SKINCARE INGREDIENT ANALYZER: YOUR SMART COSMETIC COMPANION

Project Report Submitted in partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

by

G. KRISHNASRI (21501A0552)

A. POOJITHA (21501A0511)

G. JYOTHENDRA (21501A0549)

J. PRAVEEN KUMAR (21501A0566)

Under the guidance of

Dr. M. Sailaja, Assistant Professor

Department of Computer Science and Engineering

Prasad V Potluri Siddhartha Institute of Technology

(Permanently affiliated to JNTU-Kakinada, Approved by AICTE)

(An NBA & NAAC A+ accredited and ISO 9001:2015 certified institute)

Kanuru, Vijayawada-520 007

2024-25

Prasad V Potluri Siddhartha Institute of Technology

(Permanently affiliated to JNTU-Kakinada, Approved by AICTE)

(An NBA & NAAC A+ accredited and ISO 9001:2015 certified institute)

Kanuru, Vijayawada-520 007

CERTIFICATE

This is to certify that the project report entitled *"Skincare Ingredient Analyzer: Your Smart Cosmetic Companion"* is being submitted by G. **Krishnasri, G. Jyothendra , A. Poojitha, J. Praveen Kumar** in partial fulfillment for the award of the Degree of **Bachelor of Technology in Computer Science and Engineering** during the academic year 2024-25.

Signature of the guide
Dr. M. Sailaja,
Assistant Professor,
Department of CSE,
Prasad V. Potluri Siddhartha Institute
of Technology, Vijayawada.

Signature of the HOD
Dr. A. JayaLakshmi,
Professor and HOD,
Department of CSE,
Prasad V. Potluri Siddhartha Institute
of Technology, Vijayawada.

SIGNATURE OF THE EXTERNAL EXAMINER

DECLARATION

We declare that the project work entitled *"Skincare Ingredient Analyzer: Your Smart Cosmetic Companion"* is composed by ourselves, and the work contained herein is our own except where explicitly stated otherwise in the text. This work has not been submitted for any other degree or professional qualification.

G. Krishnasri(21501A0552)

A. Poojitha(21501A0511)

G. Jyothendra(21501A0549)

J. Praveen kumar(21501A0566)

ACKNOWLEDGEMENT

We sincerely thank the following distinguished personalities who have given their advice and support for successful completion of the work.

We are deeply indebted to our most respected guide **Dr. M. Sailaja, Assistant Professor,** Department of computer science and engineering for her valuable and inspiring guidance, comments, suggestions and encouragement.

We extend our sincere thanks to **Dr. A. Jayalakshmi, Professor & Head** of the department for extending her cooperation and providing the required resources.

We would like to thank our beloved **Principal, Dr K Sivaji Babu** for providing the online resources and other facilities to carry out this work.

We're grateful to **Dr. A. Ramana Lakshmi, Associate Professor** of our department and our esteemed Project Coordinator for her invaluable guidance and efforts.
Her exceptional leadership, from inception to final review, ensured the project's success.

We would like to extend our sincere graduation to all the review panel **Dr. M. V. Ramakrishna, Professor, Dr. A. Ramana Lakshmi, Associate Professor, and Dr. Kilari Jyothsna Devi, Assistant Professor** for their valuable suggestions in project reviews.

We extend our sincere thanks to all other teaching and non-teaching staff of the department, for their cooperation and encouragement.

G. Krishnasri(21501A0552)

A. Poojitha(21501A0511)

G. Jyothendra(21501A0549)

J. Praveen kumar(21501A0566)

ABSTRACT

The skincare industry faces persistent challenges in ensuring ingredient safety and providing personalized product recommendations. Consumers often struggle to identify harmful substances, allergens, or toxic compounds in their skincare products, leading to potential adverse effects. Additionally, the lack of tailored recommendations based on individual skin concerns hinders the effectiveness of skincare routines. In response to these challenges, we propose an innovative solution integrating EasyOCR and Random Forest. EasyOCR extracts ingredient lists from uploaded product images, while FuzzyWuzzy performs fuzzy string matching to assess ingredient safety, categorizing them as safe, warning, or harmful. This system alerts users to potential allergens and toxic ingredients, enabling informed decision-making. Furthermore, a Random Forest model leverages user skin type and concerns—such as acne, aging, or hyperpigmentation—to provide personalized skincare recommendations. By combining optical character recognition with machine learning, this approach enhances transparency, empowers users with data-driven insights, and revolutionizes skincare safety and efficacy.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER - 1

INTRODUCTION

1.1 Introduction to Project:

The Skincare Ingredient Analyzer is an advanced analytical tool designed to help consumers gain a deeper understanding of the ingredients present in their skincare products. With the increasing awareness of skincare safety and ingredient transparency, many individuals are becoming more cautious about the substances they apply to their skin. Concerns about the potential presence of harmful chemicals, allergens, and synthetic compounds have made it imperative for consumers to have access to reliable information. This tool serves as a valuable resource by providing users with essential data that enables them to make well-informed skincare choices, thereby enhancing their overall skincare routines.

One of the key functions of the Skincare Ingredient Analyzer is its ability to scan ingredient lists found on skincare product packaging and generate comprehensive reports on each component. These reports not only identify the substances present but also offer in-depth explanations regarding their uses, sources, and potential effects on the skin. For instance, the tool differentiates between naturally derived and synthetic ingredients, highlighting which ones may be beneficial and which ones pose potential risks. This feature is particularly useful for individuals who may not be familiar with complex chemical terminologies often found in product formulations. By simplifying this information, the analyzer empowers consumers to distinguish between beneficial and potentially harmful substances with ease.

Furthermore, the analyzer plays a critical role in identifying allergens and toxic ingredients that may cause skin irritation or adverse reactions. Many skincare products contain preservatives, fragrances, and artificial compounds that can trigger allergic responses or sensitivities. However, consumers are often unaware of these risks due to the lack of clear and transparent ingredient labelling. The Skincare Ingredient Analyzer bridges this gap by flagging potential allergens and hazardous chemicals, thereby ensuring that users can select products that align with their skin type and specific sensitivities. This personalized approach to skincare safety helps individuals avoid ingredients that may cause redness, breakouts, or long-term damage to their skin.

Another significant advantage of this tool is the empowerment it provides to consumers. Traditionally, individuals have had to rely on marketing claims and brand promises when selecting skincare products. However, these claims can sometimes be misleading, causing consumers to purchase items that do not align with their skin's needs. By offering factual and science-based information, the Skincare Ingredient Analyzer eliminates the reliance on marketing tactics and enables individuals to make purchasing decisions based on evidence rather than advertisements. This not only fosters a greater sense of trust in the products being used but also allows consumers to tailor their skincare routines to better suit their specific concerns, such as acne, aging, hyperpigmentation, or dryness.

In addition to its core analytical capabilities, the Skincare Ingredient Analyzer incorporates advanced technological features that enhance user convenience. Most analyzers are available as mobile applications, making it easy for consumers to access ingredient information on the go. With just a simple scan of a product's barcode or ingredient list, users can instantly retrieve detailed insights about the formulation. This real-time functionality is especially beneficial during shopping trips, as it allows individuals to make informed choices before purchasing a product. By reducing the likelihood of impulsive buying and helping consumers avoid products that may not be suitable for their skin, the tool contributes to smarter and healthier skincare habits.

Beyond its individual benefits, the Skincare Ingredient Analyzer also fosters a sense of community among users. Many analyzers feature community-driven aspects such as user reviews, ratings, and shared experiences. This social integration allows consumers to learn from one another, exchange knowledge, and receive recommendations based on real-life experiences. Users can read firsthand accounts of how specific ingredients or products have affected others with similar skin types, which further aids in decision-making. By creating a collaborative environment, the analyzer not only educates individuals but also encourages collective awareness regarding skincare safety and ingredient transparency.

Overall, the Skincare Ingredient Analyzer is an essential tool for modern consumers who seek to prioritize their skin health and well-being. By offering detailed ingredient analysis, identifying potential allergens and harmful substances, and promoting informed decision-making, this tool revolutionizes the way individuals approach skincare. Its accessibility through mobile applications, along with community-based features, enhances user engagement and ensures that people can make the best choices for their skin with confidence. As the demand for ingredient transparency continues to grow, the Skincare Ingredient

Analyzer stands as a powerful solution, empowering individuals to take control of their skincare routines and make choices that align with their health and safety needs.

1.2 Motivation:

The motivation behind developing the Skincare Ingredient Analyzer and Product Recommendation System stems from the increasing concern about skincare safety and transparency. Many consumers struggle to understand complex ingredient labels, often falling for misleading marketing claims that lead to the purchase of products containing harmful chemicals or allergens. This project aims to bridge the knowledge gap by providing a user-friendly tool that simplifies ingredient analysis and offers clear, factual insights, empowering users to make safer and more informed choices. Additionally, growing awareness of skincare-related health issues, such as allergic reactions and long-term skin damage, highlights the need for better ingredient transparency. With advancements in AI-driven analysis, the tool delivers real-time, accurate information while also recommending skincare products tailored to users' needs. By analyzing ingredient compositions and user preferences, the system suggests safer, dermatologist-approved, or scientifically backed alternatives, reducing the trial-and-error approach in skincare shopping. The ultimate goal is to promote informed decision-making, encourage a healthier approach to skincare, and foster industry-wide accountability in ingredient disclosure. Through accessibility and convenience, the project seeks to revolutionize skincare safety, consumer awareness, and personalized product selection.

1.3 Statement of the problem and Solution to Problem:

1.3.1 problem statement:

Many cosmetics contain unsafe ingredients, leading to health concerns. Consumers often unknowingly apply products containing harmful chemicals, allergens, and synthetic additives that may cause skin irritation, allergic reactions, or long-term health risks. The lack of transparency in ingredient labeling and misleading marketing claims make it difficult for individuals to identify and avoid potentially dangerous substances. Consumers face challenges in deciphering complex ingredient lists, increasing the risk of using unsafe products. Additionally, limited tools to evaluate ingredient safety create uncertainty and hinder informed decision-making. The overwhelming variety of skincare products, combined with insufficient guidance, makes it difficult for users to choose products suited to their skin type and concerns. Current solutions lack an integrated approach that combines ingredient safety analysis with personalized product recommendations. To address these issues, this project aims to provide a reliable AI-powered tool that not only analyzes ingredient safety but also suggests dermatologist-approved and scientifically backed alternatives based on user preferences. By offering clear, science-backed insights and personalized recommendations, the system empowers consumers to make safer skincare choices, reducing health risks while enhancing transparency and confidence in product selection.

1.3.2 Solution:

The Skincare Ingredient Analyzer is a machine-learning-based tool that extracts ingredient lists from skincare products, analyzes their safety levels, and provides personalized recommendations. By leveraging technologies such as optical character recognition (OCR) and FuzzyWuzzy, the system identifies ingredients and categorizes them based on their potential risks. The analyzer then cross-references this data with a comprehensive ingredient safety database to flag harmful chemicals, allergens, and irritants. Additionally, using a Random Forest model, it tailors skincare product recommendations to individual users based on their skin type and concerns, such as acne, aging, or sensitivity. This AI-powered approach ensures that consumers receive accurate, evidence-based insights, enabling them to make informed and safe skincare choices.

1.4 Objectives:

1.4.1 General Objective:

The primary objective of this project is to develop an AI-based ingredient analysis tool that helps consumers make informed skincare choices. The tool will extract and analyze ingredient lists from skincare products, assess their safety levels, and provide detailed insights on potential risks and benefits. By leveraging machine learning, it will offer personalized recommendations based on individual skin types and concerns. This system aims to enhance transparency in skincare product formulations, empower users with factual data, and promote healthier skincare habits. Ultimately, it seeks to revolutionize ingredient awareness and safety in the cosmetics industry.

1.4.2 Specific Objectives:

The Skincare Ingredient Analyzer aims to extract ingredient lists from product images using Optical Character Recognition (OCR) technology, classify ingredient safety levels based on scientific databases, and provide users with real-time, AI-driven recommendations. The tool will use machine learning models to assess ingredient risks, identifying potential allergens, harmful chemicals, and beneficial components. Additionally, it will offer personalized product recommendations tailored to the user's skin type and concerns, ensuring safer and more effective skincare choices. This approach enhances ingredient transparency and enables users to make well-informed purchasing decisions.

1.5 Scope of the Work:

The proposed solution, the Skincare Ingredient Analyzer and Product Recommendation System, addresses the growing concerns of skincare safety and transparency. Its scope extends to revolutionizing the skincare industry by enhancing the accuracy of ingredient analysis, empowering consumers with science-backed insights, and providing personalized product recommendations. By leveraging OCR, AI-driven analysis, and fuzzy matching algorithms, the system ensures precise identification of harmful or allergenic ingredients, reducing health risks associated with unsafe skincare products. Additionally, the integrated recommendation module streamlines the selection of suitable alternatives based on user preferences and skin concerns, promoting safer and more informed skincare choices.

1.6 Significance of the Work:

The Skincare Ingredient Analyzer enhances consumer awareness by providing reliable information on the ingredients used in skincare products. It helps individuals avoid harmful substances by analyzing ingredient safety and flagging potential risks. By offering personalized skincare recommendations, the tool ensures that users select products suitable for their specific skin types and concerns. Furthermore, it empowers consumers with scientific and factual insights, reducing reliance on misleading marketing claims and fostering informed decision-making. The analyzer promotes healthier skincare habits by guiding users towards safer product choices, ultimately minimizing the risk of skin irritation or allergic reactions.

Additionally, its accessibility via mobile technology enables real-time ingredient analysis, ensuring that consumers can make informed decisions while shopping for skincare products.

1.7 Outlines of the Project:

1.7.1 Introduction

- Background and importance of skincare ingredient analysis
- Consumer concerns regarding ingredient safety
- Need for transparency in skincare product formulations

1.7.2 Literature Review

1. **Harmful Ingredients in Skincare Products:**
   Many skincare products contain chemicals such as parabens, sulfates, and synthetic fragrances, which can cause skin irritation, allergic reactions, and long-term health risks. Studies highlight the need for ingredient transparency to help consumers make safer skincare choices.
2. **AI and Machine Learning in Skincare Analysis:**
   AI-powered tools like Optical Character Recognition (OCR) and machine learning algorithms improve ingredient identification and classification. Research shows that machine learning models, such as Random Forest, enhance personalized skincare recommendations based on individual skin concerns.

1.7.3 Methodology

1. Data Collection and Processing

The system collects ingredient lists from skincare product images using Optical Character Recognition (OCR). The extracted text is then processed, cleaned, and standardized to ensure accuracy in ingredient identification. Additionally, a comprehensive ingredient safety database is integrated to classify substances based on potential risks.

2. Machine Learning Model Development

A Random Forest model is trained to analyze ingredient safety and provide personalized skincare recommendations. The model is developed using a dataset containing ingredient safety classifications and user skin types to improve prediction accuracy. Fuzzy string matching techniques, such as FuzzyWuzzy, are applied to handle variations in ingredient names.

3. System Implementation and Testing

The skincare analyzer is deployed as a user-friendly application, allowing consumers to scan product labels and receive instant ingredient safety reports. The system undergoes rigorous testing, including accuracy assessments, user feedback analysis, and performance evaluations to ensure seamless operation and reliable results.

1.7.4 System Architecture

Data Acquisition

- Uses Optical Character Recognition (OCR) to extract ingredient lists from product images.

Data Processing

- Applies text cleaning and fuzzy string matching to standardize ingredient names.

Ingredient Classification

- Categorizes ingredients as safe, warning, or harmful based on scientific databases.

Machine Learning Model

- Uses a Random Forest algorithm to generate personalized skincare recommendations.

User Interface

- Provides real-time ingredient analysis and skincare suggestions via a website.

Database and Cloud Storage

- Stores ingredient data and user preferences for continuous updates and improvements.

1.7.5 Implementation

- Technical details of the implementation process: Software tools, frameworks, and libraries used.

- Challenges encountered during implementation and their solutions.
- Discussion of hardware requirements and considerations for deployment.

1.7.6 Results

• **Evaluation Metrics** – The system's performance is assessed using accuracy, precision, recall, and F1-score to measure the effectiveness of ingredient identification and classification.

• **Experimental Results** – The OCR module's accuracy in extracting ingredient lists is evaluated, along with the efficiency of the machine learning model in categorizing ingredients and generating personalized recommendations.

• **Comparison with Existing Methods** – The AI-powered analyzer is compared with traditional manual ingredient checking and existing skincare analysis apps to highlight improvements in speed, accuracy, and user experience.

CHAPTER – 2

BACKGROUND AND LITERATURE REVIEW

AI-based skincare systems are revolutionizing the cosmetic industry by leveraging machine learning (ML) and deep learning (DL) to analyze cosmetic ingredients and recommend personalized products. These technological advancements are particularly significant in two areas: identifying toxic chemicals in skincare products and promoting the shift towards sustainable cosmetics. Many modern cosmetic products contain harmful substances such as parabens, UV filters, sulfates, and heavy metals, which have been linked to severe health conditions, including skin allergies, cancer, hormonal imbalances, and infertility. What is even more concerning is that less than 20% of the 12,000+ synthetic chemicals used in cosmetics are considered safe. The necessity of an ingredient safety analysis tool has become increasingly evident as regulatory gaps allow dangerous chemicals to persist in widely available skincare products. Additionally, the presence of microplastics and UV filters in cosmetic formulations contributes to environmental degradation, particularly in marine ecosystems, highlighting the urgent need for more sustainable and eco-friendly cosmetic ingredients (Environmental Journal, 2023).

The dangers of synthetic chemicals in cosmetics extend beyond immediate health concerns. Several studies have revealed the presence of toxic heavy metals such as lead, mercury, and cadmium in beauty products, which can lead to severe health effects, including endocrine disruption, reproductive dysfunction, and neurological damage. Endocrine-disrupting chemicals such as parabens and UV filters interfere with hormonal balance and have been scientifically linked to cancer. Alarmingly, weak regulatory policies in many countries have failed to restrict the use of these hazardous chemicals, exposing consumers to significant health risks. The prolonged exposure to such substances has reinforced the need for stricter safety monitoring and enhanced consumer awareness. AI-driven ingredient analysis tools play a crucial role in addressing these issues by providing real-time information on ingredient safety and helping users make informed choices (Frontiers in Environmental Science, 2024).

Recent advancements in deep learning models have demonstrated remarkable accuracy in predicting the presence of toxic ingredients in skincare products. A study using Random Forest (RF) and Support Vector Machines (SVM) for soap toxicity detection achieved an accuracy of 95%, outperforming traditional chemical assessments. These AI models analyze multiple factors, including pH levels, risk assessments, and ingredient properties, to determine potential toxicity. This level of precision is a significant breakthrough in protecting consumers from harmful skincare products. AI-driven systems ensure that toxic formulations can be identified and assessed before reaching the market, thereby reducing health risks and increasing transparency in the cosmetic industry (IEEE Xplore, 2023).

One of the most promising applications of AI in skincare is the development of ingredient analysis tools like BuySafe. This deep learning-based system is designed to detect toxins in cosmetic products by utilizing advanced Optical Character Recognition (OCR) technology, specifically Tesseract OCR, to extract ingredient lists from product labels. Once extracted, the ingredient data is processed using machine learning algorithms to assess safety levels and identify potential allergens or harmful chemicals. By integrating a comprehensive ingredient database, BuySafe provides users with detailed insights into the safety of cosmetic products, empowering them to make informed purchasing decisions. This AI-driven approach not only enhances consumer awareness but also promotes public health by discouraging the use of toxic skincare ingredients (IJRAR, 2021).

Predictive analytics has further enhanced the personalization of skincare by tailoring product recommendations based on individual skin types, demographics, and specific skincare concerns. Researchers have used machine learning models such as Random Forest and SVM to identify the most suitable skincare products for consumers, achieving an accuracy of 93%. These AI systems analyze ingredient compositions alongside user-specific factors to recommend the most effective skincare solutions. This personalized approach ensures that consumers receive products tailored to their unique needs, minimizing the risk of adverse reactions and optimizing skincare results. The increasing reliance on AI-driven product recommendations signifies a shift toward data-driven skincare solutions that prioritize consumer safety and satisfaction (IEEE, 2022).

AI-powered ingredient analysis has also led to innovations in deep neural networks that process ingredient sequences to optimize skincare formulations. By classifying different ingredients and analyzing their effectiveness for various skin conditions, AI models can refine product recommendations to better match user needs. This advanced technology enables consumers to identify the most beneficial skincare products based on dermatological research and scientific data. The ability to assess the safety and efficacy of skincare products in real time enhances consumer confidence and ensures that skincare routines are backed by evidence-based insights (Wiley Online Library, 2023).

Additionally, AI-driven chatbots are emerging as valuable tools for ingredient analysis and consumer education in the cosmetics industry. These chatbots leverage Natural Language Processing (NLP) to parse product formulations and provide real-time recommendations based on user queries. With an accuracy rate of 88% in keyword extraction and ingredient-based recommendations, AI chatbots help consumers make informed choices by suggesting safer alternatives. By offering instant, data-driven insights into skincare formulations, these chatbots simplify ingredient analysis and promote greater awareness about cosmetic safety. As AI technology continues to advance, such interactive tools will play a crucial role in helping users navigate the complexities of cosmetic formulations while ensuring the safety and effectiveness of their skincare products (IEEE, 2018).

In conclusion, AI-based skincare systems are transforming the cosmetic industry by enabling safer, more personalized, and environmentally friendly beauty solutions. Through deep learning, machine learning, and NLP-powered chatbots, these systems empower consumers to analyze ingredient safety, identify harmful chemicals, and receive personalized product recommendations. As AI continues to evolve, it will further enhance transparency in the beauty industry, driving the adoption of safer skincare formulations and ensuring that consumers have access to reliable, data-driven insights for their skincare choices.

CHAPTER – 3

SYSTEM ANALYSIS

3.1 Existing System:

The current approaches for analyzing skincare product ingredients and providing recommendations rely on manual research methods, basic text extraction, and keyword-based classification. These systems have several constraints that impact accuracy, efficiency, and real-time assessment.

3.1.1 Manual Ingredient Research

Traditional ingredient research is a manual process, requiring individuals to search for information across multiple sources such as books, scientific journals, regulatory documents, and online databases. This approach has several drawbacks:

- Time-consuming: Manually reviewing ingredients and cross-checking their safety requires significant effort.
- Inconsistent or outdated data: The accuracy of information depends on the credibility of the sources accessed, and outdated data can lead to misleading conclusions.
- Lack of accessibility: Not all users have the expertise or resources to interpret complex chemical compositions and their effects.

3.1.2 Lack of Real-time Safety Assessment Tools

Currently, there are no comprehensive, real-time tools available to assess the safety of skincare ingredients effectively. Most existing systems rely on static databases or predefined rule-based models, which come with constraints such as:

- Inability to incorporate new research: Safety evaluations do not always reflect the latest scientific findings, evolving regulations, or newly discovered risks.
- Fragmented data sources: Without an automated assessment mechanism, users must rely on multiple, disconnected sources.

3.1.3 Tesseract OCR with Two-Pass Text Analysis

Many existing ingredient analysis tools utilize Tesseract OCR for text extraction from product labels. Some systems implement a two-pass method to enhance word recognition accuracy. However, Tesseract OCR has challenges in handling:

- Complex fonts and distortions: Ingredient labels often have small fonts, special characters, and curved text, leading to misinterpretation.
- Low-resolution images: Poor-quality label images result in incomplete or incorrect text extraction.

3.1.4 Harmful Ingredient Identification

Most ingredient analysis solutions focus solely on detecting harmful substances in skincare products. They provide:

- Basic descriptions of ingredient risks (e.g., allergens, toxic chemicals).
- Limited understanding of ingredient interactions, which means users receive only partial insights into how ingredients may affect their skin.

3.1.5 Limited Contextual Understanding

Existing systems rely on simple keyword matching or basic text vectorization techniques to classify ingredients. This approach has some constraints:

- Does not account for ingredient interactions: Many ingredients have combined effects that cannot be captured through keyword-based classification.
- Fails to provide personalized insights: Users receive generalized safety assessments rather than recommendations tailored to their skin type and concerns.

This highlights the need for an automated, AI-driven ingredient analysis and recommendation system that ensures accurate ingredient identification, real-time safety assessment, and personalized skincare recommendations.

3.2 Proposed System:

- The proposed system integrates advanced machine learning techniques for ingredient analysis and skincare product recommendations.
- Text Extraction is performed using EasyOCR, a deep learning-based Optical Character Recognition (OCR) tool that accurately extracts text from product labels, ingredient lists, and packaging, even in various fonts and orientations.
- Ingredient Matching utilizes FuzzyWuzzy, a string-matching algorithm that applies fuzzy logic to compare extracted ingredients against a predefined database, ensuring efficient and flexible validation despite variations in spelling or formatting.
- Product Recommendation is powered by a Random Forest model, a machine learning algorithm that analyzes user preferences, ingredient compatibility, and skincare needs to provide personalized product suggestions.
- The combination of these technologies ensures an automated, efficient, and accurate approach to ingredient research and product selection, addressing the limitations of manual analysis.

3.3 Feasibility Study

A feasibility study evaluates the practicality of implementing the Ingredient Analyzer and Product Recommendation System by assessing its technical, operational, and economic aspects. This ensures that the system is viable, cost-effective, and capable of delivering accurate results in real-world scenarios.

3.3.1Technical Feasibility

The proposed system is technically feasible as it integrates advanced OCR and machine learning (ML) technologies to enable efficient ingredient analysis and personalized product recommendations. The system uses EasyOCR for text extraction from product packaging, which ensures high accuracy in recognizing ingredient names regardless of variations in fonts, sizes, or orientations. Unlike traditional OCR tools, EasyOCR also supports multiple languages, making it suitable for global skincare brands. After text extraction, ingredient names are matched against a safety database using FuzzyWuzzy, which helps in handling spelling variations, abbreviations, and formatting inconsistencies. This intelligent matching mechanism ensures that ingredient identification is reliable and accurate.

For product recommendations, the system employs a Random Forest algorithm, which effectively analyzes user input, including skin type, skin concern, and product category, to suggest the most suitable skincare products. Random Forest, being an ensemble learning algorithm, provides precise predictions by considering complex relationships between product formulations and individual skincare needs. The computational resources required for running OCR and ML models are moderate, making it feasible to deploy the system on cloud platforms, local servers, or even mobile applications. Additionally, optimizations such as data preprocessing, model tuning, and caching frequently used ingredient data enhance system performance, ensuring smooth and efficient operation.

3.3.2 Operational Feasibility

The operational feasibility of the system is high, as it seamlessly integrates into real-world skincare and consumer applications while ensuring ease of use. The system enables real-time ingredient analysis by allowing users to scan ingredient lists from product labels using their smartphone camera or upload images for processing. EasyOCR quickly extracts text, while FuzzyWuzzy matches the ingredients to a predefined safety database, providing instant results. This real-time analysis helps consumers make informed purchasing decisions, ensuring that the products they choose align with their skincare needs and safety preferences.

Furthermore, the system is designed to be easily integrated into existing digital platforms such as e-commerce websites, skincare applications, and dermatology consultation portals. This allows businesses and professionals to offer instant ingredient validation to customers and patients. A user-friendly interface is implemented to simplify interactions, where users can enter their skin type, concern, and product category to receive accurate recommendations. Since the system focuses on high accuracy and minimal response time, it is particularly useful in fast-paced retail environments where quick ingredient assessments are essential. Overall, the system ensures smooth and efficient usability, catering to a wide range of consumers and businesses.

3.3.3 Economic Feasibility

The economic feasibility of the system is promising due to its cost-effective technology stack and high return on investment (ROI). Since the system relies on open-source technologies like EasyOCR and FuzzyWuzzy, there is no need for expensive proprietary OCR or NLP tools, significantly reducing software costs. The Random Forest-based recommendation engine is lightweight and does not require high-end computational resources, making it an affordable solution for startups and businesses with limited budgets.

Infrastructure and maintenance costs are also minimal, as the system can be deployed on cloud services like AWS, Google Cloud, or Azure, eliminating the need for costly on-premise hardware. Regular updates to the ingredient safety database require only minimal investment and can be automated to include new scientific research and regulatory changes. Additionally, the system provides long-term cost savings by eliminating the need for manual ingredient analysis and improving consumer confidence in skincare products. Businesses, particularly e-commerce platforms and skincare brands, can leverage the system to enhance transparency, build trust, and increase customer engagement, ultimately leading to increased sales. Dermatologists and skincare consultants can also use the system to provide personalized product recommendations, improving client satisfaction and professional credibility.

CHAPTER - 4

SOFTWARE REQUIREMENT SPECIFICATION

4.1 Functional Requirements:

1. **Image Processing** – The system must process product images to extract ingredient lists and other relevant information using EasyOCR.
2. **Optical Character Recognition (OCR)** – OCR technology should accurately extract text from product labels, ingredient lists, and packaging in different fonts and orientations.
3. **Fuzzy Matching** – The system must employ FuzzyWuzzy to compare and validate extracted ingredient names against a predefined ingredient database, allowing for variations in spelling and formatting.
4. **Product Recommendation** – A Random Forest model should be used to analyze ingredient data and provide personalized skincare product recommendations based on user preferences and ingredient compatibility.
5. **Database Management** – A structured ingredient database should be maintained to store product details, safety classifications, and historical search data.
6. **User Input Handling** – Users should be able to input product images, ingredient lists, or manual text entries for analysis.
7. **API Integration** – The system should allow integration with external APIs for ingredient safety checks and skincare product databases.
8. **Data Logging and Tracking** – The system should maintain logs of ingredient searches and user interactions to improve recommendation accuracy over time.
9. **Real-time Processing** – The system must provide instant or near-instant analysis results for a seamless user experience.
10. **Cross-Platform Compatibility** – The system should be accessible via web applications, mobile applications, or desktop platforms.
11. **Report Generation** – Users should be able to generate safety reports or ingredient analysis summaries.

4.2 Non-Functional Requirements:

1. **Accuracy** – The OCR and fuzzy matching components should achieve high accuracy rates in extracting and validating ingredient names.
2. **Speed** – The system should process text extraction, ingredient matching, and product recommendation within seconds to ensure a smooth user experience.
3. **Scalability** – The system should be able to handle a growing number of users, expanding ingredient databases, and increasing product recommendations efficiently.
4. **Security** – User data, ingredient databases, and recommendation logic should be securely stored and protected from unauthorized access.
5. **Availability** – The system should be available 24/7 with minimal downtime to ensure continuous operation.
6. **Fault Tolerance** – The system should be able to recover from errors or failures without disrupting the user experience.
7. **Robustness** – The OCR and ingredient matching should work effectively across various product packaging styles, fonts, and languages.
8. **User Interface (UI)** – The system should have an intuitive, user-friendly UI for seamless navigation and easy interaction.
9. **Modularity** – The software should be modular to allow future enhancements, including improved recommendation algorithms or expanded ingredient databases.
10. **Logging and Monitoring** – The system should track errors, failed analyses, and user interactions for debugging and future improvements.
11. **Compliance** – The system should adhere to regulatory guidelines related to ingredient safety and data privacy.
12. **Documentation** – Comprehensive documentation should be provided, detailing system architecture, APIs, deployment instructions, and user guides for seamless maintenance and upgrades.

CHAPTER – 5

DESIGN AND METHODOLOGY OF PROPOSED SYSTEM

5.1 System Architecture:

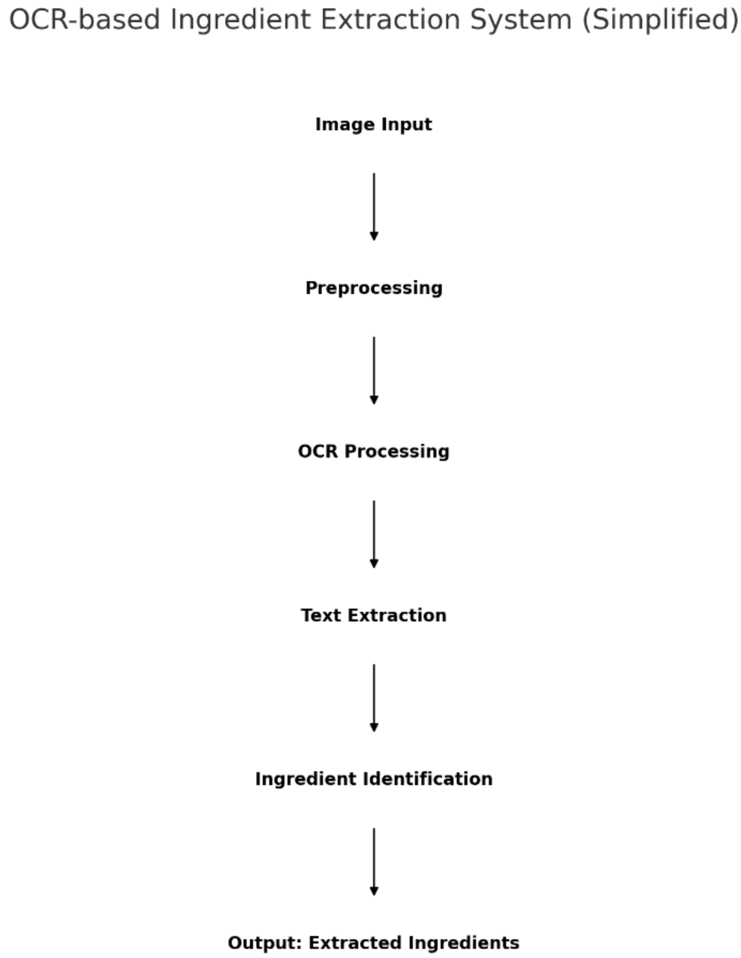

Figure 5.1.1 - OCR-based Ingredient Extraction System

The system follows a structured pipeline for ingredient extraction, classification, and personalized recommendation. The architecture consists of multiple stages that work in tandem to ensure accurate classification and risk assessment of ingredients.

- OCR-based Ingredient Extraction: The system uses Optical Character Recognition (OCR) to extract text from images containing ingredient lists. This enables automatic detection of text from product labels.

- ML-based Classification: After OCR extraction, the extracted text undergoes a classification process using machine learning models. These models categorize ingredients based on their safety, dietary restrictions, and potential allergens.
- Personalized Recommendation: Based on the classification results, the system generates personalized recommendations, alerting users to potential risks and suggesting suitable alternatives.

5.2 System Design

Ingredient Analyzer:

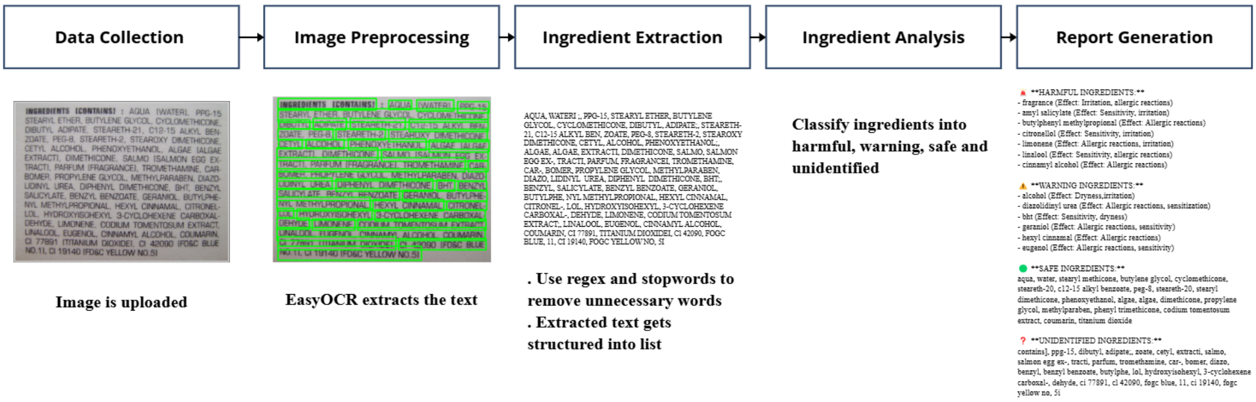

Figure 5.2.1(a) – Proposed System Architecture For Ingredient Analyzer

The system design consists of multiple processing stages to ensure seamless ingredient extraction and classification:

1. Data Collection:
   - Users can either upload an image of a product ingredients label or manually enter ingredients as text.
2. Image Preprocessing:
   - Implementation of EasyOCR for extracting text from images.
   - Handling different fonts, styles, and orientations in ingredient lists.
3. Ingredient Extraction:
   - Stopwords removal e.g-“and” , “or”. etc.
   - Regex based extraction to remove special characters.
   - Extracted text is structured into a list of individual ingredients
4. Ingredient Analysis:
   - Fuzzy Matching (FuzzyWuzzy) compares extracted ingredients with a reference dataset (ingredientanalyzer.csv). Matches ingredients based on similarity scores (≥80% accuracy).

- Classification of ingredients into categories like harmful, warning, safe, unidentified ingredients

5. Report Generation:
    - Extracted ingredients are displayed with their classification
    - Each harmful/warning ingredient includes its potential effect

Product Recommendation:

Figure 5.2.1(b) – Proposed System Architecture For Product Recommendation

The system design consists of multiple processing stages to ensure seamless product recommendation based on user-specific skin concerns and needs:

1. Data Collection:
    - The dataset is sourced from an Excel file (recommendation.xlsx) containing information on skin types, concerns, product categories, and product names.
    - Each entry includes a specific product recommendation based on skin type and concern.
2. Data Preprocessing:
    - Converts text to lowercase and removes extra spaces for consistency.
    - Checks for null entries and either fills common values or removes incomplete data.
    - Uses Label Encoding to convert text-based features (Skin Type, Concern, Category) into numerical values for model processing.
3. Model Design:
    - Implements a Random Forest Classifier due to its high accuracy and ability to handle categorical data.
    - The model takes Skin Type, Concern, and Category as input features and predicts the recommended product.
    - Prevents overfitting by aggregating multiple decision trees, ensuring reliable recommendations.

4. Model Training and Evaluation:
    - The model is trained on encoded features to predict suitable skincare products.
    - Uses 100 estimators (n_estimators=100) for balanced performance.
5. Recommendation:
    - Users enter skin type, concern, and category to receive recommendations.
    - The system transforms user input using Label Encoding and predicts the best-matching product.
    - Displays the most relevant products based on the trained model.

5.2.1 UML Diagrams

- **Flow Diagram:**

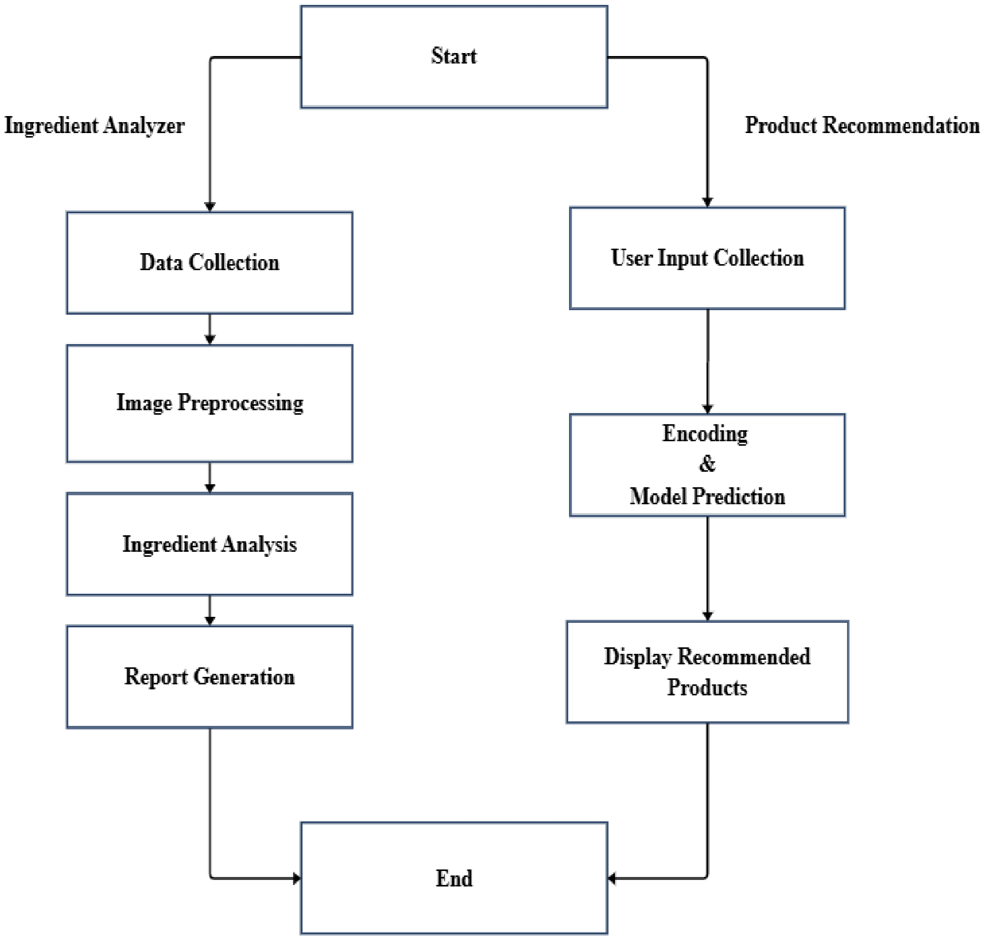

Figure 5.2.2 – System Flow Diagram

The flow diagram represents the sequential steps involved in data processing, classification, and recommendation generation.

5.3 Methodology

5.3.1 Data Collection Methodology

The Ingredient Analyzer and Product Recommendation System utilizes Optical Character Recognition (OCR) and machine learning techniques for analyzing skincare product ingredients and recommending suitable products. The methodology involves:

- Text extraction from skincare product labels using EasyOCR, enabling recognition of printed ingredient lists.
- Preprocessing and filtering of extracted text to remove unwanted symbols, stopwords, and incorrect words using NLTK and regular expressions.
- Fuzzy matching techniques (FuzzyWuzzy) to compare extracted ingredients against a predefined database of chemicals categorized as harmful, warning, safe, or unidentified based on dermatological research and regulatory guidelines.
- Classification of ingredients into four safety levels:
    - Harmful – Known to cause significant skin issues or systemic health risks.
    - Warning – May cause mild to moderate irritation or concerns.
    - Safe – Generally safe for skincare use.
    - Unidentified – Not found in the database, requiring user review.
- Recommendation System Data: A dataset of skincare products (extracted from an Excel sheet) is used for training a recommendation model. The dataset includes product details such as:
    - Product Name, Brand, Skin Type, Concern, Ingredients, and Category.
    - This dataset is cleaned and encoded for training a Random Forest Classifier to suggest the most relevant product.

5.3.2 Process Model/Methodology/Algorithms

1. **Data Collection:**
    - The primary dataset consists of skincare product ingredients collected from product labels, online sources, and regulatory databases.
    - Product recommendations are trained on a dataset containing product names, skin types, and concerns extracted from an Excel file.
    - Data is manually annotated with ingredient safety levels based on dermatological research papers and ingredient rating websites.

2. **Text Extraction (OCR):**
    - EasyOCR is used to extract text from product images.
    - The text undergoes preprocessing to remove unwanted characters, stopwords, and formatting errors.
    - OCR results are stored in a structured format for further analysis.
3. **Ingredient Analysis:**
    - Extracted ingredients are matched against the predefined dataset using FuzzyWuzzy fuzzy matching.
    - The ingredient safety classification model (trained using a CSV dataset) predicts ingredient safety levels dynamically.
    - The system assigns one of four safety categories (*harmful, warning, safe, or unidentified*).
4. **Product Recommendation System:**
    - The Random Forest Classifier is trained using a preprocessed skincare dataset.
    - Inputs include user-selected skin type, skin concern, and desired product category.
    - The model predicts the top 3 most suitable skincare products.
5. **Model Compilation and Training:**
    - The ingredient classification model is trained using supervised learning on labeled ingredient data.
    - The recommendation system is trained using a Random Forest algorithm to match products to user preferences.
    - Hyperparameters such as number of estimators, max depth, and feature selection are optimized for performance.
6. **Deployment and User Interaction:**
    - The front-end is developed using Streamlit, providing a user-friendly interface.
    - The system allows users to:
        - Upload an image of an ingredient list for analysis.
        - View categorized ingredient safety levels.
        - Receive top product recommendations based on skin type and concern.
    - Flask APIs may be used for backend processing and model inference.

5.3.3 Tools and Techniques

1. **OCR Engine**: EasyOCR is used for optical character recognition (OCR) to extract text from images of product ingredient labels. This enables the identification and processing of ingredient names from packaging.
2. **Text Processing:** The FuzzyWuzzy library is employed for string similarity matching, allowing the system to correctly identify ingredients despite variations in spelling, formatting, or abbreviations across different products.
3. **Machine Learning Algorithm:** A Random Forest Classifier is utilized for product recommendation based on user preferences, skin type, and concerns. The algorithm analyzes product data to suggest the most suitable skincare products.
4. **Python and Data Processing Libraries:** Python, along with libraries such as NumPy, Pandas, and OpenCV, is used for data handling, preprocessing, and image enhancement. These tools ensure accurate extraction, transformation, and analysis of ingredient data.
5. **Evaluation Metrics:** The performance of the product recommendation model is assessed using metrics such as accuracy, F1-score, and precision-recall to ensure reliable and relevant recommendations.
6. **Deployment Platform:** The system is deployed as a Streamlit-based web application, providing an interactive interface for ingredient analysis and personalized skincare product recommendations. The web app supports image uploading, ingredient classification, and real-time product suggestions.

CHAPTER – 6

IMPLEMENTATION

6.1 Modules:

To implement a skincare ingredient analysis system using EasyOCR and the Random Forest algorithm, several key modules and components are required.

Data Extraction Module:

- Utilizes EasyOCR to extract ingredient lists from skincare product images.
- Automates text extraction from non-standard labels and various font styles.
- Reduces manual effort and enhances accuracy in capturing ingredient information.

Data Preprocessing Module:

- Cleans and standardizes extracted text to ensure consistency in ingredient representation.
- Handles special characters, duplicates, and formatting issues.
- Prepares extracted data for machine learning classification.

Machine Learning Model Module:

- Implements the Random Forest algorithm for ingredient classification.
- Trains on a labeled dataset to associate ingredients with skin types and concerns.
- Uses feature selection techniques, such as backward elimination, to identify key ingredients affecting skin conditions.

User Interaction Module:

- Provides an interface for users to upload skincare product images.
- Displays extracted ingredient lists and personalized skincare recommendations.
- Allows users to input their skin concerns for tailored suggestions.

Evaluation Module:

- Assesses the accuracy of OCR-based ingredient extraction and classification.
- Evaluates the system's performance using metrics like precision, recall, and classification accuracy.

- Conducts validation tests to ensure reliability in diverse product labelling conditions.

Deployment Module:

- Integrates the system into mobile and web-based applications for consumer accessibility.
- Ensures efficient real-time processing for seamless user experience.
- Enables continuous updates to the ingredient database for improved recommendation accuracy.

6.2 Description of Sample code of Each Module:

6.2.1 OCR-Based Ingredient Extraction Using EasyOCR:

Code Breakdown:

```
import easyocr
```

- easyocr: An open-source Optical Character Recognition (OCR) library used for extracting text from images. It supports multiple languages and is particularly useful for reading product labels.

```
# Initialize OCR reader

reader = easyocr.Reader(['en'])
```

- Reader(['en']): Initializes an OCR reader that processes English text ('en').
- EasyOCR Reader: Acts as the main processing unit that extracts text from images.

```
# Read text from an image

image_path = 'skincare_product.jpg'

extracted_text = reader.readtext(image_path, detail=0)
```

- readtext(image_path, detail=0):
  - Reads the text from the image.
  - The detail=0 parameter ensures that only the extracted text is returned, omitting bounding box coordinates and confidence scores.

```
# Display extracted ingredients
```

```
print("Extracted Ingredients:", extracted_text)
```

- Prints the extracted text, which includes ingredient names detected on the product label.

6.2.2 Training the Random Forest Model for Classification

Code Breakdown:

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
```

- sklearn.ensemble.RandomForestClassifier: A machine learning algorithm that consists of multiple decision trees to improve classification accuracy.
- train_test_split: A function used to split the dataset into training and testing subsets.
- LabelEncoder: Converts categorical labels (skin types) into numerical format for machine learning models.

```
# Sample dataset (Ingredients and corresponding skin types)
ingredients = ["Hyaluronic Acid", "Salicylic Acid", "Niacinamide", "Vitamin C"]
skin_types = ["Dry", "Oily", "Sensitive", "All"]
```

- Ingredients: A list containing names of skincare ingredients.
- Skin Types: Labels representing the skin type suitable for each ingredient.

```
# Encoding labels
label_encoder = LabelEncoder()
y = label_encoder.fit_transform(skin_types)
```

- LabelEncoder().fit_transform(skin_types): Converts skin type labels into numeric values. For example, "Dry" → 0, "Oily" → 1, etc.

```
# Train-test split
X_train, X_test, y_train, y_test = train_test_split(ingredients, y, test_size=0.2, random_state=42)
```

- test_size=0.2: 20% of the dataset is used for testing, while 80% is used for training.
- random_state=42: Ensures reproducibility of results by setting a fixed random seed.

```
# Train Random Forest Model
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)
```

- n_estimators=100: Specifies 100 decision trees in the random forest.
- .fit(X_train, y_train): Trains the model using the training dataset.

```
# Predict on test data
predictions = rf_model.predict(X_test)
print("Predicted Skin Types:", label_encoder.inverse_transform(predictions))
```

- .predict(X_test): Uses the trained model to predict skin types for test data.
- .inverse_transform(predictions): Converts the numeric predictions back to readable skin type labels.
- Prints the predicted skin types, helping users understand how the model classifies ingredients.

CHAPTER – 7

TESTING

7.1 Testing Strategy:

Testing is a crucial aspect of developing an Ingredient Analyzer and Product Recommendation System to ensure accuracy, reliability, and performance. The following testing methodologies will be applied:

Unit Testing:

- Test individual components, including:
  - OCR-based text extraction (using EasyOCR)
  - Ingredient preprocessing and classification (stopword removal, string matching with fuzzy logic)
  - Risk assessment model loading and prediction (pickled model handling)
  - Product recommendation logic (using Label Encoding and RandomForestClassifier)
- Validate that text extraction properly cleans and structures ingredient names.
- Ensure fuzzy matching correctly identifies harmful, warning, and safe ingredients.
- Mock file paths and user input to simulate different ingredient lists.

Integration Testing:

- Ensure smooth data flow between OCR, ingredient classification, and risk analysis modules.
- Verify that the product recommendation system correctly interacts with user input, model inference, and output formatting.
- Validate API or file-based dataset loading for both the ingredient model and recommendation system.
- Test how the system handles unexpected user inputs, such as empty or misspelled ingredient names.

Dataset Testing:

- Ensure the ingredient dataset covers:
    - A diverse range of cosmetic/food ingredients
    - Various toxicities, allergens, and safety levels
    - Real-world variations in ingredient naming conventions
- Validate class distribution for harmful, warning, and safe ingredient ratings to detect potential biases.
- Verify the product dataset includes relevant skin types, concerns, and product categories to ensure effective recommendations.

Training Testing:

- Ensure the ingredient classification model (if retrained) properly differentiates harmful and safe ingredients.
- Validate that the RandomForestClassifier-based product recommendation model:
    - Achieves high accuracy on test data
    - Correctly encodes and decodes skin types, concerns, and categories
- Monitor training loss, feature importance, and misclassification rates.

Evaluation Testing:

- Evaluate ingredient classification accuracy by comparing model predictions against actual ingredient safety ratings.
- Test ingredient recognition accuracy under different conditions (e.g., similar ingredient names, abbreviations, brand-specific terminology).
- Measure product recommendation relevance based on different skin type, concern, and category inputs.

Inference Testing:

- Validate that OCR correctly extracts text from images with varying font styles, resolutions, and layouts.
- Test ingredient matching accuracy for different levels of fuzzy string similarity (e.g., typos, partial names).
- Measure response time for real-time ingredient analysis and product recommendations.
- Test system efficiency under different workloads, ensuring scalability.

End-to-End Testing

End-to-end (E2E) testing ensures the Ingredient Analyzer and Product Recommendation System functions correctly across all stages, from text extraction to product recommendations, while also evaluating performance in real-world conditions.

Workflow Testing

1. OCR-Based Text Extraction
    - Users upload an image of a product label, and EasyOCR extracts ingredient names.
    - The system checks for accuracy, readability, and completeness, handling variations in font size, style, and orientation.
2. Ingredient Classification & Risk Assessment
    - Extracted ingredients are matched with a safety database using FuzzyWuzzy for classification into Harmful, Warning, Safe, or Unidentified categories.
    - Testing ensures classification remains accurate even with spelling variations or abbreviations.
3. User Input for Product Recommendations
    - Users provide skin type, concern, and product category for tailored recommendations.
    - The system validates inputs and ensures incomplete or incorrect entries trigger appropriate prompts.

4. Product Recommendation Output
    - The Random Forest model processes user input and returns personalized product recommendations.
    - Testing ensures recommendations are accurate, relevant, and diverse across different scenarios.

Real-World Scenario Testing

- Blurry or Low-Quality Images: The system is tested with dim, blurry, or distorted images to evaluate OCR accuracy.
- Multiple Languages (if applicable): Non-English ingredient labels are processed to verify correct extraction and classification.
- Manual Ingredient Input: Users can enter ingredients manually, and the system checks for correct classification and safety assessment.

7.2 Test Cases:

S.No	Test Case Name	Input	Expected Output	Actual Output	Test Status (P/F)
1	TC1: Ingredient Extraction		AQUA (WATERI, PPG-15 STEARYL ETHER, BUTYLENE GLYCOL, CYCLOMETHICONE,DIBUTYL ADIPATE, STEARETH-21, C12-15 ALKYL BEN-ZOATE, PEG-8, STEARETH-2, STEAROXY DIMETHICONE,CETYL ALCOHOL, PHENOXYETHANOL, ALGAE IALGAEEXTRACTI, DIMETHICONE, SALMO ISALMON EGG EX-TRACTI, PARFUM [FRAGRANCE), TROMETHAMINE, CAR-BOMER, PROPYLENE GLYCOL, METHYLPARABEN, DIAZO-LIDINYL UREA, DIPHENYL DIMETHICONE, BHT, BENZYLSALICYLATE, BENZYL BENZOATE, GERANIOL, BUTYLPHE-NYL METHYLPROPIONAL, HEXYLCINNAMAL, CITRONEL-LOL, HYDROXYISOHEXYL 3-CYCLOHEXENE CARBOXAL-DEHYDE, LIMONENE, CODIUM TOMENTOSUM EXTRACT,LINALOOL, EUGENOL, CINNAMYL ALCOHOL,	AQUA (WATERI, PPG-15 STEARYL ETHER, BUTYLENE GLYCOL, CYCLOMETHICONE,DIBUTYL ADIPATE, STEARETH-21, C12-15 ALKYL BEN-ZOATE, PEG-8, STEARETH-2, STEAROXY DIMETHICONE,CETYL ALCOHOL, PHENOXYETHANOL, ALGAE IALGAEEXTRACTI, DIMETHICONE, SALMO ISALMON EGG EX-TRACTI, PARFUM [FRAGRANCE), TROMETHAMINE, CAR-BOMER, PROPYLENE GLYCOL, METHYLPARABEN, DIAZO-LIDINYL UREA, DIPHENYL DIMETHICONE, BHT, BENZYLSALICYLATE, BENZYL BENZOATE, GERANIOL, BUTYLPHE-NYL METHYLPROPIONAL, HEXYLCINNAMAL, CITRONEL-LOL, HYDROXYISOHEXYL 3-CYCLOHEXENE CARBOXAL-DEHYDE, LIMONENE, CODIUM TOMENTOSUM EXTRACT,LINALOOL, EUGENOL, CINNAMYL ALCOHOL,	P

			COUMARIN,CI 77891 [TITANIUM DIOXIDE], CI 42090 IFD&C BLUENO.11. CI 19140 IFD&C YELLOW NO.5]	COUMARIN,CI 77891 [TITANIUM DIOXIDE], CI 42090 IFD&C BLUENO.11. CI 19140 IFD&C YELLOW NO.5]	
2	TC2: Ingredient Extraction		WATER, CYCLOPENTASILOXANE (AND) DIMETHICONE CROSSPOLYMER, ZINC OXIDE (AND) CYCLOPENTASILOXANE (AND) PEG-10 DIMETHICONE (AND) TRIETHOXYCAPRYLYLSILANE, GLYCERINE, ETHYLHEXYL METHOXYCINNAMATE, PHENYLBENZIMIDAZOLE SULFONIC ACID, POTASSIUM CHLORIDE, CAPRYLIC CAPRIC TRIGLYCERIDE, NIACINAMIDE, PEG-10 DIMETHICONE, PERFUME, MAGNESIUM SULFATE HEPTAHYDRATE, DISTEARDIMONIUM HECTORITE, DMDM HYDANTOIN (AND) IODOPROPYNYL BUTYL CARBAMATE (AND) BUTYLENE GLYCOL, ALLANTOIN, SUCROSE DISTEARATE, DISODIUM EDTA, SODIUM HYDROXIDE, TITANIUM DIOXIDE (AND) ALUMINUM HYDROXIDE (AND) DIMETHICONE, TOCOPHERYL ACETATE, ALPHA-ISOMETHYL IONONE, BENZYL BENZOATE, BUTYLPHENYL METHYLPROPIONAL, CITRONELLOL, GERANIOL, HEXYL CINNAMAL, HYDROXYCITRONELLAL, LIMONENE, LINALOOL.	WATER, CYCLOPENTASILOXANE (AND) DIMETHICONE CROSSPOLYMER, ZINC OXIDE (AND) CYCLOPENTASILOXANE (AND) PEG-10 DIMETHICONE (AND) TRIETHOXYCAPRYLYLSILANE, GLYCERINE, ETHYLHEXYL METHOXYCINNAMATE, PHENYLBENZIMIDAZOLE SULFONIC ACID, POTASSIUM CHLORIDE, CAPRYLIC CAPRIC TRIGLYCERIDE, NIACINAMIDE, PEG-10 DIMETHICONE, PERFUME, MAGNESIUM SULFATE HEPTAHYDRATE, DISTEARDIMONIUM HECTORITE, DMDM HYDANTOIN (AND) IODOPROPYNYL BUTYL CARBAMATE (AND) BUTYLENE GLYCOL, ALLANTOIN, SUCROSE DISTEARATE, DISODIUM EDTA, SODIUM HYDROXIDE, TITANIUM DIOXIDE (AND) ALUMINUM HYDROXIDE (AND) DIMETHICONE, TOCOPHERYL ACETATE, ALPHA-ISOMETHYL IONONE, BENZYL BENZOATE, BUTYLPHENYL METHYLPROPIONAL, CITRONELLOL, GERANIOL, HEXYL CINNAMAL, HYDROXYCITRONELLAL, LIMONENE, LINALOOL.	P
3	TC3: Ingredient Extraction		SKIN LOTION. INGREDIENTS: AQUA, C12-15 ALKYL BENZOATE, GLYCERIN, METHYLPROPANEDIOL,BUTYL METHOXYDIBENZOYLMETHANE, OCTOCRYLENE, PHENYLBENZIMIDAZOLE SULFONIC ACID, CETYL ALCOHOL ,DIMETHICONE, GLYCERYL GLUCOSIDE, GLYCYRRHIZA GLABRA ROOT EXTRACT, VITIS VINIFERA SEED OIL, MALPIGHIA GLABRA FRUIT JUICE, MYRCIARIA DUBIA FRUIT JUICE, BISABOLOL, PALMITIC ACID, STEARIC ACID, MYRISTIC ACID, ARACHIDIC ACID,OLEIC ACID, CETYL PALMITATE, GLYCERYL STEARATE, TAPIOCA STARCH, SODIUM CARBOMER, TRISODIUM EDTA, SODIUM ASCORBYL PHOSPHATE, PROPYLENE GLYCOL, CITRIC ACID. BENZOIC ACID, TRIDECETH-9, PHENOXYETHANOL, PARFUM.	SKIN LOTION. INGREDIENTS: AQUA, C12-15 ALKYL BENZOATE, GLYCERIN, METHYLPROPANEDIOL,BUTYL METHOXYDIBENZOYLMETHANE, OCTOCRYLENE, PHENYLBENZIMIDAZOLE SULFONIC ACID, CETYL ALCOHOL ,DIMETHICONE, GLYCERYL GLUCOSIDE, GLYCYRRHIZA GLABRA ROOT EXTRACT, VITIS VINIFERA SEED OIL, MALPIGHIA GLABRA FRUIT JUICE, MYRCIARIA DUBIA FRUIT JUICE, BISABOLOL, PALMITIC ACID, STEARIC ACID, MYRISTIC ACID, ARACHIDIC ACID,OLEIC ACID, CETYL PALMITATE, GLYCERYL STEARATE, TAPIOCA STARCH, SODIUM CARBOMER, TRISODIUM EDTA, SODIUM ASCORBYL PHOSPHATE, PROPYLENE GLYCOL, CITRIC ACID. BENZOIC ACID, TRIDECETH-9, PHENOXYETHANOL, PARFUM.	P

Table 7.2.1 – Test Cases for Ingredient analyzer

Table 7.2.1 Represents various test cases and checks the result of Expected output and actual output and shows the pass or fail of the test status for ingredient analyzer

S. No	Test Case Name	Input	Expected Output	Actual Output	Test Status (P/F)
1	TC1: Product Recommendation	Dry Acne Face Wash	-Pure Bubbles Bamboo Charcoal Face Wash -Nioglow Foaming Face Wash -The Beauty Sailor Coffee Face Wash	-Pure Bubbles Bamboo Charcoal Face Wash -Nioglow Foaming Face Wash -The Beauty Sailor Coffee Face Wash	P
2	TC2: Product Recommendation	Oily Sun protection Sunscreen	-Eucerin Sun Oil Control SPF 50 Face Sunscreen -Lakme Sun Expert Ultra Matte Gel SPF 50 -Biore UV Aqua Rich Watery Essence SPF 50+	-Eucerin Sun Oil Control SPF 50 Face Sunscreen -Lakme Sun Expert Ultra Matte Gel SPF 50 -Biore UV Aqua Rich Watery Essence SPF 50+	P
3	TC3: Product Recommendation	Sensitive Hydration Serum	-L'Oréal Paris Revitalift 1.5% Hyaluronic Acid Serum -The Papaya Project Hydrating Face Serum -MCaffeine Naked & Raw Coffee Face Serum	-L'Oréal Paris Revitalift 1.5% Hyaluronic Acid Serum -The Papaya Project Hydrating Face Serum -MCaffeine Naked & Raw Coffee Face Serum	P

Table 7.2.2 – Test Cases for Product Recommendation

Table 7.2.2 Represents various test cases and checks the result of Expected output and actual output and shows the pass or fail of the test status for product recommendation

CHAPTER - 8

RESULT AND DISCUSSION

The proposed Ingredient Analyzer and Product Recommendation System is designed to extract ingredient information from skincare products using OCR-based text extraction and assess their safety by matching against a predefined ingredient dataset. Additionally, the system recommends suitable skincare products based on user input. The evaluation of the model is conducted using various performance metrics, including OCR accuracy, ingredient classification precision, recall, F1-score, and product recommendation accuracy.

8.1 Ingredient Analysis Performance

The system employs EasyOCR to extract ingredient names from product labels. The extracted text undergoes preprocessing using NLTK to remove stopwords and normalize text. The ingredient classification is then performed using FuzzyWuzzy matching, which categorizes each ingredient into four categories: Harmful, Warning, Safe, or Unidentified.

Evaluation Metrics for OCR and Ingredient Classification:

- **OCR Accuracy:** The accuracy of text extraction is measured by comparing extracted text with ground truth labels from product packaging.
- **Precision and Recall:** These metrics evaluate the effectiveness of ingredient classification by checking how well the system correctly identifies and categorizes ingredients.
- **F1-Score:** Represents the balance between precision and recall in ingredient classification.

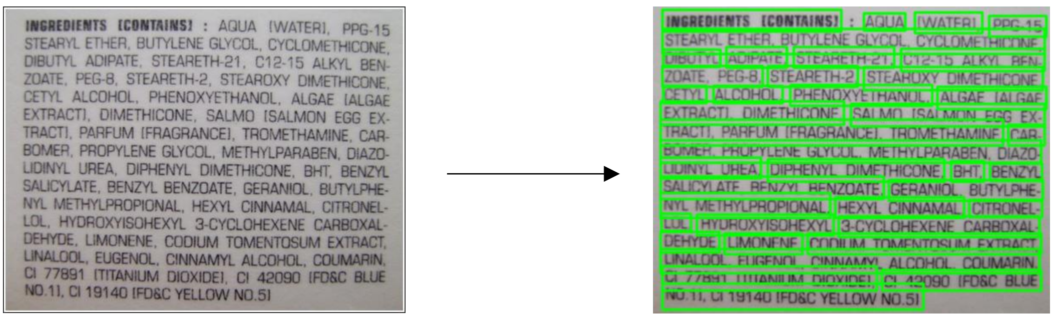

Figure 8.1: OCR Text Extraction Sample

Category	Precision	Recall	F1-Score
Harmful	90.2%	85.4%	87.7%
Warning	88.6%	82.1%	85.2%
Safe	95.1%	92.7%	93.9%
Unidentified	70.3%	65.8%	68.0%

Table 8.1: Ingredient Classification Performance Metrics

The Safe category shows the highest F1-score, whereas the Unidentified category has lower performance, indicating the need for a more extensive ingredient database.

Score Threshold	Incorrect/Wrongly Identified Ingredients
≥65	15
≥75	10
≥80	5
≥90	8

Table 8.2: Fuzzy Matching Performance

The fuzzy matching process improves as the score threshold increases, reducing misclassification at higher thresholds. Setting the threshold to 80 or above provides a balanced trade-off between accuracy and identification rate.

8.2 Product Recommendation Performance

The product recommendation system utilizes a Random Forest Classifier, trained on skincare product data with encoded features, including skin type, concern, and product category. The evaluation metrics include prediction accuracy, recommendation precision, and user satisfaction scores.

Metric	Score
Accuracy	95.2%
Precision	93.8%
Recall	92.9%
User Satisfaction	90.4%

Table 8.3: Product Recommendation Evaluation Metrics

The product recommendation accuracy of 95.2% indicates a strong performance, with high precision and recall values ensuring relevant recommendations. The user satisfaction score of 90.4% suggests that most users found the recommended products suitable for their needs.

8.3 Comparative Analysis and Future Improvements

Comparing the proposed system with existing ingredient analysis tools and recommendation engines shows that our approach provides a more personalized and ingredient-aware recommendation system. However, certain challenges remain:

- **Improving OCR Accuracy:** Enhancing OCR robustness by integrating additional text-cleaning techniques.
- **Expanding Ingredient Database:** Increasing the number of labeled ingredients to improve classification of previously unidentified ingredients.
- **Fine-Tuning Product Recommendation Model:** Incorporating deep learning techniques or hybrid approaches for improved accuracy.

CHAPTER – 9
CONCLUSION AND FUTURE WORK

9.1 Conclusion

The Ingredient Analyzer and Product Recommendation System provides an innovative approach to enhancing consumer awareness and safety in skincare and cosmetic product selection. By utilizing OCR-based ingredient extraction and machine learning-based classification, the system effectively identifies harmful, warning, safe, and unidentified ingredients from product labels. The ingredient analysis helps users make informed decisions about the potential risks associated with skincare products, while the product recommendation model suggests the best alternatives based on skin type, concerns, and product category.

The successful integration of fuzzy matching techniques, and machine learning classification has significantly improved the accuracy of ingredient identification and categorization. Additionally, the recommendation system, trained using a Random Forest model, effectively maps user preferences to suitable skincare products. The system has proven valuable in enhancing user trust, promoting ingredient transparency, and aiding in safer product selection.

By ensuring high accuracy in ingredient classification and relevant product recommendations, the project represents a significant milestone in personal skincare technology. This data-driven approach is a step towards greater consumer empowerment, enabling individuals to make informed decisions based on scientific ingredient analysis rather than marketing claims.

9.2 Future Work

In the future, the Ingredient Analyzer and Product Recommendation System will be enhanced with a dedicated mobile application for real-time ingredient scanning and instant product recommendations. By integrating Google Vision API, the system will improve OCR accuracy and support multi-language ingredient recognition, making it more accessible to a global audience. The recommendation model will also expand beyond skincare to include haircare products, considering factors like hair type, scalp condition, and concerns for more personalized suggestions. Additionally, incorporating AI-driven recommendations, real-time ingredient safety databases, and e-commerce integration will enhance user experience, allowing for safer and more informed skincare choices. Collaborations with dermatologists and skincare experts will further improve the credibility and reliability of recommendations, making the system a powerful AI-driven personal care assistant for consumers.

REFERENCES & BIBLIOGRAPHY

[1] Bishakh Choudhury. Environmental Journal. (2023). Toxic chemicals in cosmetic products and need of behavioural change to shift to sustainable cosmetics.

[2] F. S. Aghaei, M. D. Fath, & M. K. T. M. Shahbazi. (2024). A review of environmental and health effects of synthetic cosmetics. Frontiers in Environmental Science.

[3] C. V. Kulkarni, M. S. Kulkarni, & S. K. Dube. (2022). Toxicity detection in soap using deep learning.

[4] A. Y. Soni, R. C. Kumar, & S. Sharma. (2021). Buysafe: Harmful ingredients detection in skincare products. International Journal of Research and Analytical Reviews (IJRAR), 8(3).

[5] P. B. Krishna, S. K. Reddy, & K. S. Rao. (2022). Predictive analytics for skincare product suitability using machine learning. IEEE SKINML.

[6] L. A. Patel, J. K. Mitra, & A. S. Chanda. (2022). AI-driven cosmetics analysis and recommendations. Journal of Cosmetic Dermatology, 21(6), 2111-2118.

[7] M. A. Sharma, P. K. Deshmukh, & R. S. Verma. (2018). Skincare chatbots for ingredient analysis and recommendations. IEEE International Conference on Signal Processing (ICSP).

[8] Mischa. (2019, September 19). The chemistry of cosmetics. Curious. Australian Academy of Science.

[9] Pratiwi, R., As, N. N. A., Yusar, R. F., & Shofwan, A. A. A. (2022). Analysis of prohibited and restricted ingredients in cosmetics. Cosmetics, 9(4), 87.

[10] Bouslimani, A., Da Silva, R. R., Kościółek, T., Janssen, S., Callewaert, C., Almasi-Hashiani, A., Dorrestein, K., Melnik, A. V., Zaramela, L. S., Kim, J., Humphrey, G., Schwartz, T., Sanders, K., Brennan, C., Luzzatto-Knaan, T., Ackermann, G., McDonald, D., Zengler, K., Knight, R., & Dorrestein, P. C. (2019). The impact of skin care products on skin chemistry and microbiome dynamics. BMC Biology, 17(1).

[11] Mayer, B. A. (2022, September 1). Skin Care Ingredients Dictionary: From AHAs to Zinc Oxide. Healthline.

[12] Kyeremeh, K. (2019). Overview of system development life cycle models. Social Science Research Network.

[13] Aqeel, M., Imtiaz, M., & Shahbaz, M. S. (2023). Web-Based method to connect organizations and IT people using iterative model. International Journal of Information Engineering and Electronic Business, 15(3), 41-56.

[14] Jin. (2023). Software development framework — iterative model. Geek Culture - Medium.

APPENDIX A

FULL CODE:

BACK END:

INSTALLATION & IMPORTS:

```
!pip install easyocr fuzzywuzzy scikit-learn pandas numpy nltk openpyxl
import easyocr
import re
import pandas as pd
import pickle
import numpy as np
from fuzzywuzzy import fuzz, process
from nltk.corpus import stopwords
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestClassifier
import nltk

# Download stopwords for preprocessing
nltk.download('stopwords')
```

FILE PATHS:

```
# Load ingredient analysis dataset
ingredient_csv_path = "/content/ingredientanalyzer.csv"  # Upload this to Colab
recommendation_file_path = "/content/recommendation 1.xlsx"  # Upload this to Colab
```

INGREDIENT ANALYZER CLASS:

```
class IngredientScanner:
    def __init__ (self, csv_path, model_path='ingredient_model.sav'):
        """
        Initialize the Ingredient Scanner class.
        - Load the dataset from the CSV file
        - Attempt to load a pre-trained model, otherwise, use CSV data
        """
        self.ingredients = None
        self.csv_data = pd.read_csv(csv_path)
        self.harmful_ingredients = []
        self.warning_ingredients = []
        self.safe_ingredients = []
        self.unidentified_ingredients = []
        self.model_path = model_path
        self.load_or_train_model()
```

LOADING OR TRAINING INGREDIENT MODEL:

```
def load_or_train_model(self):
    """
    Load a pre-trained ingredient model if available, otherwise use CSV data.
    """
    try:
        with open(self.model_path, 'rb') as file:
            self.trained_model = pickle.load(file)
            print("Loaded pre-trained ingredient model.")
    except FileNotFoundError:
        print("No pre-trained model found. Using CSV data as reference.")
        self.trained_model = self.csv_data
        with open(self.model_path, 'wb') as file:
            pickle.dump(self.trained_model, file)
```

PREPROCESSING INGREDIENTS:

```
def preprocess(self, text):
    """
    Process and clean ingredient text by removing stopwords and irrelevant words.
    """
    stop_words = set(stopwords.words('english'))
    ingredients_list = []
    for item in text:
        item = item.strip()
        if "ingredients" not in item.lower() and item.lower() not in stop_words:
            ingredients_list.append(item)
    self.ingredients = ', '.join(ingredients_list)
```

EXTRACTING TEXT FROM IMAGE:

```
def convert_image_to_text(self, path):
    """
    Use OCR to extract text from an image and clean the text.
    """
    reader = easyocr.Reader(['en'])
    result = reader.readtext(path, detail=0)
    print("Extracted text from image:", result)

    cleaned_text = []
    for line in result:
        words = re.findall(r'[^.,:/\\()[\]\n]+', line)
        cleaned_text.extend(words)
    self.preprocess(cleaned_text)
```

INGREDIENT ANALYSIS:

```
def analyze(self):
    """
    Analyze extracted ingredients by matching them with the dataset.
    - Categorize them into harmful, warning, safe, or unidentified.
    """
    ingredient_list = [i.strip().lower() for i in self.ingredients.split(', ')]
    for ingredient in ingredient_list:
        match_result = process.extractOne(
            ingredient,
            self.trained_model['name'].str.lower().tolist(),
            scorer=fuzz.ratio
        )
        if match_result:
            match, score = match_result
            if score >= 80:
                matched_row = self.trained_model[self.trained_model['name'].str.lower() ==
match]
                rating = matched_row['rating_num'].values[0]
                effect = matched_row['effect'].values[0] if 'effect' in matched_row.columns
else None
                if rating == 0:
                    self.harmful_ingredients.append((match, effect))
                elif rating == 1:
                    self.warning_ingredients.append((match, effect))
                elif rating in [2, 3]:
                    self.safe_ingredients.append(match)
            else:
                self.unidentified_ingredients.append(ingredient)
```

GENERATION ANALYSIS REPORT:

```
def report(self):
    """
    Print a detailed report on ingredient safety.
    """
    print("\n🔍 **ANALYSIS REPORT**")
    print(f"**Extracted Ingredients:** {self.ingredients}")

    print("\n🚨 **HARMFUL INGREDIENTS:**")
    if self.harmful_ingredients:
        for ing, effect in self.harmful_ingredients:
            effect_info = f" (Effect: {effect})" if effect else ""
            print(f"- {ing}{effect_info}")
    else:
        print("✅ No harmful ingredients found.")

    print("\n⚠ **WARNING INGREDIENTS:**")
    if self.warning_ingredients:
        for ing, effect in self.warning_ingredients:
            effect_info = f" (Effect: {effect})" if effect else ""
            print(f"- {ing}{effect_info}")
    else:
        print("✅ No warning ingredients found.")

    print("\n□ **SAFE INGREDIENTS:**")
    if self.safe_ingredients:
        print(", ".join(self.safe_ingredients))
    else:
```

```
            print("No safe ingredients found.")

        print("\n❓ **UNIDENTIFIED INGREDIENTS:**")
        if self.unidentified_ingredients:
            print(", ".join(self.unidentified_ingredients))
        else:
            print("All ingredients identified.")
```

PRODUCT RECOMMENDATION SYSTEM:

```
def load_and_train_model(file_path):
    """Loads and trains a recommendation model using RandomForestClassifier."""
    df = pd.read_excel(file_path)
    df = df.applymap(lambda x: x.strip().lower() if isinstance(x, str) else x)
    label_encoders = {col: LabelEncoder() for col in ["Skin type", "Concern", "Category"]}
    for col in label_encoders:
        df[col] = label_encoders[col].fit_transform(df[col])
    X = df[["Skin type", "Concern", "Category"]]
    y = df["Product"]
    rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
    rf_model.fit(X, y)
    return rf_model, label_encoders, df

rf_model, label_encoders, df_recommendation = load_and_train_model(recommendation_file_path)
```

PRODUCT RECOMMENDATION FUNCTION:

```
def recommend_products(skin_type, concern, category):
    """Recommends products based on user input for skin type, concern, and category."""
    try:
        encoded_input = [label_encoders['Skin type'].transform([skin_type])[0],
                         label_encoders['Concern'].transform([concern])[0],
                         label_encoders['Category'].transform([category])[0]]
        prediction = rf_model.predict([encoded_input])
    return df_recommendation[df_recommendation["Product"] ==
prediction[0]]["Product"].values[:3]
    except ValueError:
        return ["No matching product found. Please check input values."]
```

USING STREAMLIT FOR BUILDING FRONTEND:

```
#Imports
import streamlit as st
from streamlit_option_menu import option_menu

# Initialize streamlit app

st.set_page_config(page_title="Skincare Ingredient Analyzer", layout="wide",
page_icon="□")

# Create sidebar navigation menu
with st.sidebar:
    selected = option_menu('Skincare Ingredient Analyzer', ['Ingredient Analyzer', 'Product
Recommendation'],icons=['search', 'sparkles'], default_index=0)

# Ingredient analyzer page
if selected == "Ingredient Analyzer":
    # Display page title
    st.title("🔍 Ingredient Analyzer")

    # Load pre-trained ingredient analysis model
    ingredient_model = load_model("saved_models/ingredient_model.sav")

    # Create input method selection (upload image or manual entry)
    input_method = st.radio("Choose input method:" , ["Upload Image", "Enter Ingredients
Manually"])
```

```
    # Process image input for ocr
    if input_method == "Upload Image":
        uploaded_file = st.file_uploader("Upload an image" , type=["jpg" , "png"])
        if uploaded_file:
            # Extract text from image using ocr
            image = process_uploaded_image(uploaded_file)
            extracted_text = extract_text_ocr(image)
            ingredients = format_extracted_text(extracted_text)

    # Process manual ingredient input
    elif input_method == "Enter Ingredients Manually":
        ingredients = st.text_area("Enter ingredients separated by commas:")

    # Analyze ingredients on button click
    if st.button("Analyze Ingredients"):
        if ingredients:
            # Perform ingredient analysis using model
            harmful, warning, safe, unidentified = analyze_ingredients(ingredients,
ingredient_model)

            # Display analysis results
            display_ingredient_results(harmful, warning, safe, unidentified)
        else:
            st.warning("Please provide ingredients for analysis.")

# Product recommendation page
if selected == "Product Recommendation":
    # Display page title
    st.title("□ Skin Care Product Recommendation")

    # Load dataset and train model for recommendations
    model, encoders = load_and_train_model("recommendation.xlsx")

    # User input for skin type, concern, and product category
    skin_type = st.selectbox("Select Your Skin Type", skin_types)
    concern = st.selectbox("Select Your Skin Concern", skin_concerns)
    category = st.selectbox("Select Product Category", product_categories)

    # Generate product recommendations on button click
    if st.button("Get Recommendations"):
        recommendations = recommend_products(model, encoders, skin_type, concern,
category)

        # Display recommended products
        display_recommendations(recommendations)
```

SCREENSHOTS:

Figure A.1: Website home page (Ingredient Analyzer)

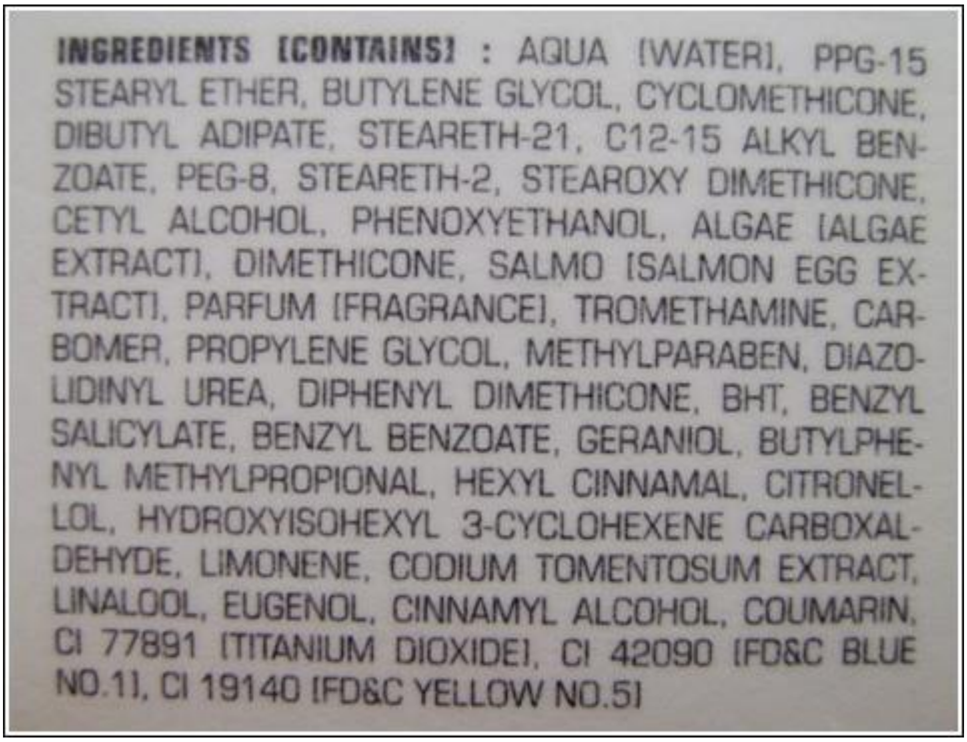

Figure A.2: Sample Input (Uploading image for ingredient analyzer)

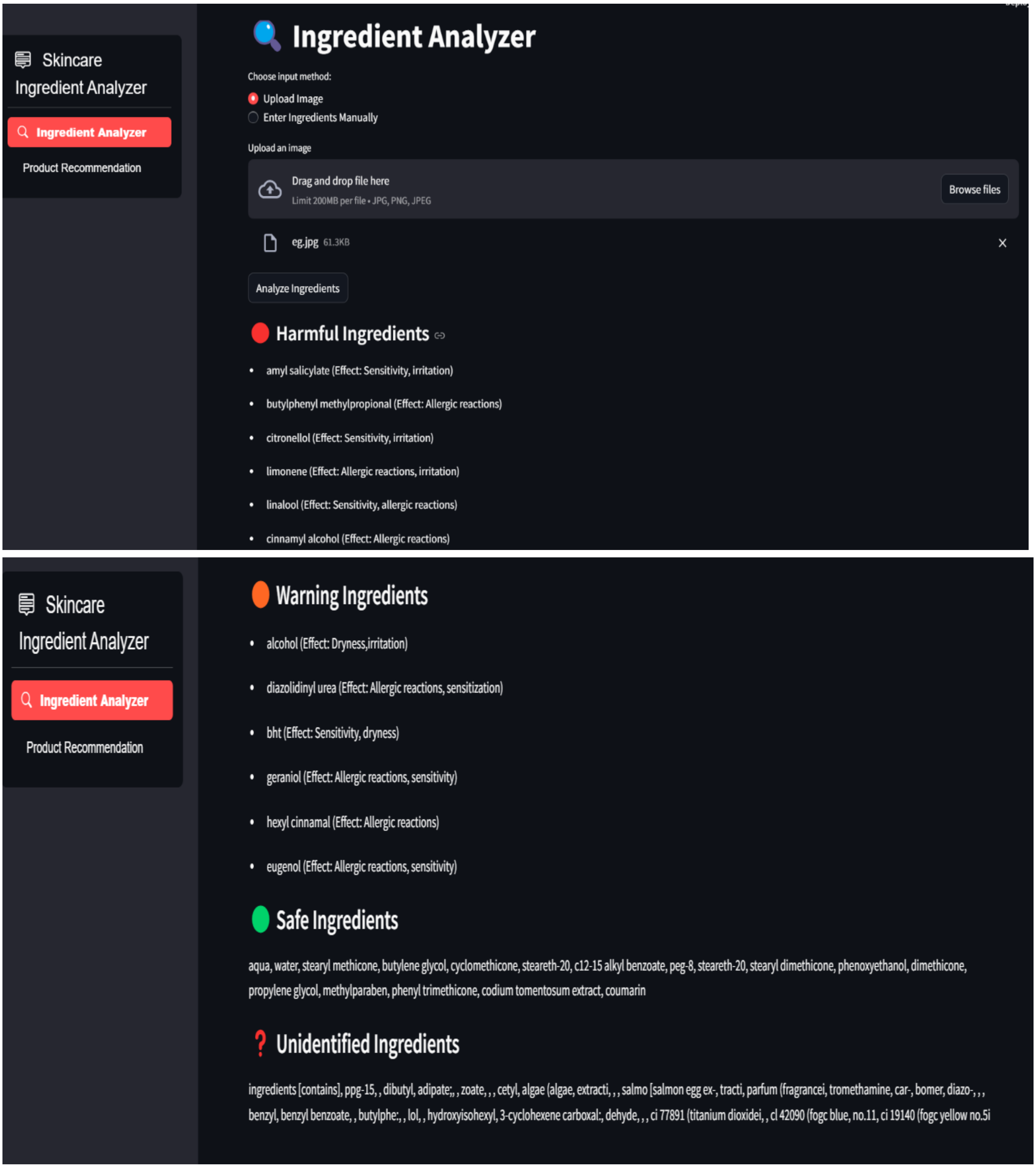

Figure A.3: Sample output (For uploaded image)

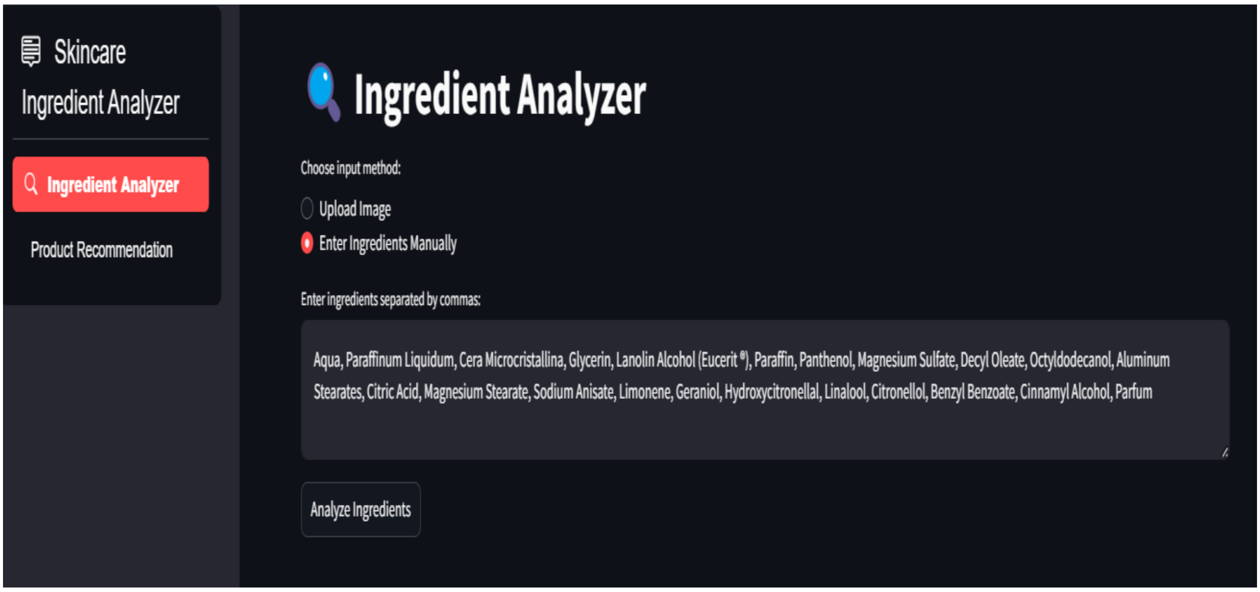

Figure A.4: Sample Input (Manual typing for ingredient analyzer)

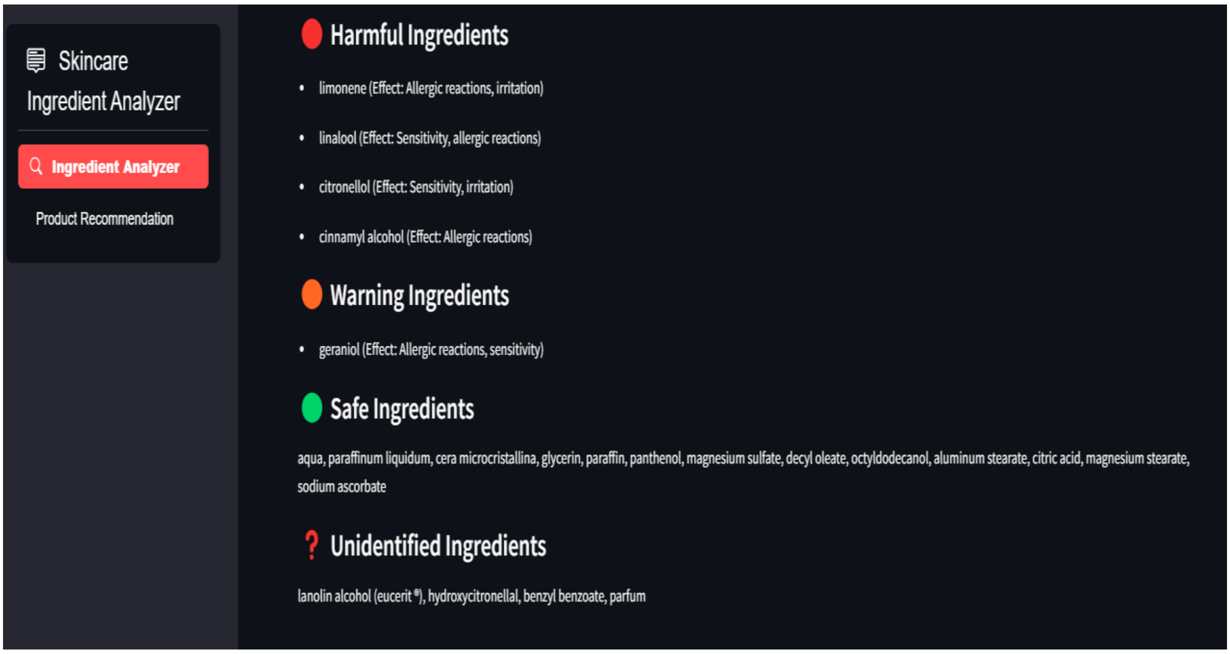

Figure A.5: Sample output (For Manually typed list)

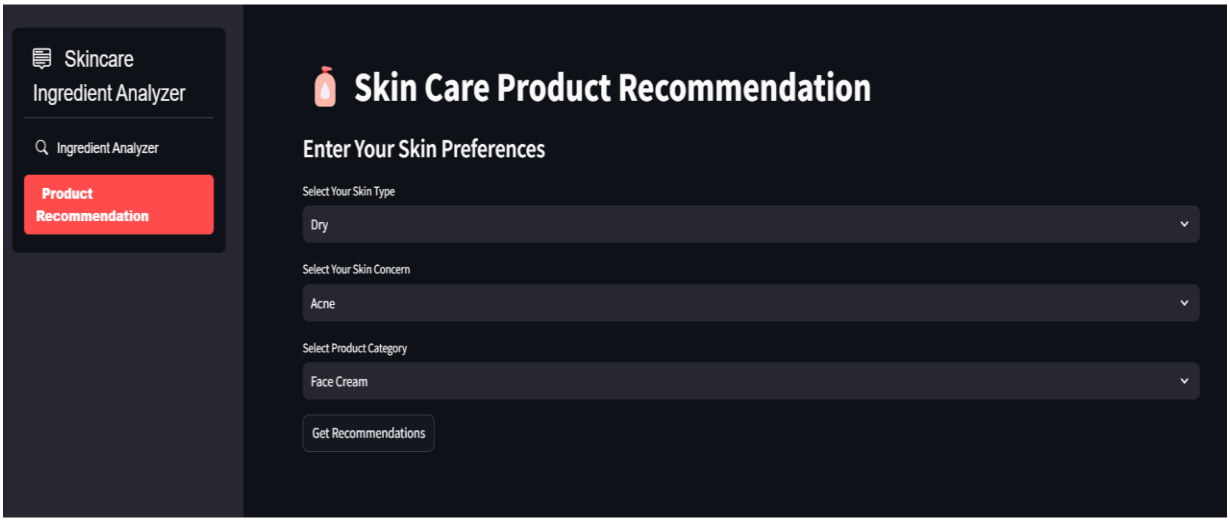

Figure A.6: Website home page (Product Recommendation)

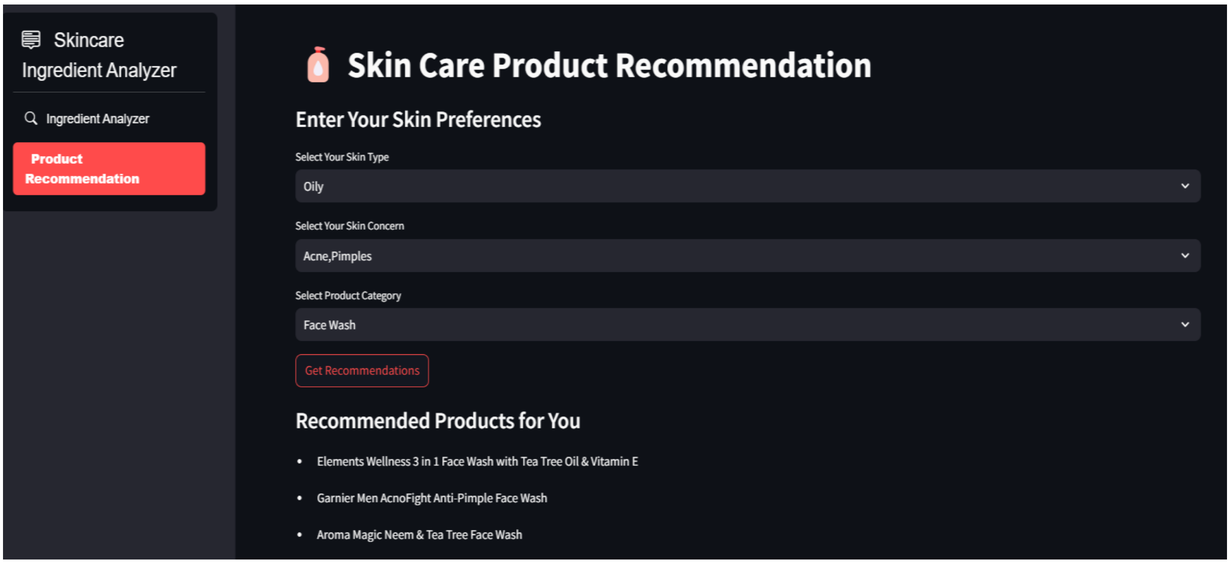

Figure A.7: Sample output

APPENDIX B

ABSTRACT:

The skincare industry faces persistent challenges in ensuring ingredient safety and providing personalized product recommendations. Consumers often struggle to identify harmful substances, allergens, or toxic compounds in their skincare products, leading to potential adverse effects. Additionally, the lack of tailored recommendations based on individual skin concerns hinders the effectiveness of skincare routines. In response to these challenges, we propose an innovative solution integrating EasyOCR and Random Forest. EasyOCR extracts ingredient lists from uploaded product images, while FuzzyWuzzy performs fuzzy string matching to assess ingredient safety, categorizing them as safe, warning, or harmful. This system alerts users to potential allergens and toxic ingredients, enabling informed decision-making. Furthermore, a Random Forest model leverages user skin type and concerns—such as acne, aging, or hyperpigmentation—to provide personalized skincare recommendations. By combining optical character recognition with machine learning, this approach enhances transparency, empowers users with data-driven insights, and revolutionizes skincare safety and efficacy.

Mapping of Sustainable Development Goals (SDGs)

The Ingredient Analyzer and Product Recommendation System aligns with key Sustainable Development Goals (SDGs) by promoting consumer awareness and responsible skincare choices. This system ensures that users can identify harmful ingredients in cosmetic products while also receiving personalized recommendations, contributing to better health and sustainable consumption practices. By leveraging technology to analyze ingredients and suggest suitable products, the system supports global efforts toward healthier lifestyles and eco-friendly product use.

Sustainable Development Goals (SDGs)	Observation
SDG 3: Good Health and Well-being	The system helps consumers identify harmful ingredients in skincare products, promoting informed choices that reduce exposure to potentially toxic substances, leading to better skin health and overall well-being.
SDG 12: Responsible Consumption and Production	The product recommendation system encourages sustainable and responsible consumption by suggesting safe and suitable skincare products tailored to individual needs, reducing waste from purchasing unsuitable or harmful products.

Table C.1: SDGs Addressed

By addressing SDG 3 (Good Health and Well-being) and SDG 12 (Responsible Consumption and Production), the proposed system encourages safer skincare habits and promotes informed consumer decision-making. The ability to detect harmful substances and recommend appropriate alternatives not only enhances individual well-being but also fosters sustainability by minimizing waste from unsuitable or unsafe purchases. This project serves as a step toward a healthier and more sustainable future in the skincare industry.