analytical approach to determining food quality and safety evaluation. MSI captures image data in specific wavelength ranges, providing spatial and spectral information of the object under analysis. Different meat qualities have different reflection intensities under different spectra [6], making it particularly useful for detecting food adulterants. It has been successfully used in various food safety applications, including the evaluation of microbial contamination in ready-to-eat vegetable salad [7], the assessment of cowpea seed health and differentiation of fungal species [8], and the evaluation of ready-to-eat pineapple quality [9]. Additionally, MSI has been used to estimate microbial spoilage in minced pork [10].

Machine learning models, such as partial least squares regression (PLSR) and support vector machines (SVM), have been applied to detect meat adulteration using MSI data [11]–[14]. However, current research on MSI for meat adulteration often utilize only limited attributes of the image data, such as mean and standard deviation, which can be a limitation in terms of accuracy and reliability. This limitation leaves room for improving the accuracy and reliability of meat quality control systems.

To address this limitation and further enhance meat quality control, the combination of MSI with deep learning techniques, such as convolutional neural networks (CNNs), has gained attention. We can improve the accuracy and reliability of meat quality control systems by using these models for image classification tasks in the food domain, specifically for meat adulteration. Although some research has explored the use of CNNs for image classification [15], there is still a gap in the literature on using these models for meat quality control systems.

A previous study developed a framework for coffee maturity classification with 15 bands of multispectral data based on CNNs and achieved a relevant high accuracy on five

classes [16], achieved up to 98% accuracy on the dataset and 100% accuracy on cross-validation. However, this approach has not been applied to the problem of meat adulteration and spoilage in MSI data.

By leveraging the comprehensive information provided by MSI and harnessing the capabilities of deep learning, we seek to develop more effective methods for preventing and detecting meat adulteration. In order to facilitate the goals of our study, it was necessary to reprocess data specifically for meat adulteration, as there was no readily available MSI image dataset for this purpose. We also adapt state-of-the-art CNN models and perform several optimizations to ensure their effectiveness in analyzing the acquired MSI data. Additionally, we explore the potential of leveraging the rich information contained in MSI data by experimenting with different pre-processing approaches.

Through our research, we aim to provide valuable information on the utilization of MSI and deep learning techniques, which can lead to the development of advanced approaches to ensure food safety and preserve the integrity of the meat supply chain. The findings of our study hold great promise for substantial advancements in current practices, leading to the development of more efficient and dependable methods for preventing and detecting food adulteration. As a result, these advancements will play a crucial role in safeguarding the safety and integrity of the food supply chain.

## II. METHODOLOGY

In this study, we conducted experiments using 180 minced meat samples from 9 adulteration classes. We extracted multi spectral images in 18 bands and encoded and resized them into the required size by a deep learning model. We started with fine-tuning SOTA CNNs models to detect patterns and features indicative of meat adulteration, but the particularities of our image dataset indicated that training from scratch might be a better option for learning relevant features. The best-performing model was selected as the baseline. Additionally, we explored three different pre-processing modalities to assess their impact on the model's performance.

### A. Datasets

The data acquisition process followed the pipeline illustrated in Fig. 1. Our study utilized a dataset consisting of MSI images depicting chicken and pork meat samples with varying levels of adulteration. The levels of adulteration spanned from 0% (indicating pure chicken) to 100% (representing pure pork), with nine intervals in between: 0%, 10%, 25%, 40%, 50%, 60%, 75%, 90%, and 100%. Chicken and pork were purchased from four different butcher shops ( $b_1, b_2, b_3, b_4$ ) in Greece. Samples from each butcher shop contained five instances per adulteration level resulting in 45 samples per butcher shop. In total, the dataset contains 180 samples from four butcher shops.

The images were acquired using the Videometer lab system developed by the Technical University of Denmark and commercialized by "Videometer A/S" (<http://www.videometer.com>).

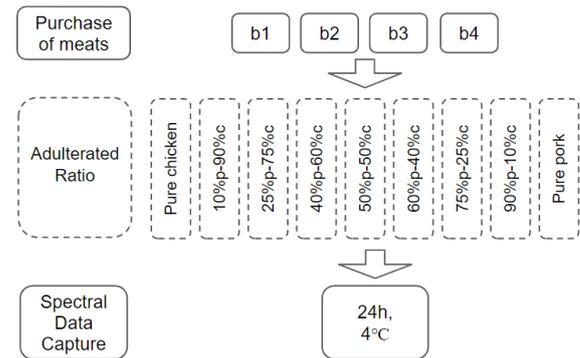


Fig. 1. Multispectral Imaging Acquisition pipeline

The MSI images consist of 18 bands for each meat sample. The samples were kept at 4°C and captured after 24 h. There are a total of 180 sample image files of various temperatures and various adulteration levels. The final size of the dataset is 3240 (180x18) grey-scale images from 180 meat samples. The dataset includes nine classes, each representing a different level of adulteration.

### B. Data Preprocessing

Proper data preprocessing is crucial as it is the foundation for subsequent data analysis. By performing appropriate data preparation techniques, we can guarantee that the analysis results are reliable and carry significant implications. Additionally, it allows us to address any potential issues or biases in the data and optimize the performance of our machine learning model.

TABLE I  
DETAILS OF IMAGE BANDS AND CORRESPONDING WAVELENGTHS USED BY VIDEOMETER LAB MSI CAMERA FOR CAPTURING MSI DATASET.

the band and wavelength details						
band	band1	band2	band3	band4	band5	band6
region	UV	Violet	Blue	Blue	Cyan	Green
wavelength	405nm	435nm	450nm	470nm	505nm	525nm
band	band7	band8	band9	band10	band11	band12
region	Green	Yellow	Red	Red	Red	Red
wavelength	570nm	590nm	630nm	645nm	660nm	700nm
band	band13	band14	band15	band16	band17	band18
region	NIR	NIR	NIR	NIR	NIR	NIR
wavelength	850nm	870nm	890nm	910nm	940nm	970nm

*a) Image Preprocessing:* Each sample in the dataset contains 18 grey-scale images of 18 non-uniformly distributed wavelengths with a size of 1200 by 1200 pixels. Each image represents a spectral feature of a sample in a particular band. Table I presents a detailed overview of the chosen wavelength bands utilized in our study. The selected wavelengths cover a spectrum ranging from 405nm to 970nm, comprising a total of 18 bands. Notably, this includes one band in the ultraviolet (UV) region and six bands within the near-infrared (NIR) region. This information provides a comprehensive understanding of the specific wavelengths employed in our

research analysis. Fig. 2 shows all 18 bands of a sample which is 10%pork-90%chicken.

In order to accommodate the large extracted images within the proposed models, we first resize the images to a standardized size of 224 by 224 pixels. This resizing ensures uniformity and compatibility across the dataset. Following the resizing step, we employ min-max scaling to encode the images as values ranging from 0 to 1. This preprocessing technique effectively normalizes the pixel values, allowing for efficient handling and analysis of the data

*b) Label Extraction:* The images were labeled based on the information contained within their names, including the adulteration level, band number, sample name, and storage condition. We converted it into integers ranging from 0 to 8. Specifically, '0' denotes pure chicken, while '8' represents pure pork. For adulteration levels between 10% to 90%, we assigned integer values from 1 to 7 to represent varying ratios of pork and chicken: 10% pork - 90% chicken, 25% pork - 75% chicken, 40% pork - 60% chicken, 50% pork - 50% chicken, 60% pork - 40% chicken, 75% pork - 25% chicken and 90% pork - 10% chicken. This scale indicates the percentage of pork and chicken present in each sample, irrespective of which meat has adulterated the other. In the case of pork-adulterated chicken, a smaller scale number denotes a higher level of adulteration. We chose to utilize a single scale to simplify the paper, instead of employing separate scales for each meat species.

While it is important to label each image accurately, we also wanted to ensure that the labels were practical for the intended application. In this case, we use one-hot encoding to encode the image labels into a numerical format, which assigns a unique numerical value to each category. This approach enabled us to quickly generate statistics and analyze the model performance based on different adulteration levels.

### C. Basic Adulteration Classification Pipeline

This study focuses on developing an automated method for detecting adulteration in meat samples using multispectral image analysis. The problem is approached as a classification task, where the performance of various deep-learning models is compared. To ensure consistency, we employ a standardized classification pipeline. This involves inputting an array with dimensions (224, 224, 18), representing the 18 bands of information in each sample, into the models. The models are then trained to predict the degree of adulteration based on the input image array.

To evaluate the effectiveness of different CNN-based models, we compare five models available in the Keras library [17]. The initial selection includes VGG16 and VGG19 [18], which serve as established benchmarks for image classification tasks. Additionally, Inception-ResNetV2 [19] and InceptionV3 [20] are chosen for their superior performance in computer vision tasks. Furthermore, we include DenseNet [21], known for its promising outcomes in similar studies.

To leverage pre-existing knowledge, transfer learning is applied to the selected CNN models. This allows us to explore

if pre-trained models can enhance the performance of our task. Fig. 3 illustrates the basic experiment pipeline of the CNN models, providing an overview of the process.

We evaluate the performance of our models using two methods. Firstly, we perform a simple train/test split by partitioning the data into two sets, with the training set containing 80% of the data and the test set containing 20%. Given the limited number of samples in our dataset, we also use 5-fold stratified cross-validation (SCV) to evaluate the models more precisely.

The models are trained using the backpropagation algorithm. This algorithm works by calculating the loss function gradient concerning the network weights and using this gradient to update the weights in a direction that minimizes the loss function. This process is repeated iteratively until the network converges on a set of weights that minimizes the loss function. The specific methods and hyperparameters depend on each model. We evaluate the models using accuracy, precision, recall, and F1 score. We chose these metrics to understand the models' performance thoroughly.

### D. Transfer Learning

To enhance the classification performance of our model, we opted to incorporate transfer learning into our training process and to evaluate its effectiveness. Given the limited nature of our dataset, transfer learning was considered a potential solution to optimize the model and improve its accuracy. The base model weights obtained from ImageNet [22], which consists of millions of images and 1000 labels, were used to initialize our model and provide a solid foundation for further training.

When dealing with datasets that contain more than three channels, such as our 18-channel multispectral data, transfer learning requires adjusting the pre-trained weights to accommodate the additional channels. Fine-tuning is performed on the first convolutional layer, configured to enable the neural network to read 18-channel images. In this process, the weights of the first convolutional layer are modified to accept the input of 18 channels. This is done by averaging the pre-trained weights of the first convolutional layer's three channels and replicating the resulting weights 18 times to accommodate all 18 input channels.

To align with our experiment's 9-class classification objective, we modified the fully connected output layer of all five models from 1000 to 9. Furthermore, to adapt transfer learning to our specific task, we made variations in the trainable layers of each utilized network. The trainable and non-trainable layers of the models are specified below.

- **VGG-16** consists of 16 trainable layers. While the first two layers remain fixed, the subsequent 14 layers are trainable. The original fully connected layers of 4096 and 4096 were replaced by adapted layers with sizes of 256 and 128, respectively.
- **VGG-19** comprises 19 trainable layers, where the first seven layers are untrained, while the remaining 12 layers are trainable. Similar modifications were made to the

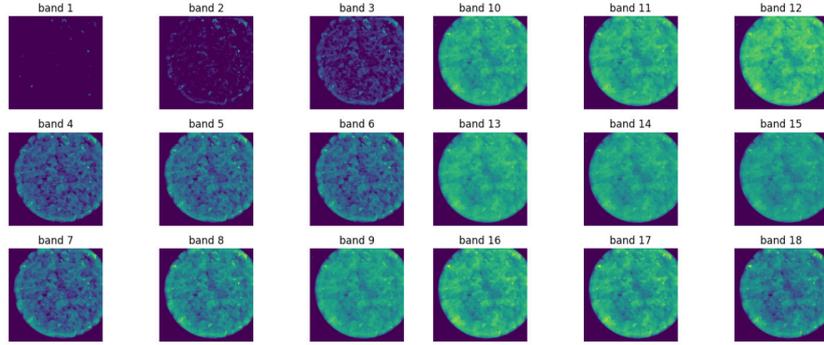


Fig. 2. The 18 bands for a 10% pork-90% chicken sample.

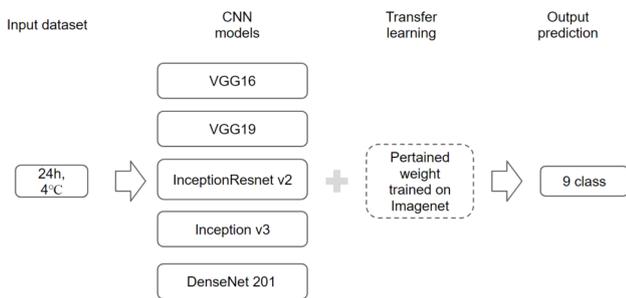


Fig. 3. Baseline CNN Classification Experiment Design

fully connected layers, adjusting their original sizes of 4096 and 4096 to sizes of 256 and 128.

- **InceptionV3** has 310 trainable layers. The first 150 layers are untrained, while the rest are made trainable. The original fully connected layer, initially sized of 2048, is replaced with two new fully connected layers sized of 256 and 128, respectively.
- **Inception-ResNetV2** The first 150 layers are untrained, while the remaining 578 are trainable. The original fully connected layers, sized of 1536, are substituted with two new fully connected layers sized of 256 and 128.
- **DenseNet201** has 706 trainable layers, with the first 150 layers left untrained and the remaining layers made trainable. Two new fully connected layers were introduced, having sizes of 256 and 128, respectively.

By adjusting the trainable layers in the specified way, the neural networks were fine-tuned to better suit our specific problem and data characteristics. The modified models were then used to conduct our experiments and to analyze their performance.

#### E. Hyperparameter Optimization

After experimenting with several optimizers, including Adam, Adamax, Adamgrad, and SGD, we selected the Adam optimizer for its superior performance. Then, we set the output layer with softmax activation function.

We utilized a learning rate scheduler function to optimize our model's performance. Our approach involved setting an initial learning rate of 0.0001 and employing an exponential decay function that reduced the learning rate by 0.095 at each epoch. The exponential decay scheme provided a smooth decay path, which was particularly effective during the initial stages of training. This strategy helped to improve the learning capacity of the model and yielded better results in our experiments.

#### F. Cross Validation

Cross-validation (CV) is a widely used technique in machine learning and data analysis to evaluate the predictive performance of models. It helps optimize hyperparameters, identify dataset issues, and prevent overfitting, ultimately improving the effectiveness of models in real-world applications. Stratified cross-validation is an essential variant of CV when working with imbalanced datasets. It ensures that each fold contains representative samples from all classes in the same proportion as the original dataset, mitigating the risk of biased assessments of model performance. By using stratified cross-validation, we can obtain more reliable estimates of the model's generalization capabilities and make better-informed decisions about its suitability for real-world applications.

For our experiments, we used a five-fold stratified cross-validation (5-SCV) approach. The data was divided into five folds, each containing an equal distribution of samples from all classes. The models were then trained on four folds and validated on the remaining one. This process was repeated five times, with each fold used as the validation set once. The stratified aspect of the cross-validation ensured that the class distribution was maintained across all folds, preventing bias in evaluating the models' performance. The final performance metrics were calculated as the average of the five iterations, providing a comprehensive and accurate assessment of the effectiveness of the models.

#### G. Model Improvements with Various Data Augmentation Configurations

To enhance the classification accuracy, we conducted various experiments on the dataset. We explored several tech-

niques, such as removing uninformative bands, augmenting the training set, and cropping the original image to gather more information. Given the amount of conditions and considered models, we chose to focus on one model (e.g. DenseNet 201 model). Stratified sampling was applied so that 80%-20% of the dataset to be used for training and testing for all experiments. Specifically, for the basic adulteration classification and uninformative bands excluded experiments, the training set consisted of 144 images with 36 testing images. For one band augmentation experiment, the training set was extended to 288 images, while the test set remained at 36 images. The purpose of this was to determine whether augmentation could help the model learn more details about the features needed to classify the original images, rather than the augmented images. In the cropping experiments, the training set is increased to 576 images (for 4 crops) and 1296 (for 9 crops), while the test set size remains constant for consistency reasons.

1) *Band Selection: Exclude Uninformative Bands using a Reflectance-based Method:* To further refine the dataset and improve the classification performance of CNN models, we conducted a band exclusion approach based on previous research by L.-C. Fengou, P. Tsakanikas, and G.-J. E. Nychas [23]. This previous study compared the mean and standard deviation of the wavelength reflectance of pure chicken and pure pork at different storage times (0 h, 24 h, 48 h) and identified the wavelength from 700 to 940 nm as uninformative, which corresponds to bands 12 to 17 in our dataset. This exclusion was based on the overlap of wavelength reflectance. It is worth noting that the band exclusion approach has improved classification accuracy in previous studies on similar datasets. By excluding uninformative bands, the amount of noise in the data was reduced, and the signal-to-noise ratio was increased, factors which could improve the performance of the models.

To evaluate the impact of the band exclusion on classification performance, we trained the CNN models on a 12-band dataset, where the uninformative bands were excluded. We selected DenseNet201 as our base model without transfer learning and trained it on the 9-class classification task.

2) *Optimizing Band Selection: Exclude Uninformative Bands using One-band Mixed Augmentation:* Data augmentation is a widespread technique in deep learning used to increase the size and diversity of the training dataset. In our experiments, we applied mixed augmentation to enhance the diversity and size of our training dataset. This technique involves applying various transformations, such as zooming, rotating, shifting, and flipping, to the original data to create new and unique images while preserving the properties of the original data. Zooming allowed us to change the scale of the images, while rotation and shearing enabled us to modify the orientation and shape of the objects within the images. The width and height shifting helped to translate the objects in the images, while the horizontal flipping created a mirror image of the original one. The mixed augmentation method selects a random combination of transformations from the set specified in the augmentation pipeline. The chosen transformations include zooming between 40% and 80% of

the original size, rotating the image up to 45 degrees, shifting the width and height of the image by up to 10%, shearing the image up to 20%, and flipping the image horizontally.

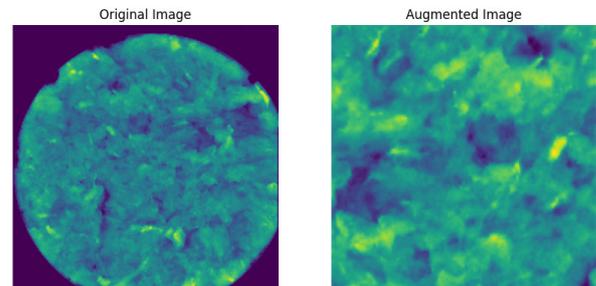


Fig. 4. Example of Augmentation

Fig. 4 shows one augmentation of one 25% pork-75% chicken sample in the dataset. The image underwent several augmentations. First, it was rotated at an angle of  $-34.97$  degrees in a counterclockwise direction. Second, it was translated horizontally by  $-0.075$  and vertically by  $11.85$  pixels. Third, it was sheared by  $-0.175$ , meaning that the object's shape in the image was distorted. Fourthly, the image was zoomed in by a factor of  $0.74$  along the x-axis and  $0.50$  along the y-axis. Finally, the image was flipped horizontally.

Multispectral images contain several bands of information, each of which may have varying contributions to the classification performance. In order to determine which bands are more informative for our models, we conduct data selection experiments. To do this, we first apply data augmentation techniques individually to each band in the dataset. Then, we combine the augmented bands and train models on each combination of bands. In this experiment, data augmentation was performed for each band in the dataset, by splitting the data into 144 samples for training and 36 for testing. We applied the described band augmentation process to the training set, which increased the training set size to 288 (for each sample, one band was augmented). Each sample had a shape of  $(224, 224, 18)$ , where 224 represents the width and height of the image, and 18 represents the number of bands. Consequently, the final input training shape became  $(288, 224, 224, 18)$ . Fig. 5 shows the augmentation pipeline of our data.

We evaluate the performance of each model and compare the results to a baseline model (DenseNet 201 without transfer learning) without augmentation. By comparing the performance of all combinations, we can identify the bands that have performed above the baseline and are therefore considered informative. This process allows us to identify the bands that provide the most useful information for classification and can help optimize the selection of bands for future experiments. Furthermore, this approach can be used to investigate the impact of data augmentation on individual bands and could help us understand the effect of each augmentation technique on the overall classification performance.

3) *Augmentation by Applying Cropping to All Bands:* To strike a balance between the amount of information conveyed

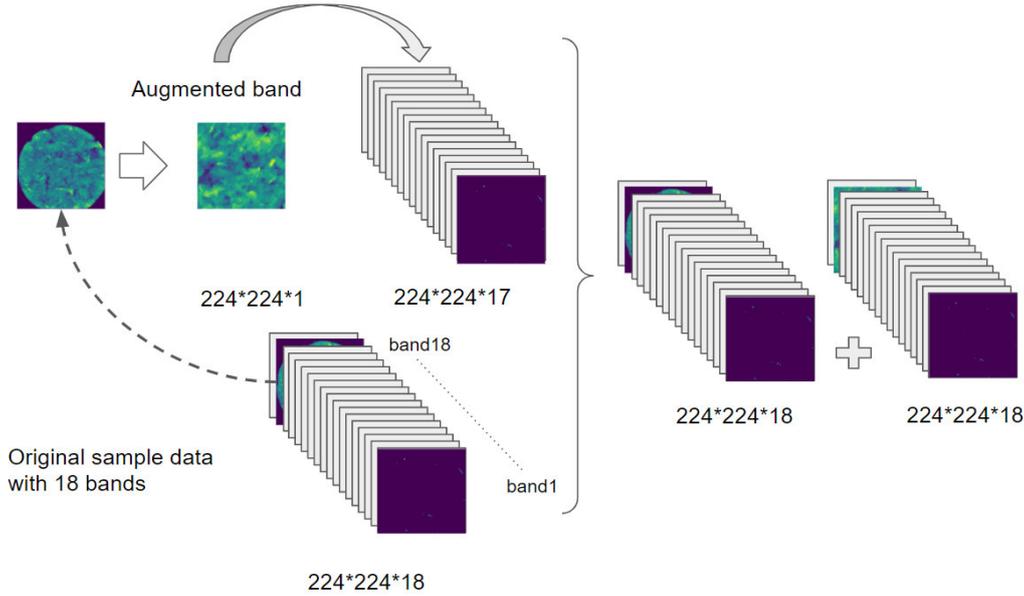


Fig. 5. Pipeline of One-band mixed Augmentation. One band of the original samples is augmented, replacing the original image. The augmented samples are combined with the original samples to form the training set.

by an image and the computational cost required for processing it, image cropping is a beneficial preprocessing technique. By cropping larger images into smaller clips, we can increase the dataset size without sacrificing crucial information, especially when obtaining additional datasets in the same field is challenging.

For the baseline classification, we resized the original images, which had dimensions of (1200 x 1200) pixels, to (224 x 224) pixels. However, this resizing process may result in a loss of information and potentially impact the accuracy of the classification results. To address this challenge, we propose an approach that involves cropping the raw images into four or nine clips. Through experimentation, we determined that four clips, each measuring (600 x 600) pixels, or nine clips, each measuring (400 x 400) pixels, were the most suitable sizes.

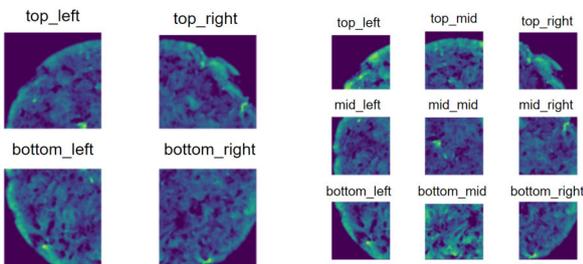


Fig. 6. Example of Image Cropping preprocessing. This image is randomly chosen from the MSI dataset and shows the 4-cropped and 9-cropped versions of the original image.

To maintain the spatial relationship between each cropped image and its original location, we incorporated position information into the extracted CSV file, alongside the corre-

sponding label. This facilitated a clear understanding of the relative location of each cropped clip throughout the data preprocessing pipeline. For the four-cropped approach, we used position information such as lt (top left), lr (bottom left), lb (bottom right), and rb (top right). For the nine-cropped approach, position information included lt (top-left), mt (top-middle), rt (top-right), lm (middle-left), mm (middle-middle), rm (middle-right), lb (bottom-left), mb (bottom-middle), and rb (bottom-right). An example of the cropped images is shown in Fig. 6.

This approach effectively increases the dataset size while preserving the essential information from the original images. Additionally, the inclusion of position information provides valuable context for interpreting and analyzing the cropped clips. Overall, cropping and incorporating position information are effective preprocessing techniques for MSI data, enhancing the analysis quality and facilitating the utilization of these data in machine learning models. In our study, we performed cropping on the original image data, generating four and nine crops, respectively.

### III. RESULTS

In this study, we present a comprehensive evaluation of the efficacy of various deep learning models.

We used well-known CNN models including VGG16 and VGG19, Inception-Resnet v2, Inception v3, and DenseNet201 to explore their performance for 9-class adulteration classification task. To further optimize model performance, we leveraged transfer learning techniques.

Furthermore, we experimented with various data augmentation configurations such as rotation, shifting, shearing, flipping and zooming.

We also explored the impact of band selection on model performance, including the exclusion of non-informative bands based on reflectance and augmentation experiments in this study.

By evaluating the performance of various models, our objective was to provide insights into the selection of the most appropriate deep-learning architecture and preprocessing techniques for meat adulteration detection.

#### A. Best-performing model identification for meat adulteration classification

Table II shows the performance of different deep learning architectures on a 9-class classification task with and without transfer learning. The details of each model is explained in II-D. The results showed that DenseNet201 without transfer learning achieved the best accuracy of 0.63 and a precision of 0.64, while DenseNet201 with transfer learning achieved the best accuracy of 0.62 and the precision of 0.61 on all data sets and combinations.

#### B. Evaluation of baseline model performance for various data augmentation configurations

As DenseNet201 was the best-performing model, it was selected as a baseline for later experiments. The details of data augmentation are explained in section II-G2. The experiments focus on investigating the effects of band selection and augmentation techniques on the model performance. The results of these experiments are summarized in Table III, demonstrating that excluding non-informative bands improves model accuracy. Furthermore, excluding uninformative bands based on augmentation is shown to enhance model performance, with varying influences observed for different bands. Additionally, the use of all-band cropping augmentation, particularly employing the nine-crop approach, leads to the best results. These findings highlight the significance of band selection and augmentation methods in improving model performance for multispectral imaging data, contributing valuable insights to the field.

1) *Band Selection: Exclude Uninformative Bands using a Reflectance-based Method:* In this experiment, we trained the baseline model on a 12-band dataset for the 9-class classification task. The bands were selected based on the mean and standard deviation of reflectance for pure classes. The results showed that using 12 bands outperformed models using 18 bands in all metrics, which achieved an accuracy of 0.69 while using 18 bands achieved an accuracy of 0.63.

2) *Optimizing Band Selection: Exclude Uninformative Bands using One-band Mixed Augmentation:* A mixed band augmentation which is detailed in section II-G2 was applied to the baseline model. Table.IV shows the mean performance of the model under varying band augmentations. Our experimental findings indicate that preprocessing the input images with different band augmentations has a considerable impact on the model's learning capacity. In order to balance computational cost and experimental accuracy, we employed two different random seeds for conducting the experiments.

The average values of the evaluation metrics are presented as the experimental results. In particular, our comparison of the results with the best baseline model introduced in Table II (achieving 0.63 accuracy for the 9-class DenseNet model without augmentation) led to a decrease in performance, including bands: 1, 2, 5, 8, 12, 13, 15 and 16. In contrast, we found that some bands, such as bands: 3, 4, 6, 7, 9, 10, 14, 17, and 18, improved the model's performance.

Fig.7 shows that band 4 (470 nm, blue), band 14 (870 nm, NIR), and band 17 (940 nm, NIR) are the top 3 most informative bands for the dataset using the DenseNet201 model (without transfer learning). The augmentation on band 17 (940nm, NIR) increases the accuracy from 0.72 to 0.81. Therefore, the choice of performing band preprocessing is critical in optimizing the model's accuracy for this classification task.

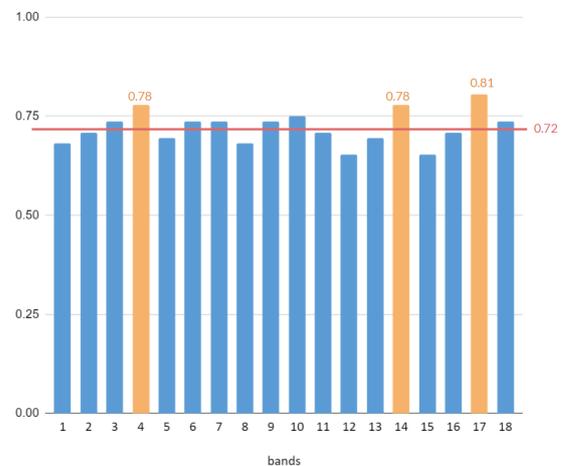


Fig. 7. Comparison of accuracies for various band augmentations on the dataset. The orange columns indicates the top 3 performing augmentation bands, the red line represents the baseline accuracy of the 9-class experiment without augmentation.

Based on the accuracies presented in Table IV, the six lowest performing bands (band1, band5, band8, band12, band13, and band15) were removed. The remaining 12 bands were stacked. The training process for this experiment adhered to the same settings as described in the baseline experiment. The 5-fold cross-validation produced average performance metrics, achieving 0.72 accuracy and 0.72 F1 score.

3) *Augmentation by Applying Cropping to All Bands:* In order to address the limitations posed by the limited size of the MSI dataset, we used the 4-crop and 9-crop approach to augment the entire 9-class data. Specifically, for the 4-cropped datasets, each class comprised 80 samples, and we used 80% of the dataset for training and 20% for testing, by making sure that none of the cropped versions of the original image would be found in both sets. We trained the baseline model in all these experiments. For the 4-cropped datasets, our experiment achieved 0.71 accuracy for 5-fold stratified cross validation. For the 9-cropped datasets, each class had 180 samples, resulting after augmentation in a total of 1620

TABLE II  
PERFORMANCE OF VARIOUS MODELS WITH 5 FOLDS CROSS-VALIDATION WITHOUT AND WITH TRANSFER LEARNING USING 18-BAND 9-CLASS MSI DATA.

CNN models	Without transfer learning				With transfer learning			
	Accuracy( $\pm$ SD)	F1 score( $\pm$ SD)	Recall( $\pm$ SD)	Precision( $\pm$ SD)	Accuracy( $\pm$ SD)	F1 score( $\pm$ SD)	Recall( $\pm$ SD)	Precision( $\pm$ SD)
VGG16	0.40 (+/- 0.16)	0.38 (+/- 0.19)	0.40 (+/- 0.16)	0.49 (+/- 0.22)	0.53 (+/- 0.09)	0.52 (+/- 0.10)	0.53 (+/- 0.09)	0.60 (+/- 0.14)
VGG19	0.47 (+/- 0.13)	0.46 (+/- 0.16)	0.47 (+/- 0.13)	0.53 (+/- 0.22)	0.52 (+/- 0.12)	0.52 (+/- 0.12)	0.52 (+/- 0.12)	0.63 (+/- 0.09)
Incep-Res v2	0.61 (+/- 0.07)	0.60 (+/- 0.09)	0.61 (+/- 0.07)	0.69 (+/- 0.08)	0.55 (+/- 0.08)	0.56 (+/- 0.07)	0.55 (+/- 0.08)	0.72 (+/- 0.03)
Inceptionv3	0.56 (+/- 0.07)	0.54 (+/- 0.09)	0.56 (+/- 0.07)	0.62 (+/- 0.12)	0.57 (+/- 0.06)	0.57 (+/- 0.06)	0.57 (+/- 0.06)	0.63 (+/- 0.04)
DenseNet201	<b>0.63 (+/- 0.12)</b>	<b>0.64 (+/- 0.12)</b>	<b>0.63 (+/- 0.12)</b>	<b>0.75 (+/- 0.09)</b>	<b>0.62 (+/- 0.07)</b>	<b>0.61 (+/- 0.08)</b>	<b>0.62 (+/- 0.07)</b>	<b>0.70 (+/- 0.10)</b>

TABLE III  
PERFORMANCE COMPARISON OF DATA CONFIGURATIONS WITH 5-FOLD STRATIFIED CROSS-VALIDATION

	Accuracy( $\pm$ SD)	F1 score( $\pm$ SD)	Recall( $\pm$ SD)	Precision( $\pm$ SD)
All Bands with Baseline (18 Bands)	0.63 (+/- 0.12)	0.64 (+/- 0.12)	0.63 (+/- 0.12)	0.75 (+/- 0.09)
Uninformative Bands Excluded based on Reflectance	0.69 (+/- 0.06)	0.68 (+/- 0.04)	0.69 (+/- 0.06)	0.76 (+/- 0.01)
Uninformative Bands Excluded based on Augmentation	0.72 (+/- 0.08)	0.72 (+/- 0.08)	0.72 (+/- 0.08)	0.78 (+/- 0.08)
All Bands, Image Cropped to 4 Parts	0.71 (+/- 0.11)	0.71 (+/- 0.13)	0.72 (+/- 0.11)	0.74 (+/- 0.10)
All Bands, Image Cropped to 9 Parts	<b>0.74 (+/- 0.08)</b>	<b>0.73 (+/- 0.09)</b>	<b>0.74 (+/- 0.08)</b>	<b>0.79 (+/- 0.10)</b>

TABLE IV  
PERFORMANCE OF THE BASELINE MODEL BASED ON ONE-BAND MIXED AUGMENTATION

	band 1	band 2	band 3	band 4	band 5	band 6	band 7	band 8	band 9
Accuracy	0.68	0.71	0.74	<b>0.78</b>	0.69	0.74	0.74	0.68	0.74
F1 score	0.69	0.70	0.76	<b>0.80</b>	0.75	0.76	0.74	0.71	0.77
Recall	0.74	0.75	0.81	<b>0.83</b>	0.78	0.79	0.78	0.76	0.84
Precision	0.71	0.71	0.82	<b>0.84</b>	0.80	0.78	0.74	0.74	0.81

	band 10	band 11	band 12	band 13	band 14	band 15	band 16	band 17	band 18
Accuracy	0.75	0.71	0.65	0.69	<b>0.78</b>	0.65	0.71	<b>0.81</b>	0.74
F1 score	0.76	0.73	0.68	0.74	<b>0.81</b>	0.68	0.73	<b>0.82</b>	0.76
Recall	0.79	0.78	0.72	0.80	<b>0.87</b>	0.72	0.77	<b>0.85</b>	0.80
Precision	0.77	0.75	0.70	0.78	<b>0.83</b>	0.71	0.76	<b>0.82</b>	0.80

samples. We used the same train-test split procedure as above and achieved improved results, with an accuracy of 0.78. The confusion matrices of the three different inputs are visualized in Fig.8. 5-fold stratified cross validation is applied on the 9-crop case, achieving an average accuracy of 0.74.

#### IV. DISCUSSION

In this study, we examine the effects of various augmentations on the MSI dataset, using CNN based deep learning models for meat adulteration detection. We also inspect the effect of transfer learning and data preprocessing on their performance. The best configuration for the 18- Band, 9 class classification is found as DenseNet201 without transfer learning with an accuracy of 0.63 and F1 score of 0.64.

First we evaluated the performance of different CNN architectures on 9-class classification tasks with and without transfer learning. Our results showed that in the 9-class classification, DenseNet201 achieved the best accuracy with and without transfer learning.

Our findings suggest that the performance of CNN architectures can be influenced by their nature and design. For example, DenseNet201 is composed of densely connected layers, where each layer within a block receives the outputs from all preceding layers within the same block. This architecture promotes feature reuse and information flow, mitigating the vanishing gradient problem.

The number of trainable layers is important because it affects the depth and complexity of the network. A deeper

network with more trainable layers has the potential to learn more complex features and patterns in the data. This may explain why Inception-Resnet v2 and DenseNet201 outperformed Inception v3 and VGG architectures in our study. The higher number of trainable layers in these architectures allows them to capture more intricate and nuanced information in the data, leading to improved performance in the classification tasks.

The achieved results indicated that transfer learning did not lead to a significant improvement in performance. This is consistent with the findings of a previous work[24], they suggested that models trained from scratch can perform just as well as those that are pre-trained, even with substantially less data.

In addition to model architecture, the data configuration was found to have essential impact on the model performance. The experiment excluding uninformative bands chosen by reflectance revealed that the reflectance-based method improves classification performance. As shown in Fig.9, the reflectance of different adulterated samples in the dataset experienced slight changes as the adulteration level increased. By removing non-informative bands (700 to 970 nm), the model achieved better performance using only 12 bands compared to using all 18 bands for the MSI dataset.

Additionally, the experiment demonstrated that the performance improvement achieved by using 12 bands was consistent across different folds of the cross-validation, as indicated by the low standard deviation of the metrics. This consistency

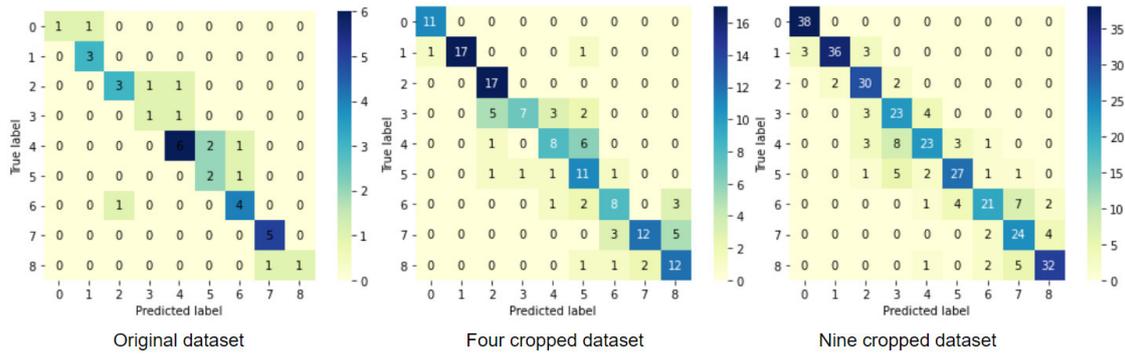


Fig. 8. Comparison of Confusion Matrices for Original, 4-Cropped, and 9-Cropped Images for the baseline model.

suggests that the exclusion of non-informative bands enhances the model’s performance, and the results are not dependent on a specific fold.

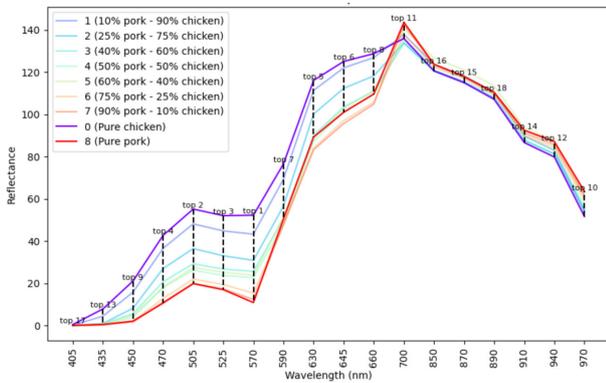


Fig. 9. Mean reflectance of samples of different adulteration level.

In the experiment on optimizing band selection using one-band mixed augmentation, the impact of different bands on the DenseNet201 model’s performance without transfer learning was investigated. The results revealed that the choice of bands for augmentation significantly influenced the model’s learning capabilities. Specifically, band 4 (470 nm, blue), band 14 (870 nm, NIR), and band 17 (940 nm, NIR) were identified as the most informative bands for the MSI dataset on the DenseNet201 model. Augmenting band 17 further improved the model’s accuracy from 0.72 to 0.81, highlighting the importance of band selection and preprocessing in optimizing performance.

Comparing the two methods, one-band mixed augmentation proved to be a better approach for band selection compared to reflectance-based feature selection. Although the performance difference between the two methods was insignificant, the augmented-based 12-band approach slightly outperformed the reflectance-based 12-band approach for the MSI dataset. This finding suggests that one-band mixed augmentation enables a more comprehensive exploration of the feature space, leading to a more robust model.

Both experiments show promising results for selecting informative bands in multispectral image classification tasks. The reflectance-based method provides a straightforward and intuitive approach, while one-band mixed augmentation allows for a more exploratory analysis, potentially uncovering new features beyond spectral characteristics alone. Future research could explore combining these two approaches to leverage their respective advantages and further enhance classification performance.

The size of an original image file, amounting to 103 MB, is a pertinent consideration in the context of the present study, which seeks to identify and analyze meat adulteration in minced chicken-pork samples. Moreover, the extraction of a (224,224,18) numpy array from the original file raises concerns about the optimal utilization of the information contained within the multispectral data. To overcome these limitations, another experiment applied cropping based preprocessing to the original (1200 x 1200) pixels image.

By employing the four-crop and nine-crop approach to augment the entire dataset for the 9-class case, significant improvements were observed. The results indicate that the nine-cropped dataset achieved the highest accuracy of 0.74, outperforming both the uncropped and four-cropped datasets. These findings show the potential of crop augmentation as an effective approach to address the challenges posed by limited dataset sizes and for maximizing the utilization of multispectral data in classification tasks.

### V. CONCLUSION

Our study highlights the potential of CNN models for detecting adulteration in minced meat samples. Among the evaluated models, DenseNet performed the best, showcasing its suitability for this task. We found that transfer learning did not significantly enhance model performance. Preprocessing data augmentation techniques, particularly our proposed one-band mixed augmentation approach, proved crucial in improving the model accuracy. Although our study had limitations, such as a small dataset and focus on a specific type of adulteration, it lays the groundwork for future research in this area. Further exploration of larger datasets and integration of

additional data types are encouraged to advance food safety practices.

## REFERENCES

- [1] C. Yang, G. Zhong, S. Zhou, Y. Guo, D. Pan, S. Wang, Q. Liu, Q. Xia, and Z. Cai, "Detection and characterization of meat adulteration in various types of meat products by using a high-efficiency multiplex polymerase chain reaction technique," *Frontiers in Nutrition*, 2022. doi: 10.3389/fnut.2022.979977
- [2] K. Edwards, M. Manley, L. C. Hoffman, and P. J. Williams, "Non-destructive spectroscopic and imaging techniques for the detection of processed meat fraud," *Foods*, vol. 10, no. 2, p. 448, 2021. doi: 10.1016/j.afres.2021.08.036
- [3] P. F. Pereira, P. H. de Sousa Picciani, V. Calado, and R. V. Tonon, "Electrical gas sensors for meat freshness assessment and quality monitoring: A review," *Trends in Food Science & Technology*, vol. 118, pp. 36–44, 2021. doi: 10.1016/j.tifs.2021.08.036
- [4] R. S. Andre, M. H. Fature, L. A. Mercante, and D. S. Correa, "Electronic nose based on hybrid free-standing nanofibrous mats for meat spoilage monitoring," *Sensors and Actuators B: Chemical*, vol. 353, p. 131114, 2022. doi: 10.1016/j.snb.2021.131114
- [5] H. J. Marvin, E. M. Janssen, Y. Bouzembrak, P. J. Hendriksen, and M. Staats, "Big data in food safety: An overview," *Critical reviews in food science and nutrition*, pp. 2286–2295, 2017. doi: 10.1080/10408398.2016.1257481
- [6] M. Peyvaste, A. Popov, A. Bykov, and I. Meglinski, "Meat freshness revealed by visible to near-infrared spectroscopy and principal component analysis," *Journal of Physics Communications*, 2020. doi: 10.1088/2399-6528/abb322
- [7] P. Tsakanikas, L.-C. Fengou, E. Manthou, A. Lianou, E. Z. Panagou, and G.-J. E. Nychas, "A unified spectra analysis workflow for the assessment of microbial contamination of ready-to-eat green salads: Comparative study and application of non-invasive sensors," *Computers and electronics in agriculture*, vol. 155, pp. 212–219, 2018. doi: 10.1016/j.compag.2018.10.025
- [8] C. H. Q. Rego, F. França-Silva, F. G. Gomes-Junior, M. H. D. d. Moraes, A. D. d. Medeiros, and C. B. d. Silva, "Using multispectral imaging for detecting seed-borne fungi in cowpea," *Agriculture*, vol. 10, no. 8, p. 361, 2020. doi: 10.3390/agriculture10080361
- [9] S. Younas, Y. Mao, C. Liu, M. A. Murtaza, Z. Ali, L. Wei, W. Liu, and L. Zheng, "Measurement of water fractions in freeze-dried shiitake mushroom by means of multispectral imaging (msi) and low-field nuclear magnetic resonance (lf-nmr)," *Journal of Food Composition and Analysis*, vol. 96, p. 103694, 2021. doi: 10.1016/j.jfca.2020.103694
- [10] L.-C. Fengou, E. Spyrelli, A. Lianou, P. Tsakanikas, E. Z. Panagou, and G.-J. E. Nychas, "Estimation of minced pork microbiological spoilage through fourier transform infrared and visible spectroscopy and multispectral vision technology," *Foods*, vol. 8, no. 7, p. 238, 2019. doi: 10.3390/FOODS8070238
- [11] M. Kamruzzaman, Y. Makino, and S. Oshita, "Rapid and non-destructive detection of chicken adulteration in minced beef using visible near-infrared hyperspectral imaging and machine learning," *Journal of Food Engineering*, vol. 170, pp. 8–15, 2016. doi: 10.1016/j.jfoodeng.2015.08.023
- [12] A. I. Ropodi, E. Z. Panagou, and G.-J. E. Nychas, "Multispectral imaging (msi): A promising method for the detection of minced beef adulteration with horsemeat," *Food Control*, vol. 73, pp. 57–63, 2017. doi: 10.1016/j.foodcont.2016.05.048
- [13] —, "Rapid detection of frozen-then-thawed minced beef using multi-spectral imaging and fourier transform infrared spectroscopy," *Meat science*, vol. 135, pp. 142–147, 2018. doi: 10.1016/j.meatsci.2017.09.016
- [14] L.-C. Fengou, A. Lianou, P. Tsakanikas, F. Mohareb, and G.-J. E. Nychas, "Detection of meat adulteration using spectroscopy-based sensors," *Foods*, vol. 10, no. 4, p. 861, 2021. doi: 10.3390/foods10040861
- [15] X. Li, X. Fan, L. Zhao, S. Huang, Y. He, and X. Suo, "Discrimination of pepper seed varieties by multispectral imaging combined with machine learning," *Applied Engineering in Agriculture*, vol. 36, no. 5, pp. 743–749, 2020. doi: 10.13031/aea.13794
- [16] M. A. Tamayo-Monsalve, E. Mercado-Ruiz, J. P. Villa-Pulgarin, M. A. Bravo-Ortiz, H. B. Arteaga-Arteaga, A. Mora-Rubio, J. A. Alzate-Grisales, D. Arias-Garzon, V. Romero-Cano, S. Orozco-Arias *et al.*, "Coffee maturity classification using convolutional neural networks and transfer learning," *IEEE Access*, vol. 10, pp. 42 971–42 982, 2022. doi: 10.1109/ACCESS.2022.3166515
- [17] F. Chollet, "Keras," <https://github.com/fchollet/keras>, 2015.
- [18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. doi: 10.48550/arXiv.1409.1556
- [19] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017. doi: 10.48550/arXiv.1602.07261
- [20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015. doi: 10.1109/CVPR.2015.7298594 pp. 1–9.
- [21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017. doi: 10.1109/CVPR.2017.243 pp. 4700–4708.
- [22] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," 2014. [Online]. Available: <https://arxiv.org/abs/1409.0575>
- [23] L.-C. Fengou, P. Tsakanikas, and G.-J. E. Nychas, "Rapid detection of minced pork and chicken adulteration in fresh, stored and cooked ground meat," *Food Control*, 2021. doi: <https://doi.org/10.1016/j.foodcont.2021.108002>
- [24] K. He, R. Girshick, and P. Dollár, "Rethinking imagenet pre-training," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019. doi: 10.48550/arXiv.1811.08883 pp. 4918–4927.