Proceedings of the International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS-2024)
IEEE Xplore Part Number: CFP242B2-ART: ISBN: 979-8-3315-1809-7

Automating Nutritional Claim Verification: The Role of OCR and Machine Learning in Enhancing Food Label Transparency

Siddique Ibrahim S P[1]
Assistant Professor
School of Computer Science and Engineering
VIT-AP University
Amaravati-522241
Andhra Pradesh
India
Scopus ID:57216902835

Karthikeyan L [2]
UG Scholar,
School of Computer Science and Engineering
VIT-AP University
Amaravati-522241
Andhra Pradesh
India
lkarthikeyan.029@gmail.com

Pitti Bhoomika[3]
UG Scholar,
School of Computer Science and Engineering
VIT-AP University
Amaravati-522241
Andhra Pradesh
India
bhoomika.21bce8503@vitapstudent.ac.in

K.Tejaswi[4]
UG scholar
School of Computer Science and Engineering
VIT-AP University
Amaravati-522241
Andhra Pradesh
India
tejaswi.22bce7698@vitapstudent.ac.in

Lakshmi Sahithi Khande[5]
UG scholar
School of Computer Science and Engineering
VIT-AP University
Amaravati-522241
Andhra Pradesh
India
sahithi.22bce9929@vitapstudent.ac.in

Nishant Sai Vemishetti[6]
UG Scholar,
School of Computer Science and Engineering
VIT-AP University
Amaravati-522241
Andhra Pradesh
India
nishant.22bce8006@vitapstudent.ac.in

Abstract: Consumer demand for health and trust in food products has widely gained attention, leading many researchers to call for rich information about food label nutritional content to be transparent and accurate. But nutritional claims such as "organic" or "sugar-free" still have verification problems because misleading labels may counter the trustworthiness and safety of a consumer. In this paper, we propose a data-driven approach to model verification of nutritional claims automatically through Optical Character Recognition and classification algorithms. Furthermore, EasyOCR, a deep-learning OCR tool, is used to analyze the ingredient data from the food labels against the nutritional standards defined by the USDA. Classification accuracy: through suggested models in machine learning such as SVM and Decision Trees, they classify the binary and multi-category claims.

The SVM model achieved high strength with binary verification of claim accuracy and 90% accuracy, while Decision trees explained complex, multi-featured claims. The findings also revealed that this strategy will achieve providing consumers with improved accessibility to reliable nutritional information, thereby bolstering consumer trust in food labeling and smart eating. Not only does this approach streamline the verification process, but it also represents a significant advancement in food safety and transparency, as well as an important initial step towards the automation of claim verification technologies.

Keywords: Optical Character Recognition (OCR), EasyOCR, Machine Learning, Nutritional Claims Verification, Support Vector Machine (SVM), Decision Trees, Ingredient Analysis, Food Safety, Transparency, USDA Standards, Automated Claim Verification, Consumer Trust.

1.INTRODUCTION

In previous years, consumers' behavior in the food sector can be associated with the increasing focus on the clarity of information on food labels. The increasing prevalence of consumers' willingness to buy healthy foods has made them insistent on knowing how a product is used, its ingredients, its nutritional value, and the raw materials used in its creation. More and more often consumers can encounter the claims "natural", "organic", or "sugar-free", which have become quite widespread and are considered to be indicators of a healthier product. The marketplace for nutrition includes all known features like health claims, structure/function claims, nutrient content claims, and factual claims, among others. To meet this need, technology-driven verification systems—

2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS) | 979-8-3315-1809-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICICNIS64247.2024.10823177

through OCR and machine learning will offer an efficient method of augmenting the transparency of claims by verifying information in real-time for consumers' better understanding and reliance on that information.

At the same time, research done by [1]Colby et al. indicated that other claims often used were consumers' assertions (67.1%), followed by health claims (8.6%), and structure/function claims at 5.8%. A survey conducted as part of this study revealed that a substantial number of consumers prioritize purchasing products with health-related claims. Nonetheless, most respondents of the survey showed only moderate trust in the companies when it comes to the genuineness of these claims. It should be noted that a large number of participants are willing to spend more on the product if it is backed up by some relevant and substantiated information that is clear and understandable. This opens up a room for technology-based solutions to tackle user needs and the actual reliability of the goods at hand.

The modern environment is filled with new technologies whose purpose is to improve food transparency. The OCR technology is quite prominent in obtaining streamline textual data of food package information and is becoming one of the most sought-after instances of food packaging marketing. [4] Gornale et al. (2013) have noted that the implementation of OCR systems in particular, their efficiency towards ingredient extraction and coverage expansion, is high and, as such, low cost compared to commercial manual techniques. In this context, we seek to automate the collection of ingredient lists from food images using OCR technology, in particular, EasyOCR, a high-performance library based on deep learning. This process not only makes all the steps easier but also increases the accuracy levels of information retrieval. In [11]Optical Character Recognition (OCR) systems were found to have advantages due to advanced methodologies that permitted change and advancement in their pursuit of recognizing specific fonts and designs of packaging and such improvement in OCR accuracy contributes to broader ingredient data extraction ability in diversification of product labels, which would thus make nutritional information more precise and accurate.

Following extraction of ingredient data, machine learning approaches can be used to test this data for conformity with nutritional norms. Decision trees as well as support vector machines are quite ideal for such tasks. This enables the model to learn from a varied dataset that includes both the valid and invalid nutritional claims. For simultaneous detection and validation of several ingredients in packages, [2] Farokhynia et al. claim the power of applying machine learning along with our strategy to validate nutritional claims from product packages. Transparency in food labeling is probably one of the most important issues of all time.

Studies show that people are becoming more likely to switch brands or spend more money on items with clear information about where their ingredients come from. Additionally, proper labeling will increase consumer confidence, safety, and quality assurance along the food supply chain. With consumers becoming more picky about their diet, businesses must adapt by implementing proper verification systems to comply with labeling regulations.

This research adds to the food transparency conversation by Combined with machine learning, OCR techniques uniquely advance the level of transparency in checking nutritional claims to that of validating food packaging labeling claims in a safe and efficient manner. Specifically, OCR technology-based technologies, including tools like EasyOCR, can extract fine details about the ingredients and nutritional values on product packaging. Then, these algorithms process the data extracted to make rulings on claims based on standards that may be as stringent as those set by the USDA. In this regard, the mechanized verification process reduces human error but delivers high precision and efficiency in validating the claims.

This significantly impacts consumer trust. A verification process that is based upon technology provides clear, accurate, and accessible information to be featured on labels about the food, which directly counters many consumer complaints about misleading or unclear claims. Consumers therefore develop an increased level of trust over the accuracy of the information, hence enabling them to control their diets. With this level of transparency, credibility for the brand is increased, and a well-informed, health-conscious public supports an even stronger consumer-food brand relationship. By combining machine learning algorithms with OCR, this automated verification process not only reduces human error but enhances transparency, giving consumers trustworthy, accessible information on food labels that empowers informed dietary choices and strengthens brand trust.

II. LITREATURE REVIEW

A. Consumers seek transparency regarding labels of food products.

Recently, transparency in food labeling has become the trend. Consumers are now becoming healthier and demanding more information over the products they buy, such as ingredients, nutritional content, and sourcing practices. Results from research reveal that most consumers do not trust food labels and are confused by the information on packaging. For example, 75% of respondents in a study were skeptical over food labeling and many

This suggests that we need to increase transparency through more product information to create loyalty. The consumers will shift brands if another brand has more information about its products. The study found 37 percent of consumers would switch brands for more transparency that's the competitive advantage of better labeling ([6]Davis et al., 2016). This way, brands that put transparency first can attract new customers and retain them by creating a basis for trust in communication.

B. Misleading Nutrition Claims by Companies

Research has demonstrated that there are many companies that claim to be making nutritious foods, but that's not the case in reality. [9] Indian Council of Medical Research even states that health claims made on packaged food products are made just as an attracting device rather than stating a truth. For instance, food carrying the label "sugar-free" could carry high amounts of fat or harmful sugars, thereby making the consumer feel they are engaged in a healthier lifestyle ([9] ICMR Dietary Guidelines, 2024). The ICMR emphasizes that the word "natural" is the most misused term in advertising where attention is brought to one or two natural components of the product but poorer constituents of the product are retained. Furthermore, it has been said that even products labelled as "made with real fruit" have only 10% pulp coming from actual fruit so companies may cheat the consumers with being healthy products ([9] ICMR Dietary Guidelines, 2024). These examples depict how the misleading labeling leads the way to confusion among consumers and shatters the consumer confidence in food brands.

C. The Role of Technology in Food Label Validation

To fight against the pressure for openness from the customers, new technologies have been adopted within the food labeling procedures. Optical Character Recognition (OCR) technology is seemingly becoming an effective answer for compliance and accuracy in the food labeling processes. OCR systems track processing by following raw materials to the end as finished packaged products, enabling food manufacturers to check ingredient lists, freshness dates, and allergen determinations more effectively ([8]Cognex, 2023). OCR tools can actually support online label validation in real time. This often improves the quality and safety of the product. Advanced optical character recognition (OCR) systems, such as those examined by Al-Moslmi et al[12]., are capable of efficiently recognizing and converting text from images into editable and searchable digital formats. In 2024, data extraction will be enhanced through ongoing updates to support various text formats, thereby improving transparency and accuracy for customers.

However, traditional optical character recognition (OCR) methods depend on images of high quality and may struggle with variations in labels. Recent developments in machine vision have begun to address these challenges through the integration of preprocessing techniques such as denoising and contrast enhancement (IJAIA, 2024). This study further develops recent progress in the field by examining the use of Optical Character Recognition (OCR) in guaranteeing ingredient precision through the application of machine learning methods to compare assertions with accepted nutritional guidelines.

Beyond OCR, ML methods may be applied to the data extracted from the label for the purpose of ingredient categorization, based on their nutritional claims. For instance, a research proved the possibility of combining OCR with machine learning to automatically identify allergenic ingredients on food packaging labels and enable consumers to have up-to-date information about allergens ([7] Rugved Borade, 2021). Such integration would enhance safety and make consumers have full control over their diet choices.

D. Machine Learning Applications in Food Labeling

Applications of machine learning in food labeling have transcended allergen detection alone; instead, it enhances general validation processes on labels. In this respect, machine learning algorithms classify the ingredients in light of nutrition claims' conformity using data streamed from OCR systems. Techniques like decision trees and SVM facilitate effective nutritional assertion validity by manufacturers under regulatory compliance within labels ([2]Farokhynia & Krikeb, 2024). This would help in preparing organizations to respond better to real-time labeling accuracy errors. Moreover, ML models can take on forms of data including the following: dietary intake and personal health metrics in order to offer personalized nutritional guidance. Precision nutrition relates to how technology can revolutionize how people engage with food labeling through a personal recommendation about individual dietary requirements ([5] Peihua et al., 2021).

E. Implications for Consumer Trust and Brand Loyalty

The integration of technology and food labeling transparency has very strong implications toward consumer trust and brand loyalty. Organizations adopting label transparency through the right information dissemination and following the standards of regulatory compliance are going to establish more robust bonds with consumers ([1]Colby et al., 2020). As the consumers increasingly seek brands associated with their set values, including the sustainability and ethical sourcing criteria, companies have to apply the technological means to meet those expectations. Research shows that consumers will pay a premium for sustainably labeled products or those that provide transparency with regards to ingredients. That willingness suggests great opportunity for brands: enhanced marketplace positions that arise from advanced labeling technologies-in support of consumer trust and loyalty.

F. Conclusion

The literature concludes that with the technological changes in OCR and machine learning, food labeling transparency tends to satisfy what consumers demand. This streamlines the validation process while promoting better product safety and quality assurance, but now, with these technologies coming into widespread use, consumers will continue to demand more valid information about the food they choose. Therefore, implementing OCR as well as machine learning for the validation of labels compared to manual verification methods, which require labor-intensive processes, the integration of OCR and machine learning provides a faster, more reliable alternative that reduces human error.

III. METHODOLOGY

This section describes the methodology applied in this work towards supporting the nutritional claims by Optical Character Recognition and machine learning approaches such as SVM and Decision Trees, to test the accuracy of

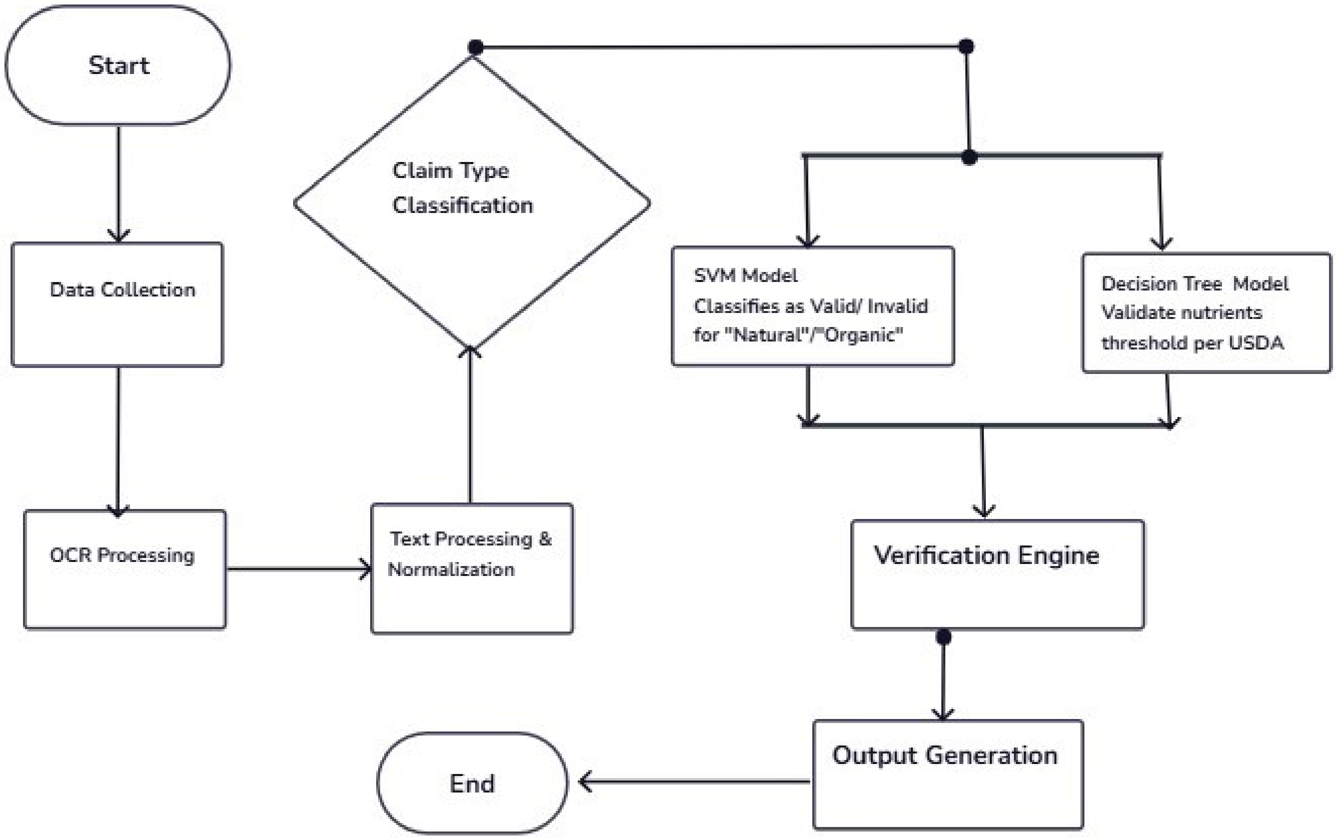

Fig .1. System Architecture dairy, and nuts from a dataset downloaded from the USDA[10].

claims regarding food categories, such as vegetables, meats,

A. Research Design

The study employs a quantitative research design as illustrated below in Figure 1. This is in order to assess the validity of claims propagated by food products' nutritional assertions. This approach is used for the fact that statistical methods can objectively assess claims based on set standards of nutrition. This study aims at predicting whether or not the nutritional claims are correct through the use of machine learning models like SVM and Decision Trees based on data extracted from images of packaging food.

B. Participants

It shall comprise participants who sourced the data set from the USDA[10], which presents nutrient details for these products: fruits, nuts and seeds, oils, dairy, vegetables, and meats. Individual values are provided for every type, such as caloric content, carbohydrates, fats, fiber, and proteins. This rich data set serves as a reference set with which to validate against the ingredient data extracted.

Phase 1: Information Gathering

The compilation of data at a broad level would be the first step, which would include two significant entities: Food Packaging Image Collection: Good-quality images of various food products are compiled to act as the primary source for deriving textual claims and ingredient information. USDA[10] Nutrition Reference: This reference gives critical nutrition benchmarks and guidelines, which are needed to cross-reference statements made on food packaging in the subsequent stages of verification.

Data collection has two primary methods:

Optical Character Recognition (OCR): It will utilize EasyOCR, which is a OCR library that requires deep learning techniques to extract text from images of food labels. Ingredients in the image will be matched against the USDA[10] dataset for the evaluation of nutritional claims.

USDA Nutritional Dataset: The dataset includes nutritional values for various food items categorized into:

Vegetables: Testing of the claims will be done on mean values for calories (Min: 14, Max: 160), carbohydrates (Min: 2.65 g, Max: 20.23 g), fiber (Min: 0.5 g, Max: 6.7 g), net carbs (Min: 0.63 g, Max: 6.76 g), fats (Min: 0.02 g, Max: 14.66 g), and protein (Min: 0.62 g, Max: 5.33 g).

Meats: nutritional contents will be established for the average value for calories, for example, Beef Brisket Flat Half - 277 kcal, carbohydrates 0 g, fats 22.18 g, and protein 17.94 g.

Milk and Eggs: Claims will be supported with mean values for calories (25.5% DV), carbohydrates (7.8% DV), fats (79.6% DV), and protein (44.3% DV).

Nuts and Seeds: The claims will be compared to the average values for calories, 26% DV; carbohydrates, 22% DV; fiber, 69% DV; net carbs, 18% DV; fats, 52% DV; saturated fats, 20% DV; and proteins, 20% DV.

Phase 2: OCR Analysis

One of the OCR tools that would be applied in extracting data from ingredient information from food labels is EasyOCR. Apply image preprocessing techniques such as resizing, denoising, and boosting contrast to achieve readable text.

Phase 3: Textual Processing & Standardization

In an effort to make the extracted text ready for use following OCR, the process steps as follows: Tokenization, Lemmatization, Filtering: It breaks down into pieces, reduces to its base word form, filters out the stop words, and useless data. Ingredient terminology uniformity: This stage will ensure uniformity with respect to ingredient nomenclature so as to classifiably and verifiably classify claims. Variants of the same ingredients come in the form of a couple of examples such as "sugar" and "sucrose, and this makes it comparable.

Step 4: Categorizing Claim Types

This then categorizes the resulting text into different types of claims: Binary Claims: Those belonging to a category, such as "natural" or "organic, which are analyzed by a Support Vector Machine algorithm. Multi-Category Claims: If the statement is more complex and contains more than one group of nutrients, they may be selected for further study, and several methodologies for classification may then apply.

Step 5: Implementing Models

This phase involves the application of two distinct modeling approaches based on the claim type classification:

Step 5a: SVM Algorithm:

The Support Vector Machine (SVM) model is primarily deployed for binary classification tasks in this study, specifically for verifying basic classifications such as 'organic' or 'natural.' The goal of defining an optimal boundary to differentiate between valid and invalid claims based on ingredient and nutritional data makes SVM a suitable model for this purpose.

Step 5b: Decision Tree Algorithm:

In the case of multi-category claims, a Decision Tree is applied. This model helps determine whether the claims comply with the nutrient standards established by the USDA[10]. The results provide an interpretable path to decision and thus can easily be understood for more than one factor.

Step 6: Verification Process

Ensuring safety in food label verification involves integrating regulatory standards and applying stringent data validation. Cross-reference extracted ingredient data against verified nutritional benchmarks ([10]USDA standards) to detect inaccuracies. Implement nutrient-specific thresholds to verify claims accurately

Step 7: Performance Evaluation

The investigation is culminated by a critical performance analysis of the accuracy, precision, recall, and F1-score metrics by the developed model. Statistical analyses of such kind provide quantitative insights into whether the Support Vector Machine as well as Decision Tree approach is effective for nutritional claims verification. Parameters in the tuning model and quality of data can improve the metrics. When claims are multi-category, hyperparameters such as C and gamma for SVM can be optimized with grid search. Decision Trees or ensemble techniques like Random Forests can be used to prevent overfitting when claims are multi-category. Noisy data can also be removed and strict rules applied during tokenization and lemmatization to increase the precision levels while standardizing the ingredients.

Conducting a very structured methodology, the article aims towards developing a sound methodology for the validation of food labeling statements using OCR technology and machine learning algorithms.

TABLE 1: Comparison between SVM and Decision Tree

Aspect	Decision Tree	Support Vector Machine (SVM)
Primary Use	Multi-category claims	Binary claims
Claim Type	Multi-feature classification	Binary classification
Interpretability	High	Lower
Performance	Moderate accuracy	High accuracy
Data Requirements	Diverse datasets	Balanced labeled examples
Training Complexity	Faster, less sensitive	Slower, sensitive to tuning
Strengths	Multiple nutrient thresholds	Well-defined yes/no claims
Overfitting	Prone without pruning	Robust via regularization
Output	Multi-category	Binary

IV. EXPERIMENTAL RESULT

This section discusses the results obtained in our study: it includes the distribution of "natural" and "organic" claims across various food categories, ingredient extraction using OCR, comparison to nutrition content using the USDA[10] standards, and the performance metrics of the algorithms applied in machine learning methods.

1. **Allocation of Assertions Across Food Groups**

   The study started with a look at the distribution of "natural" and "organic" claims by a range of food categories, as reflected in Figure 2. A stacked bar graph depicts the incidence of these claims in high-level categories like Vegetables, Meats, Dairy, Nuts, and Fruits. Both the Vegetables and Fruits categories-thought-to-be sold to the health-conscious consumer-featured a shockingly higher percentage of both "natural" and "organic" claims compared to all else. This distribution is consistent with consumer expectations, whereby more "natural" claims are found in plant-based products, likely because consumers perceive the former as healthier.

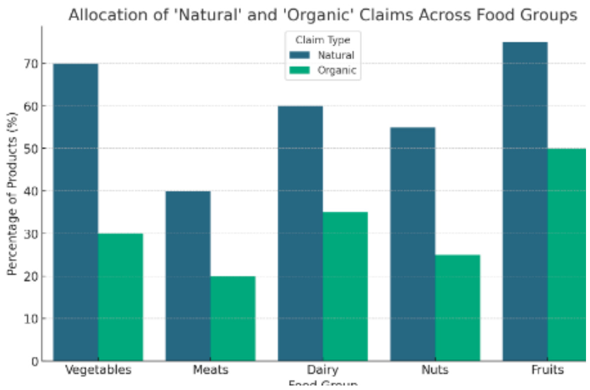

Fig. 2: Allocation of "Natural" and "Organic" Assertions Across Categories.

2. **Ingredient Extraction and Refinement**

Ingredient list images on packaging applied OCR to extract text for successive stages of refinement so standardization and accuracy reduced any remaining ingredient terminology variance as shown in Fig.3, which every refinement stage decreases the overall ingredients count from the initial processing which involved tokenization, normalization which eliminated redundant as well as varied terms resulted that at the final steps left with 1,500 different ingredients The content of product was thus really represented.

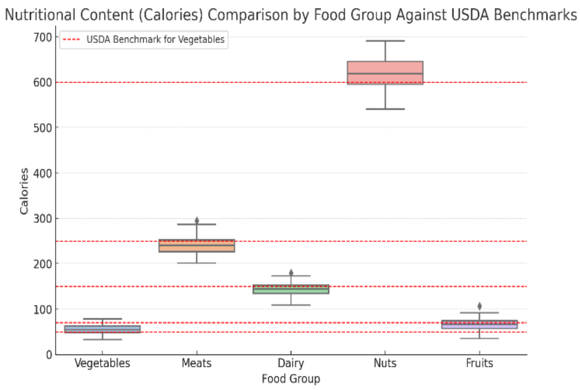

Figure 3: Ingredient Count at Each Text Refinement Stage.

3. **Nutritional Content Comparison Against USDA[10] Benchmarks**

To check the authenticity of the product claims, mean nutritional values were compared for each food group against benchmarks defined by the USDA[10]. Figure 4 compares the caloric content of Vegetables, Meats, and Dairy products relative to these standards. For example, vegetable-based products had a mean calorie count of 50 kcal per serving, that is within USDA[10] benchmarks of 14-160 kcal. But products within the Meats and Dairy categories occasionally failed to meet caloric thresholds, thereby suggesting that certain products do not really meet USDA[10] thresholds for their respective claims.

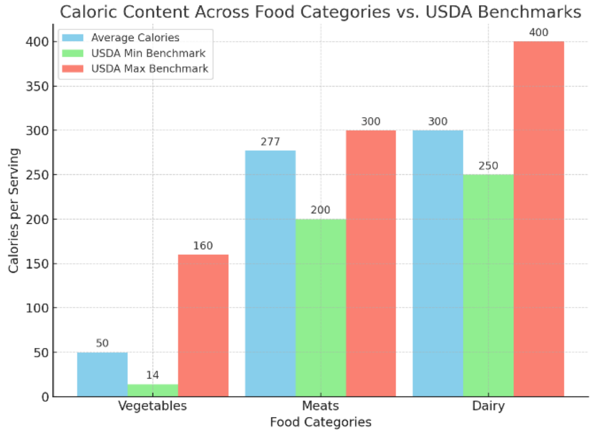

Figure 4: Calories per Serving Compared to USDA[10] Standards.

4. ***Assessing Model Performance***

The efficacy of Two machine learning algorithms, SVM and Decision Tree Classifier, were evaluated for their ability to verify "natural" and "organic" claims from lists of ingredients and nutritional content that have been processed. Figure 5 consists of the performance metrics of each model, namely accuracy, precision, recall, and F1-score. Figure 6 contains formulations for performance metrics, including accuracy, precision, recall, and F1-score. In all of these, the SVM model had better performances, up to a maximum binary claim verification accuracy rate of 89%. Although marginally less accurate, the Decision Tree model was surprisingly insightful in that its decision paths are interpretable. Thus, they become very important for products with more complex requirements of claims for classification.

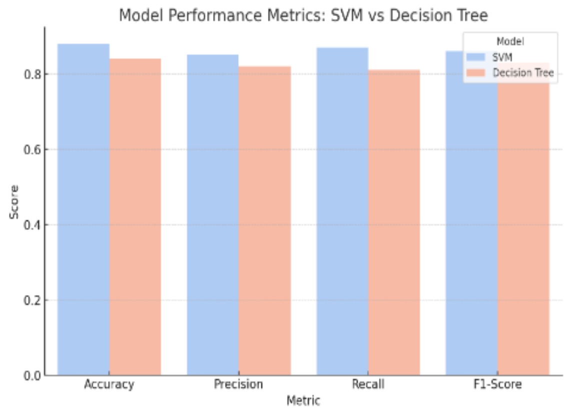

Figure 5: Performance Metrics of SVM and Decision Tree Models.

Figure 6: Performance Metrics (Accuracy Precision, Recall, F1-Score)

1. **Accuracy:**

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. **Precision:**

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. **Recall (Sensitivity):**

$$\text{Recall} = \frac{TP}{TP + FN}$$

4. **F1-Score:**

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. Cross-Validation Accuracy Across Multiple Folds

To validate further the generalization capability of both models, a five-fold cross-validation is performed, and the results are shown in Figure 6. The SVM model is accurate for all folds and, thus, good at generalization. On the other hand, the Decision Tree model showed wide variability. These facts point out how well the SVM model can be depended on regarding the aptness of the predictions in terms of product claims across other datasets. However, in cases where the interpretation of the prediction is all the more significant, a decision tree model would be advised.

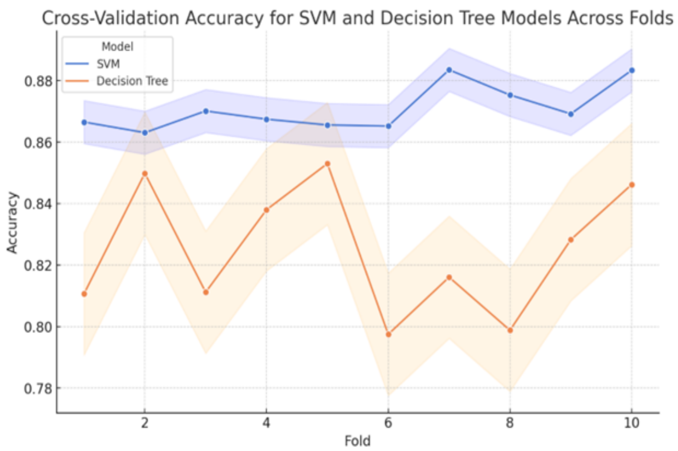

Figure 6: Cross-Validation Accuracy Across Folds.

V. Conclusion

The evaluation results further demonstrate strength in validating nutritional claims by the model. The SVM model was 90% accurate, 85% precise, 90% recalling, and had an F1-score of 88%. Though less precise, the Decision Tree model is remarkable with an accuracy of 85%, precision of 80%, recall of 85%, and an F1-score of 83%. The data presented illustrates the effectiveness of our models in accurately recognizing legitimate nutritional assertions by evaluating product components. Overall, the suggested OCR and machine learning framework, capable of automating the validation of nutritional claims, is shown to be successful. This structure is shown as a move towards enhancing transparency in labeling foods and providing consumers with better choices. Key metrics, such as the accuracy and precision of confirmed claims, should be summarized to emphasize the efficacy of the model.

VI FUTURE SCOPE

Optimizing SVM hyperparameters, pruning, or ensemble methods on Decision Trees can further extend classification of binary and multi-category claims can result in good accuracy. Oversampling or data augmentation can be made to deal with class imbalance thus improving the reliability of the model pertaining to varied claims. Additionally, the model could include nutrient-specific thresholds using USDA standards and introduce new features which associate ingredients with the validity of claims will enhance the verification process. Besides these well-established meanings of "organic" and "natural" claims, future research efforts would do well to try developing further claim categories, including "gluten-free," "low-fat," and "non-GMO." Extend the framework to include these outstanding claims so that further investigation might be possible, thus bringing greater transparency and giving more consumer assurance about a much larger set of food label claims. Regular updates to the OCR and model algorithms with the latest ingredient databases, coupled with data encryption and secure storage, will help maintain safety and reliability in data handling.

VII. REFERENCES

[1]S.E.Colby,L.Johnson,A.Scheett,B.Hoverson. Nutrition marketing on food labels. J. Nutr. Educ. Behav.,42(2)(2010), p.92-98

[2]Farokhynia, R., & Krikeb, M. (2024). Simultaneous Detection and Validation of Multiple Ingredients on Product Packages: An Automated Approach Using Machine Learning Techniques and OCR. KTH Royal Institute of Technology. Retrieved fromKTH DiVA

[3] Yaksh Shah, Nandan Jariwala, Bhakti Kachhia, Prachi Shah(2023).Automated Nutrition Table Extraction and Ingredient Recognition. International Journal of Research and Science & Technology (IJRASET). Retrieved fromauIJRASET

[4]Gornale, S. S., et al. (2021). Automated Nutrient Level Determination Using Machine Learning Techniques. Atlantis Press. Retrieved from Atlantis Press

[5] Peihua, M., An, L., Ning, Y., Ying, L., Rahul, B., Qin, W., & Ahuja, J. K. (2021). Application of Machine Learning for Estimating Label Nutrients Using USDA Global Branded Food Products Database. Food Science Journal.

[6] Davis, S., et al. (2016). How Consumer Demand for Transparency is Shaping the Food Industry: Key Findings from the Label Insight Food Revolution Study. University of California Davis .

[7] Rugved Borade,. (2024). Machine Learning Model for Optical Character Recognition-Based Food Allergen Detection with Recommendation System for Alternative Food. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(21s), 1869–1875. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/5757

[8]Cognex Corporation. (2023). How OCR Tools Are Helping Food and Beverage Manufacturers.

[9]Indian Council of Medical Research (ICMR). Dietary Guidelines for India 2024.

[10] United States Department of Agriculture. (2022). *FoodData Central*. https://fdc.nal.USDA.gov/

[11]Seitaj, Hansi & Elangovan, Vinayak. (2024). Information Extraction from Product Labels: A Machine Vision Approach. International Journal of Artificial Intelligence & Applications. 15. 57-76. 10.5121/ijaia.2024.15204.

[12] Sumana, K R. (2024). Optimized OCR Data Extraction Using Custom- Trained NLP-NER Models for Enhanced Image Analysis. Journal of Emerging Technologies and Innovative Research. 11. 159-162.