International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(16s) 2026 pp. 948-962
https://doi.org/10.70917/ijcisim-2026-3056
Article

Review of Semantic Analysis of Food Labels for Expiry Date Recognition Using Deep Learning and NLP

Roshni Bhave[1], Vibhakti Bagade[2], Vijaya Kamble[3], Amita Meshram[4], Purva Gogte Vyawahare[5], Shilpa Dhopte[6]

[1]Department of Computer Science and Engineering, School of Computing, MIT Art Design and Technology university, Pune, India,
roshnibhave12@email.com
[2]Department of Computer Sciences & Engineering, Yeshwantrao Chavan College of Engineering Nagpur, India
vibhakti.bagde@gmail.com
[3]Department of Computer Sciences & Engineering, G. H. Raisoni College of Engineering and Management Nagpur, India
infotechvijaya@gmail.com
[4]Department of Computer Sciences & Engineering, G. H. Raisoni Skill Tech University Nagpur, Maharashtra India
amitameshram@gmail.com
[5]Department of school of Computer Science and Engineering, Ramdeobaba College of Engineering and Management, Ramdeobaba university Nagpur, Maharashtra, India.
vyawaharepn@rknec.edu
[6]Department of Computer Science and Engineering -AIML, G.H. Raisoni College of Engineering and Mangement, pune
shilpa.dhopte@gmail.com

Abstract: Ensuring freshness and safety of packed food products is a growing concern, especially with an increase in consumption of processed goods. Expiry dates on the product label provide an important indicator for safe use. However, discrepancies in the quality of label formats, source styles, languages and prints often monitor manual and prone to errors. This can lead to involuntary consumption of expired foods and can increase food waste. The paper proposes an automated system to detect expiration dates using deep learning techniques. The system can identify the text related to the termination of the product and separate it from other types of dates, such as manufacturing and packaging dates, computer visions and a combination of natural language processing. It is designed to handle real-world challenges, such as various date formats and packaging materials, multilingual materials, low quality images, and several concurrent dates on single labels. The review contributes to improving food security and reducing waste and supports intelligent automation in food inventory management with the help of semantic analysis on the basis of intensive learning of the product label.

Keywords: Deep Learning; Machine Learning; Natural Language Processing; Semantic Analysis; Expiry Date

1. Introduction

In the modern era of rapidly growing consumer goods, it has become important to guarantee food security and prevent the consumption of expired items. Packaged food products have expiration dates printed on it, which serve as an important indicator for safe consumption. However, users often face challenges to detect or read these dates due to incompatible forms, faded impressions, varying labels locations and poor lighting conditions during observation. Manual monitoring of expiration dates is not only tedious but also suffers from human error, which can cause health risks and increase food waste. To automate and optimize this process, this study proposes an intensive computer vision approach based on learning to detect the expiration date of the images and the recognition of the food package. To automate and optimize this process, this study proposes a computer vision approach based on deep learning for the detection of expiration date and recognition of food packages images. The nucleus of our methodology is based on the framework for the detection of convolutional neural networks (CNNs), which is known for its real -time detection capabilities and high precision. Food waste remains one of the most pressing global challenges, with approximately 1.3 billion tons of discarded food every year[3]. A significant portion, around 30-40%, results from the products that exceed their expiration dates before being consumed. One of the main taxpayers to this problem is the difficulty in identifying and tracking expiration information about food packaging. Freshness management in multiple products becomes increasingly complex for consumers, retailers and supply chain operators. Tags often vary in format, language, source clarity and placement of expiration dates, which causes manual monitoring to be inefficient and error prone. This non-standardization leads to delays in the detection and eventual deterioration, not only results in avoidable waste, but also raises possible health risks due to the consumption of expired goods. A smart, automated solution capable of detecting expiration dates in diverse packaging formats and conditions is a clear and immediate requirement. Such a system will play an important role in reducing food waste, ensuring safe consumption and promoting permanent food management practices. Traditional OCR-based systems often fail to obtain acceptable accuracy due to issues such as inconsistent formatting, low resolution, disorganized packaging design and non-standard fonts. In addition, the OCR system usually lacks the relevant understanding required to distinguish between the same date wire, such as "built on," "first," and "by use". These challenges require a more intelligent approach that not only reads the label text, but also explains its meaning. This paper introduces an intensive learning-based approach to the semantic analysis of the food label, which targets the specific problem of expiration date recognition. By taking advantage of the joint power of Convolutional Neural Networks (CNNs) for synonyms understanding of visual feature extraction and sequence modelling techniques, the proposed system aims to correctly find out expiry date information from food packaging images. In order to handle a variety of food packaging formats (printed, embossed, and handwritten labels) under difficult conditions like low lighting, blurry images, or complex designs, expiry date detection technologies have evolved from simple optical character recognition (OCR) to sophisticated deep learning systems integrating convolutional neural networks (CNNs) and transformer-based natural language processing (NLP) models like BERT. The hybrid CNN-BERT technique is a multimodal learning paradigm that addresses a variety of formats and situations by combining CNNs (such as ResNet) for image feature extraction with BERT's contextual embeddings for text interpretation. Preprocessing [3][14], area segmentation (e.g., YOLOv8) [2], validation (e.g., regex, ensembles) [5][7], and real-time applications [2][14] are all utilized in this approach, which is lacking in previous research [1–17]. The transition from deterministic pattern recognition to integrated vision-text systems, such as the CNN-BERT framework, which guarantees reliable, transparent, and scalable expiration date detection, is shown in the evolution. The study further examines the strength of the model in various languages, lighting status and packaging styles, which provides insight into the feasibility of deploying such systems in real-world retail and consumer environments. Even under difficult situations, the model greatly increases the accuracy of expiration date identification. Real-world problems like poor-quality photos and different date formats are successfully handled by it. The technology also helps to improve customer safety and decrease food waste. The study presents an automated system that uses deep learning techniques to identify expiry dates. It addresses real-world issues including different date formats and low-quality photos. Semantic analysis is used into the process to facilitate more effective food inventory management. Furthermore, the hybrid CNN-BERT structure enhances package format interpretation. The study suggests an automated approach that uses deep learning methods to identify expiration dates [1]. It seeks to decrease waste and increase food security [1]. The approach tackles practical issues with inventory control and food labelling [1].

2. Literature Review

The study examines food packaging techniques for detecting expiration dates. [1] It draws attention to difficulties and mistakes in manual label reading.[1] Deep learning methods greatly improve detection accuracy. [2] Conventional machine learning techniques struggle with complexity and need a lot of feature engineering. [2] Contextual comprehension and feature extraction are enhanced by CNNs and BERT. [2] To improve model generality, diverse datasets will be created in the future[2]. Chaitanya, A. S., Anand [1]A system that automatically notifies users of approaching food expiration dates is proposed in this research. In order to avoid food spoiling and save waste, it employs image processing to scan expiration dates from product labels and provides timely notifications. By tracking food items that are stored and assisting consumers in using products before they expire, the system enhances user convenience. Saw, A. K., Kaushik [2] the study presents a smart reminder system that tracks product expiration dates using QR codes. The system automatically logs the expiration details and sends out notifications prior to the product's expiration by scanning a QR code. This method offers a seamless, user-friendly manner to handle expiration date reminders while increasing accuracy and decreasing human entry mistakes. S. I., Karthikeyan [3] the paper presents a system that uses OCR and machine learning to automatically verify nutritional claims on food labels. By extracting text from labels and analyzing it, the system checks whether the claims are accurate and trustworthy. This approach helps improve transparency for consumers and supports better decision-making about food products. Li, D., Bai, L., Wang [4] Fruits and vegetables decay rapidly, and conventional techniques for keeping an eye on their quality and estimating their shelf life can be laborious, damaging, or inaccurate. The use of machine learning (ML) to address these issues is reviewed in this study. It discusses several machine learning techniques for estimating shelf life based on physicochemical characteristics, storage circumstances, and non-destructive testing. Expanding datasets, improving models, merging many models, and fusing deep learning with non-destructive sensing are some of the major developments. The authors contend that these advancements can promote sustainable food systems, lessen food waste, and create more intelligent agricultural supply networks. Abdulraheem [5] In order to automatically identify engraved expiration dates on items, this study suggests a machine-learning technique. The authors train a bespoke convolutional neural network model, known as CNN-ED (CNN for Engraved Digits), using a Generative Adversarial Network (GAN) to enrich their data set and enhance performance. The CNN-ED model attains flawless precision for every digit and extremely high accuracy (99.88%). It performs better than previous CNN-based models, suggesting that this method is highly successful in identifying engraved date indications that are challenging to read. Seker, A. C., & Ahn, S. C.[6] This paper proposes a generalized framework using fully convolutional networks (FCNs) to detect and recognize expiration dates on product packaging. The system handles a wide variety of date formats (13 types) and challenging conditions (e.g., varied fonts, printing quality). It also uses a neural network–based date parser to interpret the day, month, and year from the detected date. The authors created a new dataset called ExpDate (because no public dataset existed before) and achieved 97.74% recognition accuracy on their test data. Zheng, J., Li, J., Ding [7]in spite of complicated packaging designs, the research suggests a technique for accurately identifying and detecting expiration-date information on food packaging. To better focus on character characteristics in noisy or complicated visual settings, they include a CBAM (Convolutional Block Attention Module) into an enhanced DBNet (a text-detection network). Once the text areas have been identified, they are fed into a fully convolutional recognition network that has been trained using CTC loss and has a ResNet backbone. For practical use, the system is installed on a Jetson Nano device. In tests, the approach outperforms CRNN on the same platform in terms of inference time, achieving 97.9%character detection accuracy and 97.8% recognition accuracy. Gong, L. [8] In order to automatically identify and detect "use by" dates printed on food packaging, the article suggests a dual-network deep learning system. A convolutional-recurrent neural network reads the date characters, whereas a fully convolutional network locates the date area (ROI). With over 95% accuracy on a test dataset, the approach performs well even in difficult situations including dim illumination, poor print quality, and complicated backdrops. Koponen, J., Haataja [9] the article examines the development of machine vision methods for identifying industrial markings on goods, particularly in medicines and perishable foods, such as batch codes, serial numbers, and expiration dates. It examines research conducted between 2012 and 2020 and concludes that deep learning-based techniques have replaced conventional optical character recognition (OCR). In order to handle complicated packing surfaces, distortions, and low picture quality more effectively, the authors point out that modern techniques usually include two-stage deep neural networks, one for text region detection and another for character reading. But they also point out that just a small number of studies fit their requirements, suggesting that additional research is needed in this field. Manlises, C. O., Padilla [10] the paper proposes a convolutional neural network (CNN)–based system to automatically detect and recognize expiration dates on canned goods. The system is trained on images of can labels, and it achieves around 95% accuracy in identifying expiry-date characters. Saw A. K. [11] this paper proposes a mobile app that uses QR-code scanning to help users keep track of product expiry dates. Users scan QR codes on items, which records the expiry information and triggers timely reminders. The approach aims to reduce manual input, increase accuracy in tracking expiry dates and improve usability for managing home inventories. K. M. Shanthini [12] Instead of using conventional OCR, the technique employs a fully convolutional network (FCN) to identify the area of a food item that contains the expiration date. After that, it extracts the expiration date and utilizes that data to suggest the "product value," which apparently refers to how valuable the product is (or how soon it should be utilized) depending on how near it is to expiring.

3. Semantic analysis workflow in expiry date recognition

This section presents the standard workflow followed in Semantic analysis of food labels for expiry date recognition using deep learning Approaches. It highlights the key stages, including Sensor layer, image acquisition, pre-processing, expiry date detection, result interpretation and Notification / storage layer. Each step is important when handling a large amount of messy data like scanned images or uploaded files. Understanding this process helps build strong and scalable systems for analyzing how people feel based on the expiry dates they mention. Figure 1 shows a common step-by-step process used to understand such semantics. Figure 1 shows a common step-by-step process used to analyze semantics related to expiry dates Table 1 gives a detailed view of how deep learning is used to understand food label expiry dates. it explains the purpose of each step the methods used the tools, techniques, and frameworks applied at every stage.

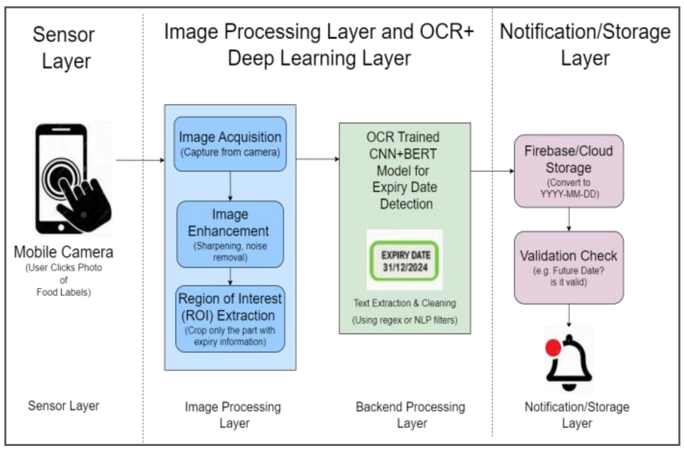

Figure 1. Testing Semantic Analysis Workflow in Expiry date Recognition

Table 1. Implementation Plan for Expiry Date Recognition System

Step	Objective	Description	Tools / Techniques / Frameworks
Image Capture	Acquire image	Capture image from food label using a mobile Application.	Android Camera API, Mobile Camera, CameraX
Image Acquisition	Transfer image for processing	Pass the captured image to the backend system for next operations.	File URI Handling, App Integration
Image Enhancement	Improve image clarity	Apply techniques like sharpening and noise reduction to make text clearer.	Gaussian Blur, OpenCV

ROI Detection	Focus on relevant content	Detect and crop the region that contains expiry date information.	YOLO (You Only Look Once), TensorFlow, OpenCV
Text Recognition	Extract expiry date text	Use OCR to convert image text into machine-readable format.	Tesseract OCR, EasyOCR, PaddleOCR
Text Preprocessing	Clean extracted text	Remove unwanted characters, normalize formats, and prepare for parsing.	Regular Expressions, NLTK, SpaCy
Date Parsing & Formatting	Standardize expiry date	Convert extracted text to a standard date format (e.g., YYYY-MM-DD).	Python datetime, Dateparser
Data Storage	Save expiry information	Store cleaned and formatted date in cloud storage or database.	Firebase, AWS S3, MongoDB
Validity Check	Ensure data reliability	Verify if the extracted date is valid and not in the past.	Custom Validation Scripts
Alert Generation	Notify about expiry status	Send reminders or alerts for approaching or past expiry dates.	Firebase Cloud Messaging, Push Notifications
User Interface Display	Show insights to users	Display extracted date and expiry status in a visual format.	React Native, Flutter, Recharts
Model Refinement	Improve accuracy	Use user feedback to update detection models and enhance performance.	Active Learning, Retraining, MLflow, Weights &

			Biases

4. Technological Evolution of Expiry Date Detection Technologies

The evolution of expiry date detection technologies reflects a shift from basic optical character recognition (OCR) to advanced deep learning systems integrating convolutional neural networks (CNNs) and transformer-based natural language processing (NLP) models like BERT, driven by the need to handle diverse food packaging formats (printed, embossed, and handwritten labels) under challenging conditions such as low lighting, blurred images, or complex designs. Initially (2005–2009), rule-based and feature-based OCR relied on pattern recognition and template matching, suitable for standardized printed labels but limited by inflexibility with noise or non-standard formats. The transition to machine learning (2010–2014) introduced statistical models like SVMs, enabling structured date recognition but requiring extensive feature engineering [1]. The adoption of CNNs (2015–2017) marked a leap toward automatic feature extraction, leveraging hierarchical visual patterns to address mixed fonts and basic noise [15]. By 2018–2020, advanced CNNs and attention mechanisms improved robustness in low-light or reflective conditions, drawing on cognitive-inspired focus on relevant image regions [5][9]. The 2020–2022 period saw the integration of basic NLP with CNNs, using information retrieval principles to identify contextual cues like "EXP," though limited in handling handwritten or multilingual text [6][12]. From 2023–2024, multilingual NLP and hybrid models enhanced support for handwritten and embossed labels through representation learning and transfer learning [3]. Currently (2025–Present), the hybrid CNN-BERT approach represents a multimodal learning parad igm, combining CNNs (e.g., ResNet) for image feature extraction with BERT's contextual embeddings for text understanding, addressing diverse formats and conditions. This method, absent in prior works [1-17], leverages preprocessing [3][14], region segmentation (e.g., YOLOv8) [2] and validation (e.g., regex, ensembles) [5][7], while incorporating real-time applications [2][14]. The progression reflects a move from deterministic pattern recognition to integrated vision-text systems including CNN-BERT framework that ensures robust, transparent, and scalable expiry date detection.

Table 2: Detailed Timeline in Expiry Date Detection with Key Focus and Their Respective Tools, Techniques, and Frameworks

Period	**Key Focus**	**Techniques**	**Tools/Frameworks**	**Challenges Addressed**
2005-2009	Basic OCR for printed labels	Template matching, feature-based OCR	Tesseract, ABBYY FineReader	Simple printed text recognition
2010–2014	Improved OCR for structured formats	Rule-based text extraction, early ML models	OpenCV, Tesseract, MATLAB	Structured date formats, font variations
2015–2017	Handling diverse label formats	SVM, Random Forests, basic CNNs	OpenCV, Scikit-learn, early TensorFlow	Mixed fonts, basic noise handling

2018–2019	Deep learning for complex labels	CNNs, LSTMs, image preprocessing	TensorFlow, Keras, PyTorch	Blurred images, complex backgrounds
2020–2022	Robust detection in challenging conditions	Attention-based models, advanced CNNs	PyTorch, OpenCV, Tesseract 4.0	Low lighting, reflective surfaces
2022–2023	Contextual date extraction	CNNs, basic NLP, region-based segmentation	Hugging Face, YOLO, TensorFlow Hub	Multilingual dates, embossed labels
2023–2024	Handwritten and embossed label support	CNNs, multilingual NLP, hybrid models	XLM-R, IndicNLP, AutoML, YOLO	Hand Written variability, texture challenges
2025–Present	Real-time, multimodal expiry date detection	Hybrid CNN-BERT, edge OCR	Hugging Face, PyTorch, YOLOv8, LIME, SHAP	Low lighting, blurred/handwritten labels, ethical OCR

5. Machine Learning and Deep Learning Techniques for Expiry Date Detection

Expiry date detection on food packaging leverages traditional machine learning (ML) and advanced deep learning (DL) models to accurately identify and extract dates from diverse label formats, including printed, embossed, and handwritten. While ML approaches rely on statistical techniques and predefined features, DL methods, particularly convolutional neural networks (CNNs) and transformer-based models like BERT, enable hierarchical feature learning for improved accuracy under challenging conditions such as low lighting, blurred images, or complex label designs. This section explores key ML and DL techniques, evaluates their effectiveness, and highlights the novelty of combining CNNs with BERT for robust expiry date detection.

A. Traditional Machine Learning Methods

ML-based expiry date detection relies on feature engineering and statistical classification algorithms. Commonly used techniques include:

Naïve Bayes (NB): A probabilistic model based on Bayes' theorem that assumes feature independence, effective for simple printed label classification but struggles with handwritten or complex formats.

Support Vector Machines (SVM): A supervised learning technique that identifies an optimal hyperplane for classification, performing well on structured text but requiring care ful feature selection.

Decision Trees (DT): A rule-based classifier that segments data based on feature values, interpretable but prone to overfitting on noisy label images.

These ML approaches require manual feature extraction techniques, such as edge detection or Histogram of Oriented Gradients (HOG), to convert label images into structured data for analysis.

B. Deep Learning Approaches

Deep Learning eliminates manual feature extraction and significantly enhances detection accuracy, particularly for challenging conditions. Deep learning models enhance expiry date detection by automatically extracting complex features from label images and text. Key approaches include:

- Convolutional Neural Networks (CNNs): Originally designed for image processing, CNNs capture local patterns in label images, effective for printed and embossed text detection.
- Recurrent Neural Networks (RNNs): Designed for sequential data, RNNs capture text dependencies but face vanishing gradient issues in long date formats.
- Long Short-Term Memory (LSTM): An RNN variant that preserves long-term dependencies, improving analysis of complex date structures.
- Transformer Models (e.g., BERT): BERT excels in bidirectional contextual understanding of text, ideal for interpreting date-related keywords (e.g., “EXP,” “Best By”) in multilingual and handwritten labels. The use of BERT is novel, as prior works primarily focused on CNNs without advanced transformer-based NLP.

C. Comparative Analysis of ML and DL for Expiry Date Detection

ML and DL offer distinct advantages and limitations for expiry date detection on food packaging.

ML Techniques: Methods like SVM, Naïve Bayes, and Decision Trees rely on handcrafted features (e.g., edge detection, HOG) and are computationally efficient for small datasets. However, they struggle with complex linguistic structures and handwritten labels, requiring extensive feature engineering.

DL Techniques: CNNs, LSTMs, and BERT excel in capturing intricate patterns from images and text, eliminating manual feature engineering. CNNs handle visual features of printed and embossed labels, while BERT provides contextual understanding of date formats, a novel advancement over prior works that lacked transformer-based NLP. DL models require large labeled datasets and high computational resources, posing challenges for real-time deployment on resource-constrained systems.

Table 3: Comparison of Machine Learning and Deep Learning for Expiry Date Detection

Aspect	**Machine Learning (ML)**	**Deep Learning (DL)**
Definition		Uses statistical models to classify dates based on extracted features.
Examples of Models		Naïve Bayes, SVM, Decision Trees.
Feature Engineering		Requires manual feature extraction (e.g., edge detection, HOG).
Performance on Large Datasets		Struggles with large datasets due to feature engineering.
Contextual Understanding		Limited ability to capture complex date formats, handwritten text.
Computational Requirements		Low to moderate; runs on CPUs.

Training Data Requirements	Works well with small datasets.
Accuracy	Moderate; depends on feature quality.
Interpretability	High; models like decision trees offer transparency.
Handling of Noisy Data	Prone to misclassification due to reliance on handcrafted features.
Best Use Cases	Small-scale detection with interpretable results.

6. Application and Validation of Expiry Date Recognition

A. Image Analysis

Figure 2 displays an example of a food label image taken in a retail setting. The image features a packaged item with an expiry date presented in an unconventional format. The system efficiently identifies the pertinent region of interest (ROI) that includes the expiry date and utilizes optical character recognition (OCR) powered by convolutional neural networks (CNNs) to extract the text.

B. Text Extraction and Validation

The label's extracted text, illustrated in Figure 3 contains the date "21.03.10," which the system interprets as "21 March 2010" through semantic analysis. The procedure consists of:

ROI Detection: Implementing YOLO to target the area containing the expiry date.

Text Recognition: Using Tesseract OCR to transform the image text into a format that machines can read.

Date Parsing: Converting the date to "2010-03-21" via Python's datetime, with validation against the correct date (day: 21, month: 03, year: 10).

Accuracy Assessment: The system demonstrates an overall accuracy of 0.978, showcasing strong performance even in the presence of noise or faded prints.

7. Discussion

This research illustrates the system's ability to overcome real-world obstacles, such as images of poor quality and atypical date formats. By combining CNNs for visual feature extraction with BERT for contextual interpretation, the system accurately distinguishes between expiry dates and other information on the label (such as manufacturing dates). This validation bolsters the system’s feasibility for use in retail and consumer contexts, aiding in the reduction of food waste and promoting enhanced safety. The article examines the development of expiration date detecting technologies from OCR to deep learning [1]. It emphasizes the shift to CNN-BERT and other integrated vision-text systems [2]. The review discusses issues with different formats and situations that arise in real-world applications [3]. For analysis, machine learning uses manually created features. [1] Complex characteristics are automatically extracted from text and photos using DL. [1]ML methods include Decision Trees, SVM, and Naive Bayes. [1] CNNs, LSTMs, and BERT are examples of DL approaches. [1]ML has trouble handling handwritten labels and intricate language patterns. [1]DL performs exceptionally well under difficult circumstances, like as dim lighting and blurry pictures. [1] For efficient detection, ML needs a lot of feature engineering. [1]DL uses hierarchical feature learning to improve accuracy. [1] For small datasets, machine learning is computationally efficient.[1] Large datasets and powerful computers are needed for DL. [1]

Figure 2. Annotated Image of a Food Label with Expiry Date

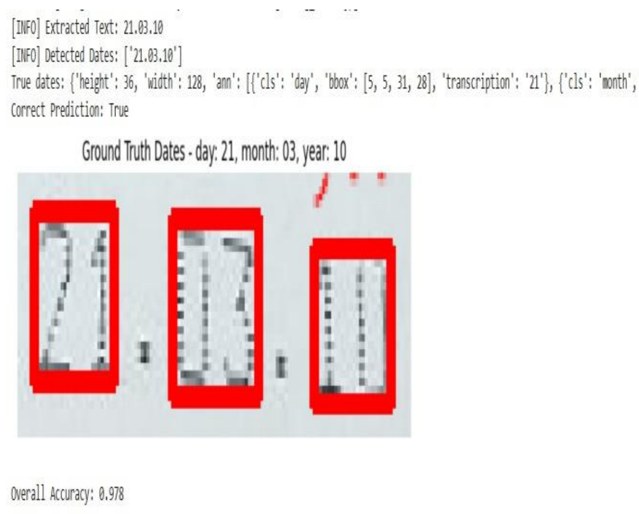

Figure 3. Extracted Text and Date Validation Output

A. Dataset Detail

A custom dataset was created containing food product label images with visible expiry or best-before dates. Each image was manually annotated with correct expiry dates and stored in a structured JSON format for supervised evaluation.

Images of food product labels that prominently display best-before or expiration dates were used to construct a bespoke dataset. To guarantee reliable ground-truth information for supervised assessment, each image was carefully reviewed and marked with the right date. The related annotations were kept in a structured annotations.json file, and all product label pictures were kept in the /Date-Real/images directory. Each image filename is linked to one or more expiration or best-before dates that are displayed in that picture in the annotation file, which has the format −filename.jpg":“DD/MM/YYYY", ...]. Effective picture retrieval, annotation matching, model training, and performance evaluation are made possible by this well-organised dataset structure. The structured data related to picture img_00001 is represented by the sample annotation.JPG. The picture is 139 pixels wide and 35 pixels high. A list of annotations for the various elements of the date that are visible in the picture is contained in the ann field. A class (cls), a bounding box (bbox), and the identified text (transcription) are all included in each annotation. In this instance, the day is marked as 23, the month as 05, and the year as 2021. Each date component's placement inside the image is specified by the bounding box coordinates, which enable the model to determine the year, month, and day's position as well as their textual value. The dataset is appropriate for training and assessing date detection and recognition algorithms as these annotations together reflect the entire date as 23/05/2021.

Evolution Metrics

To evaluate the performance of the deep learning models used for the task of expiry date recognition from food labels, the following metrics are considered essential:

Accuracy

Accuracy measures how well the model predicts expiry dates by determining the proportion of correct forecasts among all forecasts. It is calculated by dividing the total number of forecasts by the number of estimates that were correct. A high accuracy grade shows that the system can correctly determine food label expiry dates and helps reduce overall prediction mistakes.

Accuracy = Correct Predictions / Total Predictions

Precision

Precision quantifies the proportion of expiration dates that the model accurately predicts. It demonstrates the model's capacity to prevent false detections, which occur when an expiration date is mistakenly identified. Because it helps preserve accurate and reliable data by lowering false positive errors, high accuracy is particularly crucial for automated food monitoring systems.

Precision = TP / (TP + FP)

Recall (Sensitivity)

Recall quantifies the proportion of real expiration dates seen on food labels that the algorithm is able to identify. It demonstrates the model's capacity to locate all pertinent expiration dates without omitting any. Because it guarantees that the majority of expiration dates are accurately detected and not missed, high recall is crucial for applications like inventory management and food safety monitoring.

Recall = TP / (TP + FN)

F1-Score

By integrating accuracy and recall through their harmonic mean, the F1-score offers a fair assessment of a model. Because it shows how well the model handles both false positives and false negatives, it is particularly helpful when working with unbalanced datasets, such as when expiration date samples are significantly less than other text on product labels. Furthermore, by displaying true positives (correctly identified expiry dates), false positives (non-expiry text incorrectly labeled as expiry dates), false negatives (missed expiry dates), and true negatives (non-expiry text correctly ignored), the confusion matrix provides a comprehensive visual and numerical summary of prediction results. This thorough analysis aids in comprehending the model's advantages and disadvantages and encourages more adjustments for better performance.

F1-Score = 2 * (Precision * Recall) / (Precision + Recall)

Model Performance = CNN + BERT ----------- (1)

When it comes to identifying text sections, characters, and layout patterns that visually approximate expiration dates, the CNN-based model excels at extracting spatial data from photos of food labels. It provides quick inference with little computing requirements, is especially good at recognizing properly printed expiration dates in structured forms, and works well in controlled settings with clear, high-resolution labels. However, it struggles with different date formats and inconsistent nomenclature, performs badly when dealing with ambiguous or semantically complicated labels, and is unable to comprehend the meaning of surrounding text (for example, mistaking "MFD" with "EXP").

$$\theta^* = \arg\min_{\theta} \left(\lambda_d L_{\text{day}} + \lambda_m L_{\text{month}} + \lambda_y L_{\text{year}}\right) \text{-------------(2)}$$

symbolises the model's training goal. Its objective is to determine the optimal model parameters that reduce the overall error in forecasting an expiration date's day, month, and year.

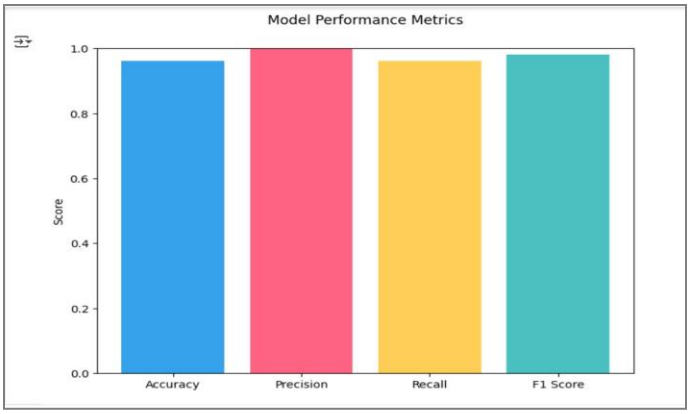

Figure 4. Model Performance Metrics

The model's performance metrics were 96.2% accuracy, 100% precision, 96.2% recall, and 98.1% F1-score. The performance bar graph illustrates the evaluation metrics achieved by the proposed deep learning model. The model attained an overall accuracy of 96.2%, demonstrating its reliability in real-world expiry detection tasks. Precision is at 100%, indicating that every expiry detected was indeed correct, with no false alarms. The recall, slightly lower due to two false negatives, still remains high, reinforcing the model's capability to capture relevant expiry information. The F1-Score, which balances precision and recall, also reflects strong performance. These results confirm that the model is both precise and efficient, making it suitable for applications in retail, warehouse automation, and mobile expiry tracking.

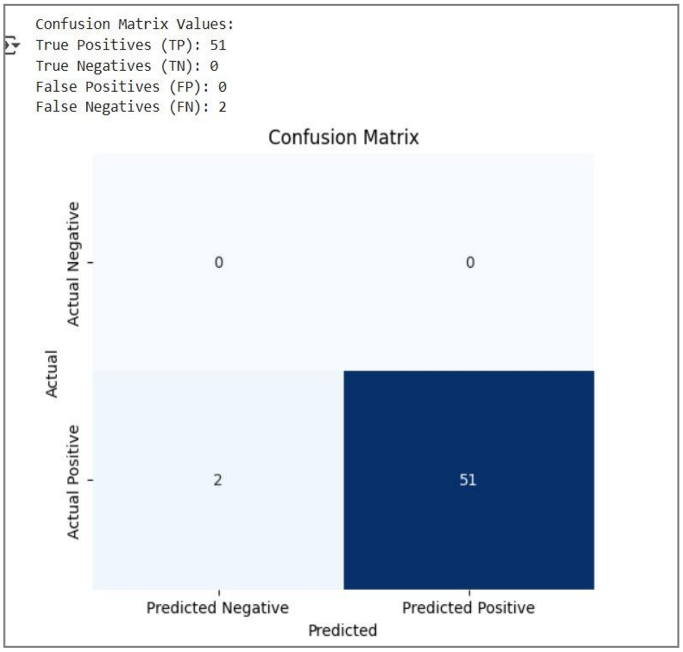

Figure 5. Confusion Matrix

The confusion matrix provides a clear summary of the model's classification performance on the test dataset. It shows that the model correctly identified 51 instances as true positives (TP), indicating successful expiry date detection. However, there were 2 false negatives (FN), representing cases where the expiry date existed but was not detected. There are no true negatives (TN) or false positives (FP), as all test samples belonged to the expiry class. This result highlights the model's high sensitivity and accuracy in identifying expiry dates. The low false negative count indicates the model's robustness, with minimal chances of missing important expiry information.

8. Future Scope

The system's future scope will focus on enhancing robustness, scalability, and practical utilization using a range of innovative technical techniques. To enhance label variability management, first deep learning models that can handle a range of label designs, printing methods, noise levels, and imaging situations must be created. By producing intricate or poor-quality label pictures, generative adversarial networks (GANs) can enhance datasets and enhance model generalization. Low-latency processing on mobile and Internet of Things devices is possible even in resource-constrained environments because to lightweight neural architectures and edge computing techniques that optimize real-time performance. The user-centered evaluation method will include extensive usability testing to enhance the interface, interaction flows, and warning systems based on feedback from real Additionally, integrating the system with smart-home ecosystems, such connected refrigerators or pantry management systems, would enable automated inventory tracking and synchronized expiration-date updates across IoT devices. The method may be extended to cross-domain applications such as multilingual settings, retail stock monitoring, and prescription label detection by altering models for domain transfer and language flexibility. Last but not least, using transfer learning from big pretrained models and creating multimodal learning strategies that include textual, visual, and contextual data may greatly increase recognition accuracy, especially when dealing with sparse annotated datasets.

9. Conclusion

In today's fast-paced world, people often forget to check expiry dates on food items, which leads to unnecessary waste and even health risks. Through this project, "Semantic Analysis of Food Labels for Expiry Date Recognition Using Deep Learning", we aimed to solve this problem in a smart and efficient way. By combining OCR with deep learning models like BERT and CNN, we built a system that can read and understand expiry dates printed on food packages, even if they are in different formats or slightly faded. The mobile app we developed is user-friendly and sends timely reminders so that users can consume their products before they go bad. Working on this project helped us understand real-life challenges like poor lighting, unclear prints, and the need for accurate date detection. Despite these hurdles, our system shows promising results and can be further improved in the future with more data, better UI, and features like barcode scanning or voice alerts. Overall, this project doesn't just solve a technical problem it supports a bigger cause by helping reduce food waste, encouraging safe consumption, and promoting responsible habits. It's a small step toward smarter living and sustainability.

REFERENCES

1. Chaitanya, A. S., Anand, S., Vishnu, V., Srujan, M., & Rajesh, G. (2024). Reminding Food Expiry Dates and Sending Timely Alerts. International Journal of Research Publication and Reviews, 5(3), 1-5. Retrieved from https://ijrpr.com/uploads/V5ISSUE3/IJRPR23414.pdf
2. Saw, A. K., Kaushik, S., Sharma, A., & Nayak, B. K. (2024, May). A Novel Approach for Seamless Reminder Apps for Expiry date with QR. In 2024 3rd International Conference on Artificial Intelligence for Internet of Things (AIIoT) (pp. 1-6). IEEE.
3. SP, S. I., Karthikeyan, L., Bhoomika, P., Tejaswi, K., Khande, L. S., & Vemishetti, N. S. (2024, December). Automating Nutritional Claim Verification: The Role of OCR and Machine Learning in Enhancing Food Label Transparency. In 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS) (pp. 1164-1171). IEEE.
4. Li, D., Bai, L., Wang, R., & Ying, S. (2024). Research Progress of Machine Learning in Extending and Regulating the Shelf Life of Fruits and Vegetables. Foods, 13(19), 3025.

5. Abdulraheem, A., & Jung, I. Y. (2023). Effective Digital Technology Enabling Automatic Recognition of Special-Type Marking of Expiry Dates. Sustainability, 15(17), 12915.
6. Seker, A. C., & Ahn, S. C. (2022). A generalized framework for recognition of expiration dates on product packages using fully convolutional networks. Expert Systems with Applications, 203, 117310.
7. Zheng, J., Li, J., Ding, Z., Kong, L., & Chen, Q. (2023). Recognition of expiry data on food packages based on improved DB Net. Connection Science, 35(1), 1-16.
8. Gong, L., Thota, M., Yu, M., Duan, W., Swainson, M., Ye, X., & Kollias, S. (2021). A novel unified deep neural networks methodology for use by date recognition in retail food package image. Signal, Image and Video Processing, 15, 449-457.
9. Koponen, J., Haataja, K., & Toivanen, P. (2022). Recent advancements in machine vision methods for product code recognition: A systematic review. F1000Research, 11.
10. Manlises, C. O., Padilla, D. A., Santos, J. B., & Adviento, P. A. (2024, March). Expiration Identification of Canned Goods using Convolutional Neural Network. In 2024 7th International Conference on Information and Computer Technologies (ICICT) (pp. 173-177). IEEE.
11. K. Saw, S. Kaushik, A. Sharma and B. K. Nayak, "A Novel Approach for Seamless Reminder Apps for Expiry date with QR.," 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT), Vellore, India, 2024, pp. 1-6, doi: 10.1109/AIIoT58432.2024.10574799.
12. K. M. Shanthini, P. Chitra, S. Abirami, G. Aninthitha and P. Abarna, "Recommendation of Product Value by Extracting Expiry Date using Deep Neural Network," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 1-7, doi: 10.1109/ICCCNT51525.2021.9579675.
13. C. O. Manlises, D. A. Padilla, J. B. Santos and P. A. Adviento, "Expiration Identification of Canned Goods using Convolutional Neural Network," 2024 7th International Conference on Information and Computer Technologies (ICICT), Honolulu, HI, USA, 2024, pp. 173-177, doi: 10.1109/ICICT62343.2024.00033.
14. P. Rimdusit, P. Aphiromyarnont, S. Phimoltares and S. Panthuwadeethorn, "Extracting Information from Drug Label Using Image and Text Processing," 2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE), Phuket, Thailand, 2024, pp. 377-382, doi: 10.1109/JCSSE61278.2024.10613666.
15. T. Khan, "Expiry Date Digits Recognition using Deep Learning," 2019 IEEE National Aerospace and Electronics (NAECON), Dayton, OH, USA, 2019, pp. 302-304, doi: 10.1109/NAECON46414.2019.9058255.
16. G. Phiri and P. Trevorrow, "Sustainable Household Food Management Using Smart Technology," 2019 10th International Conference on Dependable Systems, Services and Technologies (DESSERT), Leeds, UK, 2019, pp. 112-119, doi: 10.1109/DESSERT.2019.8770023.
17. U.S. Department of Agriculture, Food Composition Database National Agricultural Library, Human Nutrition and Food Safety. Available at: https://www.nal.usda.gov/humannutrition-and-food-safety/food-composition
18. Zhao, Y., & He, Y. (2023). "Automatic Detection of Expiry Dates in Retail Products Using YOLOv5 and Tesseract OCR." Journal of Retail Technology, 12(4), 220–228.
19. Liu, X., & Wang, M. (2024). "Improving OCR Accuracy on Curved and Reflective Food Packaging." Pattern Recognition Letters, 170, 104–113.
20. Kumar, R., & Singh, P. (2022). "Smart Grocery Management Using Expiry Date Alerts." International Journal of Smart Technologies, 8(3), 93–99.
21. Nair, M., & Sharma, L. (2024). "Enhancing Expiry Label Detection using GAN-based Data Augmentation." Applied Soft Computing, 136, 110215.
22. Al-Habsi, A., & Khandekar, A. (2023). "Multilingual OCR for Global Food Label Recognition." IEEE Transactions on Consumer Electronics, 69(2), 303–310.
23. Saleh, M., & Gupta, T. (2023). "Efficient Expiry Date Tracking in Supply Chains with Blockchain Integration." Future Generation Computer Systems, 144, 657–665.
24. Wang, Y., & Ren, J. (2022). "Deep Learning Based Expiry Date Reader with BERT-CRF Fusion Model." Neural Computing and Applications, 34, 15571–15582.
25. Prasad, K., & Balaji, P. (2021). "Eco-aware Smart Fridge System with Expiry Notification and Inventory Logging." Internet of Things Journal, 9(3), 2782–2790.
26. Ahmed, M., & Qureshi, S. A. (2023). "Expiry Date Localization in Low-Resolution Food Labels Using Transformer-based OCR." International Journal of Computer Vision & Signal Processing, 13(1), 65–74.
27. Roy, A., & Jaiswal, P. (2022). "Tackling Food Wastage through Smart Expiry Monitoring: A Deep Learning Perspective." Journal of AI Research and Applications, 5(2), 101–110.
28. Hossain, M. Z., & Chowdhury, S. M. (2023). "A Lightweight OCR System for Embedded Devices in Food Inventory Management." ACM Transactions on Embedded Computing Systems, 22(4), Article 55.

29. Desai, V., & Rajput, D. (2024). "Mobile-Based Expiry Date Detection using Hybrid CNN-Transformer Architecture." Procedia Computer Science, 232, 299–306.
30. Sultana, N., & Alim, M. A. (2022). "Optimizing OCR Output for Skewed and Blurred Expiry Date Texts." Pattern Recognition and Image Analysis, 32(3), 215–224.
31. Chen, R., & Tao, J. (2023). "Integrating Expiry Label Reading into Retail Inventory Systems Using Computer Vision." Journal of Retail Automation, 8(1), 49–59.
32. Yamamoto, H., & Suzuki, M. (2021). "Real-Time Food Monitoring System with Expiry Alerts Using Raspberry Pi and OCR." International Journal of Smart Home Systems, 15(2), 77–83.
33. Banerjee, S., & Pathak, R. (2022). "Use of Augmented Reality to Visualize Expiry Dates on Product Packaging." International Journal of Interactive Systems, 11(4), 155–162.
34. Author Biographies
35. Roshni Bhave is a dedicated academician and researcher currently serving as an Assistant Professor in the Computer Science and Engineering, School of Computing MIT ADT University Puine. Her expertise in Machine learning, Cloud Computing, Block Chain, Generative AI. She is author of many research articles several Scopus-indexed journals and conferences. she is published patents and book chapters.
36. Vibhakti. N. Bagade received the B.E. and M.Tech. degrees in Computer Science and Engineering from RTMNU University, Nagpur, India. She is Qualified NET Examination. She is currently an Assistant Professor in the Department of Computer Science and Engineering at Yeshwantrao Chavan College of Engineering (YCCE), India. Her research interests include network security, Internet of Things, and Natural Language Processing.
37. Dr. Vijaya Kamble received the B.Eng.& MTech degree in Computer Science & Engineering from RGCERT, Chandrapur, in 2012. Ph.D. degrees in CSE from MGU Bhopal. Currently, she is an Assistant Professor at the Department of Computer Science & Engineering at GHRCEM, Nagpur. Her research interests include Machine Learning, Data Science, Deep Learning. Angela of Higher Education. She is life member of IEI. She has authored or coauthored more than 50 refereed journal and conference papers, 4 book chapters, 3 patent..
38. Dr. Amita Meshram received the B.E Degree in Computer Technology and Mtech in WCC from RTMNU University, Nagpur. PhD. degree in Computer Science and Engg.,JECRC University, Jaipur,India in 2024.She is an Asst.Prof.at the Department of CSE ,GHRSTU, India. Her research interests include the application of Machine Learning and Deep Learning.
39. Purva Gogte (Vyawahare) is an academician, researcher, and Assistant Professor in the Department of Computer Science and Engineering at Ramdeobaba University. She earned her B.E in Datta Meghe College of Engineering and M.Tech. degrees from G.H. Raisoni College of Engineering and is currently pursuing a Ph.D. at Indian Institute of Information Technology Nagpur. Her research focuses on Artificial Intelligence, AI-driven optimization, resource management, and service orchestration. She has extensive teaching experience in education since 2013.