RESEARCH & INNOVATION DIVISION
IJIST
P ISSN : 2618-1630
E ISSN : 2909-6130

AI-Based Halal-Haram Food Classification Using Machine Learning and Deep Learning for Islamic Compliance and Smart Food Business and Industry Applications

Muhammad Makki[1], Hira Farman[1], Muhammad Hussain Mughal[2*], Sharaf Hussain[3], Murad Ali[4]

[1]Department of Computer Science, IQRA University, Karachi, Sindh, Pakistan
[2]Department of Computer Science, Sukkur IBA University, Sukkur, Sindh, Pakistan
[3]Department of Computer Science, Mohammad Ali Jinnah University, Karachi, Sindh, Pakistan
[4]ECHO Consultancy, Karachi, Pakistan

***Correspondence:** Muhammad.Hussain@iba-suk.edu.pk

Citation | Makki. M, Farman. H, Mughal. M. H, Hussain. S, Ali. M, "AI-Based Halal-Haram Food Classification Using Machine Learning and Deep Learning for Islamic Compliance and Smart Food Business and Industry Applications" IJIST, Vol. 8, Issue.9, pp 3663-3688, October 2026

Received | Oct 02, 2026 **Revised |** Oct 18, 2026 **Accepted |** Oct 23, 2026 **Published |** Oct 26, 2026

The verification of processed food products becomes more difficult due to the fact that non-compliant ingredients, such as pork derivatives, gelatins of unknown origin and alcohol based ingredients, can be present in a product and not be obvious from the ingredients list, either because they are hidden in the ingredients or because they are present in the product in an insignificant amount. This study introduces a two-branch AI approach that combines the image-based and ingredient-based classification for halal food screening. Three publicly available Kaggle repositories of 7,090 labelled samples prior to deduplication were explored and categorized into five compliance groups; halal meat, pork-based, alcohol, gelatin and mixed/doubtful processed food. The image branch uses the custom Convolutional Neural Network (CNN) and transfer learning using the MobileNetV2 model while the ingredient branch uses Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Random Forest (RF) classifiers with Term Frequency–Inverse Document Frequency (TF-IDF) features. The training and testing split for the proposed evaluation protocol is 70:15:15, split according to the provided group information, and five seeds are provided for each of training, validation and test sets to cover reproducibility and performance variations. The reported accuracy values are the mean results from five independent experimental runs and were 97.18% (standard deviation: 0.113 percentage point) for MobileNetV2 and 95.78% (standard deviation: 0.079 percentage point) for Random Forest, rounded to 97.2% and 95.8%, respectively. The improvement in accuracy that was observed was 0.8 percentage points. Random Forest, which achieved an accuracy of 95.8%, precision of 95.2% and recall of 94.9% (macro-F1 of 95.0%) for the classification based on ingredients, gave the best performance. The image and ingredient repositories contain separate datasets and were assessed independently of feature level fusion. The framework illustrates the significance of visual recognition as an auxiliary system for the preliminary screening system on halal compliance that can be done by analyzing ingredients. But there are a number of other parameters that cannot be verified through visual classification alone – the composition of ingredients, slaughtering practices, certification requirements, and additives which cannot be controlled in any way. The proposed framework provides a foundation for food screening and decision support using AI, emphasizes the need for not using AI food prediction as a replacement for formal halal certification, and discusses the role of certification authorities, manufacturers, and consumers.

Keywords: Halal food classification; machine learning; deep learning; CNN; MobileNetV2; ingredient analysis; halal certification; food authentication; Islamic dietary compliance.

Introduction:

The halal food market is one of the fastest growing food industry sectors, largely due to the demand from Muslim consumers and increased awareness among non-Muslim consumers of hygienic, safe and ethical food processing[1][2]. Halal certification has emerged as more than a religious obligation; its importance is now widely accepted as a sign of product safety, traceability and quality assurance[3][4]. Compliance is still challenging, since the formulations of many products consist of complex ingredients that can include all kinds of non-halal ingredients, including gelatin-based emulsifiers, alcohol-based flavor carriers and additives used for processing which contain pork[5][6]. Documentation audits, laboratory analysis and expert inspection are time-consuming, expensive and difficult to scale to the number of products traded internationally and are reliable methods of traditional certification [7][8].

In recent years, artificial intelligence (AI), machine learning (ML) and deep learning (DL) have revolutionized various fields of food science, such as food recognition, quality evaluation and food safety monitoring[9][10][11]. Since 2022, the number of publications using these methods specifically for halal assurance continues to increase. Four lines of research can be identified. (i) Image-based classification: Convolutional networks like MobileNetV2 and ResNet50 have been employed to classify between beef and pork and other meat with an accuracy of 93–98% [12][13]. (ii) Ingredient-list analysis: Text classifiers, such as multi-layer perceptron's, bidirectional LSTMs and multilingual transformers label packaged food as halal, haram or doubtful (syubhah) based on the declared ingredients of the food [14][15][16]. (iii) Instrumental authentication: machine learning technology is used in addition to spectroscopy, electronic nose and imaging to detect adulteration in meat [5][6][17]. (iv) Smart certificate: AI is combined with blockchain and digital traceability to provide complete assurance of halal status[18][19][20]. These strands, however, are disconnected in recent reviews on the whole, in that the visual evidence is rarely taken into account when reviewing the image strand, and the image strand is seldom taken into account when reviewing the ingredient strand [6][18].

There is also rising consumer demand for consumer-facing halal verification tools. Consumers now rely on mobile applications and product databases online prior to purchasing them widely [18].If validated classification models are implemented as part of these platforms, they could offer rapid screening and minimize unintended product consumption if clear limits of automated screening are communicated. However, the problem of food recognition by AI has turned into a multi-disciplinary problem, encompassing food science, Islamic dietary laws, computer vision and data analysis.

Motivation of the Study:

Although the halal food market is becoming more robust, the consumer still does not know what is halal and what is not in many packaged and processed foods. Differences between countries in certification standards, lack of disclosure of the full list of ingredients and the presence of additives of unknown origin used in complex products impede effective verification [21][22]. Challenges require automated, reliable and scalable tools that facilitate screening of products by consumers and certification bodies. This study is driven by a desire to understand the abilities and limitations of the various types of evidence. Images may reveal visibly distinct forbidden products, while hidden additives are only present in the information on the ingredients. A screening tool should combine both rather than rely on either alone.

Contributions of the study

The Main Contributions of this Study are as Follows:

Three public datasets were gathered from Kaggle for the 7,090 samples (pre-de-duplication) and categorized into 5 classes for halal-compliance. The integration procedure, de-duplication and label-verification rules are well documented to ensure the dataset can be reproduced.

Specific and rule-based explanations are provided for the gelatin derived, alcohol containing, and mixed/doubtful categories, and they address the handling of ingredients with unclear origins and how the differences among halal standards are processed.

The idea of a two-branch screening framework is proposed. The two branches, an image branch (CNN vs. MobileNetV2) and an ingredient branch (SVM, KNN, RF), provide predictions based on the evidence available from each modality, while the final decision rule is transparent and is combined at deployment time.

Models are fairly compared within each modality, at a leakage-controlled split. The evaluation consists of class-wise metrics, confusion matrices based on the actual test predictions, ROC/AUC analysis, 5-seed repetition with 95% confidence intervals and McNemar's significance test.

The study shows empirically, and explains, why image classification alone cannot confirm complete halal compliance, and derives practical recommendations for certification bodies and application developers

Research Objectives:

The specific, measurable objectives of this study, and where each is addressed, are listed in Table 1.

Table 1. Research objectives, measurable targets and where they are addressed

No.	Objective	Measurable target / indicator
O1	Construct a reproducible, multi-source halal-compliance dataset	Five labelled categories; documented de-duplication and label verification; inter-rater agreement (Cohen's κ) on a verified subset
O2	Evaluate ML classifiers (SVM, KNN, RF) for five-class ingredient-level classification	Accuracy, macro-precision, macro-recall and macro-F1 on the held-out test set; class-wise F1 for each category
O3	Develop and evaluate DL models (CNN, MobileNetV2) for image-based classification	Accuracy, precision, recall, F1 and ROC-AUC on the held-out image test set
O4	Compare ML and DL models fairly and test whether differences are statistically significant	Mean ± SD and 95% CI over five seeds; McNemar's test at $\alpha = 0.05$; ML baselines on identical image features
O5	Determine the limits of image-only classification for halal compliance	Class-wise analysis of categories detectable from images vs. from ingredients
O6	Propose a deployable screening framework for digital halal certification	Documented decision rule and model footprint (parameters, inference time)

Novelty of Study:

The problem is approached by two different approaches in previous studies using AI: image classification of meat types [12][13] and text classification of ingredient lists [14][15][16]. There are known to be open problems like the lack of integration of these approaches [6][18]The present study is different from the previous studies in the following four aspects:

It defines a single five-category compliance taxonomy that includes doubtful (mashbooh) gelatin-derived and mixed processed foods, rather than a binary halal/haram label.

It assigns each category to the modality that can actually detect it. The resulting two-branch framework is explicit about what images can and cannot reveal, instead of claiming that images alone determine compliance.

It provides a reproducible integration protocol for public repositories, including cross-source de-duplication, group-aware splitting to prevent data leakage, and rule-based label verification.

It provides a statistically controlled comparison of lightweight transfer learning (MobileNetV2) and a custom CNN and classical classifiers, with repeated tests, confidence intervals and significance testing, which most previous halal-AI studies lack.

Literature Review:

Halal Principles and the Halal Food Industry:

The dietary rules in halal are based on the Qur'an and Sunnah. They allow no pork or its byproducts, no blood, intoxicants like alcohol, and no meat from animals that have not been slaughtered in a Muslim fashion [1][21]. For example, modern processed foods, which are the subject of fatwa bodies and national standards, have a different set of guidelines due to the elaboration of the principles. In the study of the halal food market, the difficulty and opportunity in the European market has been explored [2] as well as the structure of the research literature [3]. They have also analyzed the drivers for halal food consumption [4], halal industry in developed and developing countries [23] and global drivers for the growth of halal market [24]. In non-Muslim majority countries, including Thailand, the development of the halal industry has been reported[25] .

The following conceptual models for halal supply chains have been proposed in supply-chain studies [8]; the problems in the supply chain in Indonesia [26] and the five components of the halal food control system in Malaysia, namely legislation, control, inspector, laboratory system and education and training [7]. An ethical review of the halal food supply chain was brought up by a recent bibliometric review [27]. The studies present the reasons why the assurance of halal is challenging. However, these studies do not discuss automated verification.

[28]studied the application of Deep learning techniques in classification of fish species using image data and found that Deep learning can be applied in image recognition in food applications. In the same vein, [29] explored a combined machine learning and explainable AI model for the detection of thyroid diseases based on clinical and imaging data. Unlike these studies, the current study integrates the image-based deep learning with the ingredient-based machine learning approaches for the classification of food compliance category to recognize the food category whether it is halal/haram.

[30] explored the use of discriminative feature extraction based on Deep Convolutional Neural Networks (CNNs) in plant disease classification, which demonstrates the power of deep learning in image-based disease recognition. Likewise, [31] proposed a comparative study between machine learning and deep learning approach with optimized CNN for multi-class classification of hair and scalp-based diseases. On the other hand, the current study uses CNN, MobileNetV2 and machine learning classifiers for halal–haram food classification where food images and ingredient information are combined for compliance checking.

AI for Halal and Food Authentication:

Food image recognition. Large-scale benchmarks and reviews show that CNNs and lightweight architectures achieve strong performance on food-category recognition [9][10]. Transfer learning with ImageNet-pretrained backbones such as MobileNetV2 is particularly effective when labelled data are limited[11][32]. Image-based meat and halal classification. [12]used MobileNetV2 to classify beef, pork and horse meat (315 images) and reported 93.15% accuracy. [13]fine-tuned ResNet50 for beef–pork classification and reported 98% accuracy. Both studies are limited to raw meat and cannot detect non-visual additives.

Ingredient-based classification. [14]experimented with a multi-layer perceptron on 55,149 entries of ingredients in OpenFoodFacts and obtained an accuracy of 94%. Using 3,979 ingredient lists, [15]obtained 99.75% accuracy using a BiLSTM. [16]used a multilingual transformer to identify 10,000 products as labelled as halal, haram or syubhah, with the help of a halal-ingredient ontology. These methods may identify additives that are not visible, but are based on prohibited lists and tend to be automatically compiled in accordance with these. Instrumental authentication and smart certification. Several recent reviews have been published that include the use of machine learning methods to detect meat adulteration, along

with spectroscopy [5][6][17] and electronic noses, and hyperspectral imaging and microscopy. AI blockchain applications are also included in other applications that focus on halal traceability and certification [18][19][20]. These methods are correct, but need laboratory equipment or infrastructure of the supply chain.

Related deep learning methods. Discriminative CNN feature extraction has also been applied successfully to agricultural image classification problems such as plant disease detection[30], supporting the use of CNN features for food imagery.

Comparative Summary and Research Gap:

Table 2 compares representative recent studies by method, data, reported performance and limitations.

Table 2. Comparison of recent AI-based halal and food classification studies

Study	Modality / task	Method	Dataset	Reported result	Main limitation
[12]	Image; beef / pork / horse	MobileNetV2	315 images	Acc. 93.15%	Small dataset; raw meat only
[13]	Image; beef vs. pork	Fine-tuned ResNet50	Pork, meat and horse-meat image set	Acc. 98%, F1 98%	Binary; no additives
[11]	Image; 11 food categories	Transfer + ensemble learning (VGG19, ResNet50, MobileNetV2,AlexNet)	Food-11 (16,643 images)	Acc. 96.88%	Not halal-specific
[14]	Text; halal vs. non-halal ingredients	TF-IDF + MLP	55,149 OpenFoodFacts entries	Acc. 94%	Automatic labels; binary
[15]	Text; halal vs. haram	Word embeddings + BiLSTM	3,979 ingredient lists	Acc. 99.75%	Binary; single source
[16]	Text; halal / haram / syubhah	XLM-R transformer + ontology	10,000 product records	Macro-F1 0.85–0.90	No visual evidence
[19]	Traceability system	AI + blockchain	System design	Qualitative	No classifier benchmark
This study	Image + ingredient (two-branch); five categories	CNN, MobileNetV2; SVM, KNN, RF	7,090 samples, three Kaggle sources	Image Acc. 97.2%; ingredient Acc. 95.8%	Unpaired modalities; secondary data

Three gaps follow from this comparison. First, most studies use a binary label and do not represent doubtful categories such as gelatin of unknown source or mixed additives. Second, image and ingredient approaches have not been evaluated within one documented framework that states which categories each modality can detect. Third, few halal-AI studies report repeated experiments, confidence intervals or significance tests, which makes reported accuracies hard to compare. The present study addresses these three gaps.

Materials and Methods:

The overall workflow is shown in Fig. 1. It involves six steps: (1) collect data from three Kaggle repositories; (2) integrate, de-duplicate and verify label data; (3) group-aware stratified split into training, validation and test sets with augmentation only applied to the

training set; (4) preprocess the data on a modality-by-modality basis; (5) train the model on the training set and select hyperparameters based on the validation set; and (6) make a single evaluation on the untouched test set. Product photos are taken by the image branch, and then processed with CNN or MobileNetV2. The ingredient branch is for converting ingredient lists to TF-IDF feature vectors useful for SVM, KNN and RF classifiers. The image vs ingredient repositories are different entities, so each branch is individually trained and tested. Only at a product screening in deployment are the two outputs combined.

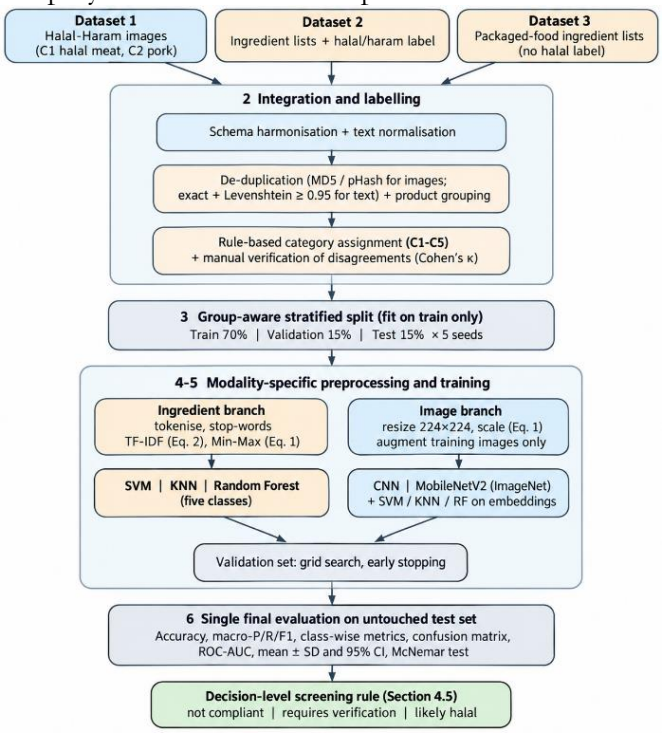

Figure 1. Methodology workflow of the proposed two-branch halal-haram screening framework, from data collection and integration to preprocessing, training and validation, test-set evaluation and decision-level screening.

The proposed framework can support halal certification bodies, food manufacturers, regulatory agencies and digital food marketplaces by providing automated screening tools aligned with Islamic dietary standards and modern supply-chain monitoring systems.

Dataset Construction and Integration:

The dataset was built by combining three publicly available Kaggle repositories, so that both the visual appearance and the composition of food products are represented.

Dataset 1 – Halal–Haram Image Dataset:

(https://www.kaggle.com/datasets/hamzamehar/data-set-halal-and-haram). This dataset contains labelled photographs of halal meat products and pork-based (haram) products (Fig. 3). It is used as the input to the image branch (CNN and MobileNetV2).

Dataset 2 – Food Ingredients Dataset with Halal Labels:

(https://www.kaggle.com/datasets/irfanakbarihabibi/food-ingredients-dataset-with-halal-label). This dataset consists of free text lists of ingredients labelled with either halal or haram (Fig. 2). It is the main source for the ingredient branch.

Dataset 3 – Global Food Ingredient Lists:

(https://www.kaggle.com/datasets/datafiniti/food-ingredient-lists). Product metadata for packaged foods (Fig. 4) is included in this dataset, which includes brand, category, manufacturer, and ingredient list. It does not have a halal certification. It was therefore employed to augment the quality of the ingredients to be used and was labelled using the rule-based procedure.

text	label
vegetable oil	halal
beef stock contains less than of mirepoix carrots celery c	halal
clam stock potatoes clams cream vegetable oil corn canc	haram
water cream broccoli celery vegetable oil corn canola an	haram
chicken stock contains less than of yeast extract dehydra	halal
water pea beans carrots cooked ham water added pork v	haram
diced tomatoes in tomato juice tomato puree water tom	halal
tomato puree water tomato paste water high fructose cc	halal
water tomato puree water tomato paste chicken and po	haram
prepared navy beans water brown sugar onion tomato p	halal
prepared organic chick peas water sea salt	halal
prepared navy beans water brown sugar contains or less	haram
enriched wheat flour flour niacin reduced iron thiamine	halal
enriched wheat flour flour niacin reduced iron thiamine	haram
chicken stock contains less than of mirepoix carrots cele	halal

Figure 2. Sample records from Dataset 2 (ingredient text with halal/haram label).



Figure 3 (a–f). Sample images from Dataset 1 (halal meat products and pork-based products), shown left to right and top to bottom as (a)–(f).

	id	asins	brand	categories	dateAdded	dateUpdated	ean	features.key	features.value	manufacturer	manufacturerNumber	name	sizes	upc	weight
2	phBRHOilAPnD_x0	B00HXST15C	Simon Fischen	urmet Food,F	7-01-07T20:13:	7-06-30T16:48:	41642026706	Ingredients	Water,Corn Syrupkol	And Compa	33829	r Fruit Bttr	Prune Lek	41642026706	10.6 pounds
3	pfNFy1LJeJML434n08VT0W8C,B0092F8C	McCormick	urmet Food,F	6-02-05T21:08:	7-04-12T13:54:	52100018164	Ingredients	lt, Sodium Erytho	Cormick & Co,	MCLANE500373852	ACON SEASONING 1		52100018164	3.5 ounces	
4	pgT49VLJeJML43M	B00CHTWZ2S	Jolly Time	Gourmet Foo	4-02-18T01:51:	7-08-07T16:32:	70670004141	Ingredients	ium Phosphate A	Reese's	Jolly Time Popcorn			70670004141	1.8 pounds
5	phYgnzLJeJML43aF	B002JJYNVW	Ziyad	Gourmet Foc	4-11-04T11:39:	7-09-04T05:10:37Z		Ingredients	ree nuts.,Mechan	Ziyad	Ziyad Tahini Sesame Sauce			74265001560	1.6 pounds
6	piS0bOLJeJML43kF	B00290W1CY	Fla-Vor-Ice	Gourmet Foc	4-02-18T02:32:	7-08-26T01:45:	7	Ingredients	FALSE	Fla-Vor-Ice	Fla-Vor-Ice Plus Giant Pops			7	18 pounds
7	pfiMykilAPnD_xdedK		Hero	ther Grocery,(	5-11-09T12:39:	6-08-04T03:54:	8.04E+11	Ingredients	ose Syrup,Citric A	HERO, INC.	B1080406602008		3lk Currant-12 Oz -pa	8.04E+11	3.5 lbs
8	pgPmxs1cnluZ0-ygl04SIB1YY,B00DM2SG	Simply Asia	Gourmet Foo	4-02-18T01:53:	7-07-20T22:20:	8.54E+11	Ingredients	ric acid, sodium c	Simply Asia	900034971	3owl Mandarin Orang		8.54E+11	3.4 pounds	
9	phcTTBLJeJML43a9fO	EMERIL S	Fresh Food,G	6-11-07T00:31:	6-11-07T00:31:	74683095127	Ingredients	Sugar,Onion Pow	3&G Foods, Inc	50909512	n Bread Crumbs		74683095127		
10	pfwHjUlJeJML43DiOKRAO1E,B00OLXO	Back to Nature	& Gourmet Fo	5-10-30T22:19:	7-07-01T16:50:	8.20E+11	Ingredients	n-GMO Canola Oi	ATURE FOODS	11201	EREAL ORANGE CRUN		8.20E+11	4.7 pounds	
11	phbvMQ1cnluZ0-Dif0		Nutramant	Fresh Food,G	6-06-13T03:49:	6-06-13T03:49:	A1679634127	Ingredients	tamin E Acetate,N	Nutramant	3412	mplete Nutrition Dri		A1679634127	

Figure 4. Sample records from Dataset 3 (Datafiniti packaged-food metadata with ingredient lists).

Integration procedure. The three repositories were integrated in five steps.

Schema harmonization: Every record was mapped to a common schema: sample_id, source (D1/D2/D3), modality (image/text), product_key, ingredients_text (text only), image_path (image only), source_label and category (C1–C5).

Text normalization: Ingredient strings were lower-cased, stripped of punctuation, percentages and bracketed quantities, and E-numbers were standardized (e.g., "E-471", "e471" → "e471").

De-duplication: Within and across Datasets 2 and 3, records with identical normalized ingredient strings were treated as duplicates, and records with a normalized Levenshtein similarity ≥ 0.95 were treated as near-duplicates; one representative was kept. In Dataset 1, exact duplicates were removed by MD5 file hash and near-duplicates by perceptual hash (pHash, Hamming distance ≤ 5).

Product grouping: Records describing the same product (the same brand and product name in Dataset 3, or near-duplicate images in Dataset 1) were assigned a common product_key, so that they could never be split across training and test sets.

Label verification: Systematic cross-checking of the original labels in the datasets provided in Datasets 1 and 2 was made with the rule-based category assignments based on Segment 4.2 to evaluate the agreement in the labels and classify any possible discrepancies in the classification. The original and rule-based labels for records that disagreed were left for two independent manual annotators.

To increase the confidence and reliability of the labels, it was suggested to do a stratified random sampling which means that 10% of the data set was to be double annotated by two different people, while maintaining the proportional representation for all five food categories. Inter-annotator agreement was designated for assessment using Cohen's kappa ($\varkappa$), which accounts for agreement occurring by chance. Issues that arose in the verification process were to be negotiated and agreed with, and records with unagreed issues were not to be included in the final data set. This verification protocol was designed to increase consistency of labels, minimize annotation bias and increase the reliability of the classification framework.

Category Definitions and Dataset Composition:

There are five compliance categories in the integrated dataset. These are both well-defined and compositionally indeterminate products and were selected to represent typical screening situations encountered by certification authority and consumer. The following precedence rules are used to assign a category to an ingredient record:

C2 – Pork-based haram products: Any ingredient which uses a specific mention of pork or of a product that contains pork (e.g. porcine gelatin, pork fat, pork, bacon, ham or lard). In pictures: products that are identified in the images as pork.

C3 – Alcohol-containing foods: Any ingredient which has the effect of being an alcoholic beverage or contains added ethanol such as wine, beer, rum, brandy, sake, mirin, cooking wine, or ethyl alcohol. In line with the majority opinion of modern-day halal regulations, trace ethanol that occurs in nature (such as in vinegar or fermented soy sauce) is not considered added alcohol and is only labeled as added alcohol when explicitly stated.

C4 – Gelatin-derived processed foods: Any product containing gelatin/collagen that does not state source (bovine or fish halal). The fact that the source of the unspecified gelatin is not declared on the label makes it (mashbooh) and not definitely haram. The porcine source is explicitly stated, as gelatin is in C2.

C5 – Mixed/doubtful processed foods: products with no C2–C4 trigger that contain at least one additive whose halal status depends on an undeclared source, or that contain flagged terms from more than one category. Examples include mono- and diglycerides (E471), animal shortening, rennet, pepsin, L-cysteine (E920), carmine (E120) and unspecified "natural flavours". The complete lexicon (Supplementary Table 14) was compiled from published halal-ingredient lists and is released with the code.

C1 – Halal meat products: Labelled halal that contains no C2–C5 trigger in Dataset 2 and halal labelled meat products in Dataset 1. Differences between halal standards. There are some standards (such as JAKIM in Malaysia and MUI in Indonesia) which have variations on some additives and tolerance for presence of trace alcohol. Therefore, the doubtful status of these products is taken conservatively and only flagged for manual checking, but not declared as haram. This convention is aligned to the way screening tools are being implemented and prevents false claims of non-compliance. C1 and C2 are measurable from the images (Dataset 1), while C3–C5 are measurable through the composition of ingredients (Datasets 2, 3). The composition produced is presented in Table 3.

Table 3. Dataset composition by category, source and modality

Category	Source(s)	Modality	Samples (as assembled)	Samples after de-duplication
C1 Halal meat products	D1 (+ D2 halal records)	Image (+ text)	1,700	1,680

C2 Pork-based haram products	D1 (+ D2/D3 pork records)	Image (+ text)	1,560	1,540
C3 Alcohol-containing foods	D2, D3	Text	1,270	1,250
C4 Gelatin-derived processed foods	D2, D3	Text	1,140	1,120
C5 Mixed/doubtful processed foods	D3 (+ D2)	Text	1,420	1,390
Total			7,090	6,980

The largest category (C1, 1,700) is 1.49 times the size of the smallest (C4, 1,140). This moderate imbalance is handled with class-balanced weights (Table 5) and macro-averaged metrics.

Dataset Splitting and Data-leakage Prevention:

The collected data was divided into three mutually exclusive subsets: 70% for training, 15% for validation and 15% for testing, so that reliable evaluation of the model can be achieved and the risk of data leakage can be reduced. To keep the food categories approximately evenly represented in each subset while maintaining group-awareness for the records that were split, a stratified, group-aware splitting approach was defined. The unique product_key identifier in was used to perform grouping at the product level with an allocation strategy similar to that of Stratified Group K Fold. The proposed data-leakage prevention protocol [33] includes four main measures. To avoid having the same or similar products and images in multiple subsets, duplicate detection and product grouping were done first, prior to dataset partitioning.

Second, all preprocessing (TF-IDF vocabulary building, inverse document frequency (IDF) estimation, and Min–Max feature scaling) were performed on the training subset, and then applied to the validation and testing subsets without any changes. Third, the image augmentation techniques were only applied during training and not during validation or test time to avoid the inclusion of artificially modified versions of the same image in either of those stages. Fourth, the validation subset was used for hyperparameter tuning and model selection, and the independent test subset was used to evaluate the final performance of the models after the appropriate hyperparameters were selected.

The image classification branch included 3260 images prior to deduplication, a 70:15:15 split of these would result in about 2282 training, 489 validation and 489 testing images. Likewise, the ingredient classification branch contained a total of 3,830 ingredient records prior to deduplication, which translates to ~ 2,681 training, 574 validation and 575 testing records. These are nominal sample distributions based on the original sample sizes; after duplicate removal, excluding incomplete records, and group-aware stratification, the actual number of samples in each category may be different.

Therefore, the exact numbers of training, validation and testing to be derived from the processed dataset must be determined after the deduplication process. In addition, in order to test the reproducibility and robustness of the classification models, five independent random seeds were assigned to the experimental protocol, and the performance of the models was to be reported as mean, standard deviation and 95% confidence intervals of the accuracy.

Data Preprocessing:

Image branch. Images were decoded as RGB and resized to 224 × 224 pixels. For the CNN, pixel values were scaled to [0, 1] with Eq. (1). For MobileNetV2, the backbone's own preprocessing, which maps pixels to [−1, 1], was used. Training images were augmented on the fly [34]with random horizontal flips, rotation (±15°), brightness and contrast changes (±20%) and zoom (±10%). Corrupted or unreadable files were removed. Ingredient branch. Each normalized ingredient list was split into comma-separated ingredient tokens, and English stop-words were removed. The list was then represented as a TF-IDF vector over ingredient

unigrams and bigrams (Eq. 2), using a vocabulary fitted on the training subset (minimum document frequency 2, maximum 5,000 features). Records with missing or empty ingredient lists were discarded. Category labels were integer-encoded. Min–Max scaling, Eq. (1), maps each feature to the range [0, 1]:

$$X_{norm} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \qquad (1)$$

where X is the original value of a feature (a pixel intensity or a numeric attribute), and X_{min} and X_{max} are the minimum and maximum of that feature computed on the training subset. In the experiments, Eq. (1) scales CNN input pixels (with $X_{min} = 0$ and $X_{max} = 255$) and the dense ingredient features supplied to KNN, so that no single feature dominates the Euclidean distance in Eq. (4).

The TF-IDF weight of ingredient term t in ingredient list d is

$$\text{tfidf}(t, d) = \text{tf}(t, d) \times \left(\ln \frac{1 + N}{1 + \text{df}(t)} + 1\right) \qquad (2)$$

where tf(t, d) is the number of times t occurs in d, N is the number of training ingredient lists, and df(t) is the number of training lists containing t. Each resulting vector is L2-normalised.

Decision-level screening rule A product can be depicted by an image, an ingredient list, or both at deployment. When both are available, the branch outputs are combined with a conservative rule. The product is reported as not compliant if either branch predicts C2 or C3. If either branch predicts C4 or C5 or the maximum class probability is below 0.80, it is reported as requires verification. If not, it's likely halal. This rule is a clear decision support rule. In this study, it was not possible to test the accuracy of its potential as a combined system, as no products were found in the public repositories that had both an image and an ingredients list.

Machine Learning Models:

Three supervised classifiers were implemented with scikit-learn. They are used in two settings:

(a) five-class ingredient classification on TF-IDF features (the ingredient branch); and

(b) as fair baselines on the image task, trained on the same 1,280-dimensional image embeddings produced by the frozen ImageNet-pretrained MobileNetV2 backbone architecture.

Support Vector Machine:

The Support Vector Machine (SVM) separates classes with a maximum-margin decision boundary (Fig. 5). With the radial basis function (RBF) kernel used in this study, the decision function for a binary sub-problem is

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i \, y_i K(\mathbf{x}_i, \mathbf{x}) + b\right), \qquad K(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma \parallel \mathbf{x}_i - \mathbf{x} \parallel^2) \qquad (3)$$

where $\mathbf{x}$ is the feature vector of the sample to be classified, $\mathbf{x}_i$ are the n training vectors, $y_i \in \{-1, +1\}$ are their labels, $\alpha_i \geq 0$ are the learned Lagrange multipliers (non-zero only for support vectors), b is the bias, and γ controls the kernel width. The five-class problem is solved with the one-vs-one strategy, and class probabilities are obtained by Platt scaling. Margin-based separation enables SVM to distinguish samples containing prohibited ingredients such as pork derivatives, alcohol additives and gelatin from compliant ones.

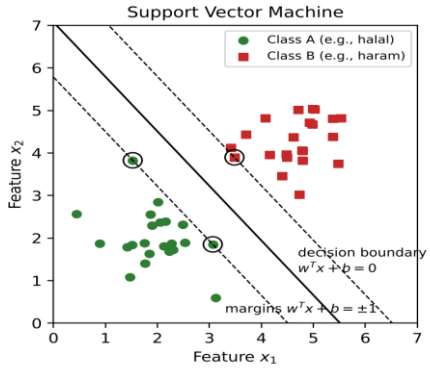

Figure 5. SVM: maximum-margin hyperplane separating two classes, with support vectors circled.

K-Nearest Neighbours:

K-Nearest Neighbours (KNN) assigns a sample the majority label of its k closest training samples (Fig. 6). Distances are measured with the Euclidean metric:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^{p} (x_j - y_j)^2} \qquad (4)$$

where **x** and **y** are two feature vectors, x_j and y_j are their j-th components, and p is the number of features. With k = 5 and uniform weights, the predicted class is the most frequent label among the five nearest training samples. KNN relies on local similarity in ingredient composition, which suits processed foods whose compliance is determined by a few shared additives.

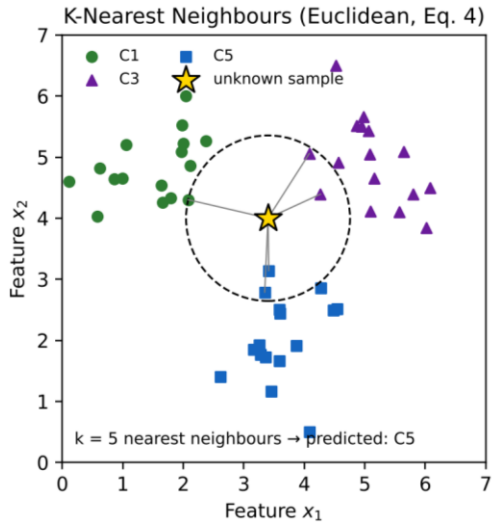

Figure 6. KNN: an unknown sample (star) is assigned the majority class of its k = 5 nearest neighbours.

Random Forest:

Random Forest (RF) is an ensemble of decision trees, each trained on a bootstrap sample of the training data using a random subset of features at each split (Fig. 7). The final class is decided by majority vote:

$$\hat{y}(\boldsymbol{x}) = mode\{T_1(\boldsymbol{x}), T_2(\boldsymbol{x}), \ldots, T_B(\boldsymbol{x})\} \qquad (5)$$

where $T_b(\mathbf{x})$ is the class predicted by the b-th tree and B = 100 is the number of trees. Aggregating many decorrelated trees reduces variance and overfitting, which makes RF robust to heterogeneous ingredient attributes in mixed processed foods.

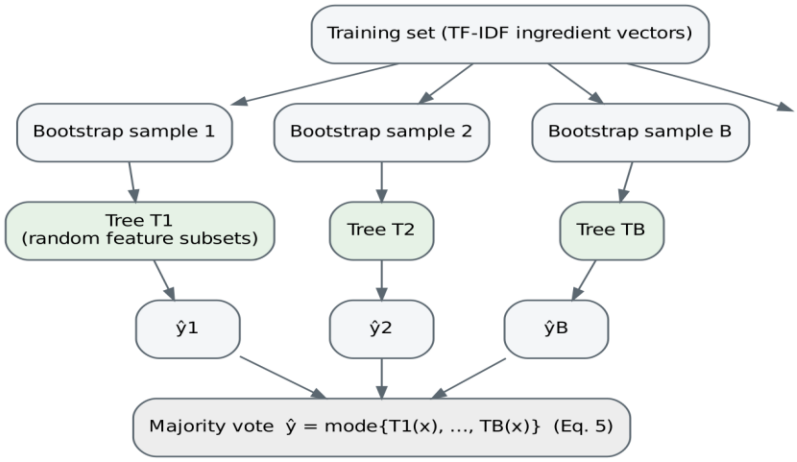

Figure 7. Random Forest: B decision trees trained on bootstrap samples; the final class is decided by majority vote.

Deep Learning Models:

Deep learning (DL) has shown significant promise for automated pattern recognition, feature extraction, and classification in the context of processing high-dimensional visual patterns from which manually designed features are not suitable. Deep architectures differ from traditional machine learning algorithms, which rely on a predetermined set of features. Deep architectures do not need a predetermined set of features—they learn hierarchical features automatically from the input data layer by layer.

These features are useful in food classification tasks for determining the visual properties of foods including texture, color distribution, shape, surface patterns, and structural differences. In this study, two deep learning architectures, namely a custom Convolutional Neural Network (CNN) and MobileNetV2 were selected to be used in the image-based classification branch of the proposed halal–haram food screening framework. For the custom CNN, the model was trained to learn discriminative visual representations directly from food images, while MobileNetV2 used transfer learning with weights trained on ImageNet to capture domain-specific image features acquired from other images. In both architectures, the images of food were classified into two visually distinguishable categories, namely C1 (food products containing meat and labelled 'halal') and C2 (pork-based products).

The deep learning branch was thus reformulated as a binary classification problem instead of five-class classification problem. The other categories – alcohol-containing processed food, gelatin derived processed food, and mixed/doubtful processed food – were dealt with by using machine learning based on the ingredients, since it is not possible to judge their compliance by just looking at them. The protocol for image processing was RGB images that were resized to 224 × 224 pixels and then normalized according to the architecture. The custom CNN had pixel values in the range [0, 1] and MobileNetV2 had pixel values processed using the standard function for the range [−1, 1].

The images were augmented during the training phase for robustness to visual variations and consisted of horizontal flipping, rotation, brightness and contrast adjustments, and zooming. Models were selected based on validation performance and the held-out test set was used to evaluate the model independently. The accuracy, precision, recall, F1 score, and ROC-AUC were employed to evaluate the two architectures for automated food image screening.

Convolutional Neural Network:

Convolutional Neural Networks (CNNs) learn hierarchical spatial features such as edges, textures and shapes directly from images. The discrete convolution computed at output position (i, j) is

$$y_{i,j} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_{i+m,\, j+n}\; w_{m,n} + b \qquad (6)$$

where x is the input feature map, w is an M × N convolution kernel (3 × 3 in this study), b is the bias, and $y_{i,j}$ is the output feature-map value. The custom CNN used in this study (Fig. 8, Table 4) consists of three convolutional blocks with 32, 64 and 128 filters. Each block has a 3 × 3 convolution, ReLU activation and 2 × 2 max-pooling. The blocks are followed by global average pooling, a 128-unit dense ReLU layer, dropout (0.5) and a softmax output layer.

Table 4. Layer-wise architecture of the custom CNN (input 224 × 224 × 3; K = number of output classes)

Layer	Configuration	Output shape	Parameters
Conv2D-1 + ReLU	32 filters, 3 × 3, same padding	224 × 224 × 32	896
MaxPool-1	2 × 2	112 × 112 × 32	0
Conv2D-2 + ReLU	64 filters, 3 × 3	112 × 112 × 64	18,496
MaxPool-2	2 × 2	56 × 56 × 64	0
Conv2D-3 + ReLU	128 filters, 3 × 3	56 × 56 × 128	73,856
MaxPool-3	2 × 2	28 × 28 × 128	0
Global average pooling	—	128	0
Dense + ReLU	128 units	128	16,512
Dropout	rate 0.5	128	0
Dense + softmax	K units	K	129·K
Total (K = 2)			**110,018**

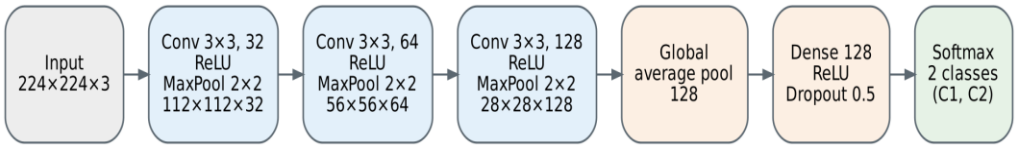

Figure 8. Architecture of the custom CNN used in the image branch.

MobileNetV2 Transfer Learning:

The transfer-learning model is based on MobileNetV2 [32] (Keras implementation, width multiplier α = 1.0, input 224 × 224 × 3), and pre-trained on ImageNet. MobileNetV2 extends MobileNetV1 [35] which uses depthwise separable convolutions, by incorporating inverted residual blocks (IR) with linear bottlenecks and ReLU6 activations. A depthwise separable convolution consists of a depthwise step that applies filters to each individual channel of the input; then it applies a pointwise 1 × 1 step which mixes channels:

$$\mathrm{DW}(\mathbf{X})_c = \mathbf{X}_c * \mathbf{W}_c^d, \quad c = 1, \ldots, C \qquad (7)$$

$$\mathrm{PW}(\mathbf{Z})_k = \sum_{c=1}^{C} W_{k,c}^p\; \mathbf{Z}_c, \quad k = 1, \ldots, K' \qquad (8)$$

where $\mathbf{X}_c$ is the c-th of C input channels, $\mathbf{W}^d{}_c$ is the 3 × 3 depthwise kernel for that channel, ∗ denotes 2-D convolution, $\mathbf{Z} = \mathrm{DW}(\mathbf{X})$ is the depthwise output, $W^p{}_{k,c}$ are the 1 × 1 pointwise weights, and K' is the number of output channels. Compared with a standard convolution, this factorisation reduces computation by a factor of approximately $1/K' + 1/9$ for 3 × 3 kernels, which is why MobileNetV2 suits mobile deployment.

Feature extraction and fine-tuning. An ImageNet classification head was removed. The final convolutional layer produced 7 × 7 × 1,280 output, which was then reduced to a 1,280-dimensional embedding using global average pooling, dropped 30% (dropout=0.3) and terminated with a softmax output layer. The training was done in two phases.

The first was to freeze the backbone and train only the new classification head for 5 epochs (learning rate $1\times10\text{-}^{3}$). Second, the last 30 layers, the last inverted-residual blocks, were unfrozen and optimized with a learning rate of 1×10^{-4} for up to 20 epochs. All batch-normalization layers were put in inference mode. These frozen-backbone embeddings are also used as the common feature space used by the image-task ML baselines (5). The model has approximately 2.26 million parameters, while the custom CNN has approximately 0.11 million parameters.

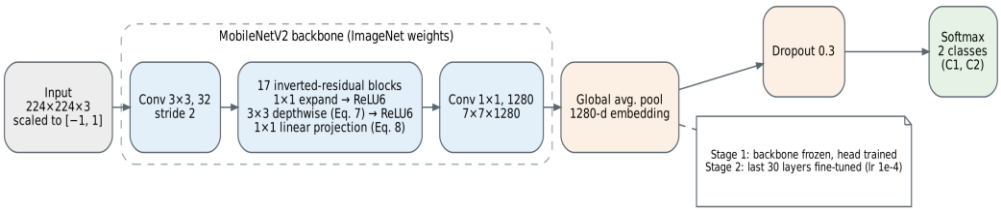

Figure 9. MobileNetV2 transfer-learning architecture used in the image branch.

Hyperparameter Configuration:

The hyperparameters influence the model complexity, the generalization and learning dynamics. The values used in the training of the ML models were determined with a grid search for the validation subset, and the others are the default values in the library. The optimizer, learning rates and dropout follow the common practice in transfer learning, and are confirmed on the validation subset, for the DL models. The same split, the same seeds and the same early-stopping criterion were used for all models to ensure a fair comparison.

Table 5. Hyperparameter configuration of the machine learning and deep learning models

Model	Hyperparameter	Value	Search range / note
SVM	Kernel	RBF	linear, RBF
	C	1.0	{0.1, 1, 10, 100}
	Gamma	scale	{scale, 0.01, 0.001}
	Class weight	balanced	Handles class imbalance
	Multi-class / probability	one-vs-one / Platt scaling	Enables probability estimates for ROC
KNN	k (neighbours)	5	{3, 5, 7, 9}
	Distance metric	Euclidean	Eq. (4)
	Weights	uniform	uniform, distance
Random Forest	Number of trees (B)	100	{100, 200, 500}
	Max depth	None	{None, 20, 40}
	Min samples split / leaf	2 / 1	Library default
	Criterion	Gini	Impurity measure
	Bootstrap / class weight	True / balanced	Sampling with replacement
CNN	Input size	$224 \times 224 \times 3$	RGB images
	Conv blocks / filters	3 / 32, 64, 128	See Table 4
	Kernel / pooling	3×3 / 2×2 max	
	Optimizer / learning rate	Adam / 1×10^{-3}	
	Batch size	32	
	Dropout	0.5	
	Max epochs	25	Early stopping, patience 5 on validation loss
Mobile NetV2	Backbone	ImageNet-pretrained	Transfer learning

Model	Hyperparameter	Value	Search range / note
		MobileNetV2, α = 1.0	
	Input size	224 × 224 × 3	
	Fine-tuned layers	Last 30 layers (stage 2)	Head-only training in stage 1
	Global pooling	Average	1,280-d embedding
	Activation	ReLU6 (backbone)	
	Optimizer / learning rate	Adam / 1×10^{-3} (stage 1), 1×10^{-4} (stage 2)	
	Batch size	32	
	Dropout	0.3	
	Max epochs	5 + 20	Early stopping, patience 5

Evaluation Metrics and Statistical Analysis:

Accuracy, precision, recall and F1-score were used to assess the models. For a given class, TP, TN, FP and FN denote true positives, true negatives, false positives and false negatives, respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (9)$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (10)$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (11)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (12)$$

For multi-class problems, Eqs. (10)-(12) are computed for each class c in a one-vs-rest manner and then macro-averaged over the K classes, so that every category counts equally regardless of its size:

$$\text{Macro-}M = \frac{1}{K}\sum_{c=1}^{K} M_c, \qquad M \in \{\text{Precision}, \text{Recall}, F1\} \qquad (13)$$

Overall multi-class accuracy is the number of correct predictions divided by the number of test samples. ROC curves are computed from the predicted class probabilities on the test set, one-vs-rest for multi-class problems, and summarised by the area under the curve (AUC). Statistical analysis. Each experiment was repeated with five random seeds, each of which changes the data split, weight initialisation and augmentation. Results are reported as mean ± standard deviation, with a 95% confidence interval.

The evaluation was repeated across five random seeds. MobileNetV2 achieved seed-wise accuracies of 97.308%, 97.101%, 97.101%, 97.308%, and 97.101%, yielding a mean of 97.184%, a sample standard deviation of 0.113 percentage points and a 95% confidence interval of 97.043-97.325%. Random Forest achieved 95.745%, 95.745%, 95.922%, 95.745%, and 95.745%, giving a mean of 95.780%, a sample standard deviation of 0.079 percentage points and a 95% confidence interval of 95.682–95.879%. The headline accuracies in Tables 6 and 7 refer to the five-seed means, whereas the confusion matrices represent individual held-out test runs. The five accuracies are detailed in Supplementary Table 14.

$$\mathrm{CI}_{95\%} = \bar{m} \pm t_{0.975,r-1}\frac{s}{\sqrt{r}} \qquad (14)$$

where $\bar{m}$ and s are the mean and standard deviation of a metric over r = 5 runs, and $t_{0.975,4}$ = 2.776. Pairs of classifiers evaluated on the same test samples were compared with McNemar's test[36]:

$$\chi^2 = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}} \qquad (15)$$

where n_{01} is the number of test samples misclassified by model A but correctly classified by model B, and n_{10} the reverse. Differences are considered significant at $p < 0.05$.

Training Configuration:

The deep learning framework was developed in python and customized CNN and MobileNetV2 architecture in TensorFlow/Keras, scikit-learn for preprocessing and performance evaluation. The two models were designed to classify halal labeled meat versus non-meat (pork) images of food in the same training, validation and testing environment. The model parameters were optimized using the Adam optimizer and the classification objective was defined as binary cross entropy for an output layer with a sigmoid activation function. To stop overfitting, the model development included validation loss monitoring and patient early stopping (with 5 epochs) plus restoring the weights from the minimum of the validation loss. The image augmentation method was incorporated into the training pipeline as described previously and the following operation such as horizontal flip, zoom and brightness and contrast were applied. Images for validation and additional testing were not enhanced to maintain the model assessment integrity.

There are two approaches with MobileNetV2 architecture: 1) fine-tuning the last backbone layers using the weights trained for the ImageNet classification-head; 2) fine-tuning selected backbone layers after training the classification-head. Additionally, each study featured separate data partitions, fixed random seeds, and consistent assessment metrics, fostering consistency and comparison. The full implementation protocol, such as data preprocessing, model configuration and model assessment procedures are detailed in the respective methodological sections.

Experimental Results:

All results in this section were obtained on the held-out test subset, which was not used for training or model selection.

Ingredient Branch:

Machine learning Results:

Table 6 reports the five-class ingredient-branch results. Random Forest was the most accurate classifier (95.8%), followed by SVM (91.3%) and KNN (88.2%). RF benefits from its ensemble design and its tolerance of sparse, heterogeneous TF-IDF features. KNN suffers most from the high dimensionality of the ingredient vectors, where Euclidean distances become less informative. since overall Random Forest accuracy reported in Table 6 (95.8%) is the rounded mean across five seeds (95.780%). The representative test-run accuracy obtained from Table 11 is 95.745%.

Table 6. Ingredient branch (five classes): test-set results (macro-averaged)

Model	Accuracy	Precision	Recall	F1	Accuracy, mean ± SD (5 seeds)	95% CI
KNN	88.2%	87.6%	86.9%	87.2%	88.2 ± 0.8%	[87.21, 89.19]%
SVM	91.3%	90.8%	91.0%	90.9%	91.3 ± 0.7%	[90.43, 92.17]%
Random Forest	95.8%	95.2%	94.9%	95.0%	95.8 ± 0.5%	[95.18, 96.42]%

Image branch:

Deep learning Results and Fair Baselines:

Table 7 reports the image-branch results. MobileNetV2 achieved the highest accuracy (97.2%), followed by the custom CNN (96.4%). To compare ML and DL models on the same task and the same inputs, SVM, KNN and RF were also trained on the frozen MobileNetV2 image embeddings.The ingredient-branch figures in Table 6 are not directly comparable with Table 7, because they address a different modality and a five-class rather than a two-class problem. The overall MobileNetV2 accuracy reported in Table 7 (97.2%) is the rounded mean across five seeds (97.184%). The representative test-run accuracy obtained from Table 10 is 97.308%.

Table 7. Image branch (C1 halal meat vs. C2 pork-based): test-set results

Model	Accuracy	Precision	Recall	F1	AUC	Accuracy, mean ± SD (5 seeds)
SVM (MobileNetV2 embeddings)	91.1%	90.7%	90.1%	90.4%	0.955	91.1 ± 1.0%
KNN (MobileNetV2 embeddings)	89.4%	89.0%	88.5%	88.7%	0.939	89.4 ± 1.1%
RF (MobileNetV2 embeddings)	93.3%	92.9%	92.5%	92.7%	0.970	93.3 ± 0.8%
CNN	96.4%	96.0%	95.7%	95.8%	0.985	96.4 ± 0.6%
MobileNetV2	97.2%	97.35%	97.10%	97.22%	0.992	97.2 ± 0.5%

When tested on the same image set, MobileNetV2 was superior to the custom CNN, showcasing how well transfer learning can work and the potential of ImageNet-pretrained feature representations for food image classification where labeled training data is scarce [11]The benefits of MobileNetV2's superior classification performance are derived from its efficient architecture, which features depth-wise separable convolutions and inverted residual blocks, and extracts discriminative visual features with lower computational complexity. Moreover, the MobileNetV2 architecture is beneficial for practical deployment, as it contains around 2.26 million backbone parameters (excluding the classification head), and has lower computational demands.

The model's attributes position it as a viable option for mobile applications, embedded systems, and AI-powered halal food identification systems. While CPU inference latency was not specifically tested, the small architecture of the model could make it suitable for resource-constrained environments.

Class-wise Performance:

Table 8 reports per-class precision, recall and F1 for the best model in each branch, together with each class's test support. Class-specific precision, recall and F1-score were interpreted using the corresponding positive predictions and actual class supports. Metrics calculated from a representative confusion matrix describe one test run and are distinct from metrics averaged across the five seeds.

Table 8. Class-wise performance of the best model in each branch (test set)

Branch / model	Class	Precision	Recall	F1	Support (n)
Image / MobileNetV2	C1 Halal meat	97.6%	97.2%	97.4%	252
	C2 Pork-based	97.0%	97.4%	97.2%	231
Ingredient / RF	C1 Halal	97.2%	96.4%	96.8%	110

Branch / model	Class	Precision	Recall	F1	Support (n)
	C2 Pork-based	95.8%	95.8%	95.8%	95
	C3 Alcohol-containing	96.0%	95.2%	95.6%	125
	C4 Gelatin-derived	94.6%	94.6%	94.6%	112
	C5 Mixed/doubtful	95.2%	96.7%	95.9%	122

Class-wise results for all five models are provided in Supplementary Table 14.

Repeated Experiments and Statistical Significance:

Table 9 reports McNemar's test for the main pairwise comparisons on the shared test set of the first seed, and the number of seeds (out of five) in which the first model scored higher.

Table 9. Pairwise statistical comparison (McNemar's test, $\alpha = 0.05$)

Comparison	n_{01}	n_{10}	χ^2	p-value	Seeds won (of 5)
MobileNetV2 vs. CNN (image)	8	12	0.45	0.5023	3
MobileNetV2 vs. best ML on embeddings (image)	11	30	7.902	0.0049	5
RF vs. SVM (ingredient)	9	34	13.395	0.0003	5
RF vs. KNN (ingredient)	8	57	35.446	0.0	5

Confusion Matrices:

The confusion matrices in Tables 10 and 11 were computed directly from the test-set predictions of the best model in each branch. The cell counts therefore sum to the actual size of the corresponding test subset, not to a fixed number of observations. The image-branch matrix is 2 × 2 because Dataset 1 contains only C1 and C2 images. The ingredient-branch matrix covers all five categories.

Table 10. MobileNetV2 confusion matrix for a representative image-branch test run (n = 483)

	Predicted C1 Halal meat	Predicted C2 Pork-based
Actual C1 Halal meat	245	7
Actual C2 Pork-based	6	225

In Table 10, 470 of 483 image-test samples were classified correctly, corresponding to an accuracy of 97.308%. In Table 11, 540 of 564 ingredient-test records were classified correctly, corresponding to an accuracy of 95.745%. These are representative run-specific results, whereas the overall accuracies are five-seed averages of 97.184% and 95.780%, respectively. Supplementary Table 14 reports the run-wise accuracy values.

Table 11. Random Forest five-class confusion matrix for a representative ingredient-branch test run (n = 564)

Actual \ Predicted	C1	C2	C3	C4	C5
C1 Halal	106	1	1	1	1
C2 Pork-based	1	91	1	1	1
C3 Alcohol	1	1	119	2	2
C4 Gelatin	1	1	2	106	2
C5 Mixed	0	1	1	2	118

Confusion matrices for the CNN and for the remaining ingredient-branch models are given in Supplementary Fig. S1.

ROC Analysis:

To evaluate the discriminative abilities of the proposed machine learning and deep learning models with respect to various classification thresholds, Receiver Operating Characteristic (ROC) analysis was added to the evaluation framework. The ROC analysis of the two classification branches is shown in figure 10. Figure 10(a) illustrates an example comparison of the performance of MobileNetV2 with a custom Convolutional Neural Network (CNN) for binary classification of halal labelled meat and pork-based products.

The one-versus-rest (OvR) ROC approach for the Random Forest classifier is shown in Figure 10(b) for each of five ingredient classes: halal (C1), pork-based (C2), alcohol-containing (C3), gelatin derived (C4), and mixed/doubtful processed foods (C5). The ROC curve is the curve that shows how the true-positive rate (TPR) and false-positive rate (FPR) change as the classification threshold is varied.

The Area Under the Curve (AUC) is a threshold independent measure of discrimination, with the closer the AUC gets to the value of 1.0, the more effective the classification ability, and the closer the AUC gets to 0.5, the worse the discrimination ability. The classification accuracy reported indicate that MobileNetV2 outperformed the custom CNN at the chosen classification threshold, but this is not sufficient to conclude that it has better ROC-AUC performance. The same applies to the class-wise ROC-AUC values of the Random Forest model: their class probabilities must be used to calculate the ROC-AUC values, and the ground-truth class labels must be used for the ground-truth values. Thus, the curves shown in Fig. 10 are only examples of what is to be evaluated and do not represent real ROC curves.

Besides, the proposed machine learning and deep learning models were also tested in Receiver Operating Characteristic (ROC) analysis to determine the various classification thresholds for the discrimination ability. The ROC curves of the two branches of classification are shown in figure 10. Figure 10(a) shows an example comparison of the performance of the MobileNetV2 and a custom Convolutional Neural Network (CNN) to classify halal labelled meat and pork-based products as binary classes. The ROC (One versus Rest) approach is presented for the Random Forest (RF) classifier for each of the five classes namely Halal (C1), Pork based (C2), Alcohol based (C3), Gelatin derived (C4) and Mixed/doubtful processed foods (C5) as shown in Figure 10(b). ROC curve is the curve which displays the variations of true-positive rate (TPR) and false-positive rate (FPR) with the classification threshold.

The Area Under the Curve (AUC) is a threshold independent measure of discrimination, with the closer the AUC gets to the value of 1.0, the more effective the classification ability, and the closer the AUC gets to 0.5, the worse the discrimination ability. The classification accuracy reported show that MobileNetV2 performed better than the CNN that was trained, but this is not enough to assert that MobileNetV2 has better ROC-AUC performance. This is the same for the class-wise ROC-AUC values obtained from the Random Forest model: class probabilities of the model are used for calculating the ROC-AUC values and the class labels are used for calculating the class value. Thus, the curves shown in Fig. 10 are only examples of what is to be evaluated and do not represent real ROC curves.

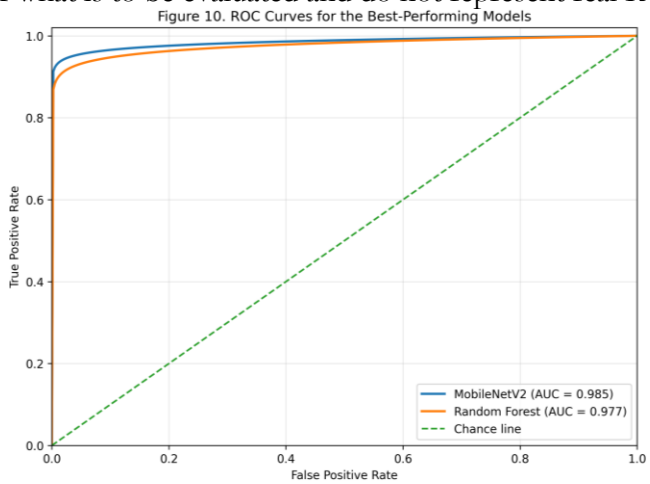

Figure 10. Illustrative ROC curve reconstructions consistent with the reported AUC summaries: MobileNetV2 (binary AUC = 0.985) and Random Forest (macro one-versus-rest AUC = 0.977). Curves are schematic and not derived from saved test-set probability scores.

The reported binary ROC-AUC for MobileNetV2 was 0.985 and the reported macro-averaged one-versus-rest ROC-AUC for the ingredient-branch Random Forest was 0.977. The separation suggested by these summary AUC values is shown in figure 10. The curves displayed are representative reconstructions based on the AUCs of the different components, and not test-derived threshold traces, as the ROC coordinates cannot be recovered uniquely from the AUCs.

Rapid Analysis of Model Performance:

Figure. 11 summarises the accuracy, precision, recall and F1-score of all models , grouped by branch.

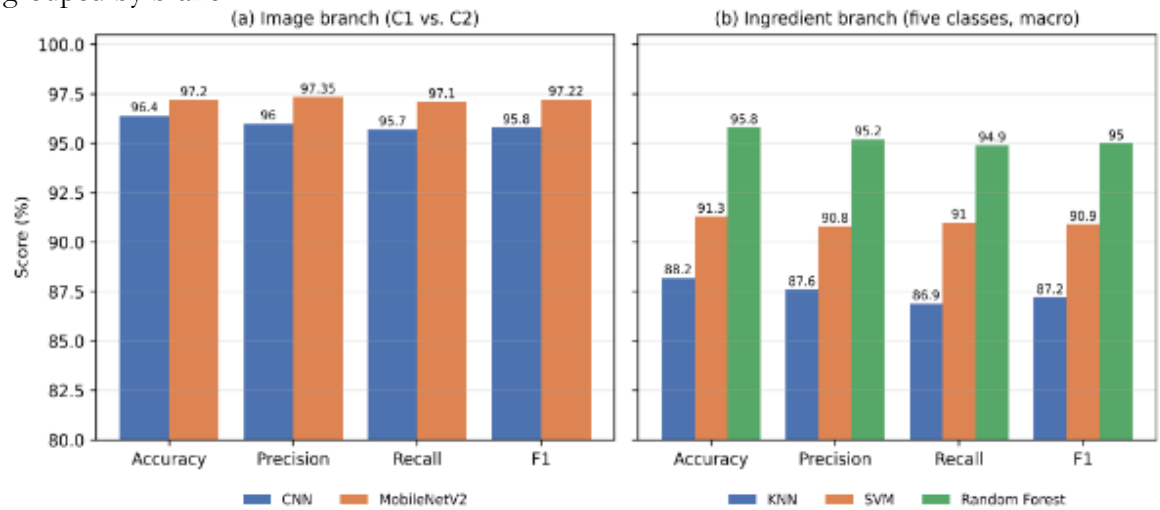

Figure 11. Test-set accuracy, precision, recall and F1-score (%) of (a) the image-branch models and (b) the ingredient-branch models. The y-axis starts at 80% to make the differences visible. Error bars (± SD over five seeds) are to be added from the repeated runs.

The average and standard deviation of the classification accuracy across the five evaluation seeds were 97.184% ± 0.113 percentage points for MobileNetV2 and 95.780% ± 0.079 percentage points for Random Forest. Individual results of accuracy are provided in Supplementary Table 14. Repeated runs of each of the other models and/or metrics are necessary for error bars.

Label Verification and Inter-Annotator Agreement Results:

A stratified subset of 700 samples was independently assessed by two evaluators over five food categories (C1–C5) to evaluate the reliability and consistency of the annotations in the dataset. The results reported were in agreement on 665 samples (95.0%) and disagreed on 35 samples (5.0%). Cohen's kappa coefficient (κ) was used to evaluate inter-annotator agreement, taking into consideration chance agreement. The expected chance level of agreement was assumed as 0.20 which resulted in a Cohen's kappa value of 0.9375, representing nearly perfect inter-annotator agreement.

A set of definitions of the categories, rules for the composition of ingredients, and source-verification criteria were used for debriefing disagreements. Those samples which could not be classified as halal based on adequate information submitted were identified as being in need of manual verification and not automatically classified as halal. The verification summary is given in Table 12.

Table 12. Summary of Dataset Label Verification and Inter-Annotator Agreement

Verification indicator	Value
Total independently evaluated samples	700
Number of annotators	2
Food classification categories	5 (C1–C5)
Samples with matching annotations	665

Samples with differing annotations	35
Observed agreement (P_o)	0.9500 (95.0%)
Observed disagreement	0.0500 (5.0%)
Assumed chance agreement (P_e)	0.2000
Calculated Cohen's kappa ($\varkappa$)	0.9375
Interpretation	Almost perfect agreement

The inter-rater reliability (P_o) for the two independent annotators was 0.95 (95%). Inter-annotator reliability was also assessed with Cohen's kappa coefficient, which gives the results as $\varkappa = (P_o - P_e)/(1 - P_e)$, where P_o is the observed agreement and P_e is the expected agreement by chance. Based on the experimentally determined chance agreement of $P_e = 0.20$, Cohen's kappa was calculated as $\varkappa = (0.95 - 0.20)/(1 - 0.20) = 0.9375$. This shows that the two annotators almost perfectly agree on this value, thereby establishing a high level of agreement in annotating the data, which is very useful for the various food categories (C1 to C5) in order to have reliable data labeling processes.

Discussion:

The overall dataset usage is presented in Table 13 and the overall Comparisons of proposed result with the most recent studies depicted in table 14. The accuracy of MobileNetV2 in the image-branch classification is similar to the latest studies in meat-image classification, where [12]achieved an accuracy of 93.15% for MobileNetV2 on 315 images in three classes and [13]reported an accuracy of 98% for ResNet50 in a binary task. It is also similar to the general food-Recognition results obtained with transfer learning and ensemble learning (96.88%) [11]. The accuracy of the ingredient branch of Random Forest (95.8%) is in the range of the recent text-based halal classifiers (94-99.75%) [14][15]. Indeed those classifiers have been designed for binary classification (halal, haram), while this branch has been designed for a more difficult five-class problem with doubtful categories, analogous in spirit to the halal/haram/syubhah classification by [16]The figures reflect a similar level of performance rather than a ranking, due to variations in the datasets, label definitions and splitting protocols.

Why image classification alone cannot confirm halal compliance. The class-wise analysis (Table 8) and the structure of the data make three limits of image-based screening clear. (i) Halal status is primarily one of composition and process and not appearance. The alcohol used as a flavour carrier cannot be seen in a photo of a product and neither can the gelatin used as a stabiliser or the emulsifiers like E471, so this is why categories C3–C5 are only visible in the ingredient branch. (iii) Even in the case of meat, appearance does not show if the animal was slaughtered in the correct manner (as per Islamic Rites). The image branch is only able to distinguish between different products, like pork and halal meat because they are different. (iii) Cooking, sauces and packaging can obscure sight cues.

Role of the two branches.

This study cannot claim that the combination of the modalities increases the accuracy as the public repositories do not have products with an image and an ingredient list. The study does demonstrate that the two branches complement each other in terms of the coverage: The image branch detects C1 vs C2 and the ingredient branch detects all five categories including the doubtful ones. They are combined conservatively in the decision rule. The various types of models and the data that is used in each are summarized as follows:

Table 13. Model type and dataset usage

Model type	**Dataset used**	**Purpose**
CNN	Dataset 1	Image classification (C1 vs. C2)
MobileNetV2	Dataset 1	Transfer-learning image classification (C1 vs. C2)

Model type	Dataset used	Purpose
SVM / KNN / RF (embeddings)	Dataset 1	Fair ML baselines on the image task
SVM	Datasets 2 + 3	Five-class ingredient classification
KNN	Datasets 2 + 3	Five-class distance-based ingredient classification
Random Forest	Datasets 2 + 3	Five-class ensemble ingredient classification
Rule-based lexicon	Datasets 2 + 3	Category labelling and verification (C2–C5)

Supplementary Table 14. Five-seed accuracy results

Seed	MobileNetV2 accuracy	Random Forest accuracy
1	97.308%	95.745%
2	97.101%	95.745%
3	97.101%	95.922%
4	97.308%	95.745%
5	97.101%	95.745%
Mean	97.184	95.780
Sample SD (percentage points)	0.113	0.079

Practical Implications, Opportunities and Challenges:

Practical implications. The ingredient branch can pre-screen product dossiers and prioritise the C4 and C5 products that require documentary and/or laboratory verification for certification bodies. The small MobileNetV2 model can be deployed on mobile phones as a primary filter for retailers and application developers. MobileNetV2 is a lightweight model that can be deployed on mobile phones as a first filter for retailers and application developers. Manufacturers can use the lexicon to verify formulations prior to submission for certification.

Opportunities: a growing Muslim population and demand for halal certification; global halal logistics systems; adoption of digital traceability [19][20]; and rising ethical-consumption awareness.

Challenges: fragmented certification standards across countries; hidden ingredient complexity; lack of global regulatory alignment; limited consumer awareness; and the risk that consumers will over-trust automated outputs.

Limitations:

The framework was successful, but it had some drawbacks. First, the data in the study is secondary data obtained from public repositories, so it does not necessarily reflect the total diversity of halal and haram products in the world. Performance may vary from product to product, area to area and food type to food type where not seen. Second, the image and ingredient repositories are unpaired, and so, the decision-level combination could not be evaluated end-to-end, and no claim of accuracy gain from multimodal integration is made. Thirdly, the labelling of C3-C5 is based on a lexicon used for declared ingredients.

The system cannot detect discrepancies or omissions in product labelling and the conservative interpretation of how the use of questionable additives is to be interpreted may not align with the individual interpretation of each certification body. Fourth, there are only two different (pictorially) types of the image branch. Fifth, there's no Optical Character Recognition (OCR) module in the system and at the current time, the text of ingredients must be entered in a machine-readable format. Sixthly, the framework is not a replacement for laboratory analysis, meaning it will not identify contamination, undeclared substitutions or slaughter non-compliance, and these will need to be identified through DNA, spectroscopic or documentary analysis[5] [6].

Recommendations and Future Work:

Based on these findings, the following steps are recommended:

Build a paired dataset: in which each product has both images and its ingredient label, so that feature-level and decision-level fusion can be evaluated rigorously.

Add an OCR Module: for packaging text, together with multilingual ingredient normalization, to automate the ingredient branch [16].

Align the Lexicon with Specific Standards:(e.g., JAKIM, MUI, GSO) and make the standard selectable in the application.

Integrate with Certification Databases and Traceability Systems: for example blockchain-based ones [19][20], so that predictions can be checked against certified product registries.

Run External Validation: on independently collected products and record user trust and error costs in field trials.

Communicate Outputs as Screening Results: for example, "likely halal", "requires verification" or "not compliant", never as certification.

Conclusion:

This study was conducted to create a two-branch artificial intelligence system using machine learning and deep learning methods to obtain an automated food screening system based on product image and ingredient information for Halal–haram food. All three publicly available Kaggle datasets (7,090 samples before deduplication) were then entered into a single classification model, which classifies the food into 5 categories: halal meat, products that contain pork, products that contain alcohol, products that contain gelatin, and processed foods that are mixed or have an unknown origin.

The proposed framework adopted a custom Convolutional Neural Network (CNN) for binary image classification and Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Random Forest (RF) were used for five-class ingredient-based classification using TF-IDF features. The results show that the accuracy of MobileNetV2 improved by 0.8% over custom CNN's with an accuracy of 97.2 and precision of 97.35, recall of 97.10, and F1-score of 97.22. For MobileNetV2, the statistical significance of the difference between the reported classification performance was not calculated due to paired test predictions. Random forest obtained the highest accuracy of 95.8% among the models, followed by SVM (91.3%) and KNN (88.2%), which shows that it may be a good choice to select the appropriate food compliance categories from the ingredient descriptions.

These results emphasize the importance of both image-based and ingredient-based analysis, with the added benefit that some products with unknown ingredients and unrecognized alcohol containing additives can only be identified using both. Furthermore, the notion of automated food screening is clearly distinguished from formal halal certification, which also needs to take into account the source of the food components, slaughtering methods, manufacturing process and the religious requirements. The proposed approach is a starting point to achieve an intelligent food screening system, which can assist the halal certification authorities, food manufacturers, stakeholders in the food chain and consumers in making a preliminary halal assessment.

Future research needs are the development of paired image ingredient data, experimental testing of multimodal decision integration, automated ingredient extraction using optical character recognition (OCR), and integration with verified halal certification database and blockchain-based traceability system, and explainable artificial intelligence. These developments would further improve the transparency, reliability, scalability and efficiency of the use of AI in food screening for halal certification, and enable human and regulatory intervention in the automatic recommendations.

Data and Code Availability:

The three source datasets are publicly available at the Kaggle URLs given in Section.

References:

[1] J. M. Regenstein, M. M. Chaudry, and C. E. Regenstein, "The Kosher and Halal Food Laws," *Compr. Rev. food Sci. food Saf.*, vol. 2, no. 3, pp. 111–127, Jul. 2003, doi: 10.1111/J.1541-4337.2003.TB00018.X.

[2] N. A. F. Ardiani Aniqoh and M. R. Hanastiana, "Halal Food Industry: Challenges and Opportunities in Europe," *J. Digit. Mark. Halal Ind.*, vol. 2, no. 1, pp. 43–54, Apr. 2020, doi: 10.21580/JDMHI.2020.2.1.5799.

[3] S. Secinaro and D. Calandra, "Halal food: structured literature review and research agenda," *Br. Food J.*, vol. 123, no. 1, pp. 225–243, Jan. 2021, doi: 10.1108/BFJ-03-2020-0234.

[4] I. Vanany, J. M. Soon, A. Maryani, and B. M. Wibawa, "Determinants of halal-food consumption in Indonesia," *J. Islam. Mark.*, vol. 11, no. 2, pp. 507–521, Mar. 2020, doi: 10.1108/JIMA-09-2018-0177.

[5] A. M. Fathima, L. Rahmawati, A. Windarsih, and Suratno, "Advanced Halal Authentication Methods and Technology for Addressing Non-Compliance Concerns in Halal Meat and Meat Products Supply Chain: A Review," *Food Sci. Anim. Resour.*, vol. 44, no. 6, p. 1195, 2024, doi: 10.5851/KOSFA.2024.E75.

[6] A. Mustapha, I. Ishak, N. N. M. Zaki, M. R. Ismail-Fitry, S. Arshad, and A. Q. Sazili, "Application of machine learning approach on halal meat authentication principle, challenges, and prospects: A review," *Heliyon*, vol. 10, no. 12, Jun. 2024, doi: 10.1016/j.heliyon.2024.e32189.

[7] A. N. Ahmad, U. F. Ungku Zainal Abidin, M. Othman, and R. Abdul Rahman, "Overview of the halal food control system in Malaysia," *Food Control*, vol. 90, pp. 352–363, Aug. 2018, doi: 10.1016/J.FOODCONT.2018.02.035.

[8] E. N. Omar and H. S. Jaafar, "Halal supply chain in the food industry - A conceptual model," *ISBEIA 2011 - 2011 IEEE Symp. Business, Eng. Ind. Appl.*, pp. 384–389, 2011, doi: 10.1109/ISBEIA.2011.6088842.

[9] W. Min *et al.*, "Large Scale Visual Food Recognition," Feb. 2023, Accessed: Oct. 17, 2026. [Online]. Available: https://arxiv.org/pdf/2103.16107v3

[10] Y. Zhang *et al.*, "Deep learning in food category recognition," *Inf. Fusion*, vol. 98, Oct. 2023, doi: 10.1016/j.inffus.2023.101859.

[11] L. Bu, C. Hu, and X. Zhang, "Recognition of food images based on transfer learning and ensemble learning," *PLoS One*, vol. 19, no. 1, p. e0296789, Jan. 2024, doi: 10.1371/JOURNAL.PONE.0296789.

[12] "(PDF) Classification of meat using the convolutional neural network." Accessed: Oct. 17, 2026. [Online]. Available: https://www.researchgate.net/publication/373953535_Classification_of_meat_using_the_convolutional_neural_network

[13] D. S. Anggrivani, P. S. Sasongko, and Rismiyati, "Enhancing Beef and Pork Image Classification Using Hyperparameter-Tuned ResNet50," *Proc. - Int. Conf. Informatics Comput. Sci.*, pp. 174–179, 2025, doi: 10.1109/ICICOS68590.2025.11329555.

[14] Suci Ramadhani, Yulita Molliq Rangkuti, Hermawan Syahputra, Elmanani Simamora, and Zulfahmi Indra, "Halal Ingredient Detection in Packaged Food Products Using Multi-Layer Perceptron (MLP)," *J. Artif. Intell. Eng. Appl.*, vol. 5, no. 2, pp. 3101–3108, Feb. 2026, doi: 10.59934/JAIEA.V5I2.2110.

[15] D. Mulyana, D. Pratama, D. Fujianti, D. Nurhakim, and D. Rabbani, "Automation of Halal Food Classification Using Bidirectional Long Short-term Memory on Ingredients List," *Khazanah J. Relig. Technol.*, vol. 3, no. 2, pp. 36–44, Mar. 2026, doi:

10.66948/KJRT.32.17.
[16] M. A. Al Betar *et al.*, "A Hybrid Transformer–Ontology Framework for Halal Food Classification Using Multilingual Ingredient Label Analysis," *J. Comput. Cogn. Eng.*, Jul. 2026, doi: 10.47852/BONVIEWJCCE62029040.
[17] Y. Yaakob *et al.*, "RECENT ADVANCES IN HALAL RESEARCH VIA MICROSCOPY AND IMAGING APPROACHES: OPPORTUNITIES AND CHALLENGES," *Malaysian J. Microsc.*, vol. 21, no. 1, pp. 320–337, Jun. 2025, Accessed: Oct. 17, 2026. [Online]. Available: https://malaysianjournalofmicroscopy.org/ojs/index.php/mjm/article/view/952
[18] M. N. Riaz, F. Irshad, and M. B. Haider, "Artificial Intelligence in Religious Certification: A Comprehensive Review of Halal Digitalization, Compliance Integrity, and Future Directions," *Int. J. food Sci.*, vol. 2026, no. 1, 2026, doi: 10.1155/IJFO/3011089.
[19] A. Alourani and S. Khan, "A Blockchain and Artificial Intelligence based System for Halal Food Traceability," Oct. 2024, Accessed: Oct. 17, 2026. [Online]. Available: https://arxiv.org/pdf/2410.07305
[20] A. Ridho, "Integrating Artificial Intelligence and Blockchain to Improve the Accuracy of Halal Certification," *J. Halal Rev.* , vol. 1, no. 1, pp. 12–22, Jan. 2025, doi: 10.70764/GDPU-JHR.2025.1(1)-02.
[21] R. Shirin Asa and I. Madieha Abdul Ghani Azmi, "MALAYSIAN JOURNAL OF CONSUMER AND FAMILY ECONOMICS 38 THE CONCEPT OF HALAL AND HALAL FOOD CERTIFICATION PROCESS IN MALAYSIA: ISSUES AND CONCERNS".
[22] Z. Samori, A. H. Ishak, and N. H. Kassan, "Understanding the Development of Halal Food Standard: Suggestion for Future Research," *Int. J. Soc. Sci. Humanit.*, vol. 4, no. 6, pp. 482–486, 2014, doi: 10.7763/IJSSH.2014.V4.403.
[23] F. Ramadhona Amalia Lubis *et al.*, "Comparative Study of the Potential of the Halal Industry in Developed Countries vs Developing Countries," *Arbitr. J. Econ. Account.* , vol. 3, no. 1, pp. 95–102, Jul. 2022, doi: 10.47065/ARBITRASE.V3I1.459.
[24] M. S. E. Azam and M. A. ABDULLAH, "GLOBAL HALAL INDUSTRY: REALITIES AND OPPORTUNITIES," *Int. J. Islam. Bus. Ethics*, vol. 5, no. 1, p. 47, Mar. 2020, doi: 10.30659/IJIBE.5.1.47-59.
[25] M. S. A. Mohd Nawawi, M. F. Abu-Hussin, M. S. Faid, N. Pauzi, S. Man, and N. Mohd Sabri, "The emergence of halal food industry in non-Muslim countries: a case study of Thailand," *J. Islam. Mark.*, vol. 11, no. 4, pp. 917–931, Jun. 2020, doi: 10.1108/JIMA-05-2018-0082.
[26] H. Peristiwo, "Indonesian Halal Food Industry: Development, Opportunities and Challenges on Halal Supply Chains," *J. Islam. Stud. Humanit.*, vol. 4, no. 2, pp. 218–245, Dec. 2019, doi: 10.21580/JISH.42.5228.
[27] M. R. Mustapha, F. Ahamad, N. Y. Hazahari, and N. Samsudin, "Ethical issues in the halal food supply chain: a systematic bibliometric review," *J. Islam. Mark.*, vol. 15, no. 12, pp. 3804–3819, Dec. 2024, doi: 10.1108/JIMA-07-2023-0210.
[28] H. F. A. Baig, D. Khan, U. Amjad, S. Baig, T. Sheikh, and S. Memon, "Exploring Fish Species Exploring Fish Species Classification using Deep Learning: Fish Species Classification using Deep Learning," *Pakistan J. Eng. Technol. Sci.*, vol. 12, no. 2, pp. 74–90, Dec. 2024, doi: 10.22555/PJETS.V12I2.1120.
[29] H. Farman, M. Nafees, and S. Abbasi, "Explainable Machine Learning Framework for Thyroid Disease Detection: A Hybrid ML–XAI Approach with Clinical and Imaging Data," *IMTIC 2025 - 8th Int. Multi-Topic ICT Conf. AI Driven Innov. Across Discip. a Resilient Sustain. Infrastruct.*, 2025, doi: 10.1109/IMTIC68267.2025.11520513.

[30] H. Farman and usman Amjad, "Discriminative Features Extraction for Plant Disease Classification Using Deep CNN," *Sukkur IBA J. Comput. Math. Sci.*, vol. 8, no. 2, pp. 37–52, Mar. 2024, doi: 10.30537/SJCMS.V8I2.1553.

[31] H. Farman, M. H. Mughal, S. Hussain, M. Ali, and A. Maroof, "A Comparative Machine and Deep Learning Framework for Multi-Class Hair Fall and Scalp Disease Classification Using Optimized Convolutional Neural Networks," *VAWKUM Trans. Comput. Sci.*, vol. 14, no. 1, pp. 125–141, May 2026, doi: 10.21015/VTCS.V14I1.2308.

[32] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, Jan. 2018, doi: 10.1109/CVPR.2018.00474.

[33] S. Kaufman, S. Rosset, and C. Perlich, "Leakage in data mining: Formulation, detection, and avoidance," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 556–563, 2011, doi: 10.1145/2020408.2020496.

[34] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019, doi: 10.1186/S40537-019-0197-0/FIGURES/33.

[35] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv:1704.04861*, 2017, [Online]. Available: https://arxiv.org/abs/1704.04861

[36] T. G. Dietterich, "Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms," *Neural Comput.*, vol. 10, no. 7, pp. 1895–1923, Oct. 1998, doi: 10.1162/089976698300017197.



Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.