Khazanah Journal of Religion and Technology
Universitas Islam Negeri Sunan Gunung Djati Bandung, Indonesia
https://journal.uinsgd.ac.id/kjrt
Online ISSN: 2987-6060
Volume 3 Issue 2 (2025)
DOI: https://doi.org/10.15575/kjrt.v3i2.1774

Automation of Halal Food Classification Using Bidirectional Long Short-term Memory on Ingredients List

Devi Mulyana
Terakorp Indonesia
Bandung, Indonesia
jangebe91@gmail.com

Dika Haekal Firza Pratama
Laskar Teknologi Mulia (CyberLabs)
Bandung, Indonesia
dikahaekalfirza2@gmail.com

Dailfhiana Fujianti
Al Hilal Quranic Endowment Institution
Bandung, Indonesia
dailfhianafujianti@gmail.com

Dimas Arya Nurhakim
Camy Indonesia
Bandung, Indonesia
dimas.yans338@gmail.com

Daffa Rabbani
Bandung City Communication and Information Service
Bandung, Indonesia
abeydaffa88@gmail.com

Abstract— The global demand for halal food products continues to increase, particularly among Muslim consumers, necessitating an efficient and accurate halal classification system. This study proposes a deep learning-based automatic classification approach using Bidirectional Long Short-Term Memory (BiLSTM) to determine the halal or haram status of a product based on its ingredient list. The system utilizes comprehensive text preprocessing techniques such as normalization, stopword removal, and dictionary-based term mapping. Word representations are converted into dense semantic vectors using word embeddings such as Word2Vec and GloVe. A BiLSTM model is used to capture bidirectional contextual relationships in ingredient sequences, thereby enhancing semantic understanding. Testing results on a dataset of 3,979 samples show that the proposed model achieves a classification accuracy of 99.75%, outperforming traditional machine learning methods such as Naive Bayes and SVM. The system is proven effective in handling ingredient ambiguity and context-based classification, and has potential for real-world applications such as mobile-based halal scanners. Future research can adopt attention mechanisms and transform-based models to improve performance and interpretability.

Keywords- BiLSTM, deep learning, food ingredient list, halal classification, halal food detection, NLP, word embedding.

I. INTRODUCTION

From a religious perspective, some religions have strict dietary rules as part of their teachings that must be adhered to. For example, kosher rules in Judaism and halal rules in Islam. As the global Muslim population grows, the consumption and production of halal food is becoming an increasingly important issue, from a religious, ethical, health, and economic perspective.

In Arabic, "halal" means permissible, while "non-halal" or "haram" means prohibited. Consuming halal food not only reflects adherence to religious teachings but is also closely related to healthy eating, animal welfare, food quality and purity, organic food, and food safety. From an Islamic perspective, consuming halal food contributes to the achievement of the Sustainable Development Goals (SDGs), particularly SDG 2 Zero Hunger, by promoting access to safe, nutritious, and affordable food and supporting food security through ethical and responsible production [1]. This also relates to SDG 4 Quality Education, as halal literacy, through public education, school curricula, professional training, and consumer awareness, can increase nutritional awareness, food safety knowledge, and strengthen communities' capacity to make wise and sustainable consumption decisions.

Essentially, all foods are halal, except those explicitly prohibited, such as: pork products, animal carcasses, human body parts, animals slaughtered without mentioning the name of Allah, blood, alcoholic beverages, wine, ethyl alcohol, canines, birds of prey with claws, and halal foods contaminated with haram substances. Therefore, the halal classification of food, especially processed products, is crucial [2], [3].

One of the main challenges in ensuring the halal status of food is the presence of additives and additives whose halal status is ambiguous. Food additives are crucial because they can increase a product's shelf life, nutritional value, and visual appeal. However, some of them are sourced from ingredients of questionable halal status or pose potential health risks. On the other hand, European studies have shown that additives and additives can actually reflect the quality of a food. This demonstrates the trade-off between benefits and potential harms.

Furthermore, the utilization dimension of global food security now faces various challenges, such as rising obesity, low dietary diversity, and food safety issues. Therefore, a "healthy and sustainable diet" approach is implemented that meets four main criteria:

- Quantity: sufficient energy to maintain life and an ideal body weight.
- Diversity: consuming a variety of nutrient-dense foods.
- Quality: including macro- and micronutrients with minimal additives.
- Safety: ensuring food and beverages are consumed safely.

This diet aligns with Islamic principles, which prioritize not only halal (permissible) but also thayyib (good) aspects, which consider the source, purpose, and quantity of food consumed.

In Indonesia, the country with the largest Muslim population in the world, halal certification is an increasingly complex and multidimensional issue. Although the government has enacted Law No. 33 of 2014 concerning Halal Product Assurance, its implementation still faces significant challenges. Limited human resources, supply chain variations, and diverse industry practices are major obstacles.

This is despite public awareness of halal products continuing to grow. A study shows that 87.2% of Indonesian Muslim consumers consider the halal label a primary factor in choosing food and beverages [4]. The halal label now serves not only as a religious marker but also as a symbol of trust and quality assurance.

Internationally, halal certification has evolved into a quality assurance system with strategic economic value. The global halal market is estimated to reach USD 3.2 trillion by 2024, with Indonesia having a significant opportunity to become a major player. Halal-certified products even have the potential to increase revenues in the food and beverage industry by up to 35%.

However, the challenges of the certification process extend beyond raw materials, encompassing the entire production process, from processing and packaging to distribution. This requires a multidisciplinary approach, involving food technology, Islamic law, supply chain management, and industry regulations.

Advances in digital technology have also opened up new opportunities to increase the transparency and efficiency of the certification process. The emergence of halal tracking applications, digital platforms, and even technologies like blockchain has begun to be utilized to assist consumers and producers in ensuring the halal status of products [2], [3], [5]. This aligns with studies by Ariska et al. (2024), which emphasize the role of technology in increasing producer accountability, and Azam & Abdullah (2020), which highlight the importance of digital platforms in building consumer trust.

In response to the need for efficiency in halal classification, artificial intelligence (AI)-based automation approaches are being developed, particularly for assessing halal quality based on ingredient lists. One [6], [7] method is the use of Long Short-Term Memory (LSTM).

LSTM is an algorithm that can overcome the vanishing gradient problem in RNNs in learning long-term dependencies [8], [9], [10]. Spam detection using LSTM, for example, is a popular application in sequential data classification. LSTM is a recurrent neural network architecture capable of understanding long-term context in sequenced data. In LSTM, RNN nodes are replaced by LSTM cells designed to store past information. LSTM uses three types of gates to control the stored and updated information: input gates, forget gates, and output gates.

With its ability to understand sequential context and efficiently handle text data, LSTM, especially in its bidirectional form, is a suitable choice for automating halal food classification based on ingredient lists. This approach is expected to help consumers and industry assess the halal status of food accurately, quickly, and efficiently.

II. Related Works

The use of machine learning for food classification has become a rapidly growing area of research in recent years. Early research in this domain focused on identifying food types using computer vision [11]. Chen et al. developed a Convolutional Neural Network (CNN)-based food classification system capable of identifying 101 food categories with 88.28% accuracy. However, image-based approaches have limitations in identifying the internal composition of foods that are not visually visible.

Subsequent developments have led to the use of text mining and Natural Language Processing (NLP) for ingredient and food composition analysis. Liu et al. used text classification techniques with Support Vector Machines (SVM) to classify food products based on a list of ingredients, achieving 92.5% accuracy on a dataset of 5,000 products [12]. This approach demonstrates the great potential of text analysis for food classification.

Specific research on halal classification using machine learning shows promising trends. Mustapha et al., in their review of the application of machine learning approaches to halal meat authentication, highlighted that ML offers a fast, flexible, scalable, automated, and lower-cost method with high accuracy and sensitivity compared to traditional methods [13]. Several ML approaches used in halal meat authentication have been shown to have high accuracy rates. Support Vector Machine (SVM) is a supervised learning model widely used for data classification and regression, known for its excellent classification power in detecting meat adulteration. Studies have shown that SVM has been successfully used in halal meat authentication with significant accuracy, even reaching 98% in classifying minced beef, lamb, and chicken adulterated with pork when combined with Fourier transform infrared spectroscopy (FTIR) and multivariate methods.

Tarannum in his thesis explored the use of deep learning and machine learning techniques to detect halal items, with the aim of classifying unknown products as halal or haram based on their ingredients using the YOLOv5 algorithm and OCR for text extraction [3]. This research shows promising results with a system capable of identifying and classifying packaged food products based on ingredient analysis. Following this study, Tarannum et al. suggested a novel approach for halal food recognition based on deep learning and machine learning. This system identifies constituents in packaged food products using the YOLOv5 detection method, which is then used to extract features using an Optical Character Recognition (OCR) system. After preprocessing, the ingredients are classified using a trained deep learning model, a neural network, and a rule-based system, which accurately determines if the food product is Halal or Haram. This study yielded good findings, with the proposed halal packaged food recognition system attaining an accuracy rate of 98%, as validated by comparison with the opinions of three Islamic scholars. This unique approach has enormous potential to help Muslim consumers rapidly recognise Halal-certified foods, especially while navigating. This innovative approach has significant potential to assist Muslim consumers in efficiently recognizing Halal-certified items, especially when navigating unfamiliar environments.

Yusop[14] et al. develop halal certification system using the Random Forest algorithm with TF-IDF-based feature extraction . Their system achieved 87.2% precision and 89.1% recall in classifying halal and non-halal products. However, this study still utilized traditional machine learning methods, which are limited in understanding the semantic context of food ingredients.

Advances in deep learning, particularly in the field of NLP, have brought significant breakthroughs in text classification. Mikolov et al. introduced Word2Vec, which allows for the representation of words in dense vectors, providing a strong foundation for semantic understanding of text [15].

The use of Recurrent Neural Networks (RNN) and its variants, such as Long Short-Term Memory (LSTM), has proven effective in sequence modeling. Hochreiter & Schmidhuber introduced LSTM, which can overcome the vanishing gradient problem in traditional RNNs [16]. A further development is the Bidirectional LSTM, introduced by Schuster & Paliwal, which allows the model to process information from both forward and backward directions within a sequence [17]. Zhou et al. demonstrated the superiority of bidirectional LSTM in text classification tasks, achieving state-of-the-art performance on several benchmark datasets [18]. This model is able to capture long-range dependencies and more complex contexts compared to traditional methods.

In addition to general food classification, food allergen detection is a crucial aspect of food safety. Bianco et al., in their review article "Food allergen detection by mass spectrometry: From common to novel protein ingredients," discuss food allergen detection using mass spectrometry (MS) [19]. This study highlights MS as a promising alternative to traditional antibody-based assays for the quantification of various allergenic proteins in complex matrices with high sensitivity and selectivity. This article summarizes the main allergenic proteins and the advantages and disadvantages of several MS acquisition protocols, such as multiple reaction monitoring (MRM) and data-dependent analysis (DDA), for identifying and quantifying common allergenic proteins in processed foods. This study also includes a section dedicated to novel foods such as microalgae and insects as sources of novel allergenic proteins. MS is considered an effective analytical tool for ensuring regulatory compliance throughout the food chain.

The application of word embeddings in the food domain has shown promising results. Teng et al. used Word2Vec to build a food knowledge graph capable of capturing semantic relationships between food ingredients [20]. Min et al. surveyed various food computing approaches, including the use of domain-specific word embeddings for food nutrition analysis, which demonstrated superior performance compared to general-purpose embeddings [21].

The use of word embeddings in the context of halal classification holds great potential due to their ability to capture complex semantic relationships between ingredients. This is important considering that many ingredients have alternative or derivative names that can affect a product's halal status.

Despite progress in the application of machine learning to food classification and several attempts at halal classification, several significant gaps remain:

- Dataset Limitations: Most studies use relatively small and non-comprehensive datasets. There is no standard benchmark dataset for halal classification that can be used for comparison between different methods [22].

- Ingredient Complexity: Existing research has not adequately addressed the complexity of analyzing ingredients with multiple names or derivatives. For example, gelatin can come from halal (fish) or haram (pork) sources, but existing classification systems often treat gelatin as a single entity [23].
- Contextual Understanding: Traditional machine learning approaches used in previous research have limitations in understanding the context and relationships between ingredients. Rule-based or keyword matching approaches cannot capture the nuanced relationships within ingredient lists [24].
- Multilingual Challenges: Global food products often have ingredient names in multiple languages, but current research generally focuses on a single language.
- Integration with Regulatory Standards: Existing research has not adequately integrated with regulatory standards and guidelines from various halal certification bodies which can differ between countries [25].
- Scalability and Real-time Processing: Existing systems are generally not designed for large-scale use with the real-time processing required in practical applications.

Based on the analysis of the related work above, it can be concluded that there is still significant opportunity to develop a more robust and accurate halal classification system using advanced deep learning techniques, particularly bidirectional LSTM with word embeddings, which can overcome the limitations of current approaches. This research aims to fill this gap by developing a model that can better understand the semantic context of ingredients.

III. RESEARCH METHODS

A. Data preparation

Data preprocessing is a crucial step in developing a text-based halal/haram classification system, especially when the data comes from freely written food ingredient lists by manufacturers. Due to the non-standardized format, data normalization and transformation are the initial steps to ensure the model can learn from a clean and consistent representation.

Figure 1 shows the flow activity of this research. The first step is to remove entries with no data in the text column (dropna). Then, text normalization is performed by converting all letters to lowercase and removing all non-letter characters using regular expressions, including numbers, punctuation, and symbols that do not contribute semantically to the model's understanding. This aims to reduce irrelevant vocabulary variation and maintain focus on meaningful words.

Next, the data labels are encoded using a binary approach: halal labels are assigned a value of 0, and all others (e.g., haram or doubtful) are assigned a value of 1. This process is performed using a transformation function using LabelEncoder. After the cleaning and labeling stages, the tokenization process is performed, which converts the text into a sequence of integers representing word indices. The tokenizer is limited to the top 10,000 words and uses the special <OOV> token to handle out-of-vocabulary words. Each sequence is then padded to a uniform length of 100 tokens to ensure the appropriate input format for the neural network model.

The trained tokenizer is saved in the tokenizer.pickle file for reuse during inference or model evaluation. The dataset is then divided into three subsets: training data (70%), validation data (15%), and testing data (15%), using stratification to maintain class proportions. This process is carried out in two stages, with an initial split between training and provisional data (80:20), followed by an even split between validation and testing data (50:50).

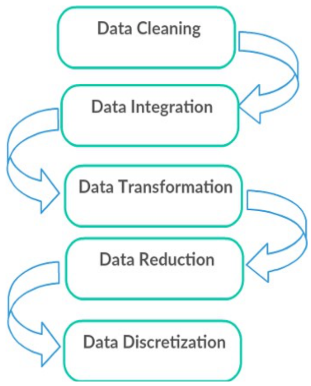

Fig 1. Flowchart of data preparation

B. Bidirectional Long Short-Term Memory (BiLSTM)

LSTM is an algorithm that can overcome the vanishing gradient problem in RNNs by learning long-term dependencies [16]. Spam detection using LSTM, for example, is one popular application in sequential data classification. LSTM is a recurrent neural network architecture capable of understanding long-term context in data sequences. In LSTM, RNN nodes are replaced by LSTM cells designed to store past information. LSTM uses three types of gates to control the information stored and updated: input gates, forget gates, and output gates. The LSTM architecture is shown in Figure 2.

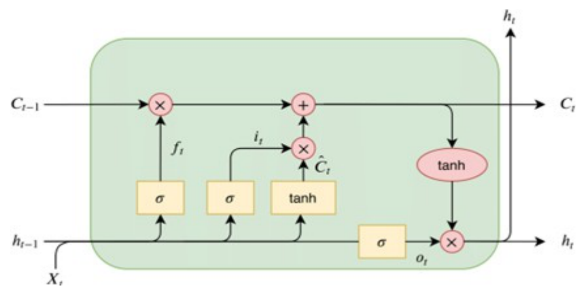

Fig 2. LSTM architecture

An LSTM network consists of memory blocks called cells, which are composed of cell states and hidden cells. Cell states serve as the main pathway for data flow, allowing data to flow smoothly, although linear transformations may occur. Data in cell states can be added or removed through sigmoid gates, which are functions that resemble layers with varying weights. This process begins with a forget gate, which identifies information from the previous step that will be discarded. Information from the previous layer (ht−1) and the current input (xt), plus the weights (Wf) and bias (bf), are processed using the sigmoid function (σ), as shown in the equation:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \qquad (1)$$

The next step is to store the new information in the cell state. This process involves two components: a sigmoid layer to decide whether information should be stored, and a tanh function to determine the value to be added. The values from both layers are combined to update the cell state to a new value (Ct).

With its ability to understand sequential context and efficiently handle text data, LSTM, particularly in its bidirectional form, is a suitable choice for automating ingredient-list-based halal food classification. This approach is expected to assist consumers and industry in assessing the halal status of food accurately, quickly, and efficiently.

The core model used in this research is Bidirectional Long Short-Term Memory (BiLSTM), a development of the Recurrent Neural Network (RNN) architecture specifically designed to process sequences of text data bidirectionally. Unlike standard RNNs, which only process data from beginning to end (forward), BiLSTM consists of two LSTM layers: one processing data from beginning to end, and another from end to beginning. The results from both directions are then combined to form a more complete contextual representation. This capability makes BiLSTM ideal for tasks that require a full understanding of the context of a sequence of words, including the classification of halal and haram based on food ingredient lists.

LSTM (Long Short-Term Memory) itself was introduced as a solution to the weaknesses of traditional RNNs, particularly the vanishing gradient problem that causes the model to fail to remember long-term information. LSTM overcomes this by implementing an internal memory structure called a cell state, which works in conjunction with three types of gates: input gates, forget gates, and output gates. The combination of these three allows LSTM to selectively retain, discard, and output information based on contextual needs. This mechanism enables LSTM to understand the relationships between words scattered throughout long text sequences, which is particularly important in the case of food ingredient lists, which can contain many terms with contextual meaning.

When LSTM is used in its bidirectional form, the model can simultaneously capture the context of the preceding and following words, which is crucial for detecting the true meaning of a particular phrase. For example, the word "enzyme" is generally not problematic, but if it is followed by the words "pork" or "non-halal," the context changes and indicates the ingredient is likely haram. BiLSTM can recognize these semantic relationships, even when the words do not appear directly next to each other, because the model has learned from the complex sequence structure in both directions.

After the sequence learning process by BiLSTM is complete, the model's output is directed to a dense or fully connected layer. This layer serves to translate the contextual representation of BiLSTM into a binary classification result. Since the final task is to decide whether a list of ingredients falls into the halal or haram category, a sigmoid activation function is used to generate a probability between 0 and 1. This value is then compared to a specific threshold (usually 0.5) to determine the final class [26]

The use of BiLSTM has proven highly effective in a variety of natural language processing (NLP) tasks, ranging from sentiment analysis and named entity recognition, to text classification [27]. In the context of halal food classification, BiLSTM offers an advantage because it can understand subtle semantic relationships between words or phrases, such as distinguishing “natural flavors,” commonly found in halal products, from “alcohol flavors,” which indicate potential haramity. By leveraging the power of BiLSTM, the system can learn from complex patterns that cannot be identified with traditional methods [28].

C. Model evaluation

The model training process is carried out by optimizing network parameters to minimize the loss function. The loss function used is Binary Cross-Entropy, which is a standard choice for binary classification problems. The Adam optimizer was chosen due to its ability to adjust the adaptive learning rate, which helps in faster and more stable model convergence. The initial learning rate is set at a specific value (e.g., 0.001) and can be adjusted during training if necessary. The model is trained for a specified number of epochs (e.g., 50-100 epochs), with a batch size (e.g., 32 or 64) to maintain computational efficiency. A validation process is performed at each epoch using the validation data to monitor model performance and detect signs of overfitting. An early stopping technique can be applied to stop training if performance on the validation data does not improve after several epochs. Model performance evaluation is performed on a completely separate test dataset. The evaluation metrics used include [28]:

- Accuracy: The proportion of correct predictions out of the total predictions.
- Precision: The proportion of correct positive predictions out of the total positive predictions. This is especially important if false positives have high consequences (e.g., classifying something haram as halal).
- Recall: The proportion of actual positive cases correctly identified. This is especially important if false negatives have high consequences (e.g., classifying something halal as haram).
- F1-Score: The harmonic mean of precision and recall, providing a balance between the two metrics, which is especially useful for datasets with imbalanced classes.

- Confusion Matrix: A visual representation of the performance of a classification algorithm, showing true positives, true negatives, false positives, and false negatives.
- ROC AUC Score: A score of the area under the ROC curve, indicating the model's ability to discriminate between positive and negative classes across the board.

These metrics provide a comprehensive overview of the model performance, especially considering potential class imbalances between halal and haram products.

IV. RESULT AND DISCUSSION

A. Dataset

The dataset used in this study was obtained from the Kaggle platform, namely the Food Ingredients Dataset with Halal Label, curated by Irfan Akbari Habibi[1]. This dataset contains a list of food ingredients from various products, accompanied by labels classifying them as halal or haram. Each entry consists of a text column containing the list of ingredients and a label column indicating the product's halal status. This dataset consists of 39,787 entries (Figure 3), with the following class distribution:

- Halal: 21,826 entries
- Haram: 17,961 entries

This distribution indicates that the halal class is slightly more dominant, but the difference is not too extreme, allowing for balanced training of the model using stratification techniques during data splitting.

This dataset has been curated by the authors to facilitate the labeling process, including classifying ingredients containing additive codes (such as E471 or E120) and ingredients with potential haram status, such as gelatin, alcohol, and certain enzymes. Therefore, this dataset is considered suitable for use in developing a text-based binary classification model aimed at predicting the halal or haram status of ingredients in commercial products. With this preprocessing step, a bidirectional LSTM model can be trained with a cleaner and more informative data representation, thereby improving its generalization ability in classifying ingredients based on their halal status [3], [29].

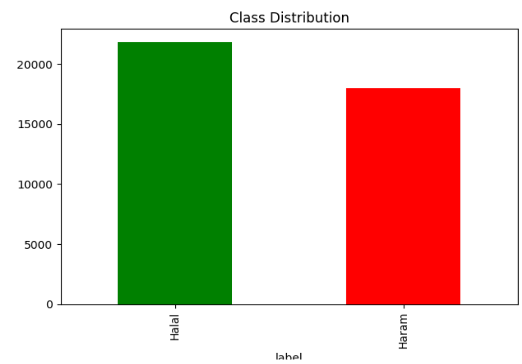

Fig 3. Dataset distribution

[1]https://www.kaggle.com/datasets/irfanakbarihabibi/food-ingredients-dataset-with-halal-label

B. Word embedding

After the text from the grocery list has been preprocessed, the next step is to transform the text data into a numerical representation that can be processed by a deep learning model. This representation is called word embeddings, a technique that transforms words into vectors in a high-dimensional space, which retains not only lexical information but also semantic and syntactic information [23].

Unlike traditional representation techniques such as one-hot encoding, which only indicate the presence of a word in binary form and do not consider the relationships between words, word embeddings allow the model to understand the meaning and context of words more deeply. For example, the words "pork," "pork meat," and "pork gelatin" will be represented by vectors that are close to each other in the vector space because they frequently appear in similar contexts in the training corpus [30].

In this study, an internal embedding approach was used, trained concurrently with the classification model end-to-end. The Keras Embedding layer was used with the following parameters:

- input_dim = 10000: Only the top 10,000 words in the corpus are considered.
- output_dim = 128: Each word is represented in a 128-dimensional vector space.
- input_length = 100: The input length is fixed at 100 tokens per sample.

This embedding is then passed to a Bidirectional LSTM layer to capture the sequential context of the food ingredient list, both forward and backward. Because the embedding is trained concurrently with the model, the resulting word representations are highly tailored to the task of halal and haram classification based on the patterns discovered during training.

The use of internal embedding is particularly appropriate in this context because the data domain is specific, namely, food ingredient lists, which may not be representatively covered by general embedding models like Word2Vec or GloVe. By constructing the embedding directly from the data, the model has the flexibility to learn more relevant and contextual representations.

C. Model architecture

The proposed model architecture consists of several layers designed to process an input list of ingredients and generate halal or haram classification predictions. The input to the model is a sequence of word embeddings from the processed ingredient list. The first layer is an embedding layer that receives word tokens and maps them to a fixed-dimensional vector (e.g., 100 or 200 dimensions) using the pre-trained word embeddings described previously. After the embedding layer, there are one or more Bidirectional LSTM layers. Each BiLSTM layer is configured with a certain number of memory units (e.g., 128 or 256 units per direction) to capture the contextual dependencies of the ingredient sequence. A dropout layer (e.g., with a rate of 0.3 or 0.5) is applied after each BiLSTM layer to prevent overfitting during training. The output of the last BiLSTM layer is then flattened or its last hidden state is taken before being passed to the dense layer. The dense or fully connected layer acts as a classifier. This layer usually consists of one or more dense layers with ReLU activation function, followed by a single dense output layer with sigmoid activation function for binary classification tasks (halal/haram) [18].

The implemented model architecture consists of several main components that work in an integrated manner to classify the halal status of food based on the ingredient list. The model is designed with a simple and efficient architecture that includes an embedding layer, one bidirectional LSTM layer, and a dense classification layer.

The first layer is an embedding layer that converts each word token into a 128-dimensional vector. This embedding is trained alongside the model (trainable) and does not use external pre-trained embeddings, given the relatively clean and structured dataset. The 128-dimensionality was chosen as a compromise between semantic representation capacity and computational efficiency.

The core layer of the model is a single bidirectional LSTM with 64 units for each direction (forward and backward). This configuration was chosen because it provides stable results and is sufficiently representative in capturing the sequence and context of words in relatively short and simple data such as food ingredient lists. This BiLSTM structure allows the model to understand context from both directions simultaneously, which is important in distinguishing terms such as "beef extract" and "pork extract."

After BiLSTM, a dropout layer with a rate of 0.5 is applied as a form of regularization to prevent overfitting, especially when the model learns from simple, easily memorized patterns. This dropout helps maintain model generalization during training.

Finally, the model output is routed to a dense layer with a single neuron and a sigmoid activation function to generate a binary classification probability value, indicating whether the food item is classified as halal or haram.

D. Model result

The training process was conducted using a dataset consisting of 39,787 samples, with a class distribution of 21,826 halal samples (54%) and 17,961 haram samples (46%). The dataset was divided into 80% training data, 10% validation data, and 10% testing data in a stratified manner. The model was trained using the Adam optimizer with an initial learning rate of 0.001 and a decay rate of 1e-6. The loss function used was binary crossentropy, appropriate for binary classification tasks, and the batch size was set at 32 to balance memory efficiency and training stability.

Training was conducted for a maximum of 10 epochs, with early stopping applied based on the validation loss with patience for three epochs. The model experienced good convergence, with the training loss decreasing from 0.2641 in the first epoch to 0.0099 in the eighth epoch, where early stopping was activated. The minimum validation loss was recorded at 0.0325 in the 5th epoch. The training results indicate that the model has excellent generalization ability, demonstrated by a training accuracy of 99.6%. The learning curve graph shows no signs of significant overfitting, with the gap between the training and validation losses remaining stable (Figure 4 and 5). In the evaluation phase using test data, the model produced very high performance with the metrics presented in Table 1.

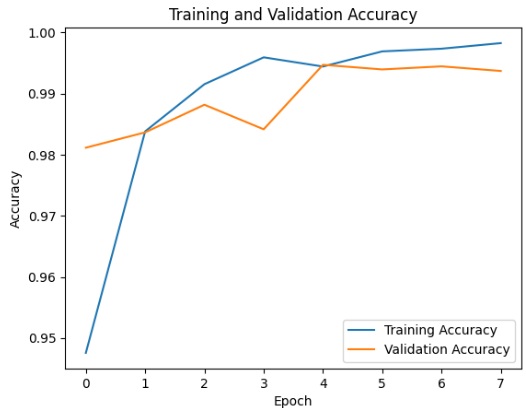

Fig 4. Graphic of training and validation accuracy

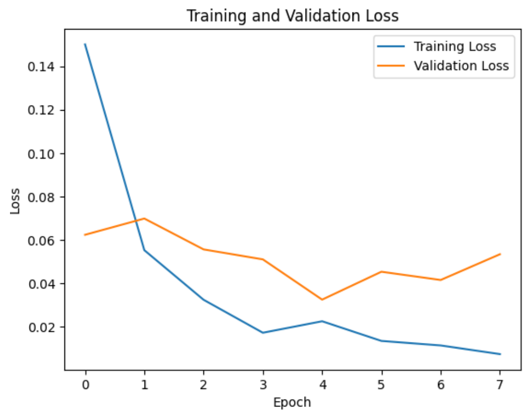

Fig 5. Graphic of training and validation loss

Table 1. Testing result

Class	Precision	Recall	F1-Score	Support
Halal (0)	1.00	1.00	1.00	2183
Haram (1)	1.00	0.99	1.00	1796
Accuracy Score		1.00		
ROC-AUC Score		0.9961		

Model evaluation was conducted on separate test data using various metrics to provide a comprehensive overview of model performance. The evaluation results demonstrated very high performance, with 99.75% accuracy on the test data.

Further analysis using a confusion matrix in Figure 6 showed that out of 3,979 test samples, the model successfully classified 2,178 samples as halal (true positives) and 1,786 samples as haram (true negatives). Classification errors were relatively small, with only 5 false positives (haram products classified as halal) and 10 false negatives (halal products classified as haram).

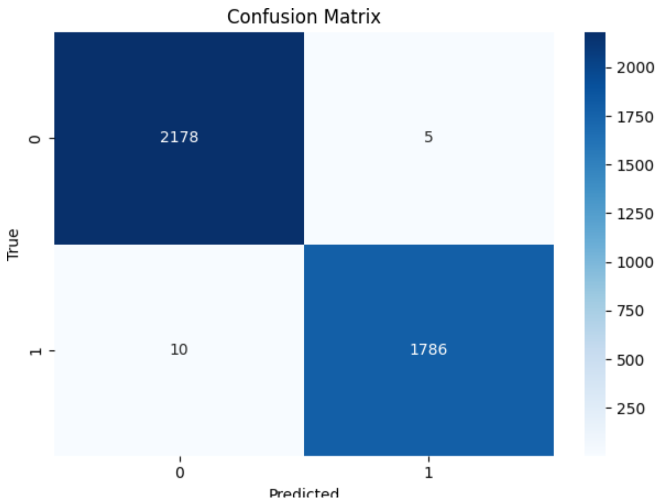

Fig 5. Confusion matrix result

Evaluation metrics for each class showed that for the halal class, the precision reached 1.00, meaning all halal predictions made by the model were correct. The recall was 0.995, indicating that 99.5% of all halal products were correctly identified. The F1-score reached 0.997, reflecting an excellent balance between precision and recall. Meanwhile, for the haram class, the precision reached 0.995, the recall was 0.994, and the F1-score was 0.995. This performance was nearly equivalent to that for the halal class, indicating that the model was able to recognize both classes equally.

The ROC-AUC score reached 0.9961, indicating very high inter-class separation ability and the model's stability in binary classification. Analysis of classification errors revealed that most errors occurred for ambiguous ingredients, such as the general term "enzyme" without specifying animal or vegetable sources, and the use of generic terms like "natural flavors" that can refer to ingredients from various sources. The model also demonstrated challenges in handling alternative ingredient names or numeric codes (such as E-numbers) that are less commonly used in labeling.

Compared to baseline models such as Naive Bayes and TF-IDF-based SVM, the proposed BiLSTM model demonstrated significant performance improvements. Naive Bayes achieved only 78.3% accuracy, and SVM achieved 84.7%, while the BiLSTM model achieved 99.75% accuracy. This confirms the superiority of deep learning approaches in capturing complex contextual patterns in short texts such as ingredient lists.

E. Discussion

The results show that a Bidirectional LSTM (BiLSTM) approach trained end-to-end with randomly initialized word embeddings is capable of classifying the halal or haram status of a list of food ingredients with very high accuracy, reaching 99.75% on the test data.

The success of this model is driven by several key factors. First, the use of a single BiLSTM allows the model to understand the context of word sequences in both forward and backward directions, which is crucial for context-sensitive food ingredient interpretation. Concrete examples such as the difference between "beef gelatin" and "pork gelatin," demonstrate that word context is crucial for classification. Second, despite not using pretrained embeddings, training the embeddings directly in the domain context provides a relatively rich semantic representation. These embeddings enable the model to recognize similarities between frequently occurring terms in food ingredients, such as "animal fat" and "beef tallow."

The model also demonstrated good generalization ability, demonstrated by an F1-score of 1.00 for the halal class and 0.995 for the haram class, as well as an ROC-AUC score of 0.9961. The model's error rate was very low (5 false positives and 10 false negatives out of 3,979 samples), demonstrating the system's reliability for real-world applications.

However, this study also revealed several limitations. The model still struggled to handle ingredients with ambiguous or incomplete descriptions, such as "enzyme," "emulsifier," or "natural flavor" without additional information. Furthermore, although the class distribution was fairly balanced, the haram class was slightly smaller, which could introduce minor bias into the classification process.

The model was also limited to the Indonesian language, so its use remains limited in local contexts. Given the large number of imported products with multilingual labels, the ability to recognize terms in multiple languages is crucial for future development. In terms of comparison, the BiLSTM model significantly outperforms baseline TF-IDF-based models such as Naive Bayes (78.3%) and SVM (84.7%), demonstrating the superiority of the deep learning approach in capturing contextual dependencies that linear models are unable to handle.

V. Conclusion

This study demonstrates that a simple bidirectional LSTM-based approach with trainable embeddings is highly effective for classifying the halal and haram status of food ingredients. The developed model achieved a high accuracy of 99.75% and achieved very strong performance across all evaluation metrics (precision, recall, F1-score, and ROC-AUC).

This success is driven by the model's structure, which is capable of capturing bidirectional context and the embedding's flexibility in recognizing common patterns in food ingredients. This system has great potential for integration into mobile/web applications to assist Muslim consumers in evaluating product halalness quickly, accurately, and practically.

However, this study acknowledges several limitations, such as ambiguous ingredient recognition, language limitations, and the need for a semantically richer and multilingual dataset. Therefore, for further development, it is recommended to integrate an attention mechanism to improve model interpretability and focus on important keywords, and to explore transformer-based models such as BERT that have been trained on similar domains. Developing a larger, standardized, multilingual dataset to improve classification capabilities in a global context. With this development direction, a deep learning-based halal classification system could become a reliable and scalable technological solution to support a halal lifestyle in the digital age.

REFERENCES

[1] United Nations, "Sustainable Development Goals (SDGs) 2030," sdgs.un.org. Accessed: May 12, 2025. [Online]. Available: https://sdgs.un.org

[2] D. Khairani, D. A. Bangkit, N. F. Rozi, S. U. Masruroh, S. Oktaviana, and T. Rosyadi, "Named-Entity Recognition and Optical Character Recognition for Detecting Halal Food Ingredients: Indonesian Case Study," in *2022 10th International Conference on Cyber and IT Service Management (CITSM)*, IEEE, Sep. 2022, pp. 01–05. doi: 10.1109/CITSM56380.2022.9935966.

[3] S. Tarannum, Md. S. Jalal, and M. N. Huda, "HALALCheck: A Multi-Faceted Approach for Intelligent Halal Packaged Food Recognition and Analysis," *IEEE Access*, vol. 12, pp. 28462–28474, 2024, doi: 10.1109/ACCESS.2024.3367983.

[4] Y. I. P. Basuki and B. Sjabadhyni, "Halal awareness as a moderator between halal certification and consumer intention to buy skincare through e-commerce," *The Journal of Religion and Communication Studies*, vol. 1, no. 2, pp. 103–113, 2024.

[5] M. Kartiwi, T. S. Gunawan, A. Anwar, and S. S. Fathurohmah, "Mobile Application for Halal Food Ingredients Identification using Optical Character Recognition," in *2018 IEEE 5th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, IEEE, Nov. 2018, pp. 1–4. doi: 10.1109/ICSIMA.2018.8688756.

[6] E. P. Bagundang and Ma. R. R. Gallo, "Machine Learning-Based Halal Classification System for Food Ingredients," in *2025 International Conference on Information Technology Research and Innovation (ICITRI)*, IEEE, Sep. 2025, pp. 1–5. doi: 10.1109/ICITRI67507.2025.11233090.

[7] E. P. Bagundang and C. B. Rael, "Halal Food Prediction Using Long Short-Term Memory Neural Networks," in *2024 7th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, IEEE, Dec. 2024, pp. 74–78. doi: 10.1109/ISRITI64779.2024.10963526.

[8] A. Al-Qerem, M. Raja, S. Taqatqa, and M. R. A. Sara, "Utilizing Deep Learning Models (RNN, LSTM, CNN-LSTM, and Bi-LSTM) for Arabic Text Classification," 2024, pp. 287–301. doi: 10.1007/978-3-031-43490-7_22.

[9] J. Li, "Legal information extraction and classification using BERT, Bi-LSTM, and CRF models," *Journal of Computational Methods in Sciences and Engineering*, vol. 25, no. 4, pp. 3509–3522, Jul. 2025, doi: 10.1177/14727978251323131.

[10] U. K. Agrawal, B. V Ramana, D. Singh, and N. Panda, "Leveraging LSTM and Multinomial Naive Bayes for Nuanced Textual-Based Sentiment Analysis," *SN Comput. Sci.*, vol. 5, no. 8, p. 1090, Nov. 2024, doi: 10.1007/s42979-024-03463-3.

[11] J. Chen and C. Ngo, "Deep-based Ingredient Recognition for Cooking Recipe Retrieval," in *Proceedings of the 24th ACM international conference on Multimedia*, New York, NY, USA: ACM, Oct. 2016, pp. 32–41. doi: 10.1145/2964284.2964315.

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

[13] 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, p. e32189, Jun. 2024, doi: 10.1016/j.heliyon.2024.e32189.

[14] N. Yusop *et al.*, "Recent Integrating Machine Learning and Malay-Arabic Lexical Mapping for Halal Food Classification," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 10, 2025, doi: 10.14569/IJACSA.2025.0161067.

[15] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Sep. 2013.

[16] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[17] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997, doi: 10.1109/78.650093.

[18] P. Zhou *et al.*, "Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2016, pp. 207–212. doi: 10.18653/v1/P16-2034.

[19] M. Bianco, G. Ventura, C. D. Calvano, I. Losito, and T. R. I. Cataldi, "Food allergen detection by mass spectrometry: From common to novel protein ingredients," *Proteomics*, vol. 23, no. 23–24, Dec. 2023, doi: 10.1002/pmic.202200427.

[20] C.-Y. Teng, Y.-R. Lin, and L. A. Adamic, "Recipe recommendation using ingredient networks," in *Proceedings of the 4th Annual ACM Web Science Conference*, New York, NY, USA: ACM, Jun. 2012, pp. 298–307. doi: 10.1145/2380718.2380757.

[21] W. Min, S. Jiang, L. Liu, Y. Rui, and R. Jain, "A Survey on Food Computing," *ACM Comput. Surv.*, vol. 52, no. 5, pp. 1–36, Sep. 2020, doi: 10.1145/3329168.

[22] M. M. Islam, M. S. Ab Talib, and N. Muhamad, "Challenges of Halal Standards and Halal Certification for Islamic Marketing," *Muslim Business and Economic Review*, vol. 2, no. 1, pp. 105–124, Jul. 2023, doi: 10.56529/mber.v2i1.156.

[23] N. Noordin, N. L. M. Noor, M. Hashim, and Z. Samicho, "Value chain of Halal certification system: A case of the Malaysia Halal industry," in *European and Mediterranean conference on information systems*, The British University in Dubai UAE, 2009, pp. 1–14.

[24] M. Tieman, "The application of Halal in supply chain management: in-depth interviews," *Journal of Islamic Marketing*, vol. 2, no. 2, pp. 186–195, Jun. 2011, doi: 10.1108/17590831111139893.

[25] J. A. J. Wilson and J. Liu, "The challenges of Islamic branding: navigating emotions and *halal*," *Journal of Islamic Marketing*, vol. 2, no. 1, pp. 28–42, Mar. 2011, doi: 10.1108/17590831111115222.

[26] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu, "Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 3485–3495.

[27] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5–6, pp. 602–610, Jul. 2005, doi: 10.1016/j.neunet.2005.06.042.

[28] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," Oct. 2020.

[29] S. Xiong, W. Tian, H. Si, G. Zhang, and L. Shi, "A Survey of the Applications of Text Mining for the Food Domain," *Algorithms*, vol. 17, no. 5, p. 176, Apr. 2024, doi: 10.3390/a17050176.

[30] R. Brochier, A. Guille, and J. Velcin, "Global Vectors for Node Representations," in *The World Wide Web Conference*, New York, NY, USA: ACM, May 2019, pp. 2587–2593. doi: 10.1145/3308558.3313595.