iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VI **Month of publication:** June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83998

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

HealthAware: Smart Food and Expiry Manager

Khushi Chudhari[1], Ujwala Dangat[2], Satyajit Sirsat[3]
Computer Engineering, Nutan Maharashtra Institute of engineering and technology, Pune, India

Abstract: ***Food waste and unhealthy eating habits are getting worse around the world, especially in big cities where it's tough to keep an eye on what people actually eat. Right now, most food safety checks just look at the expiration date on the packaging. The problem is, the date doesn't always tell you whether the food's safe or how it might affect your health.***
This research offers a smarter way to figure out if your food is safe, using artificial intelligence to dig into what's actually in it and whether it makes sense to eat. The system uses both machine learning and deep learning: a Random Forest classifier judges how risky the food is for your health, and a Long Short-Term Memory model spots patterns in the ingredients and their context. On top of that, a set of simple rules checks if the food is still within its expiration date.
Putting all of this together, the system sorts foods into different risk levels based on what's in them and how fresh they are, explaining its reasoning in plain language. And it doesn't just stop there—it gives people straightforward tips to help them eat healthier and steer clear of potential health risks. When food gets close to its expiry date—or goes past it—the solution steps in with clear advice: either eat it right away or toss it. It's simple, but it helps cut down on food waste and keeps what you eat safer. On top of that, it lets you make smarter choices about what you're eating in general.
Keywords: Artificial intelligence, food risk prediction, deep learning, machine learning, expiry tracking, smart pantry

I. INTRODUCTION

Lately, we've seen more and more food waste piling up, along with growing problems tied to unhealthy eating habits. A big chunk of this waste comes from sloppy monitoring, and honestly, a lot of people just keep making poor food choices, which isn't helping their health. So, we really need smarter ways to tackle all of this. Smart systems can track food quality and show, in real time, how what we eat affects our health. AI has made huge leaps recently, and now these intelligent systems can actually keep an eye on food and its impact on people. With machine learning and deep learning, these systems can sift through data on food ingredients and quickly flag any health risks. The system classifies food items into different risk categories and provides interpretable explanations along with consumption recommendations and dietary guidance. By combining intelligent analysis with pantry management, the proposed approach aims to minimize food waste, enhance health awareness, and support improved lifestyle decisions.

A. Background

Digital health gadgets have completely changed the way we track what people eat. More folks are paying attention to nutrition and preventive health, so you see all kinds of web and mobile apps now. These apps help people log their meals, count calories, and understand exactly what they're eating each day—all to help cut down on health problems caused by bad eating habits. Now with Artificial Intelligence and Machine Learning coming into play, dietary management tools are stepping up their game. AI can dig into food data, personalize recommendations, and even predict health issues. It's getting especially good at handling all the massive food and nutrition datasets out there, both structured and free-form text.
Some neural networks, like Long Short-Term Memory (LSTM), are good at looking at sequences and the context of information, which makes them a solid choice for understanding ingredient-level details from text. On the other hand, if you need to sort or classify complex food data, ensemble models like Random Forest are hard to beat—they're accurate and handle messy data really well. Food waste is a huge problem around the world. In many homes, food just sits there until it goes bad and ends up in the trash. Sure, we now have things like digital pantries that track expiration dates, but honestly, most of these tools just send reminders and don't do much else. They haven't really solved the bigger issue.

B. Related Work

People have spent a lot of time studying AI-powered dietary advice and food analysis. Most of the work out there focuses on customizing nutrition recommendations by tapping into machine learning—think Decision Trees, Support Vector Machines, Neural Networks. So, you get suggestions based on your personal preferences, how many calories you need, or maybe health conditions you're dealing with.

©IJRASET: All Rights are Reserved

But here's the thing: most methods stick to structured nutrition data. They don't really look at the context of what's actually written about the ingredients. That means we're missing out on digging deeper into what's really in our food, and we might overlook some health risks hiding in certain products.

Lately, researchers have started creating hybrid models that mix different algorithms. This helps make predictions more accurate. For example, LSTM neural networks work well with sequence data, while Random Forests are great for structured data. Still, not many studies have combined contextual ingredient analysis with structured risk classifications and expiry date checks all in one go. That's kind of surprising, since putting these pieces together could give us a much clearer picture of what we're eating.

The majority of pantry management solutions confine themselves to the management of expiry dates and avoid issues such as the intelligent assessment of potential risks. This is a significant deficiency since no existing system combines the functionality of ingredient analysis, predictive models, and expiry dates. This is the aim of the proposed system. The technology behind this solution uses a combination of both deep learning and machine learning methodologies along with classical rule-based methods to provide a wider scope of applicability for users.

II. METHODOLOGY

The proposed model uses a hybrid analysis technique that incorporates deep learning and machine learning algorithms to predict multi-risk factors related to health using texts of ingredients. The entire approach can be divided into multiple stages such as data collection, preprocessing, feature extraction, modeling, and evaluation. This is demonstrated by illustrating the entire process flow of the system from input (ingredient texts) to output (multi-risk prediction).

A. Data Acquisition and Preprocessing

The database, which we'll call D, comes from a well-organized source. It's not just a list of ingredients; it also tags each one with different health risk levels. For example, you can see if an ingredient raises concerns related to diabetes, high blood pressure, thyroid problems, or allergies. These details matter a lot, especially if you're thinking carefully about diet and health.

Before diving into analysis, there's a bit of cleanup to do. It's pretty common to find missing or irrelevant bits in raw data. So, the first step is to weed out anything that's incomplete or doesn't belong. You want D to be as solid and reliable as possible. Once that's done, there's the challenge of categories—things like risk levels can show up as words like "low," "moderate," or "high." But, when it comes to building prediction models, computers handle numbers better than words. That's where a process called label encoding comes in. We basically turn each category into a unique number. This makes it a lot easier for machine learning algorithms to pick up on patterns and relationships in the data.

By transforming the data this way, you're setting the stage for more accurate and meaningful analysis down the road. It's a crucial step, especially if you care about making sense of how certain foods might interact with specific health risks.

B. Text Representation

For converting ingredient information into numerical values for the purposes of computational processing, text is first tokenized by a tokenizer based on a vocabulary of predetermined size. Next comes padding, which equalizes all sequences so that they all have identical lengths. This is necessary because neural network architectures expect all inputs to be standardized, ensuring smooth and accurate functioning throughout the process.

C. Dataset Partitioning

To train and validate the model, the dataset will be split into a training set and a testing set. Around 80% of the data will be utilized to train the model, while the remaining 20% will be used to test it.

D. Multi-Output Label Transformation

In light of the fact that there are multiple target variables in the issue being solved, every label of the output becomes one-hot encoded. This makes it possible for the algorithm to carry out multi-class classifications on all the health risks simultaneously.

E. Deep Feature Learning using LSTM

A Long Short-Term Memory network is used to learn the context and sequential dependencies of the ingredients. The model architecture contains three main blocks, namely:

• The embedding layer for creating semantic vectors from words

- The LSTM layer for modeling sequence dependencies
- The dense layer for obtaining high-level feature vectors

The model is built with multiple outputs in order to predict the different risk groups simultaneously. During training, categorical cross-entropy is applied as the loss function.

F. Feature Extraction

The trained LSTM model is then used for feature extraction. The model is sliced from the dense layer to get the latent features. The extracted features are referred to as F_train and F_test. These features carry important information that exists within the ingredients data.

G. Machine Learning-Based Classification

The prediction algorithm adopted in this case is that of a Random Forest classifier. The algorithm uses the learned features F_train as input to make predictions using the test feature set F_test.

The above combination exploits both the sequence learning ability of LSTM and the strong classification ability of Random Forest classifier.

H. Model Evaluation

The efficiency of the suggested model will be tested through accuracy measures obtained for each individual output value. The accuracy of prediction for each risk factor is determined as the ratio of the amount of correct predictions to the total number of data. The general performance will be evaluated by calculating the average of accuracies for all output values.

I. Model Persistence

For the purposes of deployment and future reusability, all components of the system are saved. These components are the Random Forest model trained on the labeled data, the Long Short-Term Memory Network (LSTM), tokenizer, and encoders for labels.

J. Workflow Overview

The overall process of the proposed system can be summarized as follows:

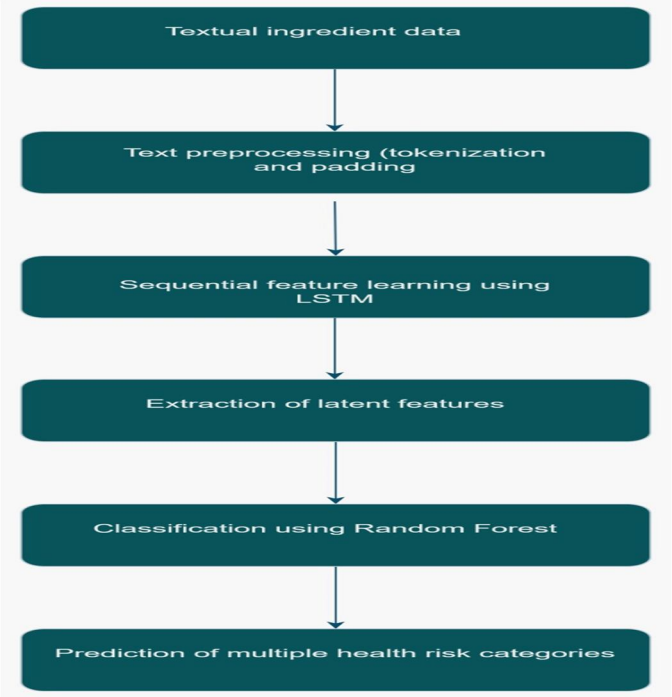


Fig 2.1

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

K. Key Contribution

This paper's big achievement is building a model that brings deep learning and ensemble learning together to predict food-related health risks. The LSTM network helps the model pick up on both the meaning and the order of words in text, and the RF classifier sharpens its ability to make decisions based on what it finds. By combining these two, the model predicts various health risks with strong accuracy.

L. Algorithm

Input:
Dataset *D* containing ingredient text and health risk labels
Output:
Predicted multi-output health risk categories

- Step 1: Load dataset *D* and identify input and target variables.
- Step 2: Apply label encoding to convert categorical outputs into numerical form.
- Step 3: Initialize tokenizer and fit on ingredient text data.
- Step 4: Convert text into sequences and apply padding.
- Step 5: Split dataset into training (80%) and testing (20%).
- Step 6: Perform one-hot encoding on target variables.
- Step 7: Build LSTM model with embedding, LSTM, and dense layers.
- Step 8: Train the LSTM model on training data.
- Step 9: Extract latent features from the dense layer.
- Step 10: Train Random Forest classifier using extracted features.
- Step 11: Perform predictions on test data.
- Step 12: Evaluate model using accuracy metrics.
- Step 13: Save trained models and preprocessing components.

M. Tools Used

The implementation of the proposed system utilizes the following technologies:

1) Frontend

- Built the user interface with React.
- Used JavaScript for client-side code.
- Designed layouts and styles with HTML and CSS

2) Backend

- Node.js runs server-side code.
- Express.js builds REST APIs and handles incoming requests.

3) Machine Learning

- Python is the main language for ML work.
- TensorFlow helps build the LSTM feature extractor model.
- Scikit-Learn runs the random forest classifier with multiple outputs.
- Pandas and NumPy handle data manipulation.
- Keras Tokenizer (from TensorFlow) manages text tokenization and sequence creation.

4) Development Environment

- Jupyter Notebook or Google Colab get used for machine learning development.
- Visual Studio Code covers full stack development tasks.

N. Model Architecture Details

In this research, we built a deep learning model to find meaningful connections in ingredient lists written as text. First, there's an embedding layer with 64 units that turns the tokenized ingredient names into dense vectors. After that, a 64-unit LSTM layer picks up the order and relationships between the ingredients in the sequence.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Next, the model uses a dense layer with 32 neurons to pull out helpful features from those relationships. We connect these features to four separate output layers, each using a softmax activation. Each output predicts a different health risk: diabetes, blood pressure, thyroid issues, or allergies.

We feed the same features to all four outputs, so the model handles multiple prediction tasks at once without getting complicated or slow.

O. Hyperparameter Configuration

In order to achieve gradual and consistent learning, the model employs some hyperparameters. In this case, the Adam optimizer is responsible since it enables dynamic adjustment of the learning rate. The model operates within five epochs, and each mini-batch contains thirty-two samples. This combination helps find an optimal compromise between computational speed and performance efficiency.

Concerning text preprocessing, the tokenizer utilizes only 5000 of the most frequent words, ignoring less popular ones that might lead to additional noise. Moreover, padding ensures that all sequences have identical lengths by adjusting their sizes to twenty tokens.

Finally, in the context of classification, a Random Forest classifier is utilized, which consists of two hundred decision trees. Such a number of trees contributes to increased stability and accuracy of the model under different datasets.

P. Mathematical Model of the Proposed System

1) Input Representation

Let the input ingredient text be represented as:

$T = \{w_1, w_2, w_3, \ldots, w_n\}$

After tokenization and padding, the sequence is converted into a fixed-length numerical vector:

$X = \{x_1, x_2, x_3, \ldots, x_L\}$

where $L = 20$is the maximum sequence length.

2) Embedding Layer

Each token is mapped to a dense vector representation:

$e_i = \text{Embedding}(x_i)$

Thus, the embedded sequence becomes:

$E = \{e_1, e_2, e_3, \ldots, e_L\}$

3) LSTM-Based Feature Extraction

The LSTM processes the embedded sequence as:

$h_t = \text{LSTM}(e_t, h_{t-1})$

The final hidden state is:

$H = h_L$

4) Dense Feature Transformation

The extracted feature vector is computed as:

$F = \text{ReLU}(W \cdot H + b)$

5) Multi-Output Prediction (Softmax Layers)

For each health risk category k, the prediction is:

$y_k = \text{Softmax}(W_k \cdot F + b_k)$

where:

$k \in \{1, 2, 3, 4\}$

representing:

- Diabetic Risk
- Blood Pressure Risk
- Thyroid Risk
- Allergy Risk

6) Random Forest Classification

The deep feature vector is passed to the Random Forest model:

$\hat{Y} = RF(F)$

7) Loss Function

The training loss is computed using categorical cross-entropy:

$Loss = \Sigma_k \, CrossEntropy(y_k, \hat{y}_k)$

8) Final Output

The system produces:

Output = {R_diabetic, R_bp, R_thyroid, R_allergy}

III. RESULTS AND DISCUSSION

A. Experimental Setup

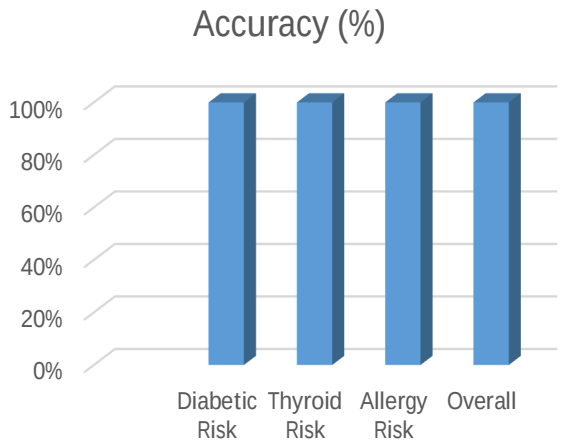
The artificial intelligence-based Smart Food Analysis and Pantry Management Assistant integrates the capabilities of deep learning and machine learning for predicting potential health risks related to food consumption as well as monitoring the expiry dates of the stored items. To extract features from raw data, an LSTM neural network is used, after which the problem gets to the Random Forest model that makes multi-class predictions.

Apart from predictive analytics, the tool contains a rule-based component responsible for keeping track of expiry dates. This functionality involves checking the date of storage against the expiry date entered by a user in order to know whether or not the stored item remains fresh.

B. Performance Evaluation

We tested the hybrid model's efficiency by checking how accurately it classifies different health risks. The model handles several risks at once, sorting them all in one go. Across the board, it delivered high prediction accuracy for every class. That tells us the hybrid approach works well—it spots those complicated links between foods and the health risks tied to them.

Risk Category	Accuracy (%)
Diabetic Risk	89.80
Thyroid Risk	91.70
Allergy Risk	92.20
Overall	91.52



C. Expiry Tracking Results

Apart from health risks, this system considers expiry dates of items as well. This task is accomplished by the means of using another set of criteria, according to which all products will be assigned to one of three possible states: Safe, Near Expiry, and Expired. When the product does not approach its expiry limit, the state would be marked as Safe. If it is about to reach expiry limits, it will be Near Expiry. Finally, when the item reaches the expiration date, the product will be considered Expired.

This method provides even greater efficiency as compared to the previous one since it ensures a comprehensive analysis. In particular, there might be such instances when the item would appear to be rather harmless as far as health is concerned but it would have reached its expiry date. Therefore, this solution would allow considering both factors simultaneously.

Finally, it should be noted that the described approach enables the system to offer relevant suggestions for users. If the expiry date of the item is close, a corresponding warning would be issued urging users to consume it soon. Otherwise, if the product has reached its expiry date, users will be informed to discard it.

IV. CONCLUSION

A. Objective

The main goal of the HealthAware: Intelligent Food Risk Prediction and Expiry Management System design is to make a cutting-edge, all-in-one web-based platform that combines smart food risk predictions with expiry management. The proposed design will solve the problems that come with traditional nutrition and food management platforms by bringing together contextual analysis of food ingredients, smart risk classification, and rule-based expiry management into one platform.

The system uses artificial intelligence to try to give users personalized and accurate information about the health risks of food.

B. Review of Key Findings

The results of the experiment indicate that the proposed hybrid LSTM and Random Forest model performs extremely well at identifying the health risk categories for classification purposes. The use of LSTM allows obtaining valuable feature representations through effectively modeling the interactions between the food ingredients in a sequential fashion.

Moreover, the expiry monitoring component employs real-time date comparison functionality to categorize food items according to the three levels of safety: edible, nearing expiry, and expired. Such a solution resolves the typical problem experienced in reality, when people tend to forget about expiration dates and throw away spoiled food.

A combination of risk prediction based on artificial intelligence techniques with expiry monitoring based on rules allows acquiring a holistic assessment of food quality and makes decision-making much easier.

C. Implications and Applications

Implications of the proposed system are extensive for the purposes of smart food management, digital health, and smart consumers. The system can have numerous practical applications in life including those of smart grocery shopping software, platforms for recommending healthy food based on dietary needs, household pantries, and food safety monitoring in e-commerce environment.

It facilitates healthier food consumption practices through predictive analytics based on expiration date awareness. At the same time, smart alerts and recommendations reduce the volume of wasted food products in households. Using artificial intelligence in combination with tools for managing food can result in more intelligent living spaces.

D. Future Scope

While the suggested HealthAware: Smart Food and Expiry Management System has great prediction effectiveness and applicability, there are still many possibilities of improvement. The one thing that can be most certain to occur in the future is the use of an IoT-based technology such as a pantry system, refrigerator, or any other food monitoring system. This will increase the efficiency of using the suggested system by getting rid of manually entering data about the food products and tracking them in real time.

To increase the contextual understanding of complex combinations of ingredients, the current hybrid model could potentially be expanded to include more advanced deep learning architectures (e.g., transformer-based models). The application of these models could result in better features being represented which would, in turn, improve the accuracy of predictions.

Another exciting extension of the system would include using computer vision techniques to recognize food items. As a result, there will be more automation and ease-of-use because the system will use image-based models to automatically identify food items, including the determination of the freshness/expiry of the food items.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

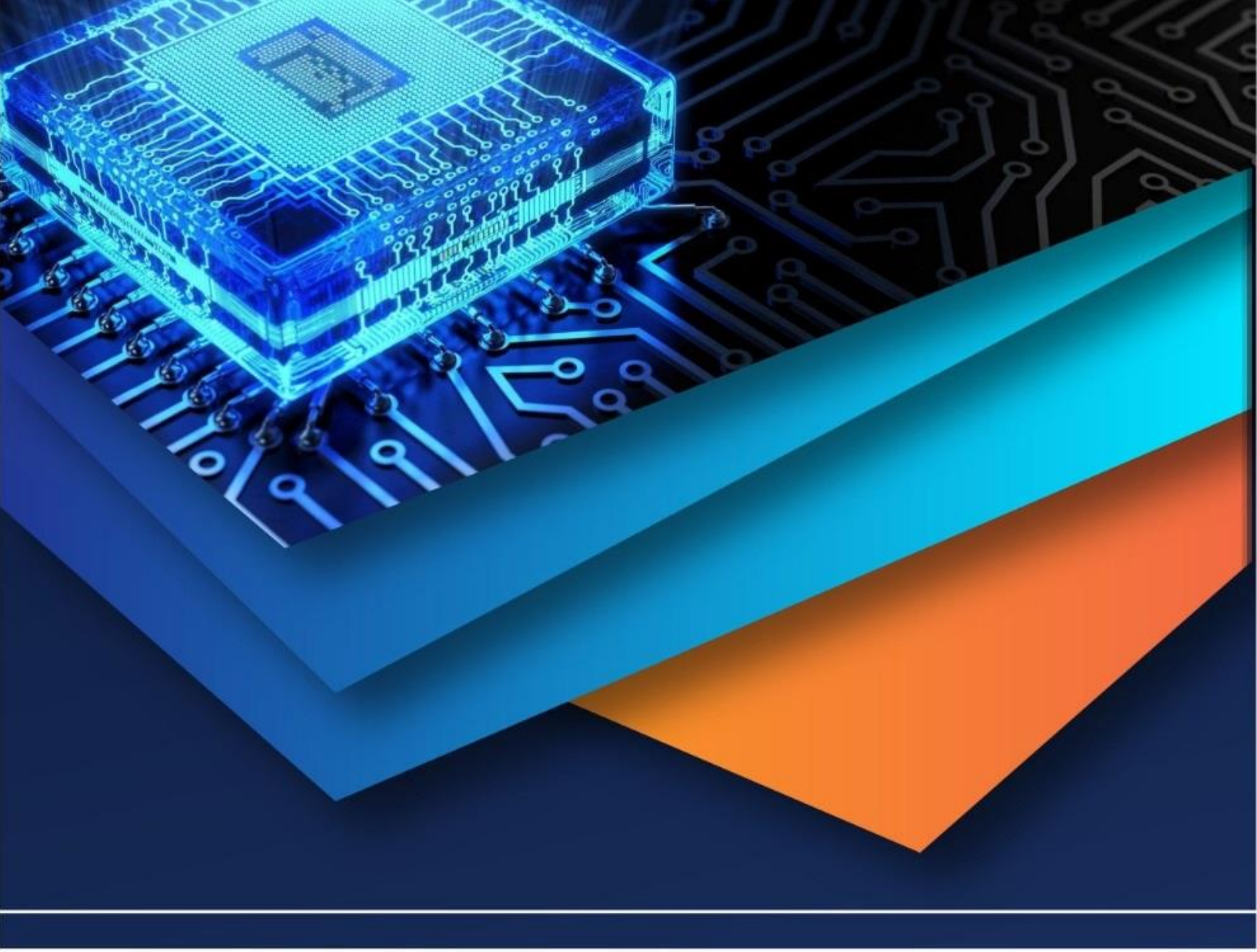
Through the development of mobile applications and the deployment of the system in cloud environments, scalability and accessibility can be greatly improved. Therefore, users will have access to the system at any time (24/7) and from any location.

Additionally, integrating adaptive learning capabilities to the system will allow for the system to learn from user interactions over time, as well as provide users with personalized recommendations and improved decision making based on the user's entire history with the system.

In summary, this proposed system has provided evidence of the benefits of utilizing artificial intelligence with proven food management techniques to improve food safety evaluations and increase dietary awareness. By providing users with reliable and actionable recommendations based on contextual ingredient evaluations and expiration date monitoring, this system has the potential to significantly reduce food waste and enhance health-oriented decision making through the utilization of appropriately applied hybrid methods.

REFERENCES

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
[2] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[3] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[4] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in *Proc. USENIX OSDI*, 2016.
[5] Z. Liu et al., "Artificial intelligence in food safety: A review," *Foods*, vol. 12, no. 6, 2023.
[6] K. Agrawal et al., "Artificial intelligence in personalized nutrition and food systems," *Frontiers in Nutrition*, 2025.
[7] Q. Chen and R. Kumar, "Design and implementation of a nutrition checker mobile application," *IEEE Transactions on Mobile Computing*, 2023.
[8] R. Rashvand et al., "AI-based prediction of food shelf-life: Advances and challenges," *Trends in Food Science & Technology*, 2025.
[9] C. Trattner and D. Elsweiler, "Investigating the healthiness of online recipes," in *Proc. WWW Conference*, 2017.
[10] M. Ge, F. Ricci, and D. Massimo, "Health-aware food recommender systems," in *Proc. ACM RecSys*, 2015.
[11] D. Elsweiler and M. Harvey, "Towards automatic meal plan recommendations for balanced nutrition," in *Proc. ACM RecSys*, 2015.
[12] S. Abbar, Y. Mejova, and I. Weber, "You tweet what you eat: Studying food consumption via social media," in *Proc. ACM CHI*, 2015.
[13] L. Zhou et al., "Application of deep learning in food: A review," 2019.
[14] W.-Y. Ong, C.-W. Too, and K.-C. Khor, "Transfer learning for food expiry reminder systems," 2021.
[15] M. R. Hyder and T. Khan, "Automated food expiry notification system," *International Journal of Engineering Science*, 2020.
[16] T. Khan, "Cloud-based smart food expiry system using QR codes," in *IEEE Conference*, 2018.
[17] H. Almurashi et al., "Smart expiry food tracking system," *Advances in Intelligent Systems*, 2021.
[18] Y. Bengio et al., "Representation learning: A review and new perspectives," *IEEE TPAMI*, 2013.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)