

Unveiling Packaged Foods to Empower Consumer Food Choices:

A Comprehensive Review of Packaged Food Analysis

Shraddha Shingne¹, Yajur Sharma², Pranav Nair³,
Yadnika Kotwal⁴, Prasad Tak⁵

Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology Pimpri, Pune, India¹

Abstract: In today's food market, consumers are often misled by deceptive marketing practices regarding the nutritional content of packaged food products. The text size on packaged food products is often small, making it difficult for consumers to read and understand the ingredients and nutritional information. Additionally, many ingredients are listed using scientific or medical terms, which can be unfamiliar and confusing to the average consumer. Furthermore, consumers are often drawn to front-of-pack labels and promotional slogans with attractive text sizes and colours, which can divert their attention from scrutinizing the ingredients list. This project, titled 'Unveiling packaged foods' addresses these issues and aims to develop a user-centric web application that promotes informed consumer food choices. The application allows users to scan the barcode or search for a packaged food product to receive detailed ingredient and nutritional analysis harnessing the power of data visualizations and web technology. Through data visualizations, we provide insights into the nutritional content of products in a quicker and easier way, helping consumers to easily identify healthier options and avoid products that may be detrimental to their health. Moreover, the project employs a metric known as 'Nutri-Score' to assess the nutritional quality of packaged food products. The project demonstrates the potential of technology to empower consumers to make healthier food choices and highlights the importance of transparent and accessible information in the food industry.

Index terms: Nutri-Score, Nutritional Analysis, Ingredient Analysis, Ensemble Learning

I. INTRODUCTION

In recent years, the landscape of consumer food choices has undergone significant transformation, driven by an increasing demand for transparency, health consciousness, and personalized experiences. As consumers become more discerning about the foods they consume, there is a growing need for innovative solutions that empower them to make informed and personalized decisions about packaged foods. In response to this demand, our research explores the development of a web application titled "Unveiling Packaged Foods" to revolutionize the way consumers interact with and understand packaged food products.

Challenges Encountered by Consumers: One of the key issues faced by consumers is the use of complex terminology and technical language on food packaging. Ingredients lists are often filled with scientific names and unfamiliar terms, making it difficult for the average consumer to decipher what they are actually consuming. This lack of understanding can lead to misconceptions about the healthiness of certain products and may contribute to unhealthy dietary choices. Front-of-pack labels often feature bold colors, enticing images and catchy slogans, which can distract consumers from paying attention to the ingredients list and nutritional information. This can be particularly problematic when these labels make exaggerated or misleading health claims, leading

consumers to believe that a product is healthier than it actually is.

Moreover, the small size of text on food packaging can also

pose a barrier to consumer understanding. Many consumers, especially those with visual impairments or reading difficulties, may struggle to read the fine print on packaging, making it challenging for them to access important nutritional information.

This review paper covers a detailed literature survey summarizing key findings in nutritional analysis and ingredient classification of packaged food products. It also includes methodologies for data collection, dataset description, and analysis techniques. Future directions and scope are discussed, focusing on areas for further research and development.

II. LITERATURE SURVEY

In the field of nutrition and food science, researchers have studied packaged food products to better understand their complexities. This literature review examines key research on packaged food analysis and nutrition. Several seminal papers offer insights and methodologies for comprehending the nutritional aspects of packaged foods, focusing on food labeling, ingredient analysis, and nutritional profiling. By synthesizing findings from these works, the review sheds

light on the challenges of promoting nutritional literacy and healthy eating habits. Through this survey, we aim to highlight key themes, methodologies, and contributions that shape our understanding of packaged foods and nutrition, informing future research and practical applications to encourage healthier food choices and lifestyles.

A. Key Insights from Studies

This section presents summaries of research papers that provide valuable insights into this domain. Each summary is crafted to encapsulate the core findings and implications of the respective research papers, offering a concise yet informative overview of the current state of knowledge in this domain.

1) *Use of food label information by urban consumers in India – A study among supermarket shoppers [1]*

Authors: Sudershan R Vemula, SubbaRao M Gavaravarapu, Vishnu Vardhana Rao Mendu, Pulkit Mathur, Laxmaiah Avula

Year: 2021

Overview: The study investigates how urban consumers in India use food label information. Despite high literacy rates, consumers prioritize manufacturing and expiry dates over nutritional content when buying pre-packaged foods. Many find nutritional information too complex to understand. Education level significantly influences engagement with food labels, indicating the need for targeted educational interventions to improve nutritional literacy. Revisiting food labeling strategies and exploring alternative formats can empower consumers to make healthier dietary choices, contributing to public health initiatives in India.

Limitations: Sample skewed towards educated consumers, potentially affecting generalizability. Potential bias in feedback from beta testers involved in system development.

Learning: The research underscores the importance of innovative technologies like VD in addressing inadequate nutrition, the need for continuous improvement, and the potential for such systems to contribute to improved public health outcomes.

2) *Nutrient Facts Analysis using Supervised Learning Approaches [2]*

Authors: J. Aravind, J. Dhalia Sweetlin

Year: 2017

Overview: In their study titled "Nutrient Facts Analysis using Supervised Learning Approaches," J. Aravind and J. Dhalia Sweetlin address the challenge of interpreting nutritional information on food labels, which often proves daunting for common consumers. Their research proposes employing supervised learning techniques to classify food products into

five levels of healthiness, aiming to simplify the understanding of nutritional information and aid consumers in making informed dietary choices. Through a sequential process of data retrieval, cleaning, labeling, and supervised learning, the study endeavors to enhance the accessibility and usability of nutritional information. By contributing insights into the application of supervised learning algorithms in interpreting nutritional food facts details, the research provides a foundation for further exploration and improvement in this area, potentially facilitating better dietary decision-making and promoting healthier lifestyles among consumers.

Limitations: The effectiveness of the predictive model in classifying food products may be sensitive to the selection of features used.

Learnings: The research highlights the potential of supervised learning algorithms to interpret nutritional information on food labels, emphasizing the need for more user-friendly nutritional labelling and suggesting further improvements in data quality and the exploration of alternative machine learning methods for food classification.

3) *NutriTrack: Android-based Food Recognition App for Nutrition Awareness [3]*

Authors: Arnel B. Oca, Jane M. Fernandez, Thelma D. Palaoag

Year: 2017

Overview: Arnel B. Oca, Jane M. Fernandez, and Thelma D. Palaoag, in this research introduce an innovative Android app designed to promote health awareness by aiding users in understanding the nutritional content of their food. The app allows users to photograph food and access its nutritional information, helping them make informed dietary decisions. Utilizing the Mifflin-St Jeor method, users can determine their daily calorie intake. Despite the app's potential, surveys conducted among users reveal a concerning lack of nutritional awareness. However, the app's framework, incorporating Clarifai and Nutritionix for food detection and nutrition calculation, demonstrates promising functionality. Overall, "NutriTrack" offers a valuable tool for enhancing dietary awareness, emphasizing the ongoing need for improved nutritional education among users.

Limitations: The accuracy and effectiveness of the NutriTrack application heavily depend on users accurately inputting their food consumption data. Inaccurate or incomplete data entry may lead to unreliable nutrition estimations and dietary monitoring.

Learning: It highlights that while innovative apps like the one developed have the potential to improve nutritional awareness, there remains a significant need to raise public knowledge about food intake and nutrition. The choice of

APIs and features can greatly influence the usability and functionality of such apps.

4) *Personalised Food Classifier and Nutrition Interpreter Multimedia Tool Using Deep Learning* [4]

Authors: M. Sundarramurthi, Nihar. M, Anandi Giridharan

Year: 2020

Overview: In their research on "Personalised Food Classifier and Nutrition Interpreter Multimedia Tool Using Deep Learning," M. Sundarramurthi, Nihar. M, and Anandi Giridharan introduce a user-friendly application aimed at addressing obesity concerns by aiding individuals in tracking their nutritional intake. The Food Classifier and Nutrition Interpreter (FCNI) utilizes Deep Learning to classify food types, providing graphical representations of nutritional values and calorie estimations, alongside multimedia audio responses for enhanced interaction. Achieving an impressive accuracy of approximately 96.81%, the FCNI tool demonstrates the efficacy of Convolutional Neural Networks (CNNs) in food image classification. This research presents a valuable tool for dietary assessment, contributing to efforts in promoting healthier lifestyles and combating obesity.

Limitations: The accuracy and reliability of the FCNI tool heavily rely on the quality of the captured food images. Poor lighting conditions, blurry images, or occluded food items may result in inaccurate classification.

Learnings: The FCNI tool demonstrates how advanced technology can aid in improving dietary assessment and promoting a healthier lifestyle.

4) *Study of Consumer Awareness on Food Labelling and Use of Pack Information for Purchase of Pre-Packaged Food Products* [5]

Authors: Samit Dutta, Deval Patel

Year: 2017

Overview: Samit Dutta and Deval Patel in this study titled "Study of Consumer Awareness on Food Labelling and Use of Pack Information for Purchase of Pre-Packaged Food Products" addresses the increasing trade and consumption of pre-packaged foods by investigating consumer awareness and utilization of food labelling information in Anand city, Gujarat, India. Through a semi-structured questionnaire administered to consumers purchasing pre-packaged foods in selected modern retail stores, the study determines the level of awareness, perception of importance, and factors influencing the reading and use of food labels. Findings indicate that while 86.7% of participants read labelling information before purchase, only a third are highly informed about food labelling. Education level and gender significantly influence awareness and perception of the importance of food labelling. Moreover, 83.3% of respondents cited price as a motivating factor for reading food labels.

Limitations: The study's sample may be skewed towards educated consumers, affecting generalizability. Feedback from beta testers involved in system development could introduce bias.

Learnings: Efforts are needed to improve food labelling and educate consumers on its importance for informed food choices. Addressing difficulties in reading labels could enhance consumer understanding and use of food labelling information.

5) *A comparison of the Health Star Rating system when used for restaurant fast foods and packaged foods* [6]

Author: Elizabeth K. Dunford, Jason H.Y. Wu, Lyndal Wellard-Cole, Wendy Watson, Michelle Crino, Kristina Petersen, Bruce Neal

Year: 2017

Overview: Investigating the applicability of the Health Star Rating (HSR) system in assessing fast foods, Elizabeth K. Dunford et al. conducted a comparative study between fast food items and packaged foods in Australia. Their research involved collecting nutrient content data from 13 major fast-food chains and calculating HSR values for each menu item, which were then compared with packaged food products. The results revealed similar distributions of HSR values between fast foods and packaged foods, although statistically significant differences were observed in seven categories. These findings suggest the potential extension of the HSR system to fast foods, offering consumers standardized healthiness indicators across all food categories. Additionally, the study highlights the importance of potential modifications to the existing HSR algorithms to better accommodate both packaged and fast foods. Overall, the research emphasizes the value of a unified approach to food labeling, facilitating healthier dietary choices across diverse food environments.

Limitations: Consumer understanding and interpretation of the HSR system, particularly in the context of fast foods, were not directly assessed, limiting the assessment of its effectiveness in influencing food choices.

Learnings: The results suggest that the HSR system can be applied to fast foods, potentially providing a standardized approach for consumers to assess the healthiness of various food choices, across fresh, packaged, and restaurant foods. The extension of HSR to fast foods could encourage a positive response from the fast food industry, similar to its impact on packaged foods.

III. METHODOLOGY

This section goes over the data collection processes, dataset description, 'Nutri-Score' calculation method and data visualization methods.

A. Data Collection Process

Data collection for this review focuses on obtaining information from the packaging of various packaged food products. The primary sources of data includes the nutritional table and the ingredients list, both of which are typically displayed on the packaging of these products. The following methods can be employed to collect this data:

- 1) *API Utilization:* The use of Application Programming Interfaces (APIs), such as the Open Food Facts API can facilitate the automated retrieval of nutritional and ingredient information from a vast database of packaged food products. This method ensures the collection of standardized and structured data. However, this approach presents a limitation, namely the scarcity of data pertaining to Indian packaged food products.
- 2) *Manual Data Entry:* In cases where API access is limited or unavailable, manual data entry can be performed. This approach too presents few limitations which are listed below:
 - *Prone to human error:* Manual data entry is susceptible to inaccuracies and omissions due to human factors, potentially compromising the quality of the collected data.
 - *Time-consuming:* This process can be lengthy, especially when dealing with a large volume of products.
 - *Labor-intensive and costly:* The manual data entry process requires dedicated human resources, increasing operational costs and resource allocation.
- 3) *Web Scraping (preferred):* Web scraping techniques can be employed to extract data from online sources, including official product websites and online retailers. This method expands the scope of data beyond what is available through APIs. Building a database from web-scraped data is advantageous for several reasons:
 - *Comprehensive Data Collection:* Web scraping allows for the collection of a wide range of data beyond what is available through APIs, providing a more comprehensive dataset for analysis.
 - *Efficiency:* Web scraping is faster and more efficient than manual data entry, enabling the collection of large volumes of data in a relatively short period.

- *Scalability:* Web scraping can easily scale to accommodate a growing database, making it suitable for projects requiring continuous data collection and expansion.
- *Improved Accuracy:* While human error can affect manual data entry, web scraping can be programmed to extract data accurately, reducing the risk of errors in the database.

- 4) *Collaboration with Data Service Providers:* Collaboration with data service providers specialized in food product data shall enable access to comprehensive datasets. This partnership shall ensure the collection of high-quality data from a reliable source.

By employing a combination of these methods, a diverse range of packaged food products can be covered.

B. Dataset Description

The dataset comprises information on packaged food products, focusing primarily on Indian packaged food products.

Each entry in the dataset includes the following attributes:

- *EAN (European Article Number):* A unique identifier for each packaged food product, facilitating accurate identification and retrieval of product information.
- *Product Name:* The name of the packaged food product, providing a clear and concise label for identification purposes.
- *Nutritional Information:* This attribute represents the nutritional information of the packaged food product. Due to its semi-structured nature, this information is stored as a JSON data type, allowing for flexibility in data representation and storage.
- *Ingredients List:* A list of ingredients used in the packaged food product, providing transparency regarding the composition of the product.
- *Category:* The category to which the packaged food product belongs, such as beverages, snacks, biscuits, chocolates, etc. This attribute helps in organizing and classifying the products for analysis and review.
- *Allergens:* Information about common allergens (e.g., gluten, nuts, dairy) present in the product.
- *Food Processing:* The degree of processing and any additives or preservatives used.

C. Nutritional Analysis: The Visual Approach

Double bar charts can be a useful visualization tool for comparing the nutritional content of two products. By displaying the nutritional values side by side, it allows for a quick and easy comparison between the two products. This can be particularly helpful for consumers who want to make informed choices about their food purchases. The chart can include various nutritional components such as calories, fats, sugars, proteins, and other relevant nutrients, making it easy to see the differences between the products at a glance.

Pie charts are commonly used in nutritional analysis to visually represent the distribution of macronutrients or micronutrients in a food product. In the context of packaged food products, pie charts can be a valuable tool for summarizing and comparing the nutritional content of different products.

Macronutrient Distribution: One of the key uses of pie charts in displaying nutritional table information is to illustrate the distribution of macronutrients (carbohydrates, proteins, and fats) in a food product.

Micronutrient Distribution: Pie charts can also be used to display the distribution of micronutrients (such as vitamins and minerals) in a food product.

Interpretation and Understanding: Pie charts provide a visually appealing way to present complex nutritional information, making it easier for readers to interpret and understand the data. The use of colours and labels in pie charts can further enhance readability and help readers quickly identify key nutritional components of a food product.

D. Nutritional Analysis: The Metric-based Approach

The Nutri-Score, also known as the 5-Colour Nutrition label or 5-CNL, is a five-color nutrition label and nutritional rating system, and an attempt to simplify the nutritional rating system demonstrating the overall nutritional value of food products. It assigns products a rating letter from A (best) to E (worst), with associated colors from green to red. Two possible approaches can be used to calculate the Nutri-Score: the first involves using the standard formula and the second, employs a machine learning approach.

Calculating Nutri-Score using the standard process: The calculation of Nutri-Score involves several steps and considerations, as outlined below:

- **Nutritional Components:** Nutri-Score takes into account both positive and negative aspects of the nutritional composition of a food product. It considers the content of energy, sugars, saturated fats, sodium, protein, fibers, fruits, vegetables, legumes, and nuts (FVLM) per 100 grams or millilitres of the product
- **Points System:** Products high in energy, sugars, saturated fats, and sodium receive positive points,

while those high in protein, fibers, and FVLM receive negative points.

- **Calculation Formula:** The total number of points is calculated by summing up the points for each component. The points are then converted into a score ranging from -15 to +40.
- **Classification:** Based on the total score obtained, the product is classified into one of five categories: A (green, most favourable) to E (red, least favourable).
- **Colour Coding:** The final step is to assign a colour to the Nutri-Score based on the category. A and B are assigned green, C is assigned yellow, and D and E are assigned red.

E. Nutritional Analysis: The Machine Learning Approach

The supervised learning approach: The machine learning approach for classifying food products into the five categories (A, B, C, D, and E) based on their nutritional quality, known as the Nutri-Score, is a multifaceted process that leverages a variety of attributes. These attributes include detailed nutritional information, identification of allergens, consideration of food processing methods, and other pertinent factors that collectively contribute to the overall healthiness of a food product.

Classifying the training data: The data used for training the machine learning models is to be classified with the assistance of nutritional experts possessing profound understanding of the intricate details of food composition, nutritional values, and health implications. This expert-driven classification process ensures that the problem of classifying food products into Nutri-Score categories (A, B, C, D, and E) follows a well-defined process and is not undecidable. By leveraging the expertise of these professionals, a comprehensive and reliable procedure can be established to determine the class of each food product based on its nutritional information, allergen content, food processing methods, and other relevant factors. This meticulous approach helps in creating a robust and accurate Nutri-Score classification system that aligns with established nutritional guidelines and standards.

Suitable machine learning algorithms: To implement this approach, a range of supervised machine learning algorithms can be employed, each with its strengths and characteristics. These algorithms include K-Nearest Neighbours, Naive Bayes, Support Vector Machines (SVM), and Decision Trees. By training these algorithms on a dataset containing labelled examples of food products with their corresponding Nutri-Scores, the system learns to recognize patterns and relationships between the input attributes and the output categories.

The training phase involves feeding the algorithms with known data and adjusting their internal parameters to

optimize performance. Once trained, the algorithms are tested using a separate dataset to evaluate their accuracy and effectiveness in predicting Nutri-Scores for new, unseen food products.

The Ensemble learning approach: The ensemble learning approach is utilized, which combines the predictions of multiple algorithms to improve overall classification accuracy and robustness.

One common ensemble method is the voting system, where each individual model in the ensemble "votes" on the predicted class for a given food product. The class that receives the most votes is then selected as the final prediction. This approach is particularly effective when the individual models have diverse strengths and weaknesses, as it allows them to complement each other and produce a more reliable prediction.

In the context of Nutri-Score classification, ensemble learning can be implemented by training multiple machine learning models using different algorithms (such as K-Nearest Neighbours, Naive Bayes, SVM, and decision trees) on the same dataset. Each model learns to recognize different patterns in the data, and by combining their predictions, the ensemble can make more accurate and robust predictions than any individual model alone.

Overcoming Overfitting and Underfitting: Ensemble learning can also involve using different subsets of the training data to train each individual model, a technique known as "bagging" or bootstrap aggregating. This approach helps reduce overfitting and improve generalization by introducing diversity into the ensemble. Overall, ensemble learning offers a powerful and flexible approach to nutri-score classification, allowing for the creation of a highly accurate and reliable system for assessing the nutritional quality of food products.

F. Ingredient Analysis

The ingredient analysis component endeavours to classify ingredients into three distinct categories—red, yellow, and green—based on their respective health impacts. Given that the precise quantities of ingredients are not explicitly provided on packaging, the primary objective is to furnish consumers with a comprehensive overview of both the quantity and health impact of each ingredient through the adept use of visualization techniques.

Consider a scenario where a food product purportedly contains a healthy ingredient, yet upon closer inspection, it becomes evident that the quantity of this ingredient is negligible compared to the presence of other, potentially unhealthy ingredients in more significant quantities. Through the application of clear and insightful visualizations that elucidate both the relative quantity of each ingredient and its corresponding health impact, ingredient analysis aims to

effectively "unveil" the true composition of packaged food products to consumers.

Moreover, ingredient analysis will not only furnish consumers with basic information pertaining to each ingredient but also endeavour to explicate any scientific terminology that may be unfamiliar to them. For instance, ingredients such as sugar often entail various scientific terms that can be unfamiliar to consumers. Through ingredient analysis, we aim to provide concise explanations for such terms, thereby empowering consumers to make well-informed decisions regarding the products they choose to consume.

IV. FUTURE SCOPE AND CONCLUSION

A. Future Scope

Integration with Health Monitoring Devices: Integrating Nutri-Score information with health monitoring devices and apps could provide consumers with real-time feedback on their dietary choices and health outcomes.

Gamification and Rewards: Implement gamification elements to make the app more engaging. Users could earn rewards or points for making healthier food choices or for using the app regularly.

Comparison with Safe Intake: Develop a feature that compares the nutritional content of a food product with the safe intake based on the user's age, weight, and other relevant factors. This can help users make more informed decisions about their food choices.

Healthier Alternatives: Use clustering algorithms to suggest healthier alternatives to the current product. This could involve analyzing the nutritional content of similar products and recommending ones that are healthier.

B. Conclusion

In conclusion, advancements in technology offer exciting opportunities to revolutionize the way we approach nutrition and food choices. By leveraging these technologies effectively, we can empower consumers to make informed decisions that positively impact their health and well-being.

REFERENCES

- [1] S. M. G. V. V. R. M. P. M. L. A. Sudershan R Vemula, "Use of food label information by urban consumers in India – a study among supermarket shoppers," 2021.
- [2] J. D. S. J. Aravind, "Nutrient Facts Analysis using Supervised Learning Approaches," 2017.

- [3] J. M. F. T. D. P. Arnel B. Oca, "NutriTrack: Android-based Food Recognition App for Nutrition Awareness," 2017.
- [4] N. M. A. G. M. Sundarramurthi, "Personalised Food Classifier and Nutrition Interpreter Multimedia Tool Us-ing Deep Learning," 2020.
- [5] D. P. Samit Dutta, "Study of Consumer Awareness on Food Labelling and Use of Pack Infor-mation for Purchase of Pre-Packaged Food Products," 2017.
- [6] J. H. W. L. W.-C. W. W. M. C. K. P. B. N. Elizabeth K. Dunford, "A comparison of the Health Star Rating system when used for restaurant fast foods and packaged foods," 2017.
- [7] J. B. M. C. C. T. Manuel B. Garcia, "Virtual Dietitian: A Nutrition Knowledge-Based System Using Forward Chaining Algorithm," 2021.
- [8] G. P. K. A. B. P. S. A. G.-R. D. P. S. M.-h. C. K. Nikhil Srinivasapura Venkateshmurthy, "A Photovoice Study to Reveal Community Perceptions of Highly Processed Packaged Foods in India," 2021.
- [9] S. T. P. Y. S. T. Natnicha Suthumchai, "FoodForCare: An Android Application for Self-Care with Healthy Food," 2016.