Halal Food Identification from Product Ingredients using Machine Learning

Sabrina Tarannum
Student Id: 012213028

A Thesis

in

The Department

of

Computer Science and Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of Master of Science in Computer Science and Engineering

United International University

Dhaka, Bangladesh

September 2023

Approval Certificate

This thesis titled "**Halal Food Identification from Product Ingredients using Machine Learning**" submitted by **Sabrina Tarannum**, Student ID: **012213028**, has been accepted as Satisfactory in fulfillment of the requirement for the degree of Master of Science in Computer Science and Engineering on September 2023.

Board of Examiners

1.

________________________________ Supervisor

Prof. Dr. Mohammad Nurul Huda
Professor & Head
Dept. of Computer Science and Engineering
United International University (UIU)
Dhaka, Bangladesh

2.

________________________________ Head Examiner

Dr. Md. Saddam Hossain
Associate Professor & Program Coordinator – BSCSE
Dept. of Computer Science and Engineering
United International University (UIU)
Dhaka, Bangladesh

3.

________________________________ Examiner-I

Prof. Dr. Md. Motaharul Islam
Professor & Director-MSCSE
Dept. of Computer Science and Engineering
United International University (UIU)
Dhaka, Bangladesh

4.

________________________________ Examiner-II

Prof. Dr. Raqibul Mostafa
Undergraduate Coordinator
Dept. of EEE
United International University (UIU)
Dhaka, Bangladesh

5.

________________________________ Ex-Officio

Prof. Dr. Hasan Sarwar
Professor & Dean SoSE
Dept. of Computer Science and Engineering
United International University (UIU)
Dhaka, Bangladesh

Declaration

This is to certify that the work entitled "**Halal Food Identification from Product Ingredients using Machine Learning**" is the outcome of the research carried out by me under the supervision of Prof. Dr. Mohammad Nurul Huda, Professor & Head, dept. of Computer Science and Engineering, United International University (UIU), Dhaka.

Sabrina Tarannum
ID: 012213028
MSCSE Program
Dept of CSE
United International University (UIU)
Dhaka, Bangladesh

In my capacity as supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

Prof. Dr. Mohammad Nurul Huda
Professor & Head
Dept. of Computer Science and Engineering
United International University (UIU)
Dhaka, Bangladesh

Abstract

Halal food plays a critical role in the Islamic faith, as it represents food that is considered lawful according to Islamic law. Muslims are encouraged to eat only Halal foods to ensure that it aligns to their religious beliefs. However, locating and verifying Halal-certified foods can be challenging, especially for Muslim travelers unfamiliar with the local food market. Muslims ensure Halal foods that ingredients are prepared in accordance with Islamic Shariah law. Indicators like the Halal emblem have been used to help Muslims identify Halal food. Unfortunately, many packaged items are not Halal-certified. To address this issue, this study presents a method for detecting Halal items using deep learning and machine learning techniques. The purpose is to determine if an unknown product is Halal (legal) or Haram (Illegal) based on its ingredients. The suggested system examines packaged food product images and identifies the ingredients using the Yolo v5 algorithm. The text on the images of the ingredients is then recognized using optical character recognition (OCR). Various machine learning algorithms, artificial neural networks, and fuzzy interference rule are applied to determine the status of the food. The final outcome is to categorize Halal and Haram food products accurately. This approach has the potential to assist Muslim consumers in identifying Halal-certified products quickly and efficiently, particularly when traveling to new locations or encountering unfamiliar products. Using intelligent technology; this study presents a new and innovative technique for detecting Halal food. The result shows that the suggested approach is effective and it might be a useful tool for Muslim consumers in ensuring that the things they buy are compatible with their religious views.

Acknowledgement

All praise is due to Allah, the most merciful and compassionate. I am grateful to Allah for blessing me with the opportunity to complete my thesis. I also wish to convey my profound gratitude and appreciation to Prof. Dr. Mohammad Nurul Huda, Professor & Head, Dept. of Computer Science and Engineering, United International University (UIU), Dhaka for his continuous guidance, invaluable suggestions, encouragement and support throughout this research.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Overview

The concept of Halal food is of great significance in Islam and has been the subject of discussion and analysis for centuries. It is an important aspect of Islamic practice and is regulated by religious law. Halal food is consistently needed for consumption by the Muslim community. Since the Muslim population does not have a Halal certificate, particularly in a Muslim nation where they are a minority, the variety of food packaging is increasing, making it challenging for them to identify Halal. To identify the raw ingredient, each packaged product generally includes a printed composition. Unfortunately, these compositions typically have obscure scientific code names assigned to them. The Muslim community is worried about choosing Halal food products since the general public is unaware of the Halal raw materials used to make food packaging. Beside these, the traditional method of determining the Halal status of food relies on expert opinions, religious texts, and subjective interpretations, which can lead to inconsistencies and misunderstandings. The growing demand for Halal food, particularly in the global market, has emphasized the need for a more systematic and standardized approach to Halal food identification. Especially for many Muslim travelers one of the most vital issues is to find out Halal food when travelling to other countries [19]. According to a survey among 66% of the Muslim tourists mentioned that Halal food is the most significant issue while travelling [33]. Many food products do not have a Halal certificate therefore it is useful to estimate the Halal status of food products.

In recent years, advances in machine learning and artificial intelligence have presented an opportunity to automate the Halal or Haram food classification process. This report presents a method using machine learning algorithms & neural network for the identification of Halal food status. The study was also evaluated the feasibility and accuracy of using machine learning techniques to classify food items as either Halal or Haram based on their ingredient list.

The developed system used machine learning algorithm to provide concrete decisions regarding Halal or Haram especially where Islamic Scholars are confused. Machine learning will provide a probabilistic value in-between 0 to1. After putting a threshold on the predicted value an object is predicted as Halal or Haram. When ingredients are hazy to read, machine learning models are useful instead of rule based because it provides a probabilistic value regarding Halal or Haram. Another advantage of using machine learning models is, in a noisy environment where Islamic Scholars are confused or image is hazy it is very difficult to write a huge number of rules.

The study involved the collection and preparation of a large dataset of food ingredients, followed by the development and testing of various machine learning algorithms. The results of the investigation are then evaluated and compared with Islamic scholar for the purpose of accurately Halal food identification. In addition, the report also discusses the challenges faced in the implementation of machine learning for Halal food identification, such as the need for accurate, diverse, and well-labeled datasets. The study's findings will give important insights into the possibilities of machine learning for Halal food recognition. The report will end with recommendations for further research, the creation of more robust and scalable algorithms, and the application of relevant rule and standards.

Overall, the application of machine learning for Halal food identification has a high potential for enhancing the classification process's accuracy, speed, and consistency. This report is intended to serve as a resource for those interested in this field and to encourage further exploration and development of this technology.

The goal of this research is to provide a solution that uses deep learning and machine learning techniques to solve the difficulties that Muslim customers encounter in discovering and validating Halal-certified items, especially while traveling to new locations. The research emphasizes the vital significance of Halal food in the Islamic religion, as well as the need of ensuring that food ingredients are Halal in accordance with Islamic Shariah law. This strategy has the potential to help Muslim customers make educated purchase decisions while also ensuring that the things they buy are compatible with their religious beliefs. Ultimately, the aim of this effort is to give a useful tool that assists Muslim consumers in ensuring that they consume only Halal foods that are compatible with their religious views.

1.2 Problem Statement

The problem occurs when Muslims travel abroad, it is difficult for them to obtain food and decide if it is Halal or non-Halal in the nation they are visiting. As a result, the number of persons affected by this issue is growing as more people go to abroad each year. To address this issue, Halal food identification based on food packaging is recommended. Muslims traveling abroad may recognize Halal products without having to look at the product contents, even if the Halal emblem is missing. Before it is included in the Halal food recognizer system, the product is first certified by several trustworthy religious experts by studying the product ingredients. As a result, the system simply has to recognize the food product by its packaging and categorize it as Halal or non-Halal food.

1.3 Research Objectives

The objectives of the research are:

- To develop a deep learning-based system identifying the ingredients from a packaged food product image.

- To extract the text from the output of the ingredient’s detection model using OCR.

- To develop a machine learning, neural network& rule-based system in order to classify Halal and non-Halal food.

- To analyze the outcome of accuracy with several trustworthy religious experts.

1.4 Outline of Thesis

This report is divided into five chapters. The following are broad descriptions of the contents of each chapter:

Chapter 1: This chapter discusses the significance of the problem area, the problem statement that leads to the notion of this thesis, the objectives of the report and the report organization.

Chapter 2: This chapter discusses the thesis summary and literature review for this work. It also includes a compendium of earlier publications' food categorization results.

Chapter 3: This chapter illustrates the method used for the system how it works to identify the status of a packaged food product.

Chapter 4: This chapter documented all of the findings from tests and experimentation, a detailed description of the results and accuracy comparison between models & scholars.

Chapter 5: This chapter contains the thesis conclusion as well as the constraints identified for this investigation. The limitations and future direction are also provided.

Chapter 2

Background and Literature Review

The Arabic word 'Halalantoyibban,' which means legal and excellent in English, appears multiple times throughout the Quran. Surah al-Baqarah chapter 2 verses 168, surah al-Maaida chapter 5 verse 88, surah al-Anfaal chapter 8 verse 61, and surah an-Nahl chapter 16 verse 114 all underscore the importance of Muslims seeking clean and Halal meals that are excellent for their health. This covers food preparation, the source of the food, the ingredients used to manufacture the food, and the manner in which animals used in the meal are slaughtered.

Halal food is permissible in Islam according to Islamic law. Consuming Halal food is an order of Allah and it is a crucial part of Islamic faith. Halal food gives assurance to the Muslim that food ingredients are produced according to the Shariah law and are clean, hygienic and environment friendly. Beside these, there are certain foods that are unlawful or prohibited and are not allow Muslims to consume. These forbidden foods are known as Haram foods according to Islamic law.

Muslims are one of the fastest growing populations in the global travel market & almost all countries of the world have at least a small Muslim community. For many Muslim travelers one of the most vital issues is to find out Halal food when travelling to other countries. The availability of Halal food is the main attributes for tourists' intentions to revisit the destination, length of stay and the type of accommodation to choose. According to a survey among 66% of the Muslim tourists mentioned that Halal food is the most significant issue while travelling [19]. So, considering the circumstances, as Haram foods are not permitted for the Muslims according to Islamic law so a system has to develop so that Muslim tourists would be sure of that the food is genuinely Halal.

2.1 ROC Curve

A Receiver Operator Characteristic (ROC) curve is a graphical representation of the diagnostic capability of a binary classifier. The true positive rate (TPR) is compared against the false positive rate (FPR) to generate a ROC curve. The true positive rate (TP/(TP + FN)) is the fraction of positive observations that were correctly anticipated to be positive out of all positive observations. Similarly, the false positive rate is the proportion of negative observations that are incorrectly projected as positive (FP/(TN + FP)). A discrete classifier delivers a single point in the ROC space if it solely outputs the predicted class. Again, for probabilistic classifiers, which offer a probability or score indicating an object belongs to one of two classes.

2.2 Confusion Matrix

The confusion matrix, also known as a contingency table, is one of the finest techniques for illustrating the performance of machine learning systems because it distinguishes between true positive, false positive, true negative, and false negative predictions.

		Actual	**Value**	**total**
		P	**n**	
	P′	True Positive	False Negative	**P′**
Prediction Outcome	**n′**	False Positive	True Negative	**N′**
	total	**P**	**N**	

Table 2.1: A Confusion Matrix

Confusion Matrix is a helpful machine learning approach for calculating Recall, Precision, Accuracy, and ROC curves. A confusion matrix visualizes and summarizes a classification algorithm's performance. A confusion matrix is shown in Table 2.1.

2.3 Precision, Recall, Accuracy and F1 Score

Precision

Precision is a statistic that calculates the proportion of true positives to total positives predicted by the model.

$$\textbf{Precision} = \frac{\textbf{True Positive}}{\textbf{True Positive}+\textbf{False Positive}}$$

A good classifier should ideally have a precision of 1 (high). Precision reaches 1 when the numerator and denominator are identical, i.e. TP = TP +FP, this also implies that FP is zero. As FP increases, the value of the denominator gets bigger than the value of the numerator, but the precision value decreases. Precision is vital in the system to ensure accurate determination of permissible (Halal) and forbidden (Haram) foods. Precise classification guarantees that religious dietary guidelines are upheld, enabling individuals to make informed choices aligned with their beliefs and dietary restrictions.

Recall

Recall concentrates on how effectively the model finds all of the positives.

$$\textbf{Recall} = \frac{\textbf{True Positive}}{\textbf{True Positive}+\textbf{False Negative}}$$

For an effective classifier, recall should ideally be 1 (high). Only when the numerator and denominator are equal, i.e. TP = TP +FN, this also implies that FN is zero. As FN grows, the denominator value becomes bigger than the numerator, and the recall value lowers. Recall holds significance as it ensures that all genuinely permissible (Halal) items are correctly identified, minimizing the risk of false negatives.

Accuracy

The Accuracy score is computed by dividing the number of correct predictions by the total number of forecasts.

$$\textbf{Accuracy} = \frac{\textbf{True Positives}+\textbf{True Negatives}}{\textbf{True Positives}+\textbf{True Negatives}+\textbf{False Negatives}+\textbf{False Positives}}$$

Accuracy is vital as it quantifies the overall correctness of the model's predictions, ensuring that both permissible (Halal) and forbidden (Haram) items are correctly identified. A high accuracy rate reflects the model's capacity to make reliable classifications.

F1 Score

F1 Score is a measure that combines recall and precision. The F1 score has a significant feature in that if any of the components (precision or recall) falls to zero, the outcome is zero. As a result, excessive negative values of any component are penalized.

$$\textbf{F1 Score} = 2 * \frac{\textbf{Precision} * \textbf{Recall}}{\textbf{Precision} + \textbf{Recall}}$$

The F1 score's significance in the system is its ability to find a balance between accurately identifying permissible (Halal) and forbidden (Haram) items.

2.4 Literature Review

Rakhmawati et al. presented a technique that employs Euclidean Distance and Cosine Similarity and lists the top five comparable items in order to determine if a product is Halal even when it does not have a Halal certification [1]. A product may be labeled as Halal if there is a resemblance between it and a product having a Halal certification of greater than 75%. Based on the closeness of the product's constituents, the algorithm can ultimately offer unidentified products from a comparable pool of Halal-certified goods [1].

For predicting the Halal status of food Rakhmawati et al uses several graph similarity algorithms after that similarity algorithms were performed on the products in the dataset and the accuracy of each algorithm is evaluated by F-measure [2]. A smartphone application created by Kartiwi et al. enables users to scan a product's ingredient list and determine using optical character recognition whether any of the stated substances may be inappropriate for consumption [3]. Fadhilah et al. uses Convolutional Neural Networks to recognize Non-Halal compositions in a product [4]. An application has been developed by Khairuddin to identify the Halal status of a product using Barcode System [5]. Akram et al. developed an application that can scan the product ingredients using OCR & then check the Halal or Haram status of a product. It also provides a product's alternatives. The software can scan a barcode on a product to check for allergy-causing substances, and if any are found, it will alert the user based on their allergy concerns [6]. Sumsudin utilized a deep learning approach based on convolutional neural networks (CNN) to identify and categorize the images into Halal and non-Halal products [7].In order to demonstrate the value of integrating Safe and Halal food in accordance with Islamic Dietary Law into the food process and supply system, Kohilavani et al. created a framework that implements SHFMS [8]. In this article, Ilyas Masudin et al. studied traceability, the Halal food supply chain, the Halal logistics, and the Halal way of life. The study focuses on the traceability mechanism provided by Halal food manufacturers and suppliers [9]. The focus of the paper is on detecting Halal meat and Halal meat chain authenticity concerns. Halal meat is defined by its origin, species, production system, manner of slaughter, and technique of meat processing. All of the issues are highlighted in this paper so that all of these traits are validated by Halal certification agencies and help Muslim customers to be protected [10]. As porcine animals are forbidden to eat in Islam so an analytical method has been presented by Mohd HafisYuswan et al. to determine that gelatin and collagen are not derived from bovine or porcine animals [11]. The purpose of this study is to provide a mechanism for dairy farm entrepreneurs to ensure food safety and halal [12]. This essay's [13] main goal is to investigate the literary roots of the term "halal-thayyib" and its use in the Qur'an, which has implications for the production of food products that are quickly evolving in the global Halal market.

In the area of halal food additives and ingredients, La Ode Nazaruddin et al. suggested some possible research areas, including ethical and sustainable sourcing, responsible consumption, consumer sovereignty, global trade, economic modeling, food security, green/sustainable supply chains, halal regulation, and product safety [33]. In order to investigate the amount of consumer awareness and the variables influencing that level of consumer knowledge of halal food items, Muhammad Anwar Fathoni et al. employed statistical descriptive analysis and structural equation modeling (SEM) analysis [34]. Regression models, binary logit models, contingent valuation models, and conjoint analysis models were used by Sulistyodewi Nur Wiyono et al. to uncover the manifestations of consumer intentions for purchasing halal food items and how this intention affected Muslim customers' decisions to buy halal food products [35]. A approach to categorize characteristics from a fruit juice-alcohol combination using machine learning and electronic nose was proposed by E. Ordukaya and B. Karlik [36]. Based on near-infrared hyperspectral imaging coupled with machine learning and the sparrow search algorithm (SSA), Binbin Fanet et al. suggested a technique for recognizing and measuring the contaminated mutton under the impact of mutton flavor essence [37].

From the above literature review it is observed that many methods have been applied for detecting Halal ingredients from food products. The following table shows the summaries and the disadvantages of the reference articles.

Ref	Summary	Disadvantage
[1]	This paper detects the status of a product which are not certified as Halal. The system generates top 5 similar products comparing Halal certified food ingredients if the similarity algorithms produce similar value higher than 75% of a Halal-certified product.	Do not cover food processing from ingredients, processing line and logistics.
[2]	The proposed system links Halal-certified products with non-Halal certified products using several graph similarity algorithms.	From 71423 products only 21239 products depict visualization relationship with Halal food products as all the products don't share the list of ingredients.
[3]	Developed an app that scans product ingredients using mobile phone camera. Then search the ingredients from database to identify the status of the product.	Application cannot read letters if lighting is bad or has dark label on the product. Can't check the status if the ingredients are unrecognized.
[4]	Convolutional neural network deep learning approach for non-Halal materials identification in food packaging. In order to determine the status of packaged foods, CNN is used for character recognition, followed by a match between the text and a list of non-Halal components.	Accuracy is only 50% once characters are combined and compared to a list of basic elements. In addition to this, the algorithm is unable to detect blank spaces, small word sizes, or uneven surfaces in images.

[5]	The system integrates a camera with its barcode recognition technology. This barcode technology is used in the HALAL food identification system. To increase identification rate, a line detection technique was also applied.	The system can only identify those HALAL foods that are certified by JAKIM. The system can read barcode only, can’t identify food status from product ingredients.
[6]	The app can guide people to find Halal food, provide alternatives, allergies & nutrition consumption.	The accuracy of this app is 80%.
[7]	MyHalal app can check name, barcode and product company name. It also checks certificate expiry date of a product and highlight it according to remaining dates.	The system can only identify those HALAL foods that are certified by JAKIM.
[8]	As porcine animals are forbidden to eat in Islam so an analytical method has been presented by authors to determine that gelatin and collagen are not derived from bovine or porcine animals.	Statistical analysis of 44 samples was used to validate the hydroxyproline, an amino acid producer for gelatin and collagen.
[9]	Traceability, the Halal food supply chain, the Halal logistics system, and the Halal way of life were all evaluated in this research. The study focuses on the traceability mechanism offered by manufacturers and suppliers in the purchase of Halal food.	No technology has been used to track the traceability system to find out the transparency of Halal food product.
[10]	The paper focuses on identifying Halal meat, Halal meat chain authenticity issues. Halal meat encompasses origin, species, production system, slaughter procedure	A system should develop to ensure that all the discussed characteristics are verified in pre-purchase stage to

	and meat processing technique. If all the features are verified by Halal certifying authorities then it will protect Muslim consumers & also promote fair trade.	provide a Halal meat to Muslim consumers.

Table 2.2: Summaries of the Review Findings

Chapter 3

Methodology

The methodology of this work consists of five parts: data acquisition, ingredients detection, image to text extraction, trains the model with Halal and non-Halal ingredients finally identify the status of the food product. The output of the system is then compared with the opinions of Islamic scholars.

One issue of employing deep learning is the needed amount of labeled data, which can be time intensive to collect. Images are collected from different sources for example, various online food store, Google and also using mobile phone cameras. The data retrieved addressed food and drinks, and it contained the product name, nutrition facts and ingredients. After data acquisition data labeling is performed to add target properties to data and label them in order to train machine and deep learning models. Object detection requires data labeling in order to execute object detection tasks.

At first images are labeled with labelImg software [20]. Then YOLO V5 [21] is used to train the model with labeled data to identify food product ingredients. After detecting the food ingredients, Optical Character Recognition (OCR) [22] is used to extract the text from the food images. Pre-processing is done, features are extracted & tested with trained machine learning models. The model is trained using a variety of techniques, including Random Forest [23], K-Nearest Neighbors [24], Naïve Bayes [25], Support Vector Machine [26], Decision Tree [27], Neural Network [28], and Rule-Based models [29]. These machine learning, neural network & rule based models are trained with Halal & Haram ingredients dataset of 1040 ingredients so that it can classify the food status as Halal (labeled as '1') or Haram (labeled as '0'). The results of the models are compared with Islamic scholars' opinion to validate that the classification of food is done accurately. The following figure 3.1 shows the overall strategy of the system.

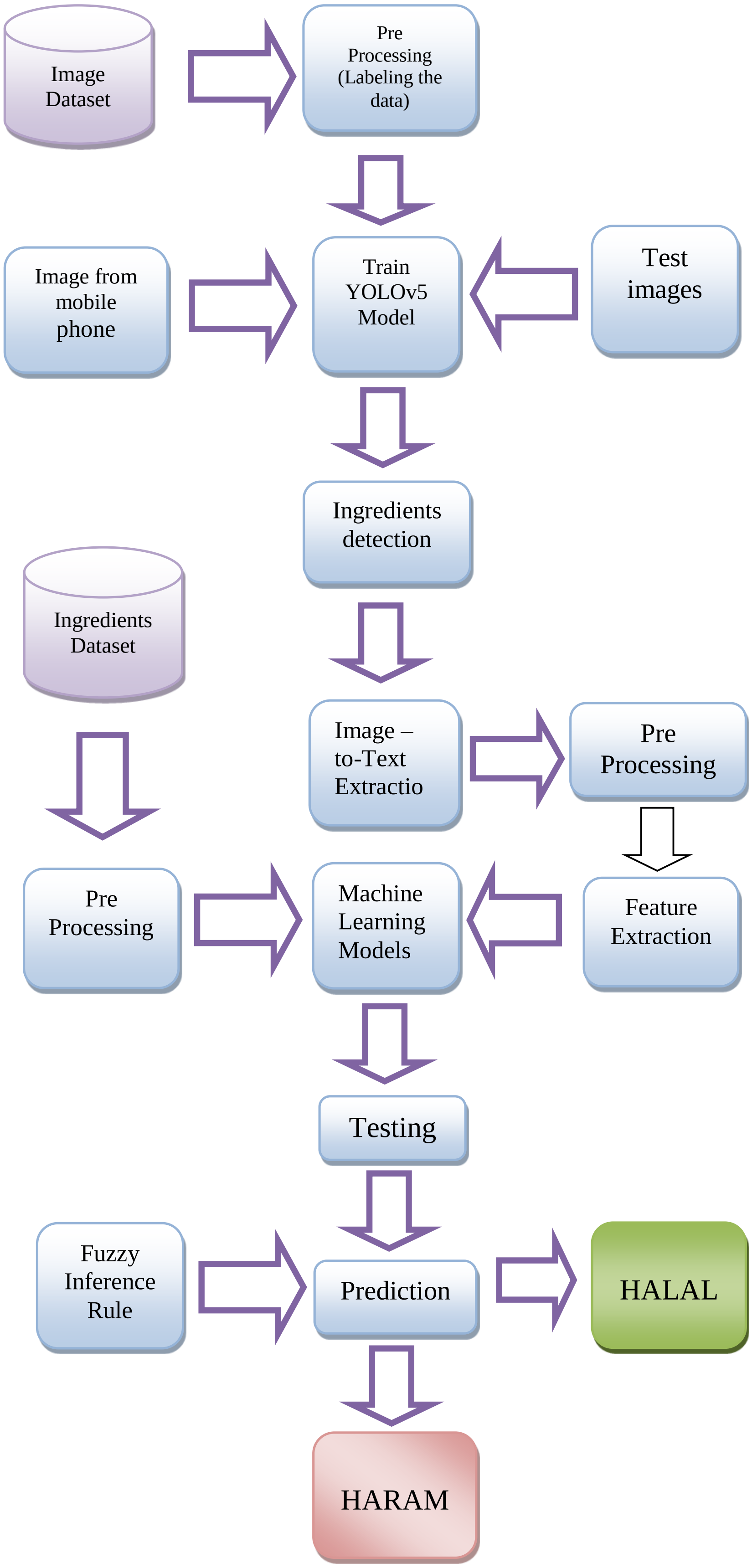

Figure 3.1: Methodology of the System

3.1 Data Acquisition

The data retrieved addressed food and drinks, and it contains the product name, nutrition facts and ingredients. Images are collected from different sources for example, various online food store, Google and also using mobile phone cameras. Total 500 labeled images are used to train the YOLOv5 model. For labeling the images for training the YOLO model labelImg software is used. For classifying Halal and Haram foods total 1040 ingredients are used for training the machine learning models. Halal & Haram ingredients are collected from Halal certification provider Islamic Food and Nutrition Council of America (IFANCA) [30] and Halal Foundation [31]. Table 3.1 shows some statistics of the dataset where Halal food is labeled as '1' & Haram food as '0'.

Ingredients	Status
Acetic acid	1 (Halal)
Adenosine 5 monophosphate	0 (Haram)
Alcohol	0 (Haram)
Bacon	0 (Haram)
Beaf	1 (Halal)
Beer	0 (Haram)
Cochineal (e120)	0 (Haram)
Crab	1 (Halal)
Emulsifier	0 (Haram)
Gelatin	0 (Haram)
Ham	0 (Haram)
Parmesan cheese	0 (Haram)
Potassium acetate	1 (Halal)
Preservative	1 (Halal)
Rainbow sprinkles	0 (Haram)

Table 3.1: Statistics of the Dataset

3.2 YOLOv5 for Ingredients Detection

A common object identification model used in computer vision tasks including segmentation, object recognition, and image classification is called YOLOv5. A dataset of 500 images is used to train the model, and 100 images are used to test the model, to identify elements from packaged food. For training, the images are labeled with the locations that contained the packaged food products ingredients. After that data validation is done to detect the ingredients in new images of packaged food products. The system will output the locations of the ingredients in the image, which will later be used to extract ingredient as text and training the machine learning models.

The following figure shows the ingredients detection from food product image using YOLOv5 model.

Figure 3.2: Ingredients Detection using YOLOv5

3.3 Image-To-Text Extraction using OCR

Optical Character Recognition (OCR) is a computer-based electronic method for converting text pictures into machine-encoded text that can subsequently be retrieved and used in text format. The goal of text extraction is to separate each word so that it may be input into the recognition step. This research extracted English language text from images so that the ingredients can be used as input for machine learning models to identify Halal and Haram foods. The following figure shows text extraction from the images of packaged foods.

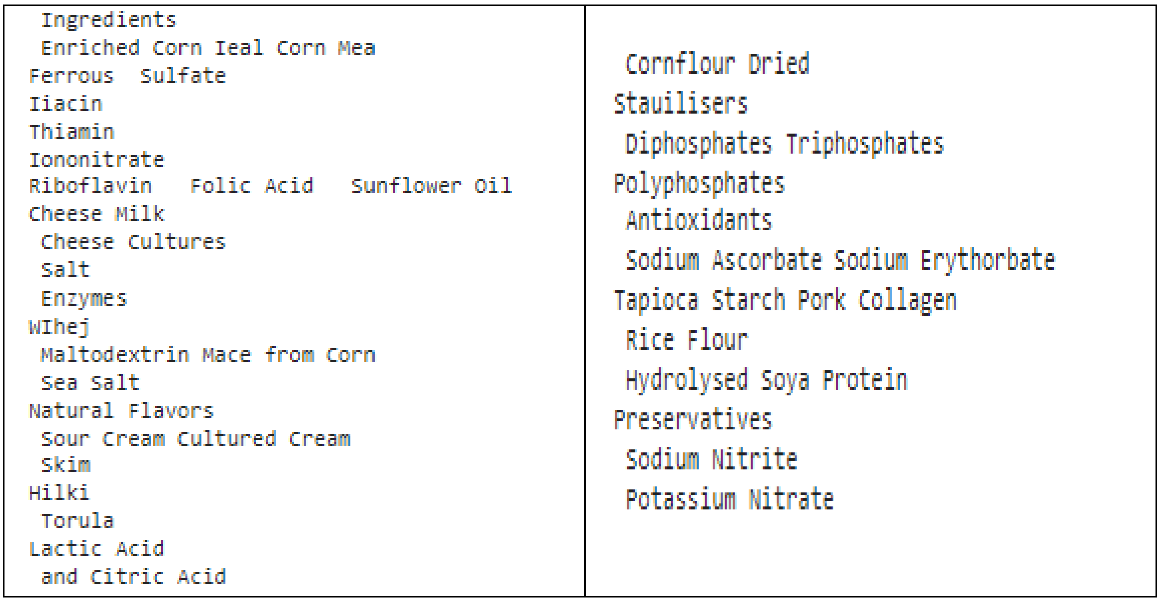

Figure 2.3: Image To Text Extraction using OCR

3.4 Data Preprocessing

The initial stage in this system is to prepare the data. Ingredients are collected from various online sources and Islamic websites [30,31]. Data are then cleaned by eliminating duplicates and missing variables. Labeling of the ingredients is done using binary values, with 1 representing Halal and 0 representing Haram ingredients. After that, the data is divided into two groups: training data and testing data. The testing data is used to evaluate the machine learning model's performance while the training data is used to train the model.

3.4.1Feature Extraction

Scikit-learn CountVectorizer[32] is used to create a vector of word or token counts from a text corpus. This involved converting categorical variables into numerical variables. A very flexible feature representation module for text, it also enables pre-processing of text data prior to constructing the vector representation.

3.5 Machine Learning Models

For identifying if the food containing any Haram ingredients or not a dataset of 1040 ingredients specifying which food is Halal and which is not is used. Halal foods are labeled as value '1' and others ingredients that are not Halal specified as '0'. Then random forest [23], k-nearest neighbor [24], naive bayes [25], support vector machine [26], decision tree [27] models are used to train with the dataset. After that, model is tested with some new data which is extracted from the food product images for validation. Then the model shows the predicted output list of Halal and Haram ingredients of the packaged food.

3.5.1 Random Forest Model

In this research, a random forest classifier has been used to predict if a food component is Halal or Haram. Several ingredients and their labels are included in the dataset. The model's hyper parameters, such as the number of trees in the forest, the deepest point at which a tree may grow, and the smallest amount of samples needed to divide a node, are selected using a grid search with cross-validation. The dataset is then used to train the random forest model. Random forest gives an accuracy of 91%.

3.5.2 K-Nearest Neighbors (KNN) Model

K-Nearest Neighbors (KNN) is a non-parametric classification and regression approach. In the case of classification, KNN determines the K-nearest neighbors of the test instance in the training set and predicts the class based on the majority class of these neighbors. K is a hyper parameter that may be altered to enhance the functionality of my model.

In my implementation, I utilized KNN model by calculating the distance between data points obtained an accuracy of 97% on testing dataset, which is a good result.

3.5.3 Naive Bayes Model

Naive Bayes is a probabilistic technique and supervised machine learning method used for classification. In my research work I have used Naive Bayes algorithm to classify the food ingredients as Halal and non-Halal. Naïve bayes model has an accuracy of 65% and after 5-fold cross validation it increases its accuracy to 69%.

3.5.4 Support Vector Machine Model

Support vector machines SVM, or supervised learning models, are used in machine learning to analyze data for classification and regression. The support vector machine technique looks for a hyperplane that can discriminate between data points in an N-dimensional space (where N is the number of characteristics). In my research, I have implemented the SVM algorithm for the classification of the food ingredients. For testing dataset, the model has an accuracy of 98%. The model is also able to make predictions on new input data.

3.5.5 Decision Trees Model

The most powerful and widely used tool for categorization and prediction is the Decision Tree. A common machine learning approach for classification and regression applications is the decision tree. In my research, I have implemented the decision tree algorithm for the classification of food ingredients. The trained decision tree model has an accuracy of 93%, suggesting that the decision tree model performed well on the test data.

3.5.6 Multi-Layer Perceptrons Model

Neural networks or multi-layer perceptrons [28], are the most useful sort of neural network. A perceptron is a model of a single neuron that serves as the foundation for larger neural networks. The objective is to create algorithms and data structures that can be used to classify properly Halal and Haram ingredients in food. The model achieved a high accuracy, with an overall accuracy of 98%.

3.5.7 Rule-Based Model

A rule-based classifier [29] is made up of a series of "IF-THEN" rules that are derived statistically from the training data. Fuzzy inference is an artificial intelligence approach that allows for approximation reasoning in settings with uncertainty or ambiguity. To identify the food status, the ingredients from a CSV file are matched to a list of Haram ingredients using rule-based model, FuzzyWuzzy library, a Python library for fuzzy string matching, is utilized for the fuzzy matching. Fuzz.token set ratio is a function in the FuzzyWuzzy library that compares two strings and returns a score between 0 and 100 based on how similar the strings are. Because the function considers the order of the words, it is ideally suited for matching ingredient lists. The loop search the list of ingredients from the CSV file and compares each one to the list of Haram ingredients using the fuzz.token_set_ratio function. If any Haram ingredients are found in the list of ingredients, then it will return "Haram". If no Haram ingredients are found, then returns "Halal".

3.6 Five-Fold-Cross Validation

In cross-validation, the data is divided into k subsets, the model is trained on k–5 subsets, and the model is then tested on the remaining subset. Each subset is utilized as the evaluation set once over the course of this operation, which is repeated k times. A final accuracy rating is then calculated by averaging the iteration outcomes. The data is split into training and testing sets using five-fold cross-validation. This method helps avoid over fitting by guaranteeing that each instance of the dataset is used for testing and training at least once. Additionally, it guarantees that every subset has the same number of samples from each class as the original dataset. Each algorithm's average accuracy was derived by averaging the accuracy values achieved in each fold. The accuracy scores obtained for each algorithm are as follows:

Models	Accuracy
Random Forest (RF)	98%
Decision Tree (DT)	98%
Naive Bayes (NB)	68%
Support Vector Machine (SVM)	68%
Multi-Layer Perceptron (MLP)	97%
K-Nearest Neighbor (KNN)	98%

Table 1.2: Accuracy of the Models after 5-Fold Validation

In terms of accuracy, greatest accuracy score is 98% for Random Forest and Decision Tree, followed closely by 98% for KNN. Multi-Layer Perceptron has a high accuracy score of 97% as well, but Naive Baiyes and Support Vector Machine have lower accuracy values of 68%. This implies that Random Forest, Decision Tree, and KNN are more appropriate for this dataset than Naive Bayes and Support Vector Machine.

3.7 Prediction using New Data

Some new packaged food product images are used to validate the models. For this purpose, close test is performed on unlabeled testing data. The trained models used to classify new food products ingredients as Halal or Haram. The ingredients are fed into the trained models and the models predicted whether the product was Halal or Haram. The accuracy of the predictions is then recorded.

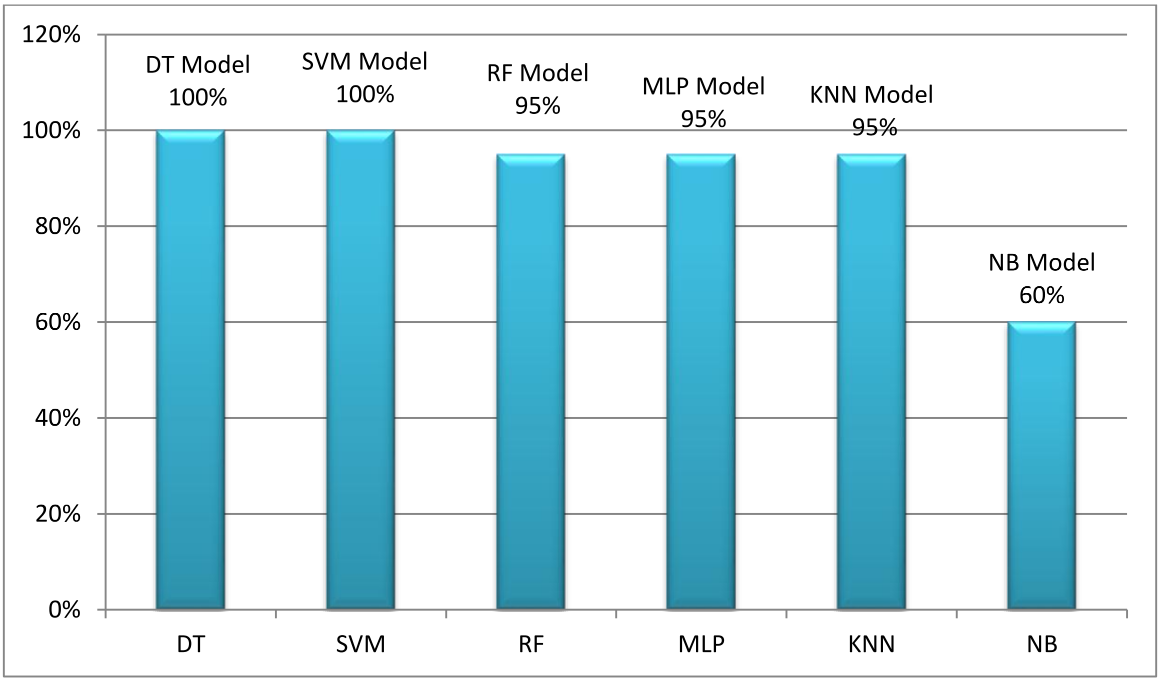

Figure 3.3: Accuracy of Models using New Dataset

The above figure 3.4 shows the accuracy comparison between different machine learning models evaluating with new unlabeled dataset. The new data classification accuracy results were recorded for each of the six models. The DT and SVM both models have accuracy of 100% using new dataset. RF, MLP and KNN got accuracy of 95%. Comparing with other models NB has a lower accuracy rate with 60%. In this research, both testing set accuracy and new dataset classification accuracy were used to evaluate the models. In testing set the status of the ingredients are already given. In new dataset the status of the ingredient is identified through machine learning models and later it is validated by three Islamic scholars.

Chapter 4

Results & Discussion

In my research work, I have used six different machine learning algorithms - Random Forest, K-Nearest Neighbors, Naive Bayes, Support Vector Machine, Decision Tree, and Multi-Layer Perceptron for the classification of packaged food product ingredients. Here is the description of the classification report for each of the models and discussing their performance.

4.1 Random Forest Model Evaluation

The model's performance is presented by a classification report where the classification report displays the accuracy, recall, and F1 score for both Halal and Haram classes.
On the testing data, the model performed well, with an accuracy of 91%, indicating a relatively high level of overall performance. The macro average F1 score was 0.53, which is the average of both classes' F1 scores. The weighted average F1 score was 0.92, which accounts for the class distribution imbalance.

	Precision	Recall	F1 Score	Support
0	0.08	0.14	0.10	7
1	0.97	0.94	0.95	196
Accuracy			0.91	203
Macro average	0.52	0.54	0.53	203
Weighted average	0.94	0.91	0.92	203

Table 2.1: Classification Report of Random Forest

4.2 KNN Model Evaluation

The evaluation results indicate that the k-NN model performed well in identifying Halal and Haram food ingredients. The model achieved an overall accuracy of 97% on the testing data, correctly classifying Halal and Haram ingredients of food products.
The average F1 score is 0.72, indicating moderate overall performance. In conclusion, the k-NN model performed well in recognizing Halal food items, with excellent precision and recall scores for the Halal class.

	Precision	Recall	F1 Score	Support
0	0.50	0.43	0.46	7
1	0.98	0.98	0.98	196
Accuracy			0.97	203
Macro average	0.74	0.71	0.72	203
Weighted average	0.96	0.97	0.96	203

Table 4.2: Classification Report of K-NN

4.3 Naive Bayes Model Evaluation

On the testing data, the model achieved an accuracy of 76%. The model demonstrated average F1 score is 0.68, precision 0.80 and precision 0.67.

	Precision	Recall	F1 Score	Support
0	0.37	0.87	0.51	30
1	0.97	0.74	0.84	173
Accuracy			0.76	203
Macro average	0.67	0.80	0.68	203
Weighted average	0.88	0.76	0.79	203

Table 4.3: Classification Report of Naive Bayes

4.4 SVM Model Evaluation

The model achieved an overall accuracy of 98% on the testing data, indicating excellent performance. The average F1 score is 0.83, which is the average of both classes' F1 scores, indicating that the model's performance on both classes was reasonably balanced.

	Precision	Recall	F1 Score	Support
0	0.80	0.57	0.67	7
1	0.98	0.99	0.99	196
Accuracy			0.98	203
Macro average	0.89	0.78	0.83	203
Weighted average	0.98	0.98	0.98	203

Table 4.4: Classification Report of SVM

4.5 Decision Trees Model Evaluation

On the testing data, the model achieved a high accuracy of 93%. The average F1-score is 0.83, which is the average of the F1-scores of both classes, and the weighted-average F1-score was 0.92, which considers the class distribution in the data. The model’s precision score is 0.96 and recall is 0.77.

	Precision	Recall	F1 Score	Support
0	1.00	0.53	0.70	30
1	0.93	1.00	0.96	173
Accuracy			0.93	203
Macro average	0.96	0.77	0.83	203
Weighted average	0.94	0.93	0.92	203

Table 4.5: Classification Report of Decision Tree

4.6 MLP Model Evaluation

The MLP Classification report shows, the binary classification model on a dataset with two classes. On the testing data, the model achieved a high accuracy, with an overall accuracy of 0.98. The average F1 score is 0.83, which is the average of both classes' F1 scores. The precision score is 0.89 and recall is 0.78.

	Precision	Recall	F1 Score	Support
0	0.80	0.57	0.67	7
1	0.98	0.99	0.99	196
Accuracy			0.98	203
Macro average	0.89	0.78	0.83	203
Weighted average	0.98	0.98	0.98	203

Table 4.6: Classification Report of MLP Classifier

In conclusion, the evaluation metrics show that the model's performance is satisfactory for classification purposes, and it can be used to classify new instances with a high level of accuracy.

The entire machine learning algorithm's performance - Random Forest, K-Nearest Neighbors, Naive Bayes, Support Vector Machine, Decision Tree, and Multi-Layer Perceptron are compared from the classification report and has been discussed below in Table 4-2.

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	91%	0.52	0.54	0.53
K-Nearest Neighbors	97%	0.74	0.71	0.72
Naive Bayes	76%	0.67	0.80	0.68
Support Vector Machine	98%	0.89	0.78	0.83
Decision Tree	93%	0.96	0.77	0.83
Multi-Layer Perceptron	98%	0.89	0.78	0.83

Table 4.7: Performance Evaluation of Different Machine Learning Models

4.7 Fuzzy Inference Rule

The fuzzy inference rule model was designed to categorize food products as halal or haram based on the presence of haram elements in the product. The dataset utilized for this model assessment was a CSV file comprising food product information as well as the presence or absence of haram substances in each product. Overall, the developed model using fuzzy inference rule was able to effectively classify food products as halal or haram based on the presence of haram ingredients. The following figure shows two lists of foods: a list of Halal and a Haram food. Among 5 arbitrary food items ingredients are compared & the numbers of matches of Halal & Haram food ingredients are counted. If any of the food ingredients contain Haram food element, then the food Status indicates as "Haram" else "Halal". This model can be used to help food regulators and halal certification organizations to ensure that food products comply with halal regulations.

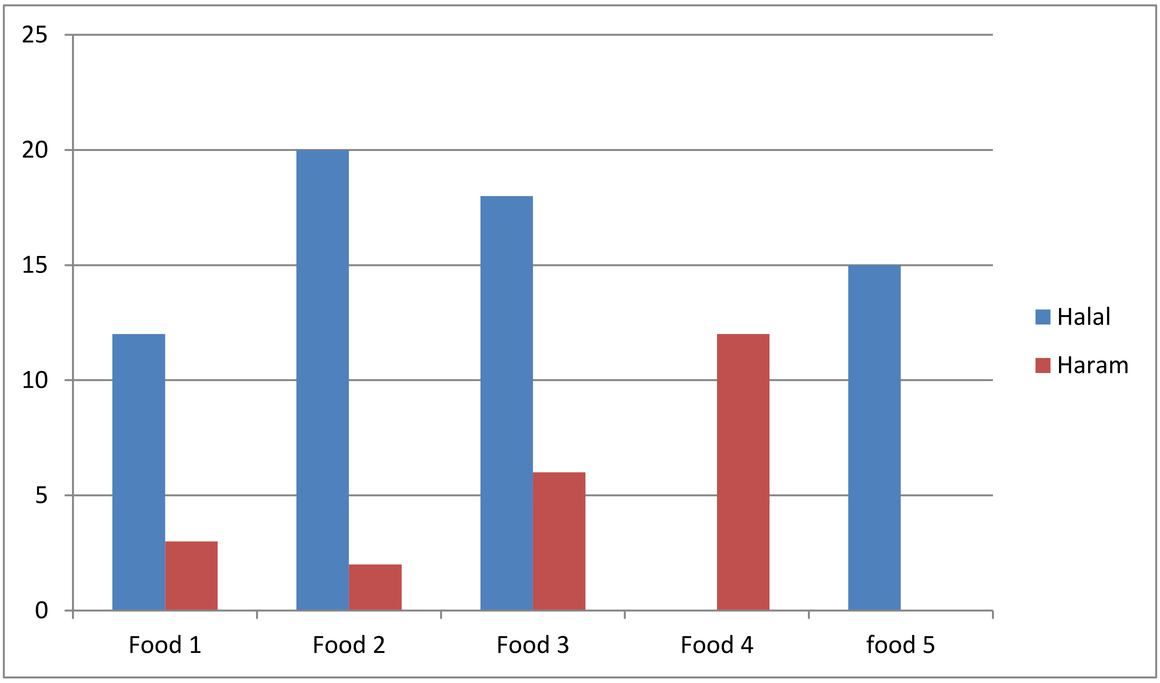

Figure 4.1: Rule-based model for food classification

4.8 Accuracy Comparison between Open and Close Test Data

In this experiment, different machine learning models is used to predict the status of food ingredients: Random Forest, K-Nearest Neighbors, Naive Bayes, Support Vector Machine, Decision Tree, and Multi-Layer Perceptron. The models were trained on a labeled dataset of dietary ingredients labeled with their Halal or Haram status. After training, the models are evaluated on a different test set and achieved high testing accuracy, as shown in the table's "Open Test Accuracy" column.

Moreover, the models are tested on an unknown dataset (unlabeled data), and their accuracy on this dataset is indicated in the table's "Close Test Accuracy" column. For the new dataset, the K-Nearest Neighbors and Support Vector Machine obtained 100% accuracy, while the Random Forest and Decision Tree models achieved 95% accuracy. For the close test, however, the Naive Bayes model had a lower accuracy of 60%.

Overall, the findings show that the K-Nearest Neighbors and Support Vector Machine models are over fitted. As the dataset is relatively small, so it is prone to overfitting. To address this, cross-validation is applied to prevent overfitting.

Models	Open Test Accuracy	Close Test Accuracy
Random Forest	91%	95%
K-Nearest Neighbors	97%	100%
Naive Bayes	76%	60%
Support Vector Machine	98%	100%
Decision Tree	93%	95%
Multi-Layer Perceptron	98%	95%

Table 4.8: Accuracy comparison between Open & Close Test data

4.9 ROC CURVE

ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. The TPR, also known as sensitivity, is defined as the ratio of accurately predicted to actual positive samples. The fall-out rate, or FPR, is defined as the ratio of mistakenly anticipated positive samples to the total number of actual negative samples. The x-axis of a ROC curve shows the FPR, whereas the y-axis represents the TPR.

The graph depicts how successfully the classifier can differentiate between positive and negative samples at various levels. A perfect classifier will have a ROC curve that crosses through the plot's top-left corner, where the TPR is 1 and the FPR is 0.

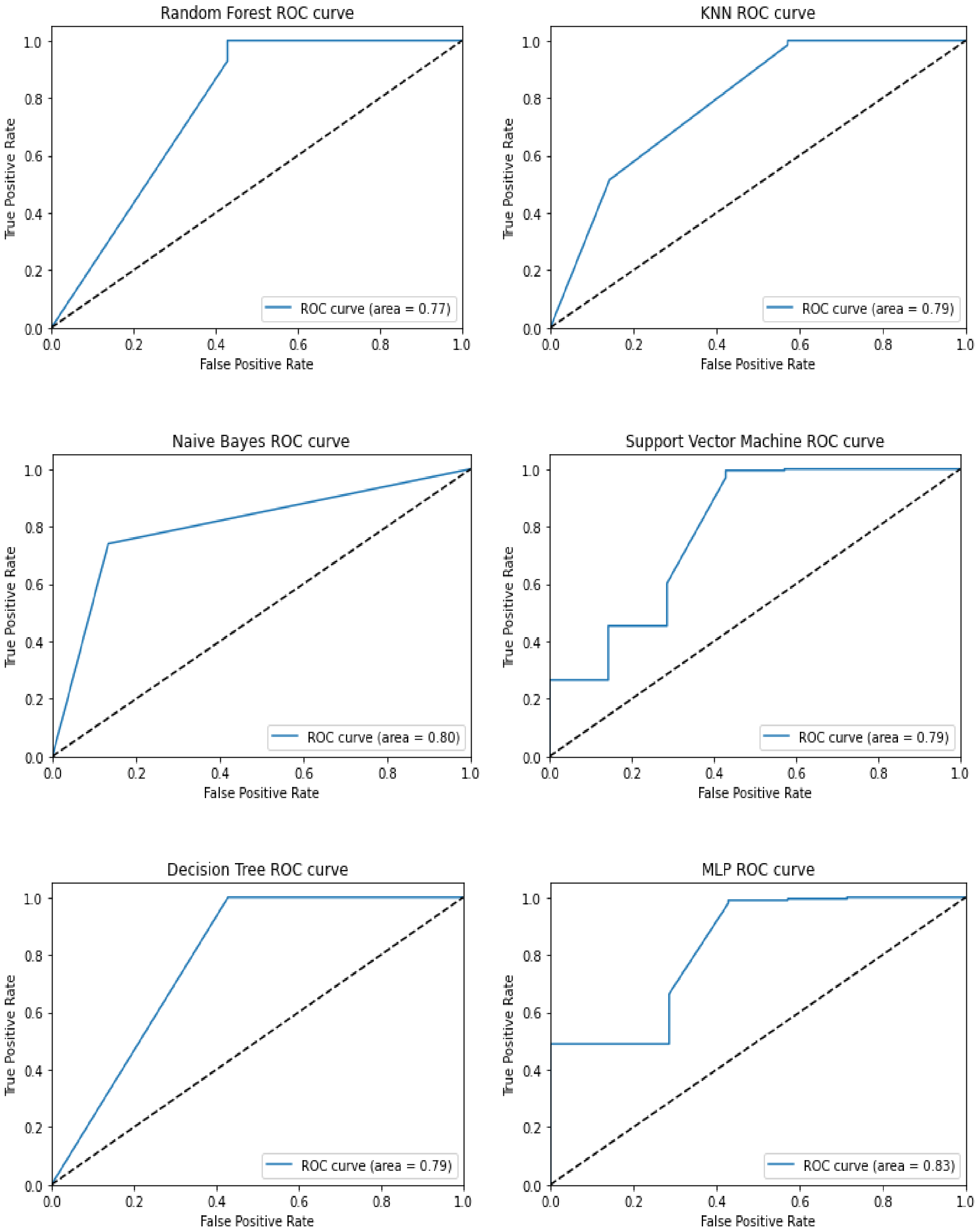

Figure 5.2: ROC Curve of Models

The plot shows the ROC curve for a binary classification models Random Forest, K-Nearest Neighbors, Naive Bayes, Support Vector Machine, Decision Tree, Multi-Layer Perceptron. The x-axis represents the False Positive Rate (FPR) and the y-axis represents the True Positive Rate (TPR). The ROC curve depicts how effectively the model distinguishes between positive and negative data at various threshold values. The dotted line indicates the classifiers, whereas the solid blue line displays the Random Forest, K-Nearest Neighbors, Naive Bayes, Support Vector Machine, Decision Tree, Multi-Layer Perceptron model's ROC curve. The area under the ROC curve (AUC) is displayed in the legend, which is a metric that summarizes the overall performance of the model. A higher AUC indicates better performance.

4.10 Confusion Matrix

The confusion matrix summarizes the observed and anticipated classes, as well as the number of samples in each class. The following Table 4.9 shows the confusion matrix for the Random Forest, K-Nearest Neighbors, Naive Bayes, Support Vector Machine, Decision Tree, and Multi-Layer Perceptron classification models.

	False Negative	False Positive	True Positive	True Negative
Random Forest	0	3	196	4
KNN	3	4	193	3
Naive Bayes	45	4	128	26
SVM	1	3	195	4
Decision Tree	0	3	196	4
MLP	2	3	194	4

Table 4.9: Confusion Matrix of Models

4.11Validation

In order to ensure that the food ingredients are classified accurately as Halal, three Islamic scholars were consulted to provide their opinions on the classification of these food ingredients. However, it was discovered that the scholars provided different opinions regarding the classification of certain ingredients. A list of these ingredients and their varying classifications by the scholars is provided below:

Ingredients	Islamic Scholar 1	Islamic Scholar 2	Islamic Scholar 3
Gelatin	Halal	Halal	Haram
Enzymes	Haram	Halal	Haram
Emulsifiers	Haram	Haram	Halal
Renin Rennet	Halal	Haram	Haram
Animal shortening	Haram	Halal	Haram
Shortening Margarine	Haram	Halal	Haram
Whey	Halal	Haram	Halal
Cochineal	Haram	Haram	Halal
Carmine color	Halal	Haram	Halal
Yeast Extract from brewer's yeast	Halal	Haram	Haram
Proteins Sausages	Haram	Haram	Halal

Table 4.10**:** Doubtful Ingredients

It is clear from the above list that there were differing opinions among the scholars regarding the classification of certain ingredients. This highlights the importance of consulting multiple scholars to ensure that the classification of ingredients is as accurate as possible.

In conclusion, the overall process of consultation and validation demonstrates a commitment to ensuring that the food is compliant with Islamic dietary laws. The following table shows the comparison between machine learning models decision & Islamic scholar decision on some random ten products ingredients.

<table>
<tr><td rowspan="3">Feature Vector</td><td colspan="8">Product-I:
(pork, water, sea salt, natural cane sugar, celery powder, cherry powder)</td></tr>
<tr><td>Islamic Scholar</td><td>RF Model</td><td>KNN Model</td><td>NB Model</td><td>SVM Model</td><td>DT Model</td><td>MLP Model</td><td>Rule Based</td></tr>
<tr><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td></tr>
<tr><td rowspan="3">Feature Vector</td><td colspan="8">Product-II:
(cauliflower, mozzarella cheese, salt, enzymes, parmesan cheese, coconut flour; dried whole egg, almond flour, tapioca starch, liquid whole egg, sodium bicarbonate; salt, xanthan gum, spices, whole eggs, water, soybean oil, whole milk, xanthan gum, salt, citric acid, bacon cured, sugar, smoke flavoring, sodium phosphate, sodium erythorbate, sodium nitrite, cheddar cheese, tree nut, almond, coconut)</td></tr>
<tr><td>Islamic Scholar</td><td>RF Model</td><td>KNN Model</td><td>NB Model</td><td>SVM Model</td><td>DT Model</td><td>MLP Model</td><td>Rule Based</td></tr>
<tr><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td></tr>
<tr><td rowspan="3">Feature Vector</td><td colspan="8">Product III:
(flour blend, brown rice, sorghum and buckwheat, brown sugar semisweet chocolate chips, cane sugar, unsweetened chocolate, cocoa butter, concentrated fruit juices, pear, grape, date paste, safflower oil, water, brown rice syrup, natural flavors , leavening baking soda, salt, vanilla extract, xanthan gum, konjac gum, rosemary extract)</td></tr>
<tr><td>Islamic Scholar</td><td>RF Model</td><td>KNN Model</td><td>NB Model</td><td>SVM Model</td><td>DT Model</td><td>MLP Model</td><td>Rule Based</td></tr>
<tr><td>Haram</td><td>Halal</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td></tr>
<tr><td rowspan="3">Feature Vector</td><td colspan="8">Product IV:
(enriched corn meal corn meal, ferrous sulfate, niacin, thiamin mononitrate, riboflavin, folic acid, sunflower oil, cheddar cheese, milk, cheese cultures, salt, enzymes, whey, maltodextrin made from corn, sea salt, natural flavors, sour cream cultured cream, skim milk, torula, lactic acid, citric acid , yeast)</td></tr>
<tr><td>Islamic Scholar</td><td>RF Model</td><td>KNN Model</td><td>NB Model</td><td>SVM Model</td><td>DT Model</td><td>MLP Model</td><td>Rule Based</td></tr>
<tr><td>Haram</td><td>Halal</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td></tr>
<tr><td rowspan="3">Feature Vector</td><td colspan="8">Product V:
(ham cured with water, salt, sodium phosphates, carrageenan, sodium erythorbate, sodium nitrite, smoke flavoring, sugar)</td></tr>
<tr><td>Islamic Scholar</td><td>RF Model</td><td>KNN Model</td><td>NB Model</td><td>SVM Model</td><td>DT Model</td><td>MLP Model</td><td>Rule Based</td></tr>
<tr><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td></tr>
<tr><td></td><td colspan="8">Product VI:</td></tr>
</table>

<table>
<tr><td rowspan="3">Feature Vector</td><td colspan="8">(roasted almond kernels, sugar, cocoa solids, cocoa butter, mango flavor, emulsifiers (322,476), glazing agents (shollac), contains added natural flavoring substances)</td></tr>
<tr><td>Islamic Scholar</td><td>RF Model</td><td>KNN Model</td><td>NB Model</td><td>SVM Model</td><td>DT Model</td><td>MLP Model</td><td>Rule Based</td></tr>
<tr><td>Haram</td><td>Halal</td><td>Haram</td><td>Haram</td><td>Halal</td><td>Halal</td><td>Haram</td><td>Haram</td></tr>
<tr><td rowspan="3">Feature Vector</td><td colspan="8">Product VII:
(enriched unbleached flour, wheat flour, malted, barley, ascorbic acid , niacin, reduced iron, thiamin mononitrate, riboflavin , folic , acid, sugar, degermed yellow cornmeal , salt, leavening baking soda, sodium acid pyrophosphate, soybean oil, powder; natural flavor , milk, eggs, soy and tree nuts flour, honey)</td></tr>
<tr><td>Islamic Scholar</td><td>RF Model</td><td>KNN Model</td><td>NB Model</td><td>SVM Model</td><td>DT Model</td><td>MLP Model</td><td>Rule Based</td></tr>
<tr><td>Haram</td><td>Halal</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td><td>Haram</td></tr>
<tr><td rowspan="3">Feature Vector</td><td colspan="8">Product VIII:
(whole wheat grain, mixed corn starch, sugar, vitamin e)</td></tr>
<tr><td>Islamic Scholar</td><td>RF Model</td><td>KNN Model</td><td>NB Model</td><td>SVM Model</td><td>DT Model</td><td>MLP Model</td><td>Rule Based</td></tr>
<tr><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td></tr>
<tr><td rowspan="3">Feature Vector</td><td colspan="8">Product IX:
(cooked marinated chicken, water, tomato puree, ginger puree, palm olein, garlic puree, yogurt powder, milk, corn flour, salt, green chilies, spices, color, paprika extract, basil, onions, single cream milk, chopped tomatoes, tomato paste, rapeseed oil, spices, ginger puree, garlic puree, cashew nuts, coriander, honey, unsalted butter, green chili puree, salt, lemon oil)</td></tr>
<tr><td>Islamic Scholar</td><td>RF Model</td><td>KNN Model</td><td>NB Model</td><td>SVM Model</td><td>DT Model</td><td>MLP Model</td><td>Rule Based</td></tr>
<tr><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td></tr>
<tr><td rowspan="3">Feature Vector</td><td colspan="8">Product X:
(jalapenos, onions, tomatillos, water, sunflower oil, apple cider vinegar, serrano peppers, sea salt, garlic, lime juice)</td></tr>
<tr><td>Islamic Scholar</td><td>RF Model</td><td>KNN Model</td><td>NB Model</td><td>SVM Model</td><td>DT Model</td><td>MLP Model</td><td>Rule Based</td></tr>
<tr><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td><td>Halal</td></tr>
</table>

Table 3.11: Comparison between Machine Learning Models & Islamic Scholar Decision

Chapter 5

Conclusion & Future Works

5.1 Conclusion

This study provides a unique approach for detecting Halal items, using deep learning and machine learning techniques. Yolo v5 algorithm is used for the suggested system to inspect packaged food product images and identifies the product's ingredients. Images are annotated for training the Yolo model. Then the trained model is tested with images so that it can recognize the ingredients from the food product image. After that, the text from the ingredient's image is then recognized using optical character recognition. The text from the ingredients images are extracted & revoking unnecessary data, these text of food ingredients are tested with various machine learning, neural network & rule based model for determining the status of the food. In case of training with different machine learning techniques, artificial neural networks, and fuzzy interference rules a dataset of food ingredients with labeled is used. Data are labeled using binary values, with 1 representing Halal and 0 representing Haram ingredients. After training, models are capable to classify the food ingredients as Halal or Haram. The outputs of the models are also validated with three Islamic scholars. The results show that the suggested approach is effective and dependable in distinguishing between Halal and Haram items. The system's accuracy rate is 98% which is high enough and might be a great tool for Muslim consumers to swiftly and easily identify Halal food items, especially when traveling to new places or meeting unfamiliar products.

5.2 Limitations

Considering the optimistic outcomes, the system has certain limitations. The limitations of the system are as follows:

- One of the main limitations is that the proposed system's accuracy rate may be affected by the quality of the images used for analysis. If the images are of poor quality, the system may not be able to recognize the text on the ingredient's image accurately.

- While using YOLOv5 to detect ingredients in packaged food products may not be completely accurate since the model may struggle to discriminate between similar-looking ingredients or detect small quantities of ingredients.
- The detection system is overfitting as the training data size is not enough and the model trains on the limited training data for several epochs. Deep neural network is prone to overfitting because they learn millions or billions of parameters while building the model. That's why deep learning has some drawbacks compared to traditional machine learning, as the need for a lot of data and computing resources to train and deploy, which is also time-consuming. It requires very large amount of data in order to perform better with advance techniques.
- The proposed system is designed to recognize text from ingredient images, but it may not be able to identify ingredients in languages other than the system's trained languages.
- The proposed system only considers Halal certification as the basis for determining the status of the food. The system does not consider other factors that may affect a Muslim's decision to consume a particular food item, such as cross-contamination with Haram foods.
- The system is capable of only distinguishing between Halal and Haram food ingredients. It does not take into account other elements that may be important to Muslim customers, such as ethical or health concerns. As a result, the system's utility is restricted to the value of Halal food identification as the major factor for purchasing food.

5.3 Future Works

- Will develop the image quality enhancement techniques, such as image denoising or image sharpening, to improve the accuracy of the system.
- Will use cutting-edge object detector YOLO model to advance the accuracy and performance of my work.
- The system will expand the language support to recognize ingredients to better serve Muslim customers from different regions.

- Will develop the system to integrate with external databases, such as Halal certification authorities, to obtain updated and reliable Halal certification information.
- Will develop the system to identify the food status where the food label doesn't contain ingredients whether it is omitted as a smaller percentage.

References

[1] N. A. Rakhmawati, A. A. Firmansyah, P. M. Effendi, R. Abdillah, and T. A. Cahyono, “Auto Halal detection products based on euclidian distance and cosine similarity,” Int. J. Adv. Sci. Eng. Inf. Technol., vol. 8, no. 4–2, pp. 1706–1711, 2018, doi: 10.18517/ijaseit.8.4-2.7083.

[2] N. A. Rakhmawati, G. K. Utomo, R. Indraswari, and I. R. Susetyo, “Halal Food Prediction Using the Similarity Graph Algorithms,” Int. J. Food Syst. Dyn., vol. 13, no. 2, pp. 165–173, 2022, doi: 10.18461/ijfsd.v13i2.B4.

[3] M. Kartiwi, T. S. Gunawan, A. Anwar, Siti, and S. Fathurohmah, “Mobile Application for Halal Food Ingredients Identification using Optical Character Recognition,” Meas. Appl., no. November, pp. 1–4, 2018.

[4] Fadhilah, Hamid, Esmeralda C. Djamal, Ridwan Ilyas, and Asep Najmurrokhman. "Non-halal ingredients detection of food packaging image using convolutional neural networks." In 2018 International symposium on advanced intelligent informatics (SAIN), pp. 131-136. IEEE, 2018.

[5] Khairudin, Nor Emilia Zetty, and Uneversiti Malaysia Pahang. "Halal Food Recognition System Using Barcode." PhD diss., UMP, 2010.

[6] T. Akram, H. H. Rizvi, S. A. Ali, S. M. Hamza, and A. Ifthikhar, “OCR and Barcode based Halal and Health Analyzer,” ICISCT 2020 - 2nd Int. Conf. Inf. Sci. Commun. Technol., 2020, doi: 10.1109/ICISCT49550.2020.9080036.

[7] Murizah Kassim, Cik Ku Haroswati, Che Ku Yahayal, Muhammad Hafiz Mazlan Zaharuddin, Zamri Abu Bakar, “A Prototype of Halal Product Recognition System,”

[8] Kohilavani, W. N. Wan Abdullah, T. A. Yang, S. A. din Sifat, and W. Zaman, “Development of Safe Halal Food Management System (SHFMS),” Food Control, vol. 127, no. April, p. 108137, 2021, doi: 10.1016/j.foodcont.2021.108137.

[9] Ilyas Masudin, Bima Bagos Rahmatullah, Muhammad Armand Agung, Inka Amalia Dewanti and Dian Palupi Restuputri “Traceability System in Halal Procurement_ A Bibliometric Review” September 2022, DOI:10.3390/logistics6040067.

[10] K. Nakyinsige, Y. B. C. Man, and A. Q. Sazili, “Halal authenticity issues in meat and meat products,” Meat Sci., vol. 91, no. 3, pp. 207–214, 2012, doi: 10.1016/j.meatsci.2012.02.015.

[11] M. H. Yuswan et al., “Hydroxyproline determination for initial detection of halal-critical food ingredients (gelatin and collagen),” Food Chem., vol. 337, no. November 2019, 2021, doi: 10.1016/j.foodchem.2020.127762.

[12] N. T. Putri, A. Kharisman, I. Arief, H. H. Abdul Talib, K. R. Jamaludin, and E. A. Ismail, “Designing Food Safety Management and Halal Assurance Systems in Mozzarella Cheese Production for Small-Medium Food Industry,” Indones. J. Halal Res., vol. 4, no. 2, pp. 65–84, 2022, doi: 10.15575/ijhar.v4i2.12996.

[13] W. Arwani, A. Jaelani, S. Firdaus, and T. F. Hanim, “Halal-Thayyib, Food Products, and the Halal Industry: A Thematic Analysis on the Verses of the Qur’an,” Al-Amwal J. Ekon. dan Perbank. Syari’ah, vol. 14, no. 1, p. 99, 2022, doi: 10.24235/amwal.v14i1.10830.

[14] S. Hasyim, “The Politics of ‘Halal’: From Cultural to Structural Shariatisation in Indonesia,” Aust. J. Asian Law, vol. 22, no. 1, pp. 81–97, 2022.

[15] M. Mortas, N. Awad, and H. Ayvaz, “Adulteration detection technologies used for halal/kosher food products: an overview,” Discov. Food, vol. 2, no. 1, 2022, doi: 10.1007/s44187-022-00015-7.

[16] P. C. Ng et al., “Recent advances in halal food authentication: Challenges and strategies,” J. Food Sci., vol. 87, no. 1, pp. 8–35, 2022, doi: 10.1111/1750-3841.15998.

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

[18] R. Sarno, K. Triyana, S. I. Sabilla, D. R. Wijaya, D. Sunaryono, and C. Fatichah, “Detecting Pork Adulteration in Beef for Halal Authentication using an Optimized Electronic Nose System,” IEEE Access, vol. 8, pp. 25–30, 2020, doi: 10.1109/ACCESS.2020.3043394.

[19] E. V. Joice Lavandoski, Adriana Brambilla, “HALAL TOURISM the importance of halal food,” J. Chem. Inf. Model., vol. 53, no. 9, pp. 1689–1699, 2013.

[20] LabelImg. (2022). (Version 1.8.0) [Computer software]. Available from https://github.com/heartexlabs/labelImg. Accessed June 24, 2023.

[21] Johnson, A. (2021). YOLOv5 (Version 5.0) [Computer software]. Available from https://github.com/ultralytics/yolov5. Accessed June 24, 2023.

[22] Smith, J. (2022). EasyOCR (Version 2.3.1) [Computer software]. Available from https://github.com/JaidedAI/EasyOCR Accessed June 24, 2023.

[23] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324.

[24] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21-27.

[25] Webb, G.I. (2011). Naive Bayes. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_576

[26] Cristianini, N., Ricci, E. (2008). Support Vector Machines. In: Kao, MY. (eds) Encyclopedia of Algorithms. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30162-4_415

[27] Fürnkranz, J. (2011). Decision Tree. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_204

[28] scikit-learn neural_network.MLPClassifier (Version 0.24.2). Available from https://scikit-

learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html. . Accessed June 24, 2023.

[29] FuzzyWuzzy (Version 0.18.0). Available from https://github.com/seatgeek/fuzzywuzzy. Accessed June 24, 2023.

[30] https://www.ifanca.org. Accessed June 24, 2023.

[31] https://www.halalfoundation.org/. Accessed June 24, 2023.

[32] CountVectorizer (Version 0.12.1). https://scikit-learn.org / stable / modules/ generated/ sklearn.feature_extraction.text.CountVectorizer.html. Accessed June 24, 2023.

[33] L. O. Nazaruddin, B. Gyenge, M. Fekete-Farkas, and Z. Lakner, “The Future Direction of Halal Food Additive and Ingredient Research in Economics and Business: A Bibliometric Analysis,” Sustain., vol. 15, no. 7, 2023, doi: 10.3390/su15075680.

[34] M. A. Fathoni, S. Suryani, and S. Sumilir, “Determinant of Consumer Awareness Toward Halal Food During The Covid-19 Pandemic: Evidence from Indonesia,” IQTISHODUNA J. Ekon. Islam, vol. 11, no. 1, p. 67, 2022, doi: 10.54471/iqtishoduna.v11i1.1319.

[35] S. N. Wiyono, Y. Deliana, E. Wulandari, and N. H. Kamarulzaman, “The Embodiment of Muslim Intention Elements in Buying Halal Food Products: A Literature Review,” Sustain., vol. 14, no. 20, pp. 1–19, 2022, doi: 10.3390/su142013163.

[36] E. Ordukaya and B. Karlik, “Fruit juice–alcohol mixture analysis using machine learning and electronic nose,” *IEEJ Trans. Electr. Electron. Eng.*, vol. 11, pp. S171–S176, doi: 10.1002/tee.22250.

[37] Fan, Binbin, Rongguang Zhu, Dongyu He, Shichang Wang, Xiaomin Cui, and Xuedong Yao. "Evaluation of Mutton Adulteration under the Effect of Mutton Flavour Essence Using Hyperspectral Imaging Combined with Machine Learning and Sparrow Search Algorithm." *Foods* 11, no. 15 (2022): 2278.