VOL. 11. NO. 3 FEBRUARY 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v11i3.7613

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

IMPLEMENTATION OF RANDOM FOREST FOR ANIMAL PROTEIN CLASSIFICATION THROUGH HYPERPARAMETER OPTIMIZATION

Ridho Ikhram[1*]; Anton Yudhana[2]; Imam Riadi[3]

Master of Informatics[1]
Electrical Engineering[2]
Information Systems[3]
Universitas Ahmad Dahlan, Yogyakarta, Indonesia[1,2,3]
https://www.uad.ac.id/[1,2,3]
2407048004@webmail.uad.ac.id*, eyudhana@ee.uad.ac.id, imam.riadi@is.uad.ac.id

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *Accurate identification of animal protein types is crucial to ensure food authenticity and safety, particularly in the context of compliance with halal principles. This study aims to implement the Random Forest (RF) algorithm to classify four types of animal protein—broiler chicken, free-range chicken, pork, and beef through hyperparameter optimization using GridSearchCV. The dataset was evaluated using 5-fold cross-validation, and feature importance analysis was conducted to identify the variables that contributed most to classification. Results showed that RF with optimized hyperparameters achieved a test accuracy of 92.81%, with macro-average precision, recall, and F1-score of 93%. The model performed best for the broiler chicken and pork classes, while the beef class exhibited a higher misclassification rate, likely due to the similarity of spectral characteristics among classes. ODOR, CO_2, H_2, NH_3, and VOC were identified as the key indicators for distinguishing animal protein types. This study contributes to halal authentication by integrating FTIR spectral data with optimized Random Forest, enabling efficient and accurate classification. Although RF proved reliable and capable of handling high-dimensional data, the study is limited by dataset size and spectral feature complexity. Future research is recommended to explore deep learning architectures, such as Convolutional Neural Networks (CNN), with larger FTIR datasets to improve model generalization and robustness.*

Keywords: *Animal Protein, Classification, Machine Learning, Accuracy, Random Forest.*

Intisari— *Deteksi jenis protein hewani yang akurat sangat penting untuk menjamin keaslian dan keamanan pangan, terutama dalam konteks kepatuhan terhadap prinsip halal. Penelitian ini bertujuan untuk mengimplementasikan algoritma Random Forest dalam mengklasifikasikan empat jenis protein hewani yaitu ayam, ayam kampung, babi, dan sapi melalui optimasi hyperparameter menggunakan GridSearchCV. Dataset diuji menggunakan evaluasi 5-fold cross-validation, dan analisis feature importance dilakukan untuk mengidentifikasi variabel yang paling berkontribusi terhadap klasifikasi. Hasil menunjukkan bahwa RF dengan hyperparameter optimal mencapai akurasi pengujian sebesar 92,81%, dengan rata-rata makro presisi, recall, dan F1-score sebesar 93%. Model menunjukkan performa terbaik pada kelas ayam dan babi, sedangkan kelas sapi mengalami tingkat kesalahan klasifikasi lebih tinggi, kemungkinan akibat kemiripan karakteristik spektral antar kelas. Variabel ODOR, CO_2, H_2, NH_3, dan VOC teridentifikasi sebagai indikator utama dalam membedakan jenis protein hewani. Penelitian ini berkontribusi pada autentikasi halal dengan mengintegrasikan data spektral FTIR dan Random Forest yang dioptimasi, sehingga menghasilkan klasifikasi yang efisien dan akurat. Meskipun RF terbukti andal dan mampu menangani data berdimensi tinggi, penelitian ini terbatas oleh ukuran dataset dan kompleksitas fitur spektral. Penelitian selanjutnya disarankan untuk mengeksplorasi arsitektur deep learning, seperti Convolutional Neural Networks (CNN), dengan dataset FTIR yang lebih besar untuk meningkatkan kemampuan generalisasi dan ketahanan model.*

Kata Kunci: *Protein Hewani, Klasifikasi, Machine Learning, Akurasi, Random Forest.*

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)
VOL. 11. NO. 3 FEBRUARI 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i3.7613

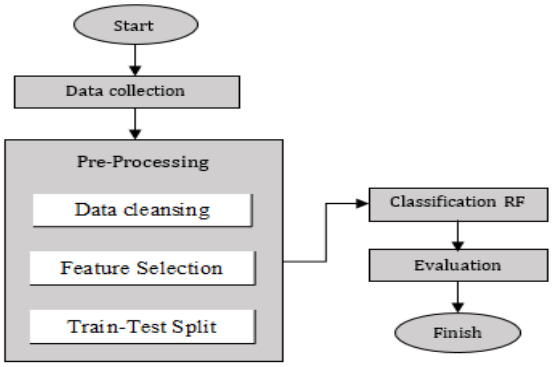
INTRODUCTION

Animal protein is one of the essential components in various food products consumed by humans, found in meat, eggs, fish, and other processed products [1][2][3]. The presence of animal protein not only affects the nutritional value but also the organoleptic properties of products. With the increasing demand for halal, vegetarian, and vegan products, accurate detection of animal protein content has become increasingly important to ensure product compliance with consumer needs and adherence to halal principles [4][5][6][7]. The accuracy of animal protein detection not only affects product quality but also public health and consumer trust. Errors in the detection process can lead to the circulation of products containing hidden ingredients, violations of halal principles, and a disruption of market integrity. In general, animal protein detection is conducted using laboratory methods such as chemical techniques, chromatography, and microscopy. Although accurate, these methods tend to be time-consuming, costly, and highly dependent on experienced operators [8]. Therefore, faster, more efficient, and scalable methods are highly needed. With technological advancements, Fourier Transform Infrared (FTIR) has emerged as a promising approach, as it enables rapid, non-destructive analysis with minimal sample preparation [9][10][11]. The integration of FTIR spectral data with machine learning algorithms such as Random Forest enables accurate classification, is robust to noise, and can handle high-dimensional data [12][13].

Several previous studies have demonstrated the success of FTIR methods combined with classification algorithms such as Support Vector Machine (SVM) and Partial Least Squares-Discriminant Analysis (PLS-DA) in detecting meat contaminants or their derivative products [14] [15]. E-nose–based approaches integrated with supervised learning have also been applied to differentiate meat types based on chemical volatiles [16]. However, most of these studies have not specifically evaluated the use of Random Forest based on FTIR spectra for classifying types of animal protein, leaving standardized classification models for halal protein authentication still limited. This study addresses this gap by developing a Random Forest model optimized through GridSearchCV for the determination of key hyperparameters, including the number of trees (n_estimators), the maximum tree depth (max_depth), and the number of features considered at each split (max_features). This optimization enables the model to capture spectral patterns specific to each protein type more accurately while minimizing classification errors. This study is also expected to serve as a foundation for future research through the development of more advanced methods, such as the application of deep learning algorithms. Deep learning approaches, particularly Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have the capability to automatically extract feature patterns from high-dimensional data such as FTIR spectra. With this capability, systems can recognize more complex chemical structures and manage higher data variability, thereby potentially enhancing classification accuracy significantly. Such innovation is crucial for addressing real-world challenges in the food industry, where testing data are often heterogeneous and subject to disturbances or noise.

MATERIALS AND METHODS

The research stages begin with data collection, followed by data understanding and data preparation. After the data collection process, system design is carried out, which is then followed by experimental modeling using the Random Forest algorithm. The next stage involves evaluating and analyzing the results. Finally, conclusions are drawn. The overall research design workflow is illustrated in Figure 1 below.



Source: (Research Results, 2025)

Figure 1. Methodology

Dataset

A dataset is a collection of data managed by researchers to support the completion of their studies [17]. The dataset in this study was obtained directly by the researcher using a Fourier Transform Infrared Spectroscopy (FTIR) machine, which is capable of detecting chemical compounds in meat by emitting infrared light onto the sample and measuring the absorbance intensity at various wavelengths. Each type of animal protein exhibits a distinct spectral pattern due to differences in chemical composition, molecular structure, and the functional groups contained within it. The dataset used in this study consists of 1,600 samples derived from four animal species, namely broiler chicken, free-range chicken, beef, and pork. The dataset was stored in CSV format, enabling direct processing with the Random Forest algorithm. Data processing was conducted using the

Google Colab platform, after which the dataset was divided into training and testing subsets [18].

Data Preprocessing

Data preprocessing is an essential stage in classification, as it is conducted to optimize the dataset for modeling and to enhance model performance [19][20]. The raw data consist of elements with varying structures and characteristics; therefore, they must be converted into a uniform structure. The main objective of pre-processing is to ensure that the data are in a format that can be processed by the Random Forest algorithm [21]. Several steps are involved in the pre-processing stage, including data cleaning, feature selection, and dataset splitting. Data cleaning is performed to reduce the file size of the dataset, with the aim of improving accuracy and enabling the algorithm to operate more efficiently. Feature selection focuses on identifying relevant features associated with animal protein in order to achieve higher classification accuracy [22]. Finally, dataset splitting is required in the Random Forest approach, where the dataset is divided into training and testing subsets. The division is carried out using an 80:20 ratio to train, test, and validate the performance of the model [23].

Classification Random Forest

Random Forest is an ensemble method based on decision trees that employs the bagging technique. The algorithm constructs multiple decision trees through a randomized process and generates predictions by aggregating (averaging) the outputs of all individual trees [24][25][26]. Hyperparameter optimization in this study was conducted using GridSearchCV with a 5-fold cross-validation scheme to ensure that the model does not experience overfitting and maintains strong generalization capability. This procedure involves constructing a grid of candidate parameters for the Random Forest algorithm to identify the optimal parameters (n_estimators, max_depth, min_samples_split, min_samples_leaf' and bootstrap). This ensemble approach offers three main advantages: (1) an inherent ability to model complex non-linear relationships, (2) enhanced robustness against overfitting through a majority voting mechanism, and (3) resilience in processing noisy datasets due to its distributed error-correction nature [27].

$$\hat{y} = \frac{1}{B}\sum_{i=1}^{B} h_i(x) \quad (1)$$

Description formula:
$\hat{Y}$: predicted value (final output)
B : number of models / number of data / number of members (batch or ensemble)
h_i : output of model i against input x

Evaluation

At this stage, evaluation is conducted for each model that has been implemented. The evaluation is performed using a confusion matrix, which represents the data generated by the algorithm and calculates the classification prediction results. This is available in the *scikit-learn* library in Python [28]. During the model evaluation stage, accuracy, precision, recall, and F1-score are obtained, which provide quantitative measures for assessing the performance of the Random Forest algorithm [29].

RESULTS AND DISCUSSION

To conduct and present the results of this research, the author employed Google Colab as the primary tool. Google Colab was selected because it is Python-based, web-based, requires no installation on a PC or laptop, and provides ease of use with integrated support for applications and datasets stored collaboratively in Google Drive.

Dataset

The dataset used in this study consists of 1,600 data samples collected directly with the assistance of the FTIR machine. From the total dataset, the data were divided into two parts with an 80:20 ratio for training and testing, respectively, and subsequently processed using the Random Forest algorithm implemented on the Google Colab platform. The dataset is presented in Table 1.

Table 1. Research Dataset

Class	C2H50H	CH4	CO2	H2	H2S	HCH0	NH3	NO2	ODOR	SMOKE	VOC
Chicken	1.268	3.044	1.404	2.834	1.547	1.612	0.018	0.274	1.594	0.695	1.530
Beef	1.364	3.292	1.386	2.866	1.728	1.810	0.020	0.269	1.734	0.724	1.662
Free-range Chicker	0.538	1.744	0.470	1.650	0.803	1.043	0.002	0.245	1.393	0.179	1.023
Pork	0.313	1.300	0.272	1.514	0.528	0.804	0.001	0.273	0.629	0.108	0.751

Source: (Research Results, 2025)

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 3 FEBRUARI 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i3.7613

Preprocessing

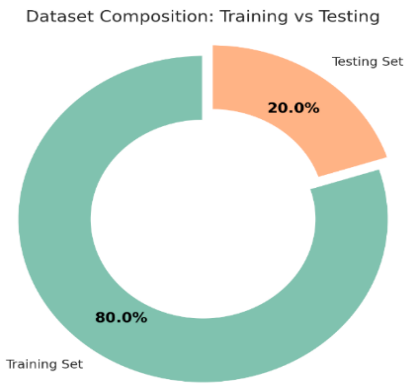
After the data collection stage, the preprocessing of the dataset was carried out. In this study, pre-processing included data cleaning (removing duplicate entries and retaining three decimal places), selection of relevant compound features for animal protein classification, and dataset splitting. The feature selection process yielded six relevant compounds: C_2H_5OH, CO_2, H_2, NH_3, ODOR, and VOC, which are directly associated with the degradation and distinctive characteristics of animal proteins. NH_3 and H_2 are produced from protein decomposition, while C_2H_5OH and CO_2 originate from microbial activity. Meanwhile, ODOR and VOC reflect the distinctive aroma of each type of meat [30] [31]. These compound profiles can be utilized to rapidly and accurately identify both the type and quality of meat. The results of the animal protein feature selection are presented in Table 2.

Table 2. Feature Selection Results

C2H50H	CO2	H2	NH3	ODOR	'OC
1.268	1.404	2.834	0.018	1.594	530
1.364	1.386	2.866	0.020	1.734	662
0.538	0.470	1.650	0.002	1.393	023
0.313	0.272	1.514	0.001	0.629	751

Source: (Research Results, 2025)

From the 1,600 labeled datasets, the data were then divided into 80% training and 20% testing subsets, with the aim of evaluating the performance of the Random Forest model in classifying types of animal proteins. The results of the dataset splitting are illustrated in Figure 2.



Source: (Research Results, 2025)
Figure 2. Dataset Division Results

Classification Random Forest

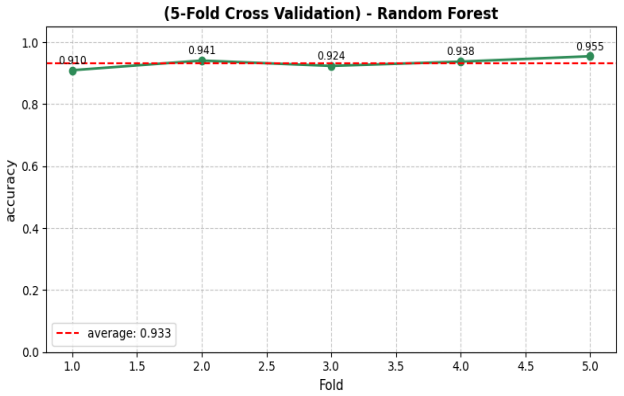
To optimize the RF algorithm in this study, GridSearchCV was applied using a 5-fold cross-validation scheme to ensure that the model did not experience overfitting and maintained strong generalization capability. This procedure involved constructing a grid of candidate parameters for the Random Forest algorithm, training the model on five different data subsets, calculating the average performance scores across all folds, and selecting the best parameter combination based on the highest accuracy. The GridSearchCV results for the RF algorithm are presented in Table 3.

Table 3. GridSearchCV results

Parameter	Value
n_estimators	200
max_depth	30
min_samples_split	2
min_samples_leaf'	1
bootstrap	False
Best Accuracy (CV)	: 93.33%

Source: (Research Results, 2025)

The integration of grid search and cross-validation was chosen because it has been proven effective in improving the stability and accuracy of artificial intelligence–based classification models. The cross-validation results for the RF algorithm are shown in Figure 3.



Source: (Research Results, 2025)
Figure 3. Cross-Validation Random Forest

In Figure 3, the 5-Fold Cross-Validation results for the Random Forest algorithm show stable and consistent accuracy across all folds, with values ranging from 91% to 95.5% and an overall average of 93.3%. The first fold recorded the lowest accuracy at 91%, indicating that the data subset in this fold had higher variability or more complex feature patterns, making classification more challenging for the model. However, the increase in accuracy in the second fold (94.1%) and the stable performance in the third (92.4%) and fourth folds (93.8%) demonstrate the model's ability to maintain performance across different portions of the dataset. The fifth fold achieved the highest accuracy at 95.5%, confirming Random Forest's capability to adapt to data variations without overfitting. Overall, this accuracy pattern

reflects that Random Forest exhibits strong generalization ability and delivers consistent performance in classifying animal protein data, with only minor fluctuations between folds, thereby supporting the model's reliability.

Model Evaluation

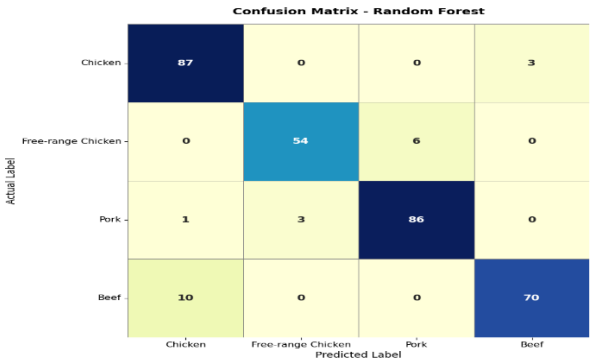
In the evaluation stage, classification was conducted using the Random Forest algorithm to assess how effectively the algorithm classified types of animal proteins. The classification results are presented in Table 4.

Table 4. RF Model Classification Results

Class	Precision (%)	Recall (%)	f1-score (%)	support
Chicken	89	97	93	90
Free-range Chicken	95	90	92	60
Pork	93	96	95	90
Beef	96	88	92	80
accuracy			93	320
macro avg	93	92	93	320
weighted avg	93	93	93	320
Akurasi model	92.81			

Source: (Research Results, 2025)

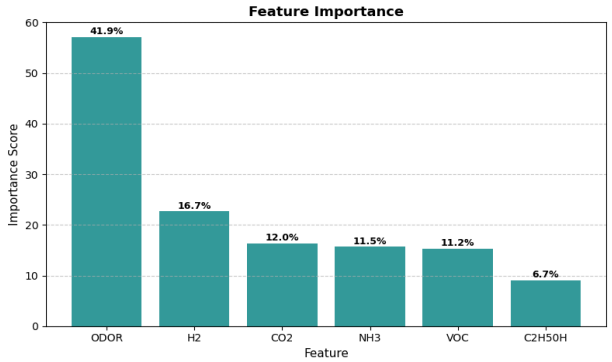
In Table 4, The evaluation results of the Random Forest classification model indicate good performance in distinguishing four categories of animal proteins, namely Chicken, Free-range Chicken, Pork, and Beef. Overall, the model achieved an accuracy of 92.81%, showing that the majority of samples were correctly classified. Class-wise metric analysis indicates that the Chicken class has a precision of 89% and a recall of 97%, meaning most Chicken samples were correctly identified despite some misclassifications. The Free-range Chicken class showed a precision of 95% and recall of 90%, indicating that the model could distinguish this class well, although a few samples were misclassified. The Pork class performed the best with a precision of 93%, recall of 96%, and an F1-score of 95%, demonstrating the model's effectiveness in accurately recognizing Pork samples. Meanwhile, the Beef class had a precision of 96% and recall of 88%, indicating some identification errors, likely due to chemical feature similarities with other classes. The macro-average F1-score of 93% confirms balanced performance across classes, while the weighted average of 93% demonstrates consistent performance considering the proportion of samples in each class. Overall, these results confirm that the Random Forest algorithm is effective in capturing the unique characteristic patterns of each type of animal protein, making it a promising tool for FTIR-based meat identification systems.



Source: (Research Results, 2025)

Figure 4. Confusion Matrix

In Figure 3, The confusion matrix shows that the Random Forest algorithm has excellent classification capability in distinguishing the four categories of animal proteins. The model demonstrates strong performance in the Chicken and Pork classes, with 87 and 86 correctly classified samples, respectively, and only minimal errors. In the Free-range Chicken class, the model accurately identified 54 samples, although six samples were misclassified as Pork. These findings indicate a similarity in spectral characteristics between the two classes in the FTIR data. Meanwhile, the Beef class is the most challenging category, with ten samples misclassified as Chicken, suggesting the possibility of overlapping spectral features that affect the model's decision boundaries. Overall, Random Forest exhibits high classification performance across all four classes, although the identified misclassifications reveal intrinsic similarities within certain protein groups that warrant further investigation in future studies.



Source: (Research Results, 2025)

Figure 4. Feature Importance

In Figure 3, The feature importance analysis shows that the ODOR feature has the most dominant contribution in the classification process, with a weight of 41.9%, indicating that odor characteristics are the most discriminative indicators for distinguishing sample types. The strong influence of

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 3 FEBRUARI 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i3.7613

ODOR is likely due to its association with the production of volatile compounds resulting from protein degradation, such as amines, volatile sulfur compounds, and ketones, which differ significantly among meat types. The CO_2 feature also holds an important position with a contribution of 12.0%, which can be explained by its relationship to microbiological activity and oxidation processes that increase carbon dioxide emissions in certain samples. Other variables such as H_2, NH_3, and VOC also provide significant contributions, reflecting the role of volatile gas components in enhancing the model's discriminative ability. Conversely, C_2H_5OH shows the lowest influence (6.7%), indicating that alcohol compounds do not exhibit sufficiently consistent variation across classes, resulting in a smaller contribution to the decision-making process. Overall, these results confirm that features directly related to volatile compounds produced from protein degradation provide the most relevant information for the Random Forest model, thereby improving classification accuracy.

CONCLUSION

The Random Forest (RF) algorithm has been proven effective in classifying types of animal proteins using FTIR spectral data, achieving a test accuracy of 92.81% and macro-average precision, recall, and F1-score of 93%. The model performed best for the Pork and Free-range Chicken classes, whereas the Beef class exhibited a higher misclassification rate, likely due to similarities in spectral characteristics between classes. Feature importance analysis confirmed that ODOR, CO_2, H_2, NH_3, and VOC were the primary indicators for differentiating animal protein types, while the 5-fold cross-validation results demonstrated stable accuracy and strong generalization capability. Although RF proved reliable and efficient in handling high-dimensional data, this study has limitations regarding dataset size and the complexity of spectral features. Future research could explore deep learning architectures, such as Convolutional Neural Networks (CNN), using larger FTIR datasets to enhance model generalization and robustness.

ACKNOWLEDGEMENTS

This research was funded by the Directorate of Research, Technology, and Community Service, Ministry of Education and Culture, Science and Technology with Contract No.: 107/ES/PG.02.00.PL/2024 dated June 11, 2024.

REFERENCE

[1] A. M. Rueda García, P. Fracassi, B. D. Scherf, M. Hamon, and L. Iannotti, "Unveiling the Nutritional Quality of Terrestrial Animal Source Foods by Species and Characteristics of Livestock Systems," *Nutrients*, vol. 16, no. 19, p. 3346, Oct. 2024, doi: 10.3390/nu16193346.

[2] V. Miller *et al.*, "Global, regional, and national consumption of animal-source foods between 1990 and 2018: findings from the Global Dietary Database," *Lancet Planet Health*, vol. 6, no. 3, pp. e243–e256, Mar. 2022, doi: 10.1016/S2542-5196(21)00352-1.

[3] T. Beal *et al.*, "Friend or Foe? The Role of Animal-Source Foods in Healthy and Environmentally Sustainable Diets," *J Nutr*, vol. 153, no. 2, pp. 409–425, Feb. 2023, doi: 10.1016/j.tjnut.2022.10.016.

[4] P. W. Harlina, V. Maritha, I. Musfiroh, S. Huda, N. Sukri, and M. Muchtaridi, "Possibilities of Liquid Chromatography Mass Spectrometry (LC-MS)-Based Metabolomics and Lipidomics in the Authentication of Meat Products: A Mini Review," *Food Sci Anim Resour*, vol. 42, no. 5, pp. 744–761, Sep. 2022, doi: 10.5851/kosfa.2022.e37.

[5] M. M. Rahman, M. S. A. Razimi, A. S. Ariffin, and N. Hashim, "Navigating moral landscape: Islamic ethical choices and sustainability in Halal meat production and consumption," *Discover Sustainability*, vol. 5, no. 1, p. 225, Aug. 2024, doi: 10.1007/s43621-024-00388-y.

[6] S. R. Aini *et al.*, "The metabolomics approach used for halal authentication analysis of food and pharmaceutical products: a review," *Food Res*, vol. 7, no. 3, pp. 180–187, Jun. 2023, doi: 10.26656/fr.2017.7(3).986.

[7] H. Zhang, N. Wisuthiphaet, H. Cui, N. Nitin, X. Liu, and Q. Zhao, "Spectroscopy Approaches for Food Safety Applications: Improving Data Efficiency Using Active Learning and Semi-Supervised Learning," Feb. 2022, doi: 10.48550/arXiv.2110.03765.

[8] R. L. A. Sitorus, "A Comprehensive Overview Of Near Infrared And Infrared Spectroscopy For Detecting The Adulteration On Food And Agro-Products—A Critical Assessment," *INMATEH Agricultural Engineering*, pp. 465–486, Aug. 2022, doi: 10.35633/nmateh-67-47.

[9] A. Yudhana, E. P. Silmina, and Sunardi, "Deteksi Kesegaran Daging Sapi Menggunakan Augmentasi Data Mosaic pada Model YOLOv5sM," *JRST (Jurnal Riset Sains*

dan Teknologi), pp. 63–71, Apr. 2025, doi: 10.30595/jrst.v9i1.24990.

[10] M. Pirhadi, N. Shariatifar, S. Pirhadi, S. M. Khodaei, and Y. Mazaheri, "Developing Infrared Spectroscopy Methods for Identification of Food Fraud and Authenticity: A Review," *Journal of Biochemicals and Phytomedicine*, vol. 3, no. 1, pp. 59–65, Jun. 2024, doi: 10.34172/jbp.2024.12.

[11] A. Kazemi, A. Mahmoudi, and M. Khojastehnazhand, "Detection of Adulteration of Ground Meat by Spectral-based Techniques and Artificial Intelligence (2020-2024)," *Iranian Food Science and Technology Research Journal*, vol. 20, no. 6, pp. 201–224, Feb. 2025, doi: 10.22067/ifstrj.2024.88158.1335.

[12] D. Ghosh and J. Cabrera, "Enriched Random Forest for High Dimensional Genomic Data," *IEEE/ACM Trans Comput Biol Bioinform*, vol. 19, no. 5, pp. 2817–2828, Sep. 2022, doi: 10.1109/TCBB.2021.3089417.

[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] M. A. Siddiqui *et al.*, "Multivariate analysis coupled with m-svm classification for lard adulteration detection in meat mixtures of beef, lamb, and chicken using ftir spectroscopy," *Foods*, vol. 10, no. 10, Oct. 2021, doi: 10.3390/foods10102405.

[15] A. Dashti *et al.*, "Assessment of meat authenticity using portable Fourier transform infrared spectroscopy combined with multivariate classification techniques," *Microchemical Journal*, vol. 181, p. 107735, Oct. 2022, doi: 10.1016/j.microc.2022.107735.

[16] L. A. Putri *et al.*, "Rapid analysis of meat floss origin using a supervised machine learning-based electronic nose towards food authentication," *NPJ Sci Food*, vol. 7, no. 1, Dec. 2023, doi: 10.1038/s41538-023-00205-2.

[17] M. R. Ardianto and R. Rushendra, "Prediksi Penyakit Diabetes Berdasarkan Perbandingan Klasifikasi Metode K-Nearest Neighbor, Naïve Bayes, Dan Decision Tree Menggunakan Rapid Miner," vol. 10, no. 2, pp. 973–985, 2025, doi: 10.29100/jipi.v10i2.6079.

[18] S. Saifullah *et al.*, "Nondestructive Chicken Egg Fertility Detection Using CNN-Transfer Learning Algorithms," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 3, pp. 854–871, Sep. 2023, doi: 10.26555/jiteki.v9i3.26722.

[19] A. Pranolo, S. Saifullah, A. Bella Utama, A. Prasetya Wibawa, M. Bastian, and C. P. Hardiyanti, "High-performance traffic volume prediction: An evaluation of RNN, GRU, and CNN for accuracy and computational trade-offs," in *BIO Web of Conferences*, EDP Sciences, Jan. 2025. doi: 10.1051/bioconf/202414802034.

[20] Rio Subandi, Herman, and Anton Yudhana, "Pre-Processing Pada Klasifikasi Citra Medis Pneumonia," *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 4, no. 1, pp. 86–93, Nov. 2023, doi: 10.51454/decode.v4i1.198.

[21] E. Ismanto, A. Fadlil, and A. Yudhana, "Analisis Perbandingan Model Fully Connected Neural Networks (FCNN) dan TabNet Untuk Klasifikasi Perawatan Pasien Pada Data Tabular Comparative Analysis of Fully Connected Neural Networks (FCNN) and TabNet Models for Patient Care Classification on Tabular Data," vol. 5, no. 3, pp. 526–532, 2024, doi: 10.37859/coscitech.v5i3.8256.

[22] E. Poerwandono and J. Perwitosari, "Penerapan Data Mining Untuk Penilaian Kinerja Karyawan Di PT. Riksa Dinar DJaya Menggunakan Metode Naïve Bayes Classification (Edhy Poerwandono 1 , Faizal Joko Perwitosari 2) Penerapan Data Mining Untuk Penilaian Kinerja Karya Di PT Riksa Dinar Djaya Menggunakan Metode Naive Bayes Classification," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, p. |pp, 2023, doi: 10.55338/saintek.v5i1.1416.

[23] L. Hakim, Z. Sari, A. Rizaldy Aristyo, and S. Pangestu, "Optimzing Android Program Malware Classification Using GridSearchCV Optimized Random Forest," *Computer Network, Computing, Electronics, and Control Journal*, vol. 9, no. 2, pp. 173–180, 2024.

[24] M. Ibrahim, "Evolution of Random Forest from Decision Tree and Bagging: A Bias-Variance Perspective," *Dhaka University Journal of Applied Science and Engineering*, vol. 7, no. 1, pp. 66–71, Feb. 2023, doi: 10.3329/dujase.v7i1.62888.

[25] M. A. Ganaie, M. Tanveer, P. N. Suganthan, and V. Snasel, "Oblique and rotation double random forest," *Neural Networks*, vol. 153, pp. 496–517, Sep. 2022, doi: 10.1016/j.neunet.2022.06.012.

[26] Nimatul Mamuriyah, Richard, and Haeruddin, "Implementation Mean Imputation And Outlier Detection For Loan Prediction Using The Random Forest Algorithm," *JITK (Jurnal*

VOL. 11. NO. 3 FEBRUARI 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i3.7613

Ilmu Pengetahuan dan Teknologi Komputer), vol. 10, no. 4, pp. 937–944, Jun. 2025, doi: 10.33480/jitk.v10i4.6437.

[27] F. Yasin, M. Raafi'u Firmansyah, D. Aldo, and M. A. Amrustian, "Multivariate Forecasting of Paddy Production: A Comparative Study of Machine Learning Models," *Jurnal Teknik Informatika (JUTIF)*, vol. 6, no. 3, pp. 2723–3863, 2025, doi: 10.52436/1.jutif.2025.6.3.4681.

[28] A. F. AlShammari, "Implementation of Model Evaluation Using Confusion Matrix in Python," *Int J Comput Appl*, vol. 186, no. 50, pp. 42–48, Nov. 2024, doi: 10.5120/ijca2024924236.

[29] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 1, Apr. 2021, doi: 10.28932/jutisi.v7i1.3216.

[30] J. Bleicher, E. E. Ebner, and K. H. Bak, "Formation and Analysis of Volatile and Odor Compounds in Meat—A Review," Oct. 01, 2022, *MDPI*. doi: 10.3390/molecules27196703.

[31] J. K. Seo, J. U. Eom, and H. S. Yang, "Application of volatilomics for meat quality, authenticity, and adulteration detection," Dec. 01, 2025, *Elsevier B.V.* doi: 10.1016/j.afres.2025.101182.