OULUN AMMATTIKORKEAKOULU

VISUAL-AWARE DEEP LEARNING-BASED MODEL FOR PREDICTING THE POPULARITY OF ONLINE FINNISH RECIPES

Mobarakeh Akbari
Master's Thesis
Fall 2024
Degree Program in Modern Software and Computing Solutions
Oulu University of Applied Sciences

Abstract

Oulu University of Applied Sciences
Degree Program in Modern Software and Computing Solutions

Author: Mobarakeh Akbari
Title of thesis: Visual-aware deep learning-based model for predicting the popularity of online Finnish recipes
Thesis supervisors: Dr. Ilpo Virtanen, Dr. Mourad Oussalah, MSc. Mehrdad Rostami
Fall 2024

The internet has emerged as a prominent platform for culinary inspiration, dining experiences, and social gatherings that revolves around food. Consequently, significant number of individuals now depend on online recipes to a greater extent than conventional cookbooks. However, worries are increasing over the healthiness of these internet recipes. This thesis explores the determinants influencing the popularity of online recipes by analyzing a dataset of over 5,000 dishes from Valio, one of Finland’s largest firms. Valio's website showcases a diverse array of culinary tastes and preferences among Finnish users. Through the analysis of visual characteristics obtained from food images (such as sharpness, contrast, RGB contrast, entropy, saturation, naturalness, and brightness), as well as other characteristics obtained by deep learning techniques and recipe attributes such as nutritional content (energy, fat, salt, etc.), cooking complexity (preparation time, number of steps, required ingredients, etc.), and user engagement (number of comments, ratings, comment sentiment, etc.), our objective is to determine the prominent factors that impact the popularity of online recipes. Our best predictor, AdaBoost, exhibits considerable accuracy (with an accuracy of 95%), outperforming other models, demonstrating that unique visual aspects of food images play a major role in their appeal. Our results show that visual characteristics and deep learning-based feature extraction and prediction can greatly boost final prediction accuracy. By providing useful insights into

contemporary cooking tastes in Finland and guiding possible future dietary policy changes, this study enhances our understanding of the variables leading to the popularity of online recipes.

Keywords: Deep learning, Visual feature analysis, Feature extraction, Online recipe analysis, Recipe popularity evaluations, Recipe rating, Prediction, Finnish food social media.

Contents

Glossary

Acc.	Accuracy evaluation metric
C	Regularization parameter
CNN	Convolutional neural network
L1	Ridge Regularization
L2	Lasso Regularization
LDA	Latent Dirichlet Allocation
MLP	Multilayer Perceptron
NLP	Natural Language Processing
RELU	Rectified Linear Unit
RQ	Research Question
SVM	Support Vector Machine

Acknowledgment

I would like to convey my deep appreciation to my supervisor, Professor Mourad Oussalah, for allowing me to be a member of his research team. His specialised knowledge and insightful guidance have been crucial in navigating me through this demanding undertaking. Furthermore, I would like to express my gratitude to MSc. Mehrdad Rostami, whose readiness to impart his expertise and provide assistance was really beneficial. The intellectual companionship we had throughout this academic trip really enhanced my overall experience.

I would like to extend my gratitude to Dr. Ilpo Virtanen, my supervisor at OAMK, for his insightful guidance and supervision, and to Dr. Mahdi Akbari for his constructive feedback during the model development phase.

I would like to acknowledge the unwavering emotional support of my parents throughout my education. Their encouragement has been a constant source of strength.

Lastly, my deepest appreciation goes to my husband, Hasan, whose unwavering love, support, patience, and understanding were my bedrock during this demanding period. His presence made all the difference.

Mobarakeh Akbari September 2024

1. INTRODUCTION

In recent years, the internet has become a key source for cooking inspiration. Platforms like Valio Ltd's recipe site, which now draws over 25 million visits annually, and Allrecipes.com in the U.S., with 7 million subscribers and 180 million recipe views, have gained significant popularity. Other food websites, such as Kochbar.de and Chefkoch.de, also serve large audiences, particularly in German-speaking regions. Surveys show that more than half of people now turn to online sources for cooking, signaling a shift away from traditional cookbooks. These platforms offer easy access to extensive recipe collections and social networking features that foster community and shared culinary experiences. Here in this research, our dataset of the study was visual and non-visual features of 5000 foods on the Valio website. The idea is to investigate how much the visual and non-visual and the combination of visual and non-visual features affect the popularity of online food recipes.

However, there are concerns about the healthiness of online recipes. Studies suggest that many popular online recipes are less nutritious, and users often find it difficult to identify healthier options (Trattner, Moesslang and Elsweiler, 2018). Ironically, these less healthy recipes tend to attract more attention. There is also evidence of a possible link between the popularity of certain online recipes and increasing obesity rates in the U.S. Understanding the factors behind the popularity of online recipes is crucial for encouraging healthier eating habits (Trattner, Parra and Elsweiler, 2017).

The main goal of this research was to use the deep learning model to predict the food popularity of online Finnish recipes. We have implemented 4 different parts for popularity prediction.

In the first part, we considered the popularity of food according to non-visual features. Then, in the second part, we investigated the food popularity according to the visual features extracted from food images. In the third part, we investigated how much the combination of visual and non-visual features can

predict food popularity. In the last part, we use different deep-learning models for extracting deep feature vectors from food images.

Traditional classification methods, such as Gradient Boosting and Logistic Regression, are often time-consuming and subjective. In contrast, deep learning techniques, particularly those using convolutional neural networks (CNNs), have become state-of-the-art for predicting food image popularity, effectively addressing many of these limitations.

1.1. Related works

This section presents an overview of particular research that makes use of food platforms in order to acquire insights on dietary behavior, social and cultural factors, and health-related issues that are present within online communities. In addition to this, it investigates research that investigates comparable topics by using information from alternative web sources, such as Twitter.com. Quite a few of these works have had a substantial impact on the path that this thesis will take and the relevance that it will have. The subject of research that investigates many aspects of human nutrition by using data obtained from the internet is a relatively young and developing topic. These studies offer several advantages in comparison to more conventional approaches in the field of nutrition science, such as questionnaires. They are less invasive and less likely to be affected by the biases connected with data gathering through self-reporting. In addition, research conducted online typically require a higher number of participants and has the potential to be scaled up globally. However, it is important to recognize potential limitations, such as assuming that searching for a particular recipe means the person will actually prepare and consume that dish. Nonetheless, this section will highlight successful studies in this area, demonstrating the significance and potential of this research field.

A research was carried out by De Choudhury and colleagues (De Choudhury, Sharma and Kiciman, 2016) to gain a better understanding of eating behavior by utilizing data from Twitter. They contended that social media sites, such as Twitter, are ideally suited for this kind of research since users routinely disclose

information about their day-to-day lives, including the foods that they consume. The researchers collected 892,000 tweets that contained phrases connected to food and then matched this information to the demographics, hobbies, and social ties of the users. They averaged nutritional information from many web sources based on particular phrases in order to predict the amount of calories included in the food. Their preliminary research found that there was a significant association (77%) between the calorie content of items that were referenced in tweets and the prevalence of obesity in all fifty states of the United States. They used this information to construct models that might predict obesity rates based on demographic characteristics and food references in tweets with the use of these algorithms. In addition, they investigated societal characteristics such as wealth and education levels, and they discovered that people with greater levels of education are more likely to tweet about and consume items that are lower in calorie density. Furthermore, they explored the social aspects of obesity by taking use of two different Twitter networks (friendship networks and mention networks), and they discovered that friends frequently have eating preferences that are similar to one another.

Abbar et al. (Abbar, Mejova and Weber, 2015) underlined the rising relevance of social media as a significant instrument for public health research, particularly in researching inequities linked to food availability and health. Their study focuses on food deserts; places defined by inadequate availability of inexpensive, healthful food alternatives. These places are commonly connected with poor food choices and health concerns such as obesity, diabetes, and heart disease. Identifying these places and comprehending their issues is a topic of substantial public interest. The authors noted that prior studies on food deserts frequently relied on surveys and self-reported data, which were deficient in rigorous research methodologies and adequate sample sizes. To address these constraints, their research employed Instagram, a swiftly expanding social media network where members regularly exchange photographs. The researchers obtained data using Instagram's official API, enabling them to access public photographs and relevant metadata, including food-related hashtags. In a similar vein, Fried et al. (Fried *et al.*, 2014) carried out a study with the purpose of determining whether or not it is possible to use tweets to forecast demographic

traits that are associated with dietary habits and behaviors. During the course of their investigation, they concentrated on tweets that had particular hashtags associated with food. These tweets were gathered from October 2013 to May 2014, resulting in a substantial dataset consisting of 3.5 million tweets. The effective implementation of prediction tasks based on the linguistic content of these tweets was a significant accomplishment that contributed to the success of their research. The majority of baseline investigations were exceeded by their predictive models, which utilized around 30 million words from the tweets in order to create predictions. These jobs encompassed a wide range of demographic parameters, including the geographical locations of users (city, region, state), as well as aspects at the state level, such as the prevalence of persons who are overweight, the rates of diabetes, and even political inclinations. Two distinct groups of characteristics were utilized by the researchers in order to build these prediction models. As a first step, they utilized lexical characteristics by locating food-related terms throughout food glossaries. Following that, they utilized topic modeling in order to unearth previously concealed theme patterns within the text. The results of their tests, particularly those that focused on diabetes prediction, revealed important insights. In addition, the researchers utilized this data in order to generate a variety of visualizations, such as temporal histograms, geo-referenced heatmaps, and word clouds. Through the use of these visualizations, complicated worldwide patterns of food consumption were brought to light, therefore providing a more in-depth comprehension of the links that exist between Twitter conversations, demographic features, and eating preferences.

Wagner and Aiello (Wagner and Aiello, 2015) did a quantitative study to evaluate gender-based disparities in food-related content and associated stereotypes in the media. The data for the study came from the social media site Flickr. Not only did they consider food to be an essential requirement, but they also considered it to be a means by which individuals in modern society might express part of who they are. The purpose of this study was to determine whether or not there were gender-specific patterns of uploading food content and to gain an understanding of the elements that are responsible for these patterns. They gathered a big dataset consisting of around 15 million Flickr photographs from one million users between the years 2005 and 2014, with male users accounting for 41% of the

total collection. With the use of an online dictionary of popular food-related phrases, they eliminated entries that were not linked to food and maintained only those that had at least one food-related tag and gender information that was accessible to the general public. According to the findings of their investigation, there was statistical evidence indicating that particular categories of food were primarily uploaded by one gender. By way of illustration, beer was 41% more popular among males than it was among women, with 24% of men sharing at least one photo of beer, whilst only 17% of women did so. As part of a further investigation, they looked at the top one hundred pictures that were found in the search engine results for phrases such as "eating meat" or "eating fish." The workers in the crowd determined which gender was more likely to consume the meals that were displayed, whether it was males or females, adults or younger individuals. The results of this poll showed several fascinating trends, such as the fact that alcohol is highly popular among males, despite the fact that it is frequently advertised to women. On the other hand, meals such as milk and fast food were considered to be gender-neutral, but sweets, and coffee were more frequently linked with females in the media. The conclusion that Wagner and Aiello came to was that their method may be used in conjunction with conventional surveys on dietary choices and food intake. This finding highlights the possibility of employing data from social media platforms for research of this kind. The cross-sectional study done by Chunara et al. (Chunara *et al.*, 2013) aimed to examine the relationship between social networks and the prevalence of obesity. The aim of their research was to ascertain the extent to which the interests of Facebook users may reliably forecast the incidence of obesity in the United States. A selection of users was done on the basis of their interests, which were assessed as either favorably or adversely connected to obesity. To provide an example, activities such as "watching television" were considered inactive and associated with obesity, whereas "outdoor fitness activities" represented a lifestyle that was both physically active and beneficial for health. Modeling the activity levels of users was performed by the researchers through the use of linear regression and k-fold cross-validation. By routinely splitting the data into training and test sets, k-fold cross-validation, which is especially beneficial for small datasets, boosts prediction accuracy and ensures statistical significance. This is performed by continually splitting the data into multiple sets. Following the

conclusions of the study, it was determined that Facebook users who had interests associated to physical activity had a predicted prevalence of obesity that was 12% lower across the United States and roughly 7.2% lower in New York districts. On the other hand, the rate of obesity was 27.5% in New York City neighborhoods where inhabitants pursued hobbies such as watching television. Throughout the study, it was demonstrated that there exists a significant correlation between sedentary pastimes and obesity. On the other hand, the scientists stressed that additional study is necessary to thoroughly appreciate these sorts of interactions.

The authors Said and Bellogìn (Said and Bellogín, 2014) conducted an investigation on the ways in which social interaction has the potential to improve food recommendation systems, particularly in relation to online recipes. They brought attention to the fact that, in contrast to other product suggestions such as music or films, food recommendations have potentially substantial implications for one's health. Because of this, any system that has the potential to affect the health of a user has to continue with caution, regardless of the provider's various financial objectives. In addition to this, they stressed the need of taking into account the geographical location of the user, since some locations are more likely to have health issues that are associated with food. The research made use of a dataset obtained from Allrecipes.com in October of 2013. This dataset contained information from 170,000 individuals, 54,000 recipes, 8,400 ingredients, and 17 million ratings. Data pertaining to health, with a particular emphasis on obesity rates, was acquired from County Health Rankings, which includes information from more than 3,400 counties in the United States. As a result of the fact that location information was supplied in freeform language, it was difficult to match users to their respective counties. This issue was solved by the researchers through the process of manually matching and refining the dataset. The first step in their investigation was to evaluate 10 counties, which had both the lowest and greatest obesity rates, by examining the frequency of ingredients based on user evaluations. By analysing the patterns of ingredient utilization, the study was able to successfully identify between these country groups. It was stated by the authors that this data might be beneficial for the development of future customized meal recommendation systems. These

systems could encourage healthier recipes or adapt suggestions based on the individual's risk of obesity. Although the authors acknowledge that there are some limitations, such as the lack of confidence regarding the quantities of items that are used in meals and the difficulties that arise when attempting to identify substances from user-generated freeform language, they believe that this first study is a potential starting point for further research.

An investigation was carried out by Trattner, Elsweiler, and Howard (Trattner, Moesslang and Elsweiler, 2018) to evaluate the nutritional value of recipes obtained via the internet, ready-made meals, and recipes found in cookbooks. This was done in light of the growing concern over nutrition and its influence on health. They brought up the fact that bad eating habits have been connected to health problems, and programs such as ChooseMyPlate in the United States and Change4Life in the United Kingdom advocate home cooking as a better option. On the other hand, they stated that the healthiness of a meal is dependent on the items that are used and the methods that are used to cook it. In their study, they conducted a statistical analysis of three different types of meals that are commonly seen in contemporary diets: recipes found online, ready-made meals, and recipes found in cookbooks. The researchers analysed 100 recipes from cookbooks, 100 ready-made meals, and online recipes from Allrecipes.com, which totaled 5,237 recipes collected from the internet between the years 2000 and 2010. The investigation concentrated on main courses that had sufficient amounts of nutritional information, including carbohydrate content, salt content, calorie content, and fat content. To determine whether or not these dishes were healthy, they utilized two international standards: the recommendations established by the World Health Organization (WHO) and the "traffic light" system established by the Food Standards Agency (FSA) of the United Kingdom. The World Health Organization (WHO) score evaluated the presence of seven required nutrients, but the Food Safety Administration (FSA) method classified recipes according to four primary macronutrients, calling them green for healthy and red for harmful. Their preliminary research revealed that just six of the recipes available online were in complete compliance with the WHO criteria. The recipes on Allrecipes.com were, overall, less nutritious than other recipes. They frequently did not reach the guidelines for fat, saturated fat, and fiber that they

included, although they did, in general, fulfill the requirements for protein. Additionally, they discovered that dishes found in cookbooks had the lowest salt level, followed by those available online and meals that were already prepared. Recipes found in cookbooks and those found online frequently matched the criterion for sugar content in an equal manner, whereas ready-made meals typically did the best possible. Consistent findings were discovered in subsequent research that looked at patterns of change over time. Although recipes obtained via the Internet might not be as healthful as one might anticipate, the authors concluded that their research had certain limitations. Considerations that need to be taken into account include variances in real cooking processes, changes in the nutritional values that are listed on product labels, and variations in the methodologies that are used to calculate nutrient content.

An investigation was carried out by Kusmierczyk and Nørvåg (Kusmierczyk and Nørvåg, 2016) with the purpose of recognizing trends in the names of internet recipes and investigating the practical implications of these findings. The primary objective of their study was to investigate the connections that exist between the words that are used in the names of dishes and the nutritional composition of those foods. They made the observation that, despite the fact that users convey the majority of their communication through text, there is a dearth of study on the relationship between textual content and health-related characteristics. In order to solve this issue, they conducted an analysis on a dataset consisting of 204,000 recipes that were found on the website Allrecipes.com. Following the elimination of recipes that did not contain adequate nutritional information, they were left with around 58,000 recipes. Preprocessing was necessary for recipe titles since they are brief and sometimes contain free-form language. This necessitated the removal of special characters, digits, and stopwords so that the titles could be processed. Following the use of stemming and the retention of words that occurred at least twice, they were able to resolve ambiguities and spelling mistakes, which resulted in 4,679 distinct words. The first experiment that they carried out consisted of doing a statistical study of the distribution of nutritional values for each word that was placed in the names of recipes. Through the use of information acquisition, they evaluated the ways in which specific food-related terms altered the amount of nutrients present. As a result of this study,

connections between certain food phrases and nutrients were discovered, as well as associations between other nutrients. In their second experiment, they developed a thorough and interpretable model based on the findings of the first experiment by employing a unique strategy that merged Latent Dirichlet Allocation (LDA) with linear regression. This method was utilized to construct the model. In order to model the components of recipes, LDA was utilized, and linear regression was utilized to establish a connection between these components and their respective nutritional values. It was determined through validation that this particular model generated the most precise outcomes. In the third and last experiment, they attempted to determine the nutritional content of dishes by relying just on the phrases that were included in the titles of the meals.

Kusmierczyk, Trattner, and their research team (Kusmierczyk, Trattner and Nørvag, 2015) investigated virtual food online communities. The writers underscored the value of innovation in guaranteeing the long-term success of restaurants and chefs. However, they pointed out that there has been no comprehensive study undertaken in the virtual arena of this topic. An analysis was conducted using a dataset acquired from Kochbar.de, consisting of over 400,000 recipes that were published from 2008 to 2014. This dataset includes information such as preparation instructions and classifications. Moreover, the collection comprised data on 230 distinct recipe categories, 200,000 individuals, and more than 7 million recipe reviews. However, a small number of 5,000 persons regularly submit recipes, with each user contributing more than 10 menu items. The website Kochbar.de was chosen for its extensive metadata and other recipe details, including components and nutritional values, that were crucial for their analytical investigation focused on component combinations. An impediment they faced was that the specified components were shown as unstructured text, requiring preprocessing and filtering. A standard statistical filtering procedure was performed, maintaining component names that occurred more than 100 times and substituting those that occurred less than 200 times with more frequent alternatives to decrease uncertainty. By applying a filter, the original list of constituents was reduced to 2,208 distinct objects from an initial 334,000.

The initial investigation analysed community trends and quantified creativity and complexity by utilizing three components: two related to entropy and conditional entropy, and a third innovation factor metric derived from Jaccard similarity. These characteristics were specifically designed to compare recipe ingredients. Evidence revealed that while the quantity of components was constant, there was a progressive rise in creativity within the community as time progressed. A hypothesis was formulated that users amalgamated common components to generate novel recipes, however the pace of creativity was progressively diminishing, indicating a possible plateau in the future. Furthermore, innovation exhibited seasonal and temporal trends, characterized by minor variations throughout the year, reaching its highest point at the start of the year and following summer, maybe suggesting heightened creativity during these periods.

In the subsequent study, the researchers examined innovation tendencies at the individual user level. To enhance the dependability of the findings, they excluded individuals with less than 10 recipes, resulting in a remaining sample size of around 5,000 users. They distinguished between two categories of users: those exhibiting lower innovation factors and those displaying higher levels of innovation. Through the use of linear regression, the researchers examined the innovation factors of users over time and observed that the degrees of innovation were very consistent for the majority of users across the years. Furthermore, they conducted an investigation into the influences on innovation and found that the location of the user had the greatest effect, as quantified by the amount of information gained. This surprising finding suggested a necessity for more investigation into the impact of location on innovation.

A comprehensive investigation was undertaken by Ahn et al. (Ahn *et al.*, 2011) to reveal the underlying principles governing the pairings of ingredients in gastronomic traditions worldwide. Their major purpose was to investigate whether there are measurable and reproducible factors that control why some ingredient combinations are chosen and others are avoided in culinary techniques. The analysis focused on the "shared flavor compounds" concept, which posits that substances containing similar flavor compounds are more prone to enhance one another's taste. To illustrate this, they provided examples such

as the combination of white chocolate and caviar in restaurants, which is attributed to the presence of the organic chemical trimethylamine in both components. The researchers constructed a bipartite graph connecting substances with their corresponding taste compounds, revealing that the majority of foods generally consist of roughly 51 such compounds. This ingredient-compound network was essential for developing and evaluating their hypothesis by analysing topological characteristics.

Three distinct online recipe websites: Allrecipes.com, Epicurious.com (both based in the United States), and Menupan.com (a Korean site)—were used to gather data for their analysis. To avoid any Western bias in their findings, the Korean website was included. The collection consists of 381 separate components and 1,021 different taste compounds, with an average of eight elements per dish. Their first experiment's analytical findings revealed that Western European and North American cuisines tended to blend ingredients with a greater degree of similar taste components. Asian cuisine, on the other hand, often emphasizes combinations with notable taste differences. A second experiment, which evaluated the possibility that certain meals share more chemicals than others and are characteristic of different cuisines, supported this finding even more.

To study these patterns further, the researchers discovered crucial components driving these tendencies. They observed that a small number of regularly used chemicals greatly affected these results. For example, North American food often emphasized components like eggs, cream, cocoa, butter, and milk, but East Asian cuisine depended on ingredients such as onions, ginger, pork, and chicken. In a third experiment, Ahn et al. evaluated diverse cuisines and discovered that South European and Latin cuisines are more comparable to Asian food than to Western European cuisine, as they employ components that do not share as many taste compounds. The research conducted by Ahn and his team offers valuable insights into the intricate dynamics of component pairings in many cuisines, therefore emphasizing the interconnectedness of tastes, ingredients, and culinary traditions on a global scale. Dominik conducted in-depth study using statistical analysis of information from two well-known food community websites,

Kochbar.de and Allrecipes.com, which represent various Western culinary traditions. The goal of comparing these systems was to get a more comprehensive understanding. Key components of the investigation were the social networks inside these websites and the characteristics of the recipes, or "features." Previous research on the popularity of online content and eating habits influenced the creation of these components. Predictive modeling studies were conducted to verify the statistical findings and assess the use of these features. The results showed that specific aspects of recipes may often be used to predict how popular they would become in the future. For the Kochbar.de dataset, user activity metrics like ratings, comments, and the number of uploaded recipes were particularly important predictors. On Allrecipes.com, on the other hand, innovation-related factors including recipe originality, ingredient popularity, and visual characteristics (such saturation and picture entropy) had a greater impact on recipe popularity.

1.2. Research questions

This master's thesis aims to demonstrate the extent to which visual factors influence the popularity of online recipes on the Valio website (https://www.valio.fi/reseptihaku/). We will answer the following research queries (RQ) as this thesis develops:

- **RQ1:** How do non-visual features such as nutritional content, cooking complexity, and user engagement metrics contribute to recipe popularity?
- **RQ2:** Do visual features affect the popularity of food recipes?
- **RQ3:** How does the combination of visual and non-visual features enhance the predictability of a recipe's popularity?
- **RQ4:** Can a deep learning model effectively use these features to develop a food recommender system that considers health?

1.3. Structure of the thesis

The thesis is structured into six main chapters. The Introduction outlines the background, motivation, and objectives of the study, focusing on feature extraction in food images using deep learning, and highlights its significance in food-related applications. The Literature Review explores relevant research, starting with food social media and the psychology of food choices, then delving into feature engineering, texture and visual features, and classifiers like Logistic Regression, Random Forest, Support Vector Machine, and Gradient Boosting, as well as deep learning techniques such as CNNs and Autoencoders. The Methods chapter details the dataset, feature selection process, and experimental phases, with an explanation of the evaluation metrics used, such as accuracy and F1-score. The Results present the findings from the experiments, comparing the performance of various models and features and including comparisons to similar studies. The Discussion interprets these results, examining the strengths and limitations of the models and their broader implications. Finally, the Conclusion summarizes the key findings, addresses the research questions, and suggests directions for future research and practical applications.

2. LITERATURE REVIEW

2.1. Food social media

The challenge of sustaining healthy diets is a substantial obstacle for adolescents and young adults (Arnett, 2007). National health surveys indicate that over 80 percent of adolescents fail to fulfil the guidelines for a nutritious diet (Shay *et al.*, 2013). Additionally, dietary habits tend to deteriorate throughout early adulthood when young individuals transition into living independently (Niemeier *et al.*, 2006). Progressively, this may result in increased susceptibility to cardiovascular disease, diabetes, and other long-term illnesses (Mikkilä *et al.*, 2005). Although traditionally neglected in conventional nutrition interventions and seen a relatively healthy phase of individuals' lives, this developmental stage is now gaining considerable attention due to the emergence of epidemiological evidence indicating negative outcomes for weight gain, physical activity, and dietary intake (Nelson *et al.*, 2008)(Arnett, 2000). Interventions aimed at adolescents and young adults, often carried out in high school and university environments, have shown limited efficacy via behavior-focused instruction, assessments with feedback, and the involvement of peers (Hoelscher *et al.*, 2002).

2.1.1. The psychology of food choice

People generally make about 200 meal decisions each day. Food selection is an intricate process affected by several contextual elements at biological, personal, situational, societal, and socio-economic levels. Primary topics examined in food literature are flavor or sensory appeal, health implications, ethical issues, convenience, pricing, and weight management (Steptoe, Pollard and Wardle, 1995). Food choices typically mirror mood, with people eating for emotional comfort, to boost mood, remember prior experiences, or try something new (Zandstra, De Graaf and Van Staveren, 2001). The value placed on these traits differs across individuals, based on criteria such as age, gender, race, lifestyle, financial level, cultural background, and education (Glanz *et al.*, 1998; Prescott *et al.*, 2002). However, data reveals that for most individuals, the top drives are the flavor of food and its sensory appeal, followed by worries about health, weight control, nutritional content, and cost (Stafleu *et al.*, 1991). Recent research on internet recipes confirms these findings (Harvey, Ludwig and Elsweiler, 2013; Elsweiler, Trattner and Harvey, 2017). Food selection can exhibit bias in several manners; for instance, individuals tend to make suboptimal choices while experiencing hunger and being exposed to meals high in calories, or when their emotional state (Macht, 2008) or stress levels (Oliver, Wardle and Gibson, 2000) are high. Moreover, the social environment has an impact on behavior, as obese persons are more prone to have obese acquaintances (Christakis and Fowler, 2007) and tend to consume much more food while dining in groups compared to when eating alone. Forecasting the popularity of internet recipes is connected to food recommender systems and theoretical research on the decision-making processes behind meal selection. These study topics imply that forecasting recipe popularity is tough. However, knowing the elements driving food preferences gives insights into qualities that may aid in anticipating popularity. In the context of online recipes, we employ these findings to direct our investigations as we conduct experiments with popularity prediction.

2.2. Feature Engineering

Feature engineering is the process of transforming unprocessed data into significant information that may be directly used in machine learning models. It is essentially about developing characteristics that facilitate the construction of prediction models. A feature, or dimension, is an input variable utilized for prediction generation. Since model performance significantly depends on the quality of the data used during training, feature engineering is a key preprocessing step that includes picking the most important parts of raw training data for the given predicting task and model type.

2.2.1. Feature Extraction

Today, Machine Learning is increasingly utilized to analyze vast and complex datasets. Over the past decade, deep learning has significantly improved the efficiency of learning models. Many Machine Learning tasks focus on classification problems, starting with the extraction of features from input data to create a new data representation for the task at hand. A classification system is then trained on these features. Once trained, the system should accurately predict the class label of unseen data. Traditionally, extracted features were handcrafted, tailored specifically to the input data and task, often linked to particular data types, and not robust to changes. A different approach involves using Machine Learning to learn a feature extractor. Instead of designing a system to classify images directly, a learning system is developed to extract features from the input. For images, this involves the network learning higher-level features from the input pixels. This method is considered superior to handcrafted features for several reasons. Training a model on each dataset allows it to adapt to various input types, whereas handcrafted features often require fine-tuning for each dataset. Additionally, this approach eliminates the need for expert knowledge of the images being analyzed.

Feature extraction refers to various methods for creating combinations of variables to address problems while accurately representing the data. It specifically involves identifying distinct image features. Techniques like Average

RGB, Color Moments, Cooccurrence, Local Color Histogram, Global Color Histogram, and Geometric Moments are employed to extract features from test images.

Average RGB: This feature seeks to first filter out photos with bigger disparities in multi-feature queries. The selection of this method is based on its efficient use of data to describe the feature vector and its lower computational requirements in comparison to alternative approaches. Color moments refer to metrics employed to distinguish between photographs by analyzing their color characteristics. The essential principle is that the color distribution of a picture may be read as a probability distribution

The three color moments can be characterized as:

Mean refers to the arithmetic average of the color values recorded in a picture.

$$E_i = \sum_{N}^{j=1} \frac{1}{N} P_{ij} \qquad (1)$$

Standard Deviation: The standard deviation is the square root of the variance of the distribution

$$\sigma_i = \sqrt{\left(\frac{1}{N}\sum_{N}^{j=1}\left(P_{ij} - E_i\right)^2\right)} \qquad (2)$$

Skewness: Skewness can be understood as a measure of the degree of asymmetry in the distribution.

$$s_i = \sqrt[3]{\left(\frac{1}{N}\sum_{N}^{j=1}(P_{ij} - E_i)^3\right)} \qquad (3)$$

To improve system performance, a number of additional image qualities have been included, such as maximum probability, dissimilarity, cluster prominence,

autocorrelation, contrast, energy, entropy, homogeneity, sum variance, sum average, and difference entropy.

In order to represent texture in photographs, the Grey Level Co-occurrence Matrices (GLCM) are often used. It tallys the frequency with which a given characteristic (such as a Gray level) occurs in a certain spatial connection with another feature.

The color histogram Because color is so easy to interpret and extract, it is the most often used feature. A color histogram uses a series of bins to show the color distribution of a picture.

Geometric moments are fundamental features in image processing that permit the derivation of area (or total intensity), centroid, and orientation. Integrating this functionality with other features, such as co-occurrence, might potentially provide enhanced outcomes for the user.

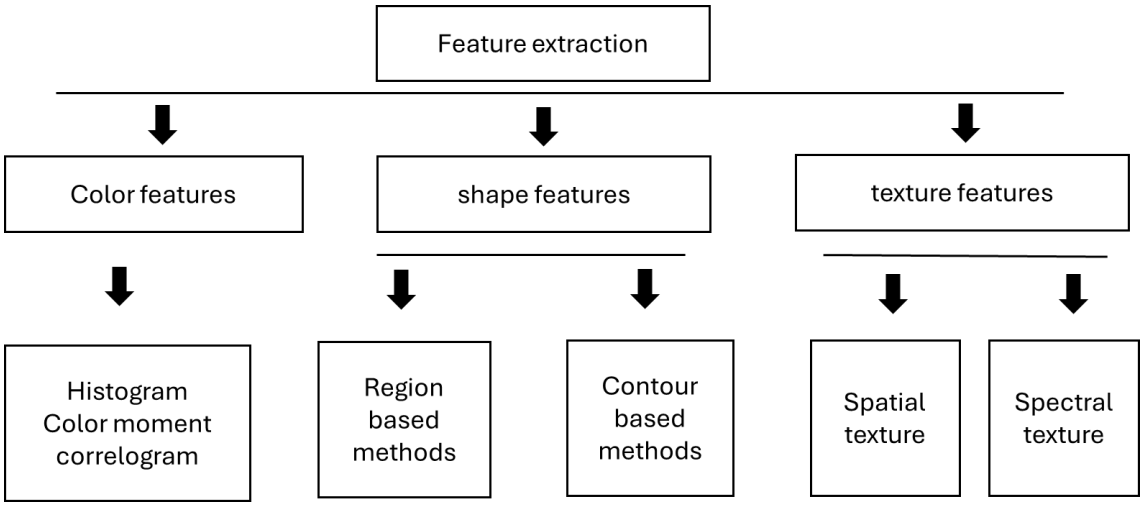

FIGURE. 1. Block Diagram of Feature Extraction. Adopted from (Kavya and Harisha, 2015)

Colour is a very important feature in images. Colour attributes are accurately defined inside designated colour spaces or models. Several colour spaces, including RGB, LUV, HSV, and HMMD, have been used in academic research. The extraction of colour features from pictures or particular areas becomes feasible with the selection of the colour space. Prominent colour characteristics

explored in scholarly works include colour histogram, colour moments (CM), colour coherence vector (CCV), and colour correlogram. Sophisticated colour descriptors have been proposed to enhance the ability to differentiate between colours across certain photometric variations. The observable colour of an item is primarily dictated by two physical parameters: (1) the spectral power distribution of the light source and (2) the surface reflectance properties of the object. Contemporary developments in colour descriptors may be classified into two main groups: innovative colour descriptors derived from histograms and colour descriptors derived from SIFT. Within the HSV colour space, it is well-established that the hue exhibits instability in close proximity to the grey axis. (Van De Weijer, Gevers and Bagdanov, 2006) conducted an examination of error propagation in the colour transformation operation. Experimental results indicate an inverse relationship between the trust in a colour and its saturation. Thus, to enhance the resilience of the hue histogram, each sample of the hue is weighted based on its saturation. Accordingly, the H colour model is both scale-invariant and shift-invariant in relation to light intensity. Given that the intensity channel is a mix of the R, G, and B channels, the SIFT descriptor is not invariant to light colour changes. (Van De Weijer, Gevers and Bagdanov, 2006) introduced the combined visualisation of the hue histogram with the SIFT descriptor, which exhibits both scale-invariance and shift-invariance. In reference (Abdel-Hakim and Farag, 2006), the use of colour invariants as an input to the SIFT descriptor resulted in the creation of a scale-invariant CSIFT descriptor that integrates light intensity. Table 2 presents a thorough overview of many colour techniques published in the literature, emphasizing their strengths and weaknesses. The dominant colour descriptor, colour structure descriptor, and scalable colour descriptor are denoted as DCD, CSD, and SCD, respectively. Additional detailed information on colour descriptors may be found in the reference paper by (Zhang, Islam and Lu, 2012).

TABLE 1. Contrast of different color descriptors (Kavya and Harisha, 2015).

Color method	Advantages	Disadvantages
Histogram	Simple to compute, intuitive	High dimension, no spatial info, sensitive to noise
CM	Compact, robust	enough to describe all colors, no spatial info
CCV	Spatial info	High dimension, high computational cost
Correlogram	Spatial info	Very high computational cost, sensitive to noise, rotation, and scale
DCD	Compact, robust, perceptual meaning	Need post_processsing for spatial info
CSD	Spatial info	Sensitive to noise, rotation, and scale
SCD	Compact on need, scalability	No spatial info, less accurate if compact

When attempting to encode simple mathematical patterns, such as straight lines in a variety of orientations, the shape is an important visual feature that plays a role in identifying and recognizing real-world items. The contour-based and region-based techniques are the two broad categories that may be used to generally classify the many approaches to the extraction of shape features. When attempting to describe the shape of an object, region-based techniques make use of the whole surface area of the object, while contour-based approaches rely only on the information that is included inside the contour outline of the object. The use of local space-time characteristics as a representation for action recognition and visual detection has significantly increased in popularity in recent years. Because these features encode different prominence and motion patterns in video, they provide an independent portrayal of events in terms of the spatial-temporal alterations, scales, background noise, and many motions that are present in the image.

There have been several alternative techniques to feature localization and description that have been reported in the literature. These approaches have shown promising results for action categorization and other detection tasks (Wang *et al.*, 2009). According to Ke et al. (Ke, Sukthankar and Hebert, 2005), the use of volumetric attributes for the purpose of event detection in video sequences was investigated. They extended the concept of Haar-like features to include three-dimensional spatiotemporal volumetric features, which was an expansion of the two-dimensional box features (Viola and Jones, 2001). A contour-motion feature descriptor was proposed by Liu et al. (Liu *et al.*, 2009) for robust pedestrian detection. This descriptor makes use of space-time contours as the low-level representation of the pedestrian. After that, a three-dimensional distance transform is carried out to expand the one-dimensional contour into the space of three dimensions.

2.3. Texture features

Another essential characteristic of pictures is their texture, which is often exploited by human visual systems for the purposes of recognition and interpretation. To quantify characteristics such as smoothness and regularity, it evaluates the intensity variation of a surface. Texture can differentiate between textured and nontextured photos and may be supplemented with other visual features such as color to improve retrieval efficiency. Texture alone is not sufficient to discover images that are like one another. There are two primary categories that may be used to classify a variety of ways for representing textures: structural and statistical. Statistical methods such as Fourier power spectra, co-occurrence matrices, shift-invariant principal component analysis (SPCA), Tamura features, Wold decomposition, Markov random field, fractal model, and multi-resolution filtering techniques like Gabor and wavelet transform can be employed to characterise texture by analysing the statistical distribution of image intensity. A multitude of diverse approaches have been proposed to extract texture features.

Methods for extracting texture features may be categorised into two groups: spatial and spectral approaches. The categorisation is determined by the domain

from which the texture characteristic is obtained. Spectral approaches operate by converting an image into the frequency domain and then extracting features from the resulting picture. On the other hand, spatial methods extract texture characteristics by either computing pixel statistics or identifying local pixel patterns in the original image domain. Two methodological techniques are used for the extraction of texture characteristics. It is crucial to acknowledge that both spatial and spectral features have their own advantages and disadvantages.

TABLE 2. Summary of advantages and disadvantages of texture method types (Kavya and Harisha, 2015).

Texture method	Advantages	Disadvantages
Spatial texture	Meaningful, easy to understand, can be extracted from any shape without losing information	Sensitive to noise and distortions
Spectral texture	Robust, need less computation	No semantic meaning, need square image regions with sufficient size

2.4. Visual features

These features were initially found to be effective for photographs on the Flickr platform, but a recent research shows that a subset of these features is also effective in evaluating the attractiveness of photographs related to online recipes (Elsweiler, Trattner and Harvey, 2017). Specifically, the derived features include sharpness, contrast, saturation, colorfulness, entropy, and naturalness, all of which are formally defined below.

Sharpness: This metric evaluates the clarity and detail level of an image, linked to the brightness contrast at the edges within the image as mentioned in (Trattner, Moesslang and Elsweiler, 2018). The algorithm utilizes the images Laplacian, divided by the locale average luminance (μ_{xy}) around pixel (x, y):

$$\text{sharpness} = \sum_{xy} \frac{L(x,y)}{\mu_{xy}}, \qquad with\ L(x,y) = \frac{d^2I}{dx^2} + \frac{d^2I}{dy^2} \qquad (4)$$

Sharpness Variation: Like saturation variation, sharpness variation is determined by the standard deviation of all pixel sharpness values.

Contrast: Contrast refers to the relative difference in brightness or color among local features in an image. Contrast, as stated in the reference (Hermann, 2001), is the evaluation of the disparity in visual characteristics between two or more sections of a field when observed either simultaneously or consecutively. Various metrics exist for assessing contrast, however, root mean square contrast (RMS-contrast) is typically employed to compare images (Pedro and Siersdorfer, 2009).

We calculate RMS-contrast as follows (Trattner, Moesslang and Elsweiler, 2018):

$$\text{contrast} = \frac{1}{N} \sum_{x,y} \left(I_{xy} - \bar{I}\right) \qquad (5)$$

where I_{xy} is the intensity of a pixel, $\bar{I}$ represents the arithmetic mean of the pixel intensity and *N* is the number of pixels in the image.

RGB Contrast: The RGB contrast is identical to the fundamental contrast calculation stated before but is extended to the three-dimensional RGB color space.

Saturation: The International Commission on Illumination (Hermann, 2001) defines picture saturation as the "colorfulness of an area judged in proportion to its brightness" (Trattner, Moesslang and Elsweiler, 2018). It indicates the quality or vividness of the color impression. An RGB approximation may be used to estimate saturation in the HSV color space.

$$\text{saturation} = \frac{1}{N} \sum_{x,y} s_{xy}, \qquad with \qquad (6)$$

$$S_{xy} = max\left(R_{xy}, G_{xy}, B_{xy}\right) - min\left(R_{xy}, G_{xy}, B_{xy}\right) \qquad (7)$$

where *N* is the number of pixels in an image and $R_{xy}, G_{xy}\ and\ B_{xy}$ are the coordinates of the color of the pixel in sRGB space.

Saturation Variation: This method evaluates saturation variation by calculating the sample standard deviation of all pixel saturation values in the image (Trattner, Moesslang and Elsweiler, 2018).

$$\text{saturation\_variation} = \sqrt{\frac{\Sigma_{x,y}\left(s_{xy} - \bar{s}\right)^2}{N-1}} \qquad (8)$$

Where N is the number of pixels, s_{xy} is the list of pixel saturations, and $\bar{s}$ represents the average pixel saturation.

Brightness: The average brightness of a picture indicates the perceived visual energy output of a light source. To extract the brightness of the sample pictures, the AvgBrighness class was used with the default NTSC weighting system and without any mask. It uses a conventional brightness algorithm (Trattner, Moesslang and Elsweiler, 2018):

$$\text{brightness} = \frac{1}{N}\sum_{xy} Y_{xy}, \qquad \text{with}$$

$$Y_{xy} = \left(0.299 * R_{xy} + 0.587 * G_{xy} + 0.114 * B_{xy}\right) \qquad (9)$$

where Y_{Xy} represents the luminance value, and N is the number of pixels in the image. R_{xy} , G_{xy} and B_{xy} are the three RGB color channels of the pixel at (x,y).

Colorfulness: The International Commission on Illumination defines colorfulness as an "attribute of a visual perception that determines how chromatic a color appears to be." Colorfulness can be calculated using the individual color distances of the pixels. Therefore, the image must be converted to the sRGB color space with

$$rg_{xy} = R_{xy} - G_{xy} \ and \ yb_{xy} = \frac{1}{2}(R_{xy} + G_{xy}) - B_{xy} \qquad (10)$$

Consequently, colorfulness can be measured as (Trattner, Moesslang and Elsweiler, 2018).

colorfulness = $\sigma_{rgyb} + 0.3 \times \mu_{rgyb}$ with

$$\sigma_{rgyb} = \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2} \qquad (11) \qquad \mu_{rgyb} = \sqrt{\mu_{rg}^2 + \mu_{yb}^2} \qquad (12)$$

where R_{xy} , G_{xy} and B_{xy} represent the color channels of the pixels, σ denotes the standard deviation, and μ signifies the arithmetic mean.

Entropy: In information theory, entropy measures the randomness or information content provided by a source. Image entropy often indicates how much information needs encoding by a compression algorithm. For instance, an image of moon craters with high edge contrast has high entropy, making it difficult to compress. This property makes entropy useful for measuring an image's texture. We used Shannon entropy as follows: First, the image was converted to grayscale, giving each pixel an intensity value. Next, the occurrences of each distinct value were counted. Finally, we applied the following formula (Trattner, Moesslang and Elsweiler, 2018).

$$\text{entropy} = -\sum_{x\epsilon[0..255]} p_x \ . \log_2^{(p_x)} \qquad (13)$$

where p_x is the probability of a pixel having the grayscale value x among all pixels in the image.

Naturalness: The idea of naturalness refers to how closely an image aligns with human visual perception of reality, particularly in terms of colorfulness and dynamic range. Though it is subjective, naturalness is a crucial metric for

assessing image quality in color image design (Huang, Wang and Wu, 2006). According to San Pedro and Siersdorfer (Pedro and Siersdorfer, 2009), it can be evaluated by the following method: First, convert the image color space to HSL if it hasn't been done already. Then, consider only the pixels within the thresholds of 20 ≤ L ≤ 80 and S ≥ 0.1. Next, categorize these pixels into one of three groups: 'Skin,' 'Grass,' or 'Sky,' based on their H (hue) value. To determine the naturalness of each group, use the average saturation value (μ_s) of the pixels within that group (Trattner, Moesslang and Elsweiler, 2018).

$$N_{skin} = e^{-0.5\left(\frac{\mu_s^{skin}-0.76}{0.52}\right)^2} \qquad if\ 25 \le hue \le 70 \qquad (14)$$

$$N_{Grass} = e^{-0.5\left(\frac{\mu_s^{Grass}-0.81}{0.53}\right)^2} \qquad if\ 95 \le hue \le 135 \qquad (15)$$

$$N_{sky} = e^{-0.5\left(\frac{\mu_s^{sky}-0.43}{0.22}\right)^2} \qquad if\ 185 \le hue \le 260 \qquad (16)$$

In the final step, the naturalness index is determined using

$$naturalness = \sum_i w_i N_i \qquad i\epsilon\{'Skin','Grass','Sky'\} \qquad (17)$$

where w_i denotes the proportion of pixels belonging to a specific group within the entire image. The naturalness index ranges from 0 (completely unnatural) to 1 (completely natural).

2.5. Classifiers

2.5.1. Logistic Regression

Cox's logistic regression (Cox, 1972), first in 1972, is a fundamental statistical technique extensively used in many domains such as machine learning, epidemiology, and the social sciences for the purpose of addressing binary classification issues (Ahn, 2011; Levenson *et al.*, 2016). The core assumption of the Logistic Regression approach is the depiction of the likelihood that an input data item belongs to one of two categories. Logistic regression uses the logistic or sigmoid function to transform a linear combination of input data into a numeric probability value ranging from 0 to 1. In contrast to linear regression, which generates predictions only based on continuous numerical data. The logistic curve, represented by a S shape, functions as the structural foundation of the model. The output of the logistic function, for the given input parameters, is the estimated probability of the positive class, often represented by the letter "1" in binary classification. The likelihood estimate in logistic regression is generated by applying the logistic function to the linear combination, which is the estimated weighted sum of the input features. This particular combination is often known as the linear combination. The algorithm undergoes training using a dataset containing predetermined results, and its parameters (weights and intercept) are adjusted to optimise the probability of faithfully replicating the observed data. One of the notable features of Logistic Regression is its high level of interpretability. A comprehensive assessment of the coefficients given to each input characteristic allows for a clear evaluation of the impact it has on the probability of belonging to the positive class. Furthermore, Logistic Regression offers insights into the odds ratio, which quantifies the extent to which the probability of the positive result vary when each variable increases by one unit. Logistic Regression is particularly valuable for explanatory modelling and hypothesis testing due to its ability to give informative insights into the odds ratio.

Within the domain of logistic regression, the symbol "C" denotes the regularization parameter, often known as the "inverse of regularization strength" in certain circles. This hyperparameter serves to achieve a balance between

maximizing the model's convergence to the training data and preventing overfitting.

2.5.2. Random Forest

The Random Forest algorithm (Breiman, 2001) is a flexible and effective ensemble learning approach commonly applied in machine learning studies (Hsu *et al.*, 2017; Huang *et al.*, 2018; Karthika, Murugeswari and Manoranjithem, 2019; Aufar, Andreswari and Pramesti, 2020). It is effective for both classification and regression issues, demonstrating great versatility across numerous academic subjects. In essence, Random Forest is a compilation of decision trees that demonstrates exceptional performance in handling large information and generating precise predictions. The fundamental benefit of this strategy is its capacity to handle the common shortcomings of individual decision trees, such as the propensity to overfit the data. Standalone decision trees frequently acquire noise and distinctive patterns in the training data, leading to poor performance on new, undiscovered data. To address this issue, Random Forest employs a method called "bagging" or "Bootstrap Aggregating." The bagging technique involves constructing several decision trees, each trained on separate subsets of the data. To include diversity in the training process, these subsets are generated by randomly picking the original dataset with replacement. Additionally, at each decision tree node, a random subset of features is assessed for splitting, further improving the model's randomization and durability. The primary strength of Random Forest lies in its ensemble nature, which combines the predictions of several decision trees to avoid overfitting and improve accuracy. Each tree in the ensemble produces an output during the prediction process, which may be either a class prediction for classification objectives or a numerical value for regression tasks. The eventual prediction is then decided by a majority vote for classification tasks or by averaging for regression assignments. The versatility, interpretability, and capability of Random Forest to effectively manage high-dimensional data and complex variable interactions make it a very indispensable tool in many research applications.

2.5.3. Support vector machine

A fundamental machine learning technique extensively used in social media data analysis research is Support Vector Machine (SVM) (Cortes and Vapnik, 1995) (Al-Zoubi *et al.*, 2018)(Khanday, Khan and Rabani, 2021). This package is very adaptable and resilient, specifically designed for both classification and regression tasks. The efficacy of Support Vector Machines (SVM) resides in their ability to process data that lacks linear separability, thereby conferring significant benefits in tackling intricate research enquiries. Fundamentally, Support Vector Machines (SVM) aim to identify an ideal hyperplane, which is a multidimensional decision boundary, inside the feature space that divides data points into distinct classes. A fundamental attribute of Support Vector Machines (SVM) is its emphasis on determining the hyperplane that maximises the margin, represented by the distance between the decision border and the nearest data points from each class, known as "support vectors." The goal of this method is to improve the confidence in classification by striving for a larger margin, therefore minimising the influence of noisy or outlier data points. Should linear separation not be achievable, Support Vector Machines (SVM) utilises the "kernel trick." Using this novel mathematical approach, Support Vector Machines (SVM) may convert data into higher-dimensional spaces where linear separation becomes feasible. The core kernel functions, which include linear, polynomial, and radial basis function (RBF) kernels, enable Support Vector Machines (SVM) to efficiently tackle a wide range of nonlinear problems encountered in many academic disciplines. This approach serves as a robust and flexible machine learning tool, with the ability to handle data with a large number of dimensions and complex interconnections among variables. Owing to its capacity to efficiently tackle both binary and multiclass classification issues, it is highly suitable for use in many research settings. Furthermore, the regularisation parameter (C) of Support Vector Machines (SVM) enables accurate modification of the balance between maximising margin and minimising error, therefore ensuring that the model conforms to given analytical objectives.

2.5.4. Gradient boosting

Gradient Boosting (Friedman, 2001) is a powerful ensemble machine learning approach recognized for its extremely effective prediction skills and versatility in handling numerous data-driven issues. (Athanasiou and Maragoudakis, 2017; Abdurrahman, Irawan and Setianingsih, 2020; Neelakandan and Paulraj, 2020) Its suitability for both regression and classification problems makes it an indispensable tool for extracting insights and generating precise forecasts. Fundamentally, Gradient Boosting is a methodology that combines several weak learners, often decision trees, to construct a resilient prediction model. The fundamental innovation of Gradient Boosting lies in its sequential training methodology. It repeatedly generates a succession of decision trees, with each one focusing on the errors or residuals of the prior tree. The process starts by constructing a basic decision tree, often known as a "shallow tree" or "stump." The first tree produces forecasts that are then compared directly to the actual target values. The disparities between these predictions and the actual numbers indicate the errors or residual deviations. Subsequent decision trees are developed expressly to remedy these errors, adding additional significance to the data points that were previously misclassified or those whose predictions were the furthest from the actual values. As fresh trees are added to the ensemble, Gradient Boosting constantly modifies and refines its predictions, thereby decreasing errors and boosting accuracy. The final prediction is generated by combining the results of all the unique trees, where each tree contributes a weighted vote to the end conclusion. An inherent characteristic of Gradient Boosting is its capacity to tolerate diverse data sources and effectively address both regression and classification tasks. In addition, it provides essential insights into the importance of features.

2.6. Deep learning

In this section, we discuss different feature extraction techniques in deep learning.

Feature extraction plays a key role in image processing, particularly in applications like food recognition or classification. Features are distinctive elements or patterns that convey essential information in an image, such as color, texture, and shape. Recently, deep learning has emerged as a highly effective method for automatically extracting these features. Unlike traditional techniques, deep learning identifies relevant features directly from the data, without the need for manually designed rules.

Conventional feature extraction techniques depend on preset algorithms to identify edges, colors, textures, and shapes. Although these methods can be helpful, they face challenges when dealing with complex images like food. Food images often include various components, such as plates, backgrounds, and multiple ingredients, which can make it difficult for traditional methods to capture all significant features. Deep learning, particularly using convolutional neural networks (CNNs), has proven to be highly successful in addressing these issues.

2.6.1. Convolutional Neural Networks (CNNs)

Deep Learning has become a key approach for addressing self-perception challenges, such as interpreting images, human speech, and robotic exploration of the environment. The proposed methodology aims to apply the concept of Convolutional Neural Networks (CNNs) to image recognition. The primary focus of the proposed model is to understand CNNs and utilize them for image recognition tasks. CNNs extract feature maps from 2D images using filters, emphasizing the spatial relationships between pixels rather than relying on fully connected layers of neurons. CNNs have proven to be highly effective in image processing and have outperformed previous methods in computer vision tasks like handwriting recognition, natural object classification and image segmentation. CNNs are designed to recognize visual patterns directly from pixel images with minimal preprocessing. Most CNN architectures adhere to a common design framework: they apply convolutional layers to the input, periodically down sample (Max pooling) the spatial dimensions and increase the number of feature maps. In addition to convolutional layers, CNNs also include fully connected layers, activation functions, and loss functions (such as cross-

entropy or softmax). However, the most crucial operations in CNNs are the convolutional layers, pooling layers, and fully connected layers (Hossain and Alam Sajib, 2019).

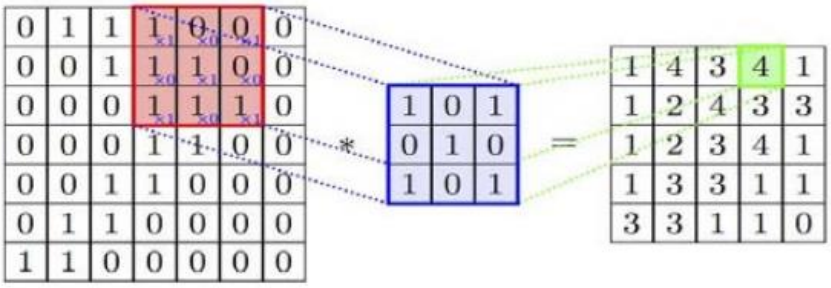

FIGURE 2. Convolution operation (Hossain and Alam Sajib, 2019).

When building a CNN, it's standard practice to add pooling layers after each convolution layer to decrease the spatial dimensions of the representation. This reduction helps lower the number of parameters, thereby reducing computational complexity. Additionally, pooling layers assist in mitigating the risk of overfitting. By choosing a pooling size, we can minimize the number of parameters by selecting the maximum, average, or sum of the values within the selected pixels.

Figure 3 illustrates the max pooling and average pooling operations.

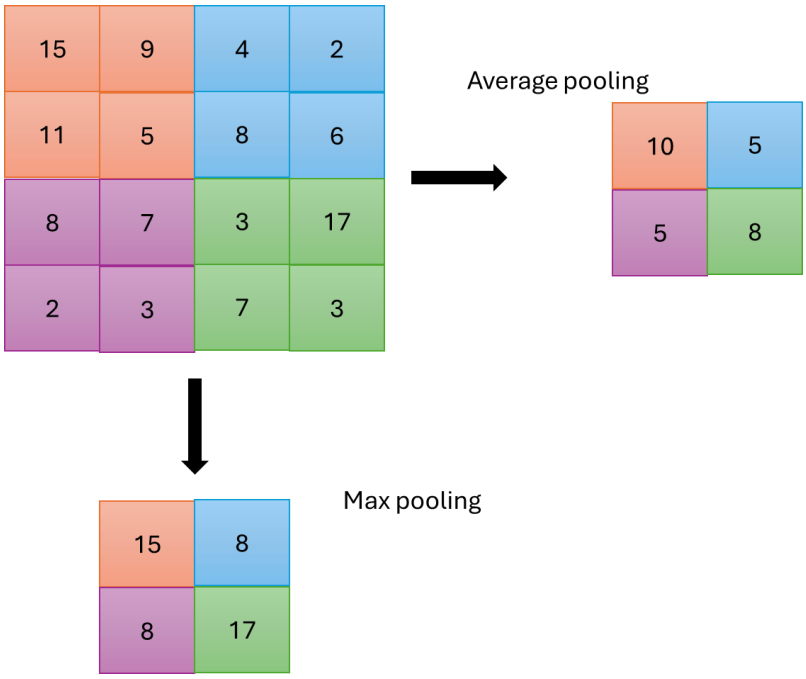

FIGURE 3. Max pooling and Average pooling operation, adopted from (Hossain and Alam Sajib, 2019)

A fully connected network is a type of architecture where every parameter is connected to every other parameter, allowing the network to determine the relationship and impact of each parameter on the labels. By incorporating convolution and pooling layers, we can significantly reduce the time and space complexity. Finally, a fully connected network can be constructed at the end to classify the images.

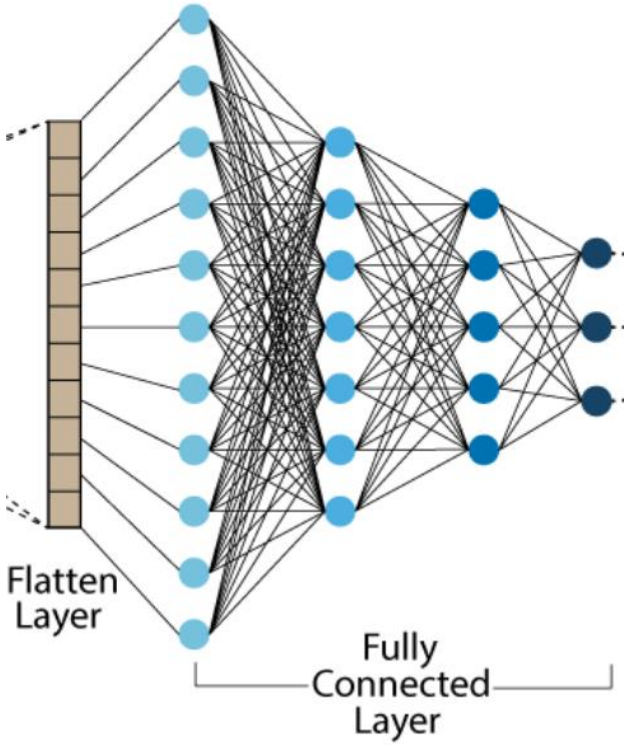

FIGURE 4. Fully connected layer (Rguibi et al., 2022).

2.6.2. Autoencoder

An autoencoder is a kind of neural network designed for unsupervised learning, which helps in learning efficient data representations. It is frequently used for tasks like dimensionality reduction, feature extraction, and data compression. Unlike supervised methods that require labeled data, autoencoders work with unlabeled data and train the network to replicate the input data.

Autoencoders play a vital role in processing food images due to their ability to automatically extract key features from complex visual data. By learning to capture essential aspects such as texture, color, and shape in a compressed latent space, they enable more efficient classification and clustering of different food items. Additionally, autoencoders are highly effective for data compression, reducing the dimensionality of large food image datasets while preserving crucial information. This is particularly useful for storing or transmitting food images without significant loss of quality. Denoising autoencoders are also beneficial in cleaning up noisy or blurred images, enhancing the clarity and quality of the data.

Unlike traditional methods like PCA, autoencoders can handle non-linearities in the data, allowing for deeper and more meaningful feature extraction, which is critical when dealing with the complexities of food image analysis. An autoencoder is made up of several key components:

1. Encoder: The encoder's role is to compress the input data into a lower-dimensional form, known as the latent space. This compression reduces the data's dimensionality, focusing on the most relevant features. The encoder transforms the input data X into a hidden representation h, which is usually smaller in size, using the function:

$$h = f(Wx + b) \qquad (18)$$

where, x is the input data, W represents the weights, b is the bias, f is an activation function (e.g., ReLU, sigmoid) and h is the compressed representation of the data.

2. Latent Space Representation: This is the compressed, lower-dimensional version of the input produced by the encoder. It captures the most essential features of the input data in a compact form, which is crucial for feature extraction.

3. Decoder: The decoder reconstructs the original input from the latent space. It performs the reverse operation of the encoder, expanding the compressed representation back to its original size:

$$x' = g(W'h + b') \quad (19)$$

Where, x' is the reconstructed data, W' and b' are the decoder's weights and biases, g is the activation function used by the decoder.

The primary goal of training an autoencoder is to reduce the difference between the original and the reconstructed input. This difference is measured by a loss function, typically “Mean Squared Error (MSE)”, which is defined as:

$$\{\mathcal{L}\}(x, x') = \left||x - x'|\right|^2 \qquad (20)$$

The training process of an autoencoder involves several steps. First, during the feedforward phase, the input data is passed through the encoder, which produces a latent representation, and then through the decoder to reconstruct the original input. Next, the loss function calculates the difference between the original input x and the reconstructed output x', to minimize the reconstruction error. Following this, the model uses backpropagation to update its weights and biases, adjusting based on the error and using optimization techniques like gradient descent. Over multiple iterations or epochs, the autoencoder continues to refine its performance, gradually reducing the reconstruction error as it becomes more effective at recreating the original input from the latent space.

There are several types of autoencoders, each designed for specific tasks. An “Undercomplete Autoencoder” uses a smaller latent space than the input, which forces the model to focus on the most important features during the compression process. In contrast, a “Sparse Autoencoder” may have a larger latent space, but it enforces sparsity by adding a penalty to the loss function, ensuring that only a small number of neurons are active, resulting in sparse representations. The “Denoising Autoencoder” is trained to clean up noisy input data, making it particularly useful in tasks such as noise reduction for images or speech. Lastly, the “Variational Autoencoder (VAE)” is a more advanced model that learns a probabilistic distribution of latent variables. This allows it to generate new data samples by drawing from this distribution, making VAEs particularly valuable in generative modelling tasks.

3. METHODS

This project is structured into four distinct phases, each focusing on different aspects.

1. Phase One: This phase is dedicated to non-visual features.

2. Phase Two: This phase focuses exclusively on visual features.

3. Phase Three: This phase integrates both non-visual and visual features.

4. Phase Four: This final phase employs a deep learning model to extract feature vectors from each image.

Figure 5 provides a summary of each phase. We explore each phase and its scenarios in detail as we progress.

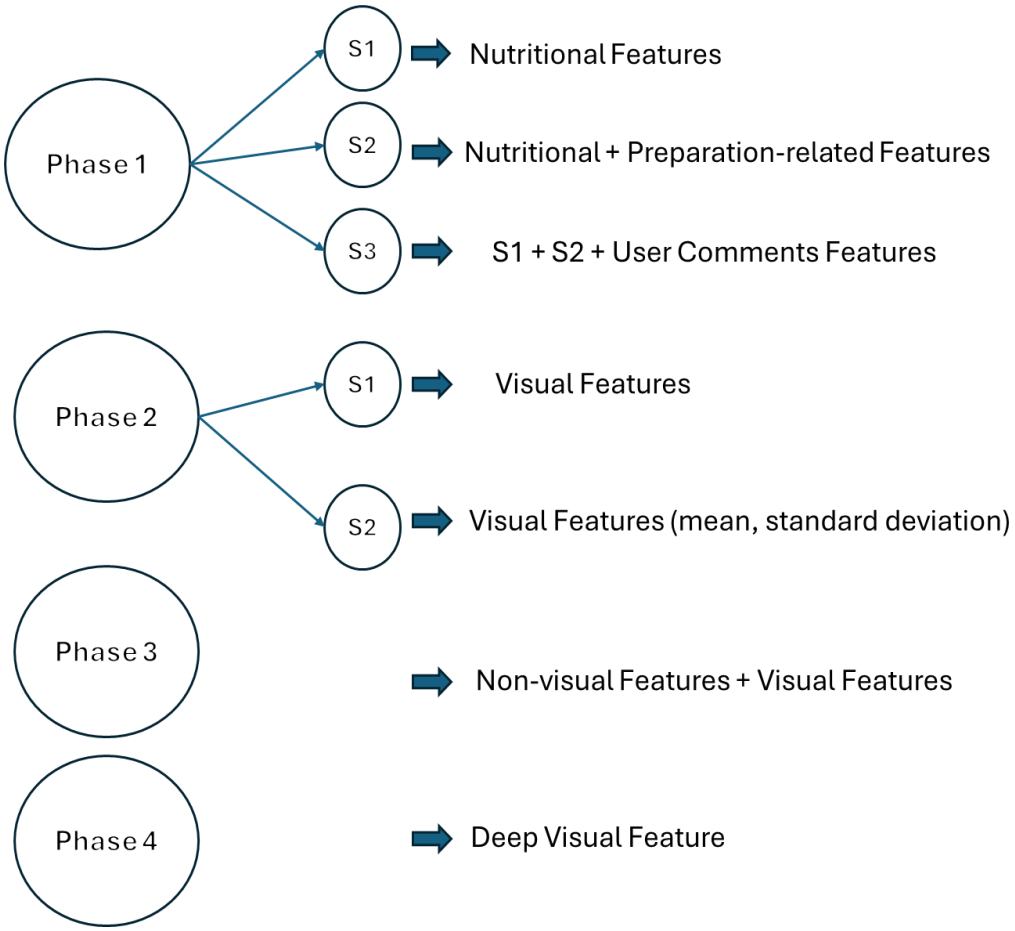

FIGURE 5. Illustration of different phases and scenarios of the study

3.1. Dataset

We constructed our dataset by collecting information from www.valio.fi (Figure 3), which was selected because it is one of the most well-known and biggest social media networks that focuses on food and recipes with more than 25 million visits annually. Using our web crawling strategy, we were able to collect 5,472 recipes that were published between the years 2010 and 2022. We extracted a variety of attributes for each recipe, including the Recipe Name, the Date and Time of Publication, the Ingredients, the Time Required for Preparation, the Difficulty Level, Tags, Ratings from Users, Comments from Users, and nutritional information per 100 grams (which included the amount of energy, protein, carbohydrates, fat, saturated fat, dietary fibre, and salt). The fundamental information about this dataset is included in Table 1, along with a summary of the items that were extracted. We only included recipes that had readily accessible

ingredients and nutritional information to ensure that they were in line with the objectives of our research. Throughout the crawling process, we eliminated 663 recipes that did not have any nutritional information, which led to the compilation of a final list of 4,833 meals that fulfilled our standards.

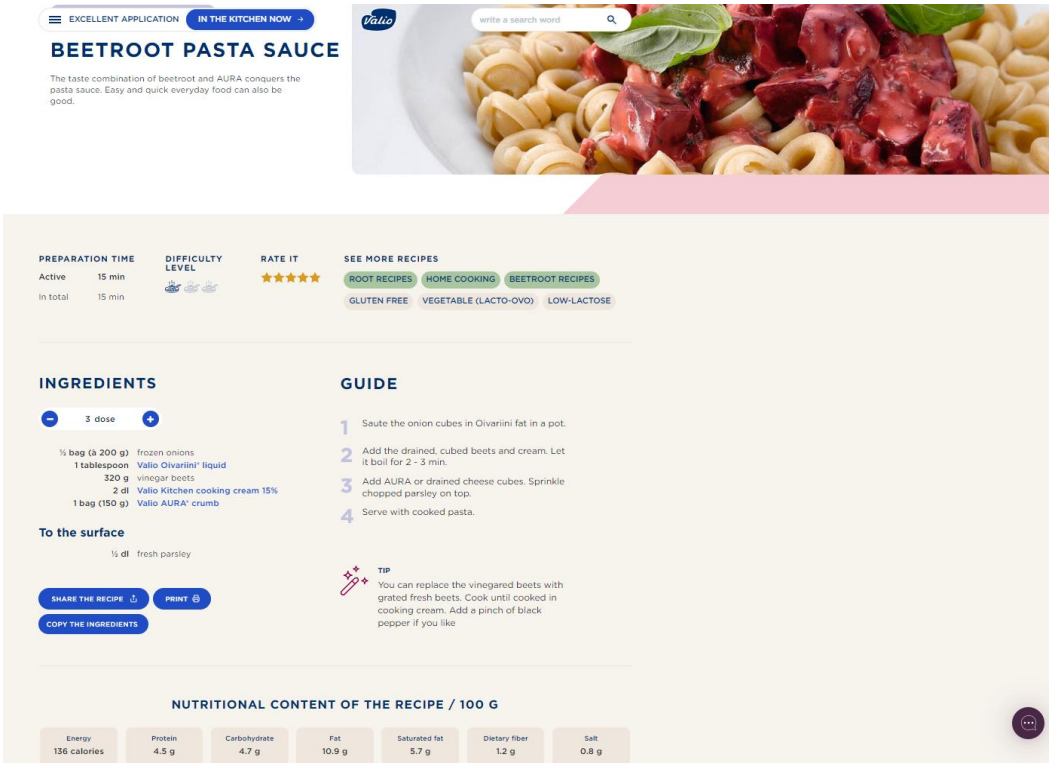

FIGURE 6. Food image and its non-visual feature and texture features from Valio website.

3.2. Feature Selection Scenarios

Creating traits that may assist in predicting the popularity of meals and dishes can be done in a variety of different ways. Several aspects, including flavor, texture, nutritional content, physical surroundings, attitudes, motivations, individual preferences, and accessible information, are discussed in Scheibehenne et al.'s (Scheibehenne, Miesler and Todd, 2007) study on food choices. These factors include taste, texture, nutritional content... In light of this, several predictive elements have been applied, by the findings of cognitive psychology, the features utilized by earlier researchers (such as Elsweiler et al.

(Elsweiler, Trattner and Harvey, 2017) and Rokicki, Herder, and Trattner (Rokicki *et al.*, 2016)), and the information gained from the examination of popularity.

3.3. Phase 1

In general, this thesis is divided into four phases, the first phase includes only non-visual features, the second phase includes only visual features, and the third phase is a combination of visual and non-visual features. In the non-visual features phase, we have three different scenarios.

Several different methods may be used to identify characteristics that might be of assistance in predicting the popularity of recipes and foods. In the study that they conducted on meal selections, Scheibehenne et al. (Scheibehenne, Miesler and Todd, 2007) highlighted the huge number of factors that influence human decisions. These factors include flavor, texture, nutritional content, the physical surroundings, attitudes, motives, individual preferences, and information. Because of this, a variety of predictive features have been applied. These characteristics include concepts from cognitive psychology, characteristics that have been examined in the past by researchers such as Elsweiler et al. (Elsweiler, Trattner and Harvey, 2017) or Rokicki, Herder, and Trattner (Rokicki *et al.*, 2016) , and the information that has been gathered via the study of popularity trends. This thesis improves on these earlier efforts by developing three scenarios, each of which has a unique set of characteristics, to illustrate the dynamics of the popularity of recipes found online. Scenario 1 focuses on nutritional issues, Scenario 2 emphasizes characteristics linked to preparation challenges, and Scenario 3 includes additional user-related components that were not addressed in the two scenarios that came before it. Further to the point:

Scenario 1:

The first scenario is one in which the nutritional aspects of the dishes are the primary focus. The numerical distribution of the key nutritious components per 100 grams is used to calculate the nutritional content of each recipe, and this presentation is shown in the form of a numerical representation. The following

seven aspects of nutrition are examined in this scenario: energy, proteins, carbohydrates, fat, saturated fats, dietary fibre, and salt. For the research, these variables are taken from the recipe website and used as characteristics.

Scenario 2:

The emphasis is on preparation-related elements. Preparation Time reflects the average time necessary to make a given meal. The Difficulty amount quantifies the amount of difficulty of the preparation, ranging from 1 (the easiest) to 3 (the most complicated). In the context of culinary preparation, ingredients refer to the fundamental constituents employed in the creation of a meal. The website offers comprehensive information on the required amount of each ingredient for the given recipe (Number of Ingredients). This scenario also contains the Number of Steps in preparing each recipe, as the website specifies.

Scenario 3:

This scenario integrates user input by incorporating comments and reviews, which contribute user viewpoints on the content of the recipes. This can strengthen the advice process for other users. A cumulative sum of 14,081 ratings and 24,630 comments were gathered for all individual recipes. This scenario includes (1) the total number of ratings (Rating Count) for each dish, (2) the total number of comments (Comment Count) for each recipe, (3) the sentiment of the comments, and (4) the number of tags for each food.

Recipe rating annotations

We classified recipes into "Rating Groups" according to their mean user ratings, distributed on a continuous scale from 0 to 5. Figure X illustrates that the bulk of recipes fall into the "good" Rating Group, defined by ratings over 3.5, which represents 55% of the total. Approximately 30% of the recipes are classified as belonging to the "poor" Rating Group. In the next part, we investigate the categorization findings resulting from these three established rating classes.

We categorized recipes into differentiated Rate Groups based on their mean ratings, which offer a valuable understanding of how people evaluate the quality of these dishes. The Rate Groups are defined as follows:

- Class 'bad': Recipes marked with ratings less than 1.
- Class ' normal': Recipes with ratings between 1 and 3.5.
- Class ' good ': Recipes with ratings over 3.5.

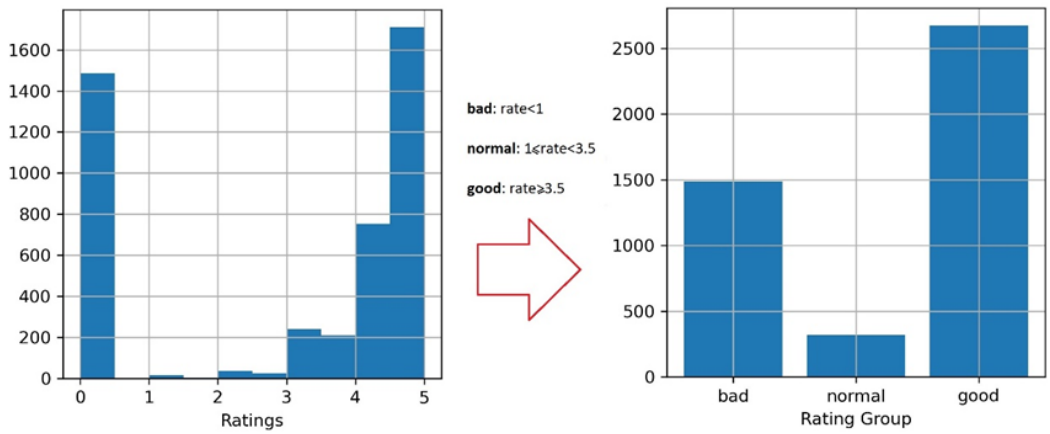

FIGURE 7. Illustration of the various categories assigned to each recipe according to the rates provided by users.

In the first step of section one, we try to preprocess our data set. we tried to solve missing value problems. We used some techniques for removing the missing value such as mean and median before predicting the popularity of food.

Insight into Data and Correlation Heatmap:

In data analysis, understanding the relationships between different variables is crucial. A correlation heatmap is a powerful visualization tool that helps in this process by displaying the pairwise correlation coefficients between a set of variables in a matrix format. The major purpose of this component of the research is to forecast a recipe's Rating Group using numerous variables. The Rating Group is established based on user ratings. Hence, a more robust link between these characteristics and the rating might offer useful insights into the factors that impact the ratings. As shown in Figure 3, the features most significantly connected with the Rating (about 30%) are the sentiment of comments, the total number of comments, and the total number of ratings, all of which exhibit a positive association. This suggests that for every given mood, remark count, or rating count, the ratings of the recipe are expected to rise in a linear manner.

Additionally, the number of stages in the recipe and the quantity of items utilized show a pretty substantial link (approximately 20%). The largest association identified (about 90%) is between a recipe's fat content and energy, demonstrating that dishes higher in fat content often have greater energy levels. Moreover, the number of stages, the number of components, and the recipe's complexity level are also substantially associated (about 45%).

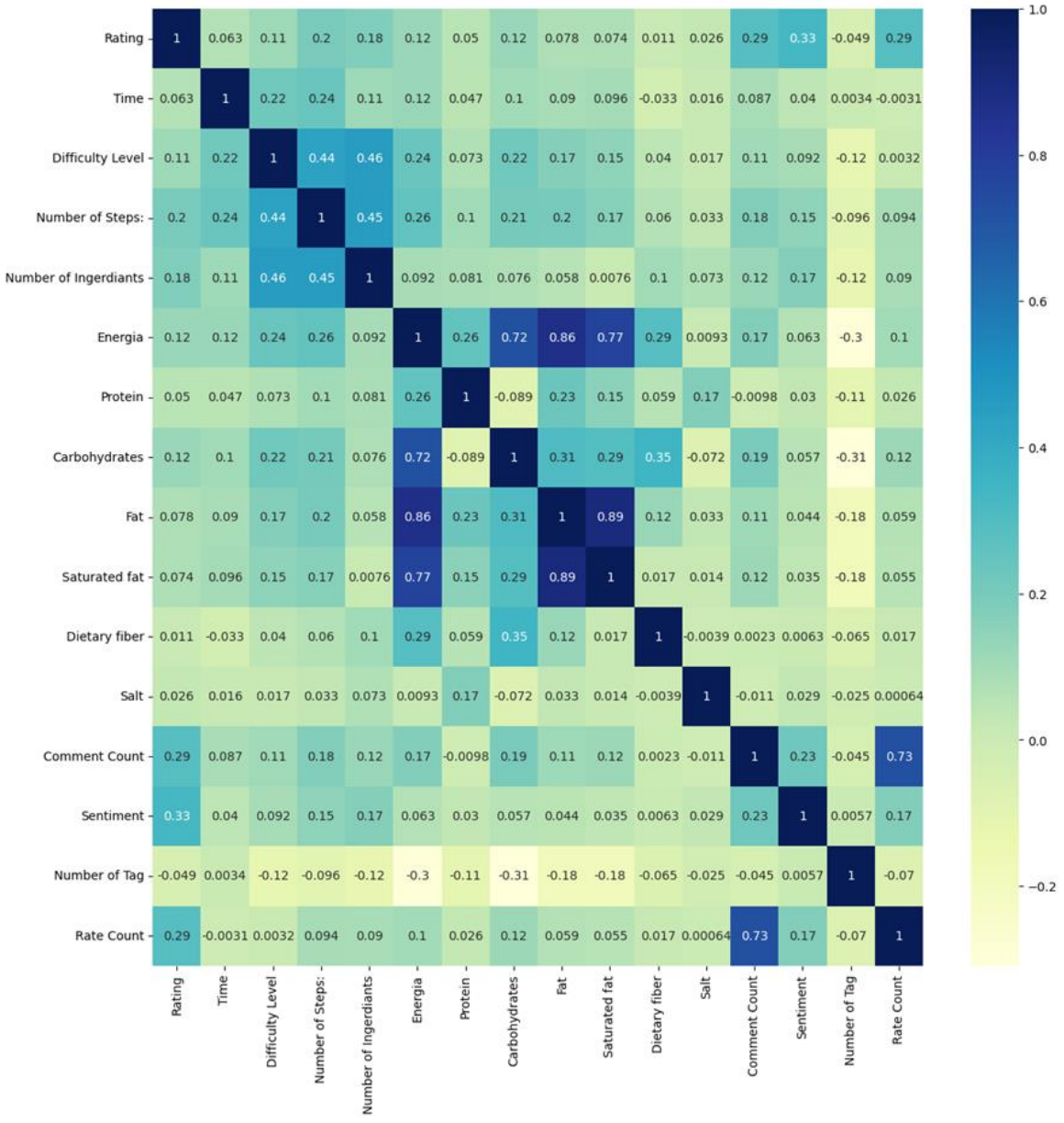

FIGURE 8. Correlation heat map between different non-visual features and target value

As shown in the correlation heatmap, the features with the strongest correlations to the Rating Group (around 12% to 16%) include the RGB contrast mean, sharpness standard deviation and brightness mean, all of which show a positive correlation. This indicates that as the RGB contrast mean, sharpness variability, or brightness increases, the rating group tends to increase in a linear manner, albeit weakly. Additionally, other features like sharpness mean and entropy also show slight positive correlations with the Rating Group (approximately 10%). The highest correlations observed within the image features (approximately 96%) are between the RGB contrast mean and brightness standard deviation, indicating that these aspects are closely related in images. Furthermore, there are significant correlations (around 94%) between sharpness mean and sharpness standard deviation, reflecting consistency in image sharpness.

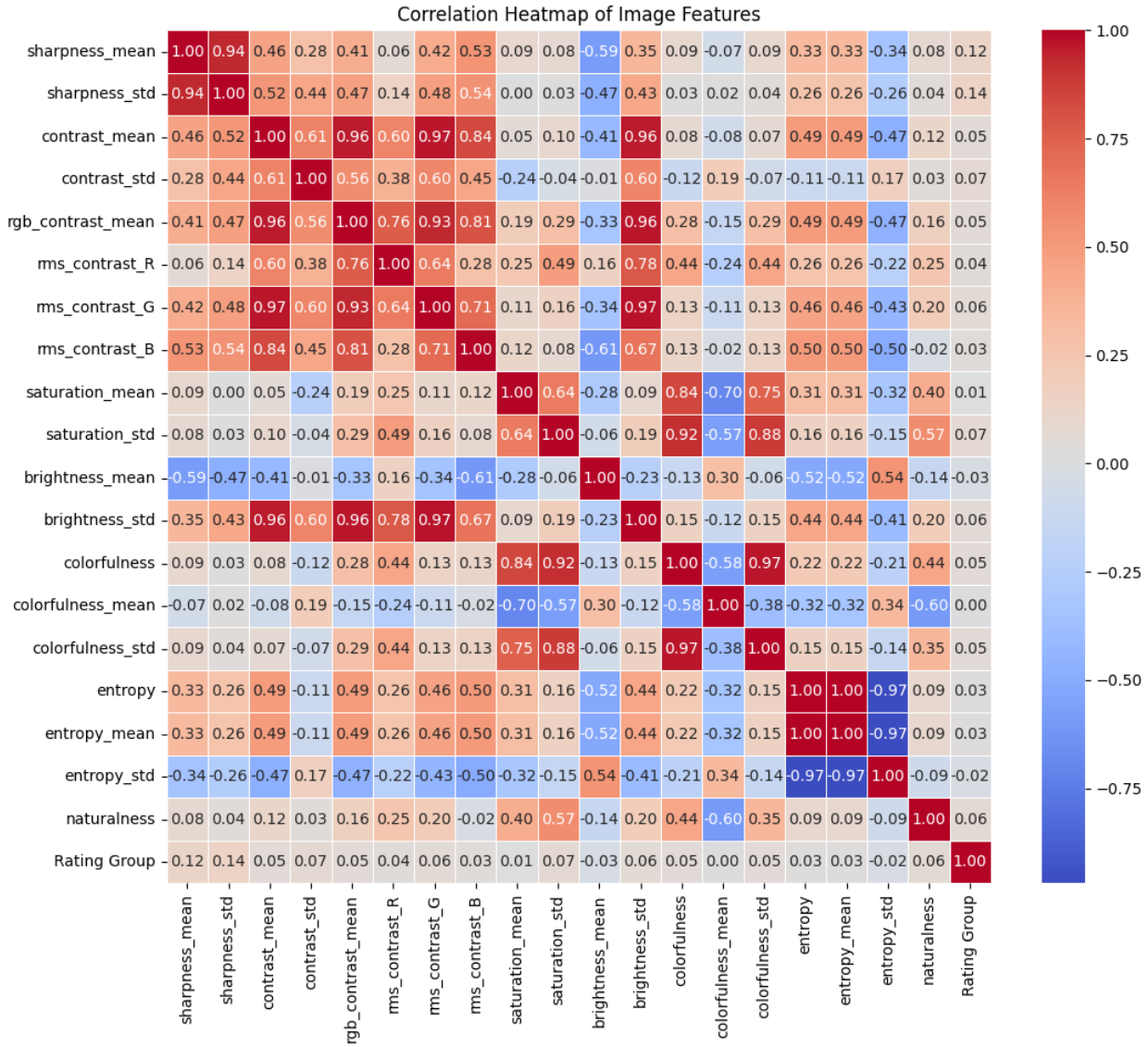

FIGURE 9. Correlation heat map between different visual features and target value.

3.4. Phase 2

This phase is divided into two distinct scenarios. In the first scenario, we predicted food popularity based on several visual features extracted from food images following image processing. These features include sharpness, sharpness variation, contrast, RGB contrast, brightness, entropy, and naturalness, all of which are detailed in section 2.4. In this initial phase, the number of visual features is limited. However, in the second scenario, in addition to the primary visual features, we also calculated the mean and standard deviation for each feature. Our analysis revealed that incorporating the mean and standard

deviation for each feature did not significantly impact the accuracy of our predictions.

3.5. Phase 3

In this phase both phase 1 and phase 2 are combined in a way food popularity is predicted based on both non-visual features (phase 1) and visual features (phase 2). To increase the accuracy of prediction the normalized feature values within a range of 0-1 were used.

3.6. Phase 4

In this phase, we used deep learning methods to extract feature vectors from food images using an autoencoder. Then we predicted the food popularity with different traditional classification methods separately then, we combined these deep feature vectors with our non-visual features and investigated the popularity again.

Evaluation protocol:

In this section, we split the dataset into two segments: 80% for training and 20% for testing, using a fixed random state to ensure consistent results. This method is commonly used in machine learning to guarantee that the model is properly trained and evaluated, resulting in a more reliable and accurate assessment of its performance.

3.7. Evaluation metrics

We employed accuracy (Acc.) and F1-score as suitable evaluation metrics to gauge the model's performance across various hyperparameter combinations. While both F1 score and accuracy are standard metrics for assessing classification models, they highlight different facets of the model's effectiveness.

3.7.1. Accuracy

Accuracy (Acc.) is a fundamental metric used to evaluate the overall correctness of a classification model. This metric is derived by dividing the count of correctly predicted cases (including both true positives and true negatives) by the total count of occurrences in the dataset. Despite its simplicity and comprehensibility, this metric may not always be the optimal selection, particularly for datasets with imbalances. A high accuracy score may be misleading when one class is dominating, since the model may primarily predict the majority class and perform poorly in the minority class.

$$Acc. = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \quad (21)$$

3.7.2. F1-score

The F1-score is a metric that balances quality by considering both accuracy and recall. Precision evaluates the ratio of correct positive predictions to all positive predictions produced by the model, whereas recall measures the ratio of actual positive occurrences that were forecasted correctly by the model. Integrating these two metrics, the F1-score provides a unified result that represents the overall performance of the model.

$$F1 = \frac{2 \cdot Precision \cdot Recall}{(Precision + Recall)} \quad (22)$$

Where:

$$Precision = \frac{True\ Positives}{(True\ Positives + False\ Positives)} \quad (23)$$

And:

$$Recall = \frac{True\ Positives}{(True\ Positives + False\ Negative)} \quad (24)$$

In imbalanced datasets, the F1-score is particularly advantageous as it considers both false positives and false negatives, therefore providing a more reliable statistic in such particular cases. A higher F1-score indicates superior model performance in terms of optimising the balance between accuracy and recall.

4. RESULTS

In the first phase of our project, we investigated three distinct scenarios, each of which used a different set of features for classification. For the first scenario, just the nutritional qualities were taken into consideration, but for the second scenario, both the nutritional qualities and the challenges of preparation were taken into consideration. The accuracy of classification provided by both situations was almost comparable, with scenario 2 showing a little advantage.

However, the highest accuracy was achieved in the third scenario (Figure 10), where user engagement features were added alongside nutritional and preparation difficulty features, resulting in the most effective classification approach.

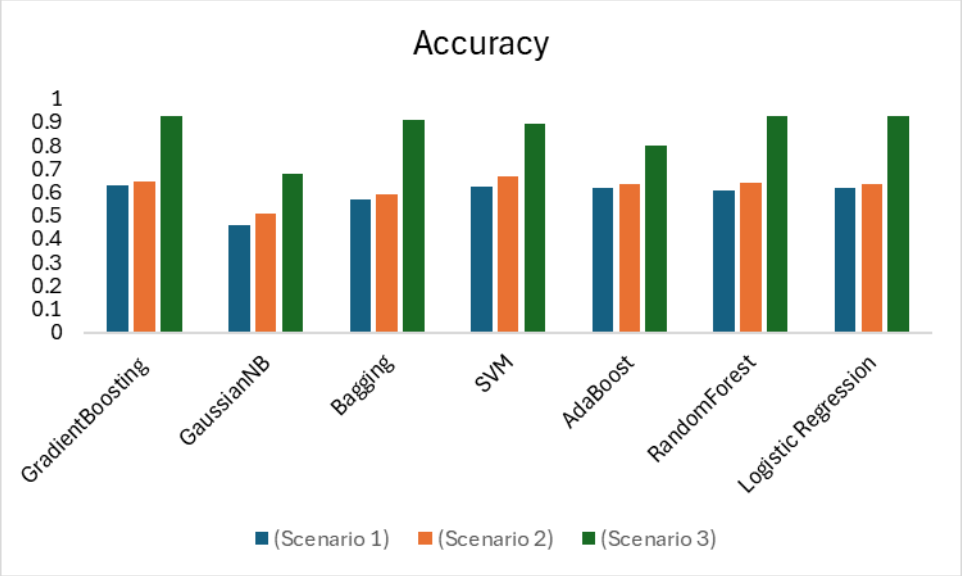

FIGURE 10. Accuracies for different classifiers and scenarios.

Table 3 shows the accuracy of the different scenarios using different classification methods. It has been shown that the accuracy values in scenario 3 are significantly higher than in scenario 1 and scenario 2. This could be because in scenario three includes more features including the total number of ratings (Rating Count) for each dish, the total number of comments (Comment Count) on

each recipe, the sentiment of the comments, and the number of tags for each food.

TABLE 3. Accuracy of different classification methods based on non-visual features.

Classifier:	Scenario 1	Scenario 2	Scenario 3
Gradient Boosting	0.63	0.644	0.9245
Gaussian NB	0.46	0.51	0.679
Bagging	0.57	0.59	0.91
SVM	0.624	0.67	0.8965
AdaBoost	0.62	0.633	0.80
Random Forest	0.61	0.64	0.926
Logistic Regression	0.62	0.636	0.9255

In the second phase, to maintain equal harmony among visual features and be comparable, we also calculated mean and standard deviation, which even by adding the mean and standard deviation for each image, in the accuracy value didn't make any significant changes. The table below shows the accuracy values for two different scenarios.

TABLE 4. Accuracy of different classification methods based on Visual features.

Classifier	Visual Features	Visual Features (mean, std)
Gradient Boosting	0.59	0.60
Gaussian NB	0.575	0.50
Bagging	0.57	0.55
SVM	0.577	0.62
AdaBoost	0.58	0.60
Random Forest	0.60	0.61
Logistic Regression	0.58	0.62

Table 5 shows the accuracy values using different classification methods for different phases. In Phase 1 we investigated the accuracies only for non-visual features, in Phase 2, we calculated the accuracy only for visual features. However, in phase 3, both non-visual and visual features were used to predict the food popularity. As shown in Table 4, the accuracy for phase number 3 is higher than for phase one and phase two. This means that the visual features affect the food's popularity and user rating.

TABLE 5. Accuracy values of different classification methods for different phases.

Classifier	Visual Feature	Visual Feature Mean, std	Non_ visual Scenario 3	Visual + Non_visual	Combine Deep Visual Feature
Gradient Boosting	0.59	0.60	0.924	0.93	0.944
Gaussian NB	0.575	0.50	0.68	0.71	0.63
Bagging Classifier	0.57	0.55	0.91	0.92	0.62
SVM	0.577	0.62	0.8965	0.78	0.932
AdaBoost Classifier	0.58	0.60	0.80	0.86	0.945
Random Forest	0.60	0.61	0.926	0.931	0.95
Logistic Regression	0.58	0.62	0.9255	0.926	0.85

In phase 4 we used deep learning techniques to extract feature vectors from images, and as you see the table 6 shows the accuracy value for these two models. We can see that the accuracy by using deep learning is higher than traditional model and this indicates that deep learning methods can be more accurate than traditional feature extraction methods.

TABLE 6. Accuracy of different classification methods based on Combinations of Features with Handy and Deep Feature Vector

Classifier	Visual + Non_visual	Deep Visual Feature
Gradient Boosting	0.93	0.944
Gaussian NB	0.71	0.63
Bagging Classifier	0.92	0.62
SVM	0.777	0.932
AdaBoost Classifier	0.855	0.945
Random Forest	0.931	0.95
Logistic Regression	0.926	0.85

4.1 Comparison to Similar Studies

An analytical research compared Valio, a Finnish internet-based food social network, with comparable platforms from other nations: Allrecipes.com, which represents American culinary culture, and Kochbar.de, which represents German culinary culture. A comparative analysis, derived from Mößlang's study (Trattner, Moesslang and Elsweiler, 2018), was conducted to assess the predictive capability of various characteristics and validate statistical findings by predictive modeling experiments. The findings indicated the presence of recipe attributes that have wide applicability and have a substantial impact on the future popularity of online recipes across all web platforms. Nevertheless, there were noticeable variations in popularity patterns across the three websites.

The research conducted by Mößlang (Trattner, Moesslang and Elsweiler, 2018)examined recipes from Kochbar and Allrecipes and demonstrated that the characteristics derived in the study had robust predictive capabilities. The primary objective was to predict whether a recipe would surpass the expected level of popularity within a certain period. For this purpose, three classifiers were employed: Random Forest, Naive Bayes, and Generalised Linear Models. The experiment had positive results, as some setups achieved an accuracy rate of up to 89%. Significantly, models trained with Kochbar.de data routinely achieved the very greatest level of accuracy. Among the classifiers, Random Forest had the highest level of effectiveness, while no one classifier consistently surpassed the others.

These findings validated earlier hypotheses, establishing that generally relevant variables have a substantial impact on the future popularity of online recipes. These criteria included the previous engagement of the user in posting the recipe, the level of presentation excellence, and the originality of the dish's idea. Nevertheless, user engagement indicators, such as the quantity of comments or ratings exchanged, in addition to the quantity of recipes published, had a more significant impact on the popularity of recipes on Kochbar.de. Conversely, variables associated to creativity, such as the originality of the recipe, the ranking

of components in terms of popularity, and the characteristics of the picture such as saturation and image entropy, had a more significant influence on the popularity of recipes on Allrecipes.com. User interaction elements, such as the quantity of ratings and comments, were identified as the primary determinants of a recipe's ranking on Valio, the Finnish platform. Furthermore, depending on the method of classification used, nutritional factors like as energy and fat content were also identified as significant. The aforementioned results underscore the significance of cultural disparities and divergences in user behaviour across several online food networks. The authors emphasise the significance of tailoring recommendation systems and analytical approaches to align with the distinct features and preferences of each platform.

TABLE 7. Comparison of different study models and corresponding accuracy

Item	Data	Attributes	Best Methods	Accuracy	Key Finding
This study	Valio	Nutritional Difficulty User Engagement	Auto Encoder	0.95	Deep Feature Vector and visual feature extraction
Mahdi et al.	Valio	Nutritional Difficulty User Engagement	Logistic Regression Random Forest	0.93	User engagement features improve prediction substantially
Mößlang et al.	Allrecipes Kochbar	Nutritional Difficulty User Engagement Recipe novelty	Random Forest	0.89	Innovation-related features and image features have a greater influence on the popularity User engagement features were more pronounced on popularity

5. DISCUSSION

Our study indicated that non-visual features such as nutritional content, cooking complexity, and user engagement metrics significantly contribute to recipe popularity by addressing diverse user preferences and needs. Nutritional content appeals to health-conscious users, with recipes emphasizing balanced nutrition or specific dietary requirements often attracting more attention. Cooking complexity influences popularity by catering to different skill levels, with simpler recipes being more accessible to a broader audience, while more complex recipes may appeal to experienced cooks seeking a challenge. Additionally, user engagement metrics like comments, ratings, and shares provide critical insights into a recipe's social validation, often driving further visibility. This corresponds to RQ1. Deep learning models can analyze these non-visual features to predict how they impact overall recipe popularity, enabling more tailored and effective recommendations.

Visual features play a critical role in the popularity of food recipes on social media platforms. High-quality images with appealing aesthetics, such as vibrant colors, clear resolution, and well-balanced contrast, can significantly boost user engagement, as they create a visually stimulating experience that entices viewers to interact with the content. Elements like food presentation, composition, and lighting contribute to making a recipe visually appealing, which in turn can influence users to like, share, and comment on the post. Deep learning models can analyze these visual features to uncover patterns that correlate with higher engagement, thus demonstrating that visual appeal is a powerful factor in driving recipe popularity. When combined with textual content, these visual features amplify the overall attractiveness of a recipe, making it more likely to capture user attention and spread across social media, highlighting the importance of visual elements in content success which answers to our research question RQ2.

Incorporating visual features such as image contrast, resolution, color balance, and composition can significantly impact user behavior on social media and, correspondingly, the popularity of recipes. High-quality images with sharp resolution and appealing contrast can draw more attention, making users more likely to engage with the content through likes, shares, and comments. Visual appeal often serves as a first impression, influencing whether users will explore the recipe further. Factors like vibrant colors, well-lit images, and professional food styling can create a more enticing presentation, encouraging users to try the recipe and boosting its visibility on social media platforms. This underscores the importance of combining both textual and visual elements for predicting and driving recipe popularity by addressing RQ3.

Deep feature extraction can significantly enhance the prediction of recipe popularity by capturing complex and nuanced patterns from both textual and visual data. Deep learning models, such as convolutional neural networks (CNNs), can automatically extract high-level visual features like image resolution, contrast, texture, and color distribution, which are critical in influencing user engagement on social media. By leveraging deep feature extraction, these models can uncover hidden correlations between the visual appeal of images and the richness of textual descriptions, leading to more accurate and robust predictions of how well a recipe will perform in terms of likes, shares, and comments. Combining deep features from both modalities enables a more comprehensive understanding of the factors driving recipe popularity, outperforming traditional feature extraction methods. Therefore, a deep learning model can effectively leverage visual and textual features to develop a food recommender system that not only predicts recipe popularity but also considers health factors. By incorporating deep feature extraction from food images and recipe descriptions, the model can assess both the visual appeal and nutritional aspects of recipes. By aligning these deep features with user behavior patterns, such as preferences for healthier meals or engagement with specific diet-friendly recipes, the model can provide personalized recommendations that optimize both popularity and health considerations. This approach answers our research question RQ4 and also enhances user satisfaction by balancing visual appeal

with health-conscious choices, making it possible to promote healthier eating habits on social media platforms while maintaining user engagement.

The limitations of this research can be observed in several areas. Firstly, the study relies heavily on data from the Valio website, which primarily represents Finnish culinary preferences. This may limit the generalizability of the findings to other regions or cultures where dietary habits and food trends differ significantly. Furthermore, the dataset may not fully capture the diversity of global culinary preferences, restricting the applicability of the conclusions to a broader context. Another limitation is the focus on visual and non-visual features like nutritional content, preparation difficulty, and user engagement metrics, but the potential influence of external factors such as seasonal trends, marketing campaigns, or the role of influencers in shaping recipe popularity is not considered. These social factors could play a significant role in driving engagement but were not included in the analysis. Additionally, while deep learning models were used to analyze visual features, the research could have benefitted from a more diverse set of advanced machine learning techniques. Relying on AdaBoost and other models might not fully uncover complex patterns, and the study could have experimented with other neural network architectures to improve the prediction accuracy even further.

Future developments in recipe popularity prediction could involve the integration of large language models (LLMs), such as GPT-based architectures, to enhance the analysis of recipe text. LLMs can be used to better understand the context, creativity, and uniqueness of recipe descriptions, as well as user-generated content like reviews and comments. These models can generate more sophisticated textual features by identifying subtle language patterns that resonate with users, such as appealing descriptions, humor, or engaging narratives. This advanced textual analysis can contribute to more accurate predictions of how the language used in recipes influences their popularity. Additionally, LLMs can be employed to personalize recommendations by generating recipe summaries or even suggesting modifications based on user preferences and trends. Another promising direction for improving recipe popularity prediction is the incorporation of emotional insight into the model. By

analyzing user reactions to recipes, such as sentiments expressed in comments, reviews, or social media posts, deep learning models can gain a better understanding of the emotional responses' recipes evoke. This can include detecting positive emotions like excitement or satisfaction, or negative emotions such as frustration with complex instructions. Understanding these emotional cues can help the model refine its predictions of recipe popularity by capturing not just what users engage with, but why they engage. Emotional insight can also help tailor recommendations to align with users' mood states, creating a more empathetic and effective recommender system. Furthermore, future research could involve incorporating social influence and trend dynamics into recipe popularity prediction models. Recipe popularity on social media is often driven by viral trends, influencer endorsements, and peer recommendations. By integrating social network analysis, the model can capture how interactions between users, influencers, and communities impact recipe visibility and engagement. Additionally, tracking the evolution of food trends—such as seasonal ingredients, cultural food movements, or popular diets—can help the model stay updated on shifting user preferences. This social context can be combined with visual and non-visual features to provide more dynamic and timely predictions of recipe popularity, allowing for more effective recommendations that capitalize on emerging trends.

6. CONCLUSION

Providing a brief overview of the results and addressing any constraints in the methodology, this chapter functions as the thesis's conclusion. Furthermore, it offers a perspective on possibilities for improvement and the direction of future research. The primary aim of this master's thesis is to provide deeper understanding of the hidden patterns and processes that control the popularity of internet recipes. Gaining a thorough comprehension of these sociodynamic processes has the capacity to enable the creation of sophisticated, health-oriented recommender systems. Such technologies have the potential to tackle the increasing food-related health issues that are widespread in modern society. The methodology used in this study was a statistical examination of datasets obtained from the Valio website. This platform embodies the unique culinary traditions of Finland. A more comprehensive picture of the mechanisms influencing recipe popularity was obtained via comparison analysis. The assessments were based on recipe attributes (nutritional and preparation difficulty) as well as user involvement functions (such as ratings and comments) and visual representations associated with food. On order to evaluate the ability of these traits to make accurate predictions and confirm the statistical results, we carried out predictive modelling studies.

This research offers valuable insights into the factors that affect the popularity of online recipes, with a specific focus on data from the Valio website. By analyzing both visual features from food images and non-visual features like nutritional content, cooking complexity, and user engagement, the study emphasizes the importance of these factors in determining the appeal of recipes to users. The findings reveal that visual aspects, particularly those linked to food presentation, play a significant role in attracting users and increasing engagement on digital platforms, suggesting that high-quality images are key to improving a recipe's visibility and appeal.

Additionally, the study shows that non-visual elements, such as nutrition and ease of preparation, also have a significant impact on recipe popularity. Healthier or easier-to-prepare recipes tend to attract a broader audience, highlighting users' preference for convenience and nutrition. Moreover, user engagement metrics—such as the number of comments, ratings, and sentiment expressed in reviews—strongly influence how popular a recipe becomes. When these features were combined, prediction models achieved better accuracy, with AdaBoost and other machine learning techniques showing strong performance.

The research also highlights the potential of deep learning methods to further improve prediction accuracy. By applying deep feature extraction, the study demonstrates how advanced models can capture complex relationships in both visual and non-visual data, leading to more accurate predictions of recipe popularity. Particularly in phase 4 of the study, the use of deep learning produced the highest accuracy, indicating that these models are effective in identifying connections between visual appeal and user interaction.

Despite these positive results, the research has some limitations. The focus on data from Finland’s Valio website may limit the applicability of the findings to other cultural contexts. Additionally, the study does not account for social factors such as viral trends or influencer recommendations, which could have influenced the analysis. Future studies should address these limitations by using a broader range of data and investigating how social influence affects recipe popularity.

In conclusion, this research demonstrates that both visual and non-visual factors are crucial in determining the popularity of online recipes. By employing advanced machine learning and deep learning techniques, the study provides a better understanding of the factors driving user engagement with food content. These results have important implications for the creation of food recommender systems that promote healthier eating habits while meeting users' preferences for visually appealing and easy-to-prepare dishes. Future research should explore the integration of emotional analysis, social influence, and large language models to further improve recipe prediction models and deliver more personalized recommendations.

7. REFERENCES

Abbar, S., Mejova, Y. and Weber, I. (2015) 'You tweet what you eat: Studying food consumption through twitter', in *Conference on Human Factors in Computing Systems - Proceedings*. doi: 10.1145/2702123.2702153.

Abdel-Hakim, A. E. and Farag, A. A. (2006) 'CSIFT: A SIFT descriptor with color invariant characteristics', in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. doi: 10.1109/CVPR.2006.95.

Abdurrahman, M. H., Irawan, B. and Setianingsih, C. (2020) 'A Review of Light Gradient Boosting Machine Method for Hate Speech Classification on Twitter', in *ICECIE 2020 - 2020 2nd International Conference on Electrical, Control and Instrumentation Engineering, Proceedings*. doi: 10.1109/ICECIE50279.2020.9309565.

Ahn, J. (2011) 'Digital divides and social network sites: Which students participate in social media?', *Journal of Educational Computing Research*, 45(2). doi: 10.2190/EC.45.2.b.

Ahn, Y. Y. *et al.* (2011) 'Flavor network and the principles of food pairing', *Scientific Reports*, 1. doi: 10.1038/srep00196.

Al-Zoubi, A. M. *et al.* (2018) 'Evolving Support Vector Machines using Whale Optimization Algorithm for spam profiles detection on online social networks in different lingual contexts', *Knowledge-Based Systems*, 153. doi: 10.1016/j.knosys.2018.04.025.

Arnett, J. J. (2000) 'Emerging adulthood: A theory of development from the late teens through the twenties', *American Psychologist*, 55(5). doi: 10.1037/0003-066X.55.5.469.

Arnett, J. J. (2007) 'Emerging Adulthood: What Is It, and What Is It Good For?', *Child Development Perspectives*, 1(2). doi: 10.1111/j.1750-8606.2007.00016.x.

Athanasiou, V. and Maragoudakis, M. (2017) ‘A novel, gradient boosting framework for sentiment analysis in languages where NLP resources are not plentiful: A case study for modern Greek’, *Algorithms*, 10(1). doi: 10.3390/a10010034.

Aufar, M., Andreswari, R. and Pramesti, D. (2020) ‘Sentiment Analysis on Youtube Social Media Using Decision Tree and Random Forest Algorithm: A Case Study’, in *2020 International Conference on Data Science and Its Applications, ICoDSA 2020*. doi: 10.1109/ICoDSA50139.2020.9213078.

Breiman, L. (2001) ‘Random forests’, *Machine Learning*, 45(1). doi: 10.1023/A:1010933404324.

De Choudhury, M., Sharma, S. and Kiciman, E. (2016) ‘Characterizing dietary choices, nutrition, and language in food deserts via social media’, in *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. doi: 10.1145/2818048.2819956.

Christakis, N. A. and Fowler, J. H. (2007) ‘The Spread of Obesity in a Large Social Network over 32 Years’, *New England Journal of Medicine*, 357(4). doi: 10.1056/nejmsa066082.

Chunara, R. *et al.* (2013) ‘Assessing the Online Social Environment for Surveillance of Obesity Prevalence’, *PLoS ONE*, 8(4). doi: 10.1371/journal.pone.0061373.

Cortes, C. and Vapnik, V. (1995) ‘Support-Vector Networks’, *Machine Learning*, 20(3). doi: 10.1023/A:1022627411411.

Cox, D. R. (1972) ‘Regression Models and Life-Tables’, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 34(2). doi: 10.1111/j.2517-6161.1972.tb00899.x.

Elsweiler, D., Trattner, C. and Harvey, M. (2017) ‘Exploiting food choice biases for healthier recipe recommendation’, in *SIGIR 2017 - Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. doi: 10.1145/3077136.3080826.

Fried, D. *et al.* (2014) ‘Analyzing the language of food on social media’, in *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014*. doi: 10.1109/BigData.2014.7004305.

Friedman, J. H. (2001) ‘Greedy function approximation: A gradient boosting machine’, *Annals of Statistics*, 29(5). doi: 10.1214/aos/1013203451.

Glanz, K. *et al.* (1998) ‘Why Americans eat what they do: Taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption’, *Journal of the American Dietetic Association*, 98(10). doi: 10.1016/S0002-8223(98)00260-0.

Harvey, M., Ludwig, B. and Elsweiler, D. (2013) ‘You are what you eat: Learning user tastes for rating prediction’, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi: 10.1007/978-3-319-02432-5_19.

Hermann, C. (2001) ‘The International Commission on Illumination - CIE: What It Is and How It Works’, *Symposium - International Astronomical Union*, 196. doi: 10.1017/s0074180900163831.

Hoelscher, D. M. *et al.* (2002) ‘Designing effective nutrition interventions for adolescents.’, *Journal of the American Dietetic Association*. doi: 10.1016/s0002-8223(02)90422-0.

Hossain, M. A. and Alam Sajib, M. S. (2019) ‘Classification of Image using Convolutional Neural Network (CNN)’, *Global Journal of Computer Science and Technology*. doi: 10.34257/gjcstdvol19is2pg13.

Hsu, C. C. *et al.* (2017) ‘Social media prediction based on residual learning and random forest’, in *MM 2017 - Proceedings of the 2017 ACM Multimedia Conference*. doi: 10.1145/3123266.3127894.

Huang, F. *et al.* (2018) ‘Random forest exploiting post-related and user-related features for social media popularity prediction’, in *MM 2018 - Proceedings of the 2018 ACM Multimedia Conference*. doi: 10.1145/3240508.3266439.

Huang, K. Q., Wang, Q. and Wu, Z. Y. (2006) ‘Natural color image enhancement and evaluation algorithm based on human visual system’, *Computer Vision and Image Understanding*, 103(1). doi: 10.1016/j.cviu.2006.02.007.

Karthika, P., Murugeswari, R. and Manoranjithem, R. (2019) ‘Sentiment Analysis of Social Media Network Using Random Forest Algorithm’, in *IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing, INCOS 2019*. doi: 10.1109/INCOS45849.2019.8951367.

Kavya, R. and Harisha (2015) ‘Feature Extraction Technique for Robust and Fast Visual Tracking : A Typical Review’, *International Journal of Emerging Engineering Research and Technology*, 3(1).

Ke, Y., Sukthankar, R. and Hebert, M. (2005) ‘Efficient visual event detection using volumetric features’, in *Proceedings of the IEEE International Conference on Computer Vision*. doi: 10.1109/ICCV.2005.85.

Khanday, A. M. U. D., Khan, Q. R. and Rabani, S. T. (2021) ‘SVMBPI: Support Vector Machine-Based Propaganda Identification’, in. doi: 10.1007/978-981-16-1056-1_35.

Kusmierczyk, T. and Nørvåg, K. (2016) ‘Online food recipe title semantics: Combining nutrient facts and topics’, in *International Conference on Information and Knowledge Management, Proceedings*. doi: 10.1145/2983323.2983897.

Kusmierczyk, T., Trattner, C. and Nørvag, K. (2015) ‘Temporal patterns in online food innovation’, in *WWW 2015 Companion - Proceedings of the 24th International Conference on World Wide Web*. doi: 10.1145/2740908.2741700.

Levenson, J. C. *et al.* (2016) ‘The association between social media use and sleep disturbance among young adults’, *Preventive Medicine*, 85. doi: 10.1016/j.ypmed.2016.01.001.

Liu, Y. *et al.* (2009) ‘Contour-motion feature (CMF): A space-time approach for robust pedestrian detection’, *Pattern Recognition Letters*, 30(2). doi: 10.1016/j.patrec.2008.03.007.

Macht, M. (2008) ‘How emotions affect eating: A five-way model’, *Appetite*. doi: 10.1016/j.appet.2007.07.002.

Mahdi, A (2023) 'Recipe popularity prediction in Finnish social media by machine learning models', https://urn.fi/URN:NBN:fi:oulu-202310133120.

Mikkilä, V. *et al.* (2005) ‘Consistent dietary patterns identified from childhood to adulthood: The Cardiovascular Risk in Young Finns Study’, *British Journal of Nutrition*, 93(6). doi: 10.1079/bjn20051418.

Neelakandan, S. and Paulraj, D. (2020) ‘A gradient boosted decision tree-based sentiment classification of twitter data’, *International Journal of Wavelets, Multiresolution and Information Processing*, 18(4). doi: 10.1142/S0219691320500277.

Nelson, M. C. *et al.* (2008) ‘Emerging adulthood and college-aged youth: An overlooked age for weight-related behavior change’, *Obesity*. doi: 10.1038/oby.2008.365.

Niemeier, H. M. *et al.* (2006) ‘Fast Food Consumption and Breakfast Skipping: Predictors of Weight Gain from Adolescence to Adulthood in a Nationally Representative Sample’, *Journal of Adolescent Health*, 39(6). doi: 10.1016/j.jadohealth.2006.07.001.

Oliver, G., Wardle, J. and Gibson, E. L. (2000) ‘Stress and food choice: A laboratory study’, *Psychosomatic Medicine*, 62(6). doi: 10.1097/00006842-200011000-00016.

Pedro, J. S. and Siersdorfer, S. (2009) ‘Ranking and classifying attractiveness of photos in folksonomies’, in *WWW’09 - Proceedings of the 18th International World Wide Web Conference*. doi: 10.1145/1526709.1526813.

Prescott, J. *et al.* (2002) ‘Motives for food choice: A comparison of consumers from Japan, Taiwan, Malaysia and New Zealand’, *Food Quality and Preference*, 13(7–8). doi: 10.1016/S0950-3293(02)00010-1.

Rguibi, Z. *et al.* (2022) ‘CXAI: Explaining Convolutional Neural Networks for

Medical Imaging Diagnostic', *Electronics (Switzerland)*, 11(11). doi: 10.3390/electronics11111775.

Rokicki, M. *et al.* (2016) 'Plate and prejudice: Gender differences in online cooking', in *UMAP 2016 - Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*. doi: 10.1145/2930238.2930248.

Said, A. and Bellogín, A. (2014) 'You are what you eat! Tracking health through recipe interactions', in *CEUR Workshop Proceedings*.

Scheibehenne, B., Miesler, L. and Todd, P. M. (2007) 'Fast and frugal food choices: Uncovering individual decision heuristics', *Appetite*, 49(3). doi: 10.1016/j.appet.2007.03.224.

Shay, C. M. *et al.* (2013) 'Status of Cardiovascular Health in US Adolescents', *Circulation*, 127(13). doi: 10.1161/circulationaha.113.001559.

Stafleu, A. *et al.* (1991) 'A review of selected studies assessing social-psychological determinants of fat and cholesterol intake', *Food Quality and Preference*. doi: 10.1016/0950-3293(91)90033-B.

Steptoe, A., Pollard, T. M. and Wardle, J. (1995) 'Development of a measure of the motives underlying the selection of food: The food choice questionnaire', *Appetite*, 25(3). doi: 10.1006/appe.1995.0061.

Trattner, C., Moesslang, D. and Elsweiler, D. (2018) 'On the predictability of the popularity of online recipes', *EPJ Data Science*, 7(1). doi: 10.1140/epjds/s13688-018-0149-5.

Trattner, C., Parra, D. and Elsweiler, D. (2017) 'Monitoring obesity prevalence in the United States through bookmarking activities in online food portals', *PLoS ONE*, 12(6). doi: 10.1371/journal.pone.0179144.

Viola, P. and Jones, M. (2001) 'Rapid object detection using a boosted cascade of simple features', in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. doi: 10.1109/cvpr.2001.990517.

Wagner, C. and Aiello, L. M. (2015) 'Men eat on mars, women on venus? An

empirical study of food-images', in *Proceedings of the 2015 ACM Web Science Conference*. doi: 10.1145/2786451.2786505.

Wang, H. *et al.* (2009) 'Evaluation of local spatio-temporal features for action recognition', in *British Machine Vision Conference, BMVC 2009 - Proceedings*. doi: 10.5244/C.23.124.

Van De Weijer, J., Gevers, T. and Bagdanov, A. D. (2006) 'Boosting color saliency in image feature detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1). doi: 10.1109/TPAMI.2006.3.

Zandstra, E. H., De Graaf, C. and Van Staveren, W. A. (2001) 'Influence of health and taste attitudes on consumption of low- and high-fat foods', *Food Quality and Preference*, 12(1). doi: 10.1016/S0950-3293(00)00032-X.

Zhang, D., Islam, M. M. and Lu, G. (2012) 'A review on automatic image annotation techniques', *Pattern Recognition*, 45(1). doi: 10.1016/j.patcog.2011.05.013.