JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
ICASET
2024

Research Article
Collection: Intelligent Computing of Applied Sciences and Emerging Trends

Enhancing Food Safety: A Machine Learning Approach for Accurate Detection and Classification of Food Allergens

Hafsah Shaukat[1], Arooj Sultan[1], and Humayun Salahuddin[1*]

[1]Department of Computer Sciences,NCBA&E (Sub-Campus), Multan, 60000, Pakistan.
*Corresponding Author: Humayun Salahuddin. Email: humayun.mul@ncbae.edu.pk

Academic Editor: Salman Qadri **Published:** April 01, 2024

Abstract. Food allergies can severely impact quality of life for patients in certain situations. Diagnosing food allergies before exposure is expensive and risks over diagnosis. The identification of allergenic foods reveals significant connections within food allergy data, ultimately benefiting patients. The model utilize machine learning technique for detection and classification of food allergens. Food allergen detection model is proposed that predicts if foods contain allergens. The model facilitates doctors and nutritional experts in formulating a food list by promoting the retention of non-allergenic foods, thereby mitigating the risk of nutritional deficits. The model scores food descriptors and tests their classification performance. Random forest, support vector machine, and k-nearest neighbors models achieved training and validation accuracies of 96.8%, 93.54%, and 95.16%, respectively. Decision tree achieved the highest evaluation accuracy at 98.4%. This proposes an approach to understand food allergens' nature. Conducting an in-depth analysis of allergen prevalence allows to acquire valuable insights into the occurrence of different allergens in various food products. The system has the potential to empower personalized recommender systems for individuals with dietary restrictions or allergies, enhancing their ability to make informed and safe food choices.

Keywords: Food allergy; Prediction; Machine Learning; Random Forest; Decision Tree; Allergy detection; K-Nearest Neighbors.

1. Introduction

Food allergies are a serious health risk that affect people all over the world in a time when culinary diversity has no boundaries. A thorough analysis of the underlying mechanisms, contributory variables, and far-reaching repercussions for affected individuals and society at large has become necessary due to the increasing prevalence of allergic reactions to different dietary items [1]. The significant effect that food allergies have on patients' quality of life has become a critical issue in the complex field of healthcare. In addition to being extremely expensive, the conventional method of identifying food allergies prior to exposure carries a high chance of over diagnosis, making it an extremely difficult task [2].

A food allergy patient's quality of life may be greatly impacted. Avoiding foods to which they are allergic is still the safest course of action currently available. Nutritional problems may occur as a result of removing food allergens, proteins in foods, or compounds that trigger aberrant immune responses [3]. There is often a lack of information regarding food allergies, which forces patients and their caregivers to create their own safety plans and increases social isolation and discrimination. Allergies in toddlers, adolescents, or adults can occasionally result in deadly reactions [4].

A wide range of tests are used to discover food allergies in order to pinpoint the precise allergens that cause negative reactions. Commonly used are skin prick tests, in which an allergen is applied topically in modest doses, and any ensuing skin reactions are noted. Immunoglobulin E (IgE) and other particular antibodies are measured in blood tests, which shed light on how the immune system reacts to allergens. Advanced laboratory techniques are utilized to diagnose allergies, especially food allergies. These procedures include the BAT (Basophil Activation Test) and MAT (Mast Cell Activation Test). The assessment of certain immune cells' activation in response to allergens is the main objective of both tests. Mast cell activation syndrome (MCAS), which can present with a variety of allergy-like symptoms, is one illness that the MAT test can diagnose. Under strict physician supervision, oral food challenges entail introducing potential allergens progressively in order to watch for negative reactions [1-6]. Identification of trigger foods is aided by elimination diets, in which possible allergens are methodically eliminated and then reintroduced. Additionally, by locating certain protein components inside allergens, component-resolved diagnostics (CRD) offer a more thorough investigation. All of these tests are essential for correctly identifying food allergies, which helps medical practitioners to customize treatment plans and gives people the power to make educated food choices [8-9].

Food allergies can cause reactions that range in severity from mild to severe, and in rare cases, they may even be fatal. Localized symptoms like itching, hives, or minor swelling are common in mild responses and usually appear on the skin or in the vicinity of the face. Gastrointestinal symptoms such as nausea, vomiting, or abdominal discomfort can also be considered moderate responses [10-24]. There may also be manifestations of respiratory problems such nasal congestion or wheezing. The most serious reactions, referred to as anaphylaxis, can cause a broad, fast immune response that impacts several organ systems. When anaphylactic reactions are not treated with adrenaline right once, they can be fatal and cause breathing difficulties, a reduction in blood pressure, and loss of consciousness [21-22].

The environment of food allergies is mostly shaped by a small number of allergens that are consistently the most common causes of unfavourable immune responses. Known as the "top food allergens," As shown in figure 1 this small group consists of fish, shellfish, eggs, milk, peanuts, tree nuts, and wheat. Since these allergens are known to be widely present in a variety of food products, it might be difficult for those who are sensitive to them to avoid them. Allergy reactions can range in intensity from minor symptoms to potentially fatal anaphylaxis [5].

Estimating the risk of severe reactions is one of the most important areas of unmet knowledge in the management of food allergy patients. It is rare for severe food allergies to cause almost deadly reactions. With a confidence interval of 0.94 to 3.45, the estimated incidence of mortality from food-related allergic reactions is 1.81 per million person-years. Though still rare, cases of almost fatal anaphylaxis necessitating intensive care support occur around ten times more frequently. [6]. But these kind of extreme responses are erratic. The majority of individuals who pass away from fatal food anaphylaxis only had minor reactions in the past. Because we can't tell who is more likely to experience severe reactions, persons with food allergies are sometimes treated as though they have an equal chance of dying from their allergies. This may result in unneeded worry, severe food restrictions, and worsened health [7].

Due to the immature state of their immune systems and digestive tracts, children are more susceptible to food allergies. Most food allergies, like those to milk or eggs, first manifest in childhood and then go away as the child becomes older. Food allergies in children are a dangerous, potentially fatal illness that significantly lowers quality of life for both sufferers and those who care for them [8-9]. The academy for allergy and clinical immunology in Europe (EAACI) reports that during the preceding ten years, the proportion of children with allergies has doubled [10]. It was estimated that 8% of children in the US suffered from food allergies in 2011; among these, approximately 40% had experienced severe responses in the past [11]. Thirty percent of children with dietary allergies are also allergic to multiple foods. [12]. In the United States, every three minutes an allergic food reaction sends a sufferer to the emergency room [13]. The World Allergy Organization's White Book of Allergies states that 20–30% of people worldwide currently suffer from one or more allergies due to the rise in the incidence of allergic disorders [14-15].

Figure 1. Top 8 food allergens.

In Lebanon, a survey was carried out to find out about food allergies and allergens among those with a diagnosis of food allergies. According to reports, the most common food allergen is wheat (15.7%), and the most common symptoms are rash and itchy skin (71.4% and 68.6%, respectively). Of the participants, only 34 (48.6%) said they carried medication, 58 (70.7%) received a diagnosis by blood testing, and 22 (31.4%) said they knew a great deal about the subject. The average knowledge score of the participants was 84.2 ~ 11.5%. The average score for participants in terms of best practices was 47.8 ~ 28.3% [16].

For food allergies, there is currently no known cure, and the mainstay of management is to avoid the particular food(s) that cause the allergy. Strict avoidance is the cornerstone of treatment for patients with top food allergies, including those to peanuts, tree nuts, milk, eggs, soy, wheat, fish, and shellfish. This entails reading food labels closely, having open lines of contact with restaurant employees, and being hyperaware of any allergy sources.

Over the last few years, it has been demonstrated that the most frequent unfavorable response to dietary components is food allergies. Accurate and comprehensive information on food ingredients is crucial because the only method to prevent allergic phenomena from occurring is to avoid allergenic foods. Within this framework, the provision of quick, economical, accessible, and trustworthy analytical techniques for the identification of particular allergen contents in food items is required and essential.

A system is introduced in order to prevent food-allergic reactions in daily life. The concept encourages the retention of non-allergenic foods, reducing the likelihood of nutritional deficiencies and helping physicians and nutrition specialists create a food list for patients. The restaurants and food sector accommodate food allergens for identification of specific allergic ingredients for safe food choices. Practitioners are frequently asked to advise patients in advance about managing food allergies in public settings. The purpose of this work is to present a thorough analysis and evaluation of the relevant research as well as a thorough synopsis of certain preventative measures that practitioners may suggest and that patrons of restaurants and patients may use. Different machine learning techniques used for validation and testing random forest achieved the highest evaluation accuracy at 98.4%.

2. Literature Review

Food allergies have a profound impact on the dynamic field of healthcare, greatly influencing patients' quality of life, particularly in emergency situations. The traditional approach to food allergy identification is preventive and expensive, which puts financial strain on the system and raises the risk of over diagnosis. With the use of machine learning, many researchers introduced innovative model provides a promising answer to the difficult problem of precisely identifying and categorizing food allergies. Its importance comes not just from identifying foods that trigger allergies but also from illuminating deep relationships within the intricate network of information about food allergies.

The most common chronic pediatric disorders in Europe are allergy-related diseases, which are the result of intricate interactions between hereditary and environmental factors. Applying machine learning (ML) techniques provides an innovative way to analyze this complex data. Thus, the goal of this pilot research was to find predictors for the occurrence of parental-reported allergy at 4-6 years of age using feature selection in machine learning. The machine learning research employed a total of 248 features and 130 youngsters, 116 without and 14 with parental reports of allergies. The recursive ensemble feature selection approach (REFS) yielded twenty characteristics, with the Multi-layer Perceptron Classifier showing the highest area under the curve (AUC) of 0.86 (SD 0.08). The use of tobacco during pregnancy, atopic parents, gestational age, days of fever, diarrhea, coughing, and rash in the first year of life, and previous exposure to antibiotics, The traits that were indicative of allergies were resistin, IL-27, MMP9, CXCL8, CCL13, Vimentin, IL-4, CCL22, GAL1, IL-6, LIGHT, and GMCSF [1].

Due to the shortcomings of current clinical testing, Oral Food Challenges (OFCs) are crucial for precisely diagnosing food allergies. However, some patients are unwilling to undergo OFCs, and those who are willing may find it difficult to visit allergists in rural or community healthcare settings [2]. From 1,112 individuals with 1,284 OFCs, retrospective data on clinical characteristics including serum-specific immunoglobulin E (IgE), total IgE, skin prick tests (SPTs), comorbidities, sex, and age was gathered. These attributes served as the basis for the development of multiple machine learning models that forecasted the results of OFCs for milk, eggs, and peanuts—three frequent allergens. The best-performing model for each allergy was an ensemble of random forest (egg) or Learning Using Concave and Convex Kernels (LUCCK) (peanut, milk) models, with an Area under the Curve (AUC) of 0.91, 0.96, and 0.94 for predicting OFC results for the corresponding allergens. Additionally, the sensitivity and specificity of each of these models were higher than 89%. Certain IgE and the wheal and flare values from SPTs are highly predictive of the results of OFCs, according to the model's interpretation utilizing SHapley Additive exPlanations (SHAP).

The results of the investigation suggest that ensemble learning might be able to predict OFC outcomes and pinpoint relevant clinical traits. [3].

The population-based Melbourne HealthNuts and SchoolNuts projects used oral food challenges (OFC), the gold standard for detecting food allergies (FA), to quantify food allergy. The study found that the prevalence of FA was 4%–5% in older children and early teens and 10% in neonates. Based on cross-sectional population-based studies carried out in the United States, as depicted in Figure 2, the estimated prevalence of FA in adults is 11%, whereas in children it is around 8% [21]. Based on data from the EuroPrevall-iFAAM birth cohort, prevalence in children was shown to be much lower across Europe, ranging from 1.4% to 3.8%. Aside from having a significant financial impact, food allergies also significantly lower quality of life. Even though the first approved treatment for peanut allergy is now available, the current standard of care for other food allergies remains strict avoidance. The development of biologics and other state-of-the-art medications, together with advancements in the field of allergen immunotherapy, have all contributed to the search for safe and effective treatment alternatives for food allergies [9].

Foods that the patient is not allergic to should be avoided since diagnosing food allergies before the patient is exposed to the allergen or allergens is expensive and can lead to over diagnosis. By identifying their food allergens and preventing the need for expensive diagnostics, patients would benefit from the correlations between characteristics of food allergy data. If the food sIgE analysis or SPT result is positive for one or more items in the list generated using the association criteria, then the molecular test performance should be discarded. Since the molecular test is typically far more expensive than SPT and far more expensive than blood testing, this benefits the patient monetarily. Refrain from removing items that don't hurt them and stop a nutritional deficit from occurring, which will enable medical assistance. Use apriori technique on test results from patients with food allergies and additional information about other foods that trigger allergic reactions will be added to the training data. With this information, the patient is advised to use the CRD tool to obtain a diagnosis if, following the SPT test or blood test for a particular food IgE, they also test positive for other food elements that aren't on the list compiled using the association rules [4].

Numerous studies have suggested a link between an infant's gut microbiota and the development of allergies. Using Long Short-Term Memory (LSTM) networks, longitudinal gut microbiome profiles may be utilized to predict food allergies in the early years (0–3 years). Using the DIABIMMUNE dataset, researchers show that our approach outperforms Hidden Markov approach, Multi-Layer Perceptron Neural Network, Support Vector Machine, Random Forest, and LASSO regression in terms of predictive ability. Additionally, the usefulness of employing smaller representations of microbiological traits in LSTM network training is evaluated. We considered classical feature selection techniques based on variance and Minimum Redundancy Maximum Relevance (mRMR) before training LSTM networks. Additionally, we used a sparse autoencoder to extract potential latent representations. The thorough analysis shows that when compared to the other machine learning models that have been examined, LSTM networks with the mRMR specified features perform noticeably better [5].

Asthma is a frequent concurrent illnesses, according to case studies including food-induced death or near-death incidents [6].As a result, asthma has been linked to risk, which makes sense considering that respiratory impairment is the primary mechanism causing catastrophic consequences in food allergies. Asthma is prevalent, though, and more than 50% of people with food allergies have it. Because of this and the limited frequency of fatal responses, asthma has little predictive value for severe reactions. More than 99.9% of people with asthma who are allergic to foods will never have a reaction that is actually fatal [7].

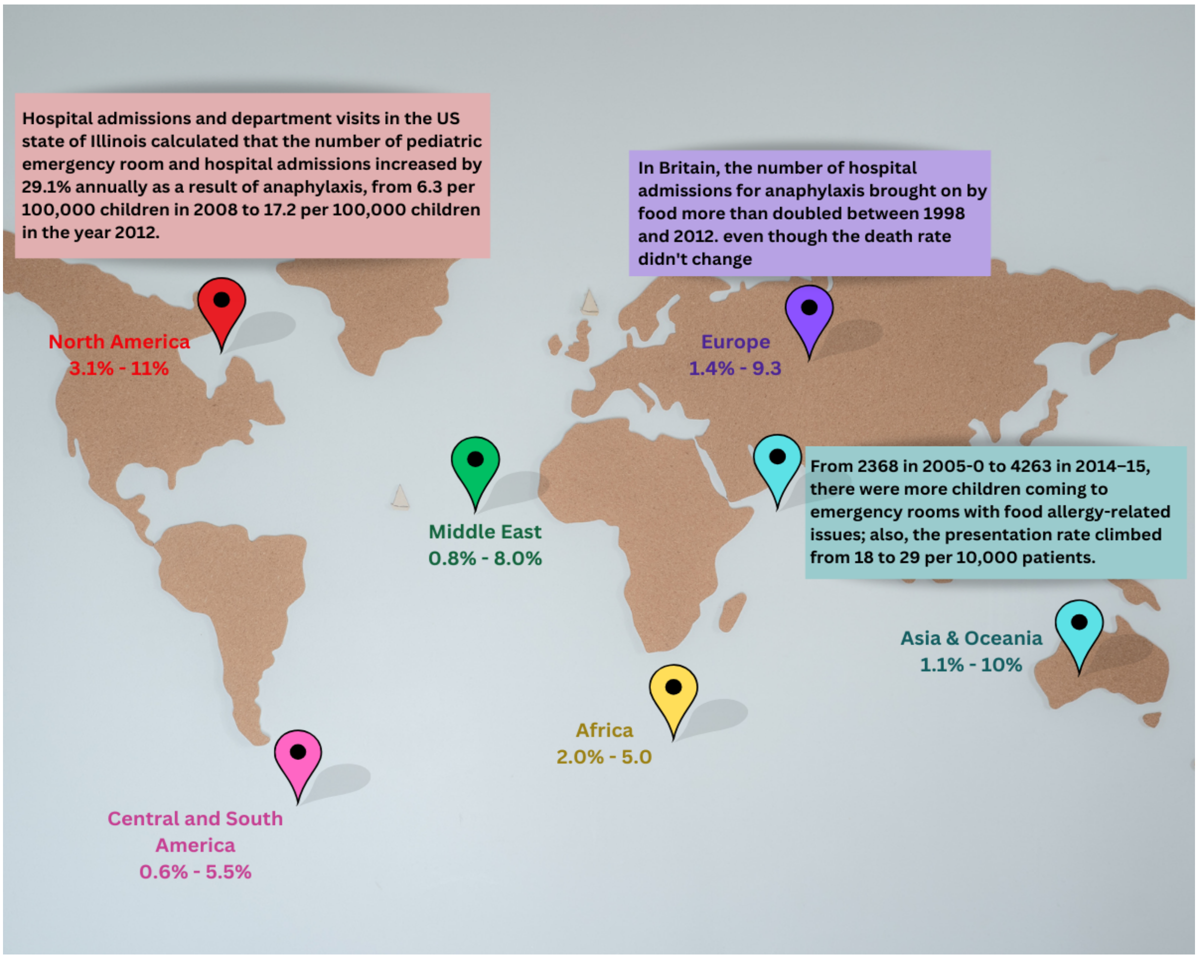

Figure 2. Prevalence of food allergy varies globally.

Researchers propose to apply novel ensemble learning models, eXtreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM), in conjunction with the pre-training BERT deep learning model to predict the allergic reactions of dietary proteins. Excellent findings are obtained from extensive experiments. They were better than the earlier research. The findings showed that the accuracy was 0.9310 and the AUC value of BERT (the top performer) was 0.9578. The tests were carried out in order to evaluate and compare the various models' attributes and to offer recommendations for relevant circumstances. This study is the first to be published that uses the aforementioned technique to determine the allergenicity of food proteins [8].

It is necessary to conduct more research on the connections between food allergies and obesity and psychiatric comorbidities such as autism, depression, and attention deficit/hyperactivity disorder [10]. The main causes of food allergies are certain foods and food products, which are frequently linked to causing allergic reactions [23]. That group of food products are listed in figure3.

Doctors in Canada found that 2.53% of children had food allergies (95% Confidence Interval, 2.48%-2.59%). The most common food allergy reported by children was peanuts (0.8%), which was followed by tree nuts (0.6%), eggs (0.3%), fruit (0.2%), finned fish (0.2%), and shellfish (0.2%). Only 33.7% of kids with dietary allergies got a prescription for an auto injectable epinephrine pump. A logistic regression analysis found that children with food allergies were more likely to have an atopic comorbidity (OR, 2.20; 95% CI, 2.06-2.35) and less likely to be overweight (OR, 0.84; 95% CI, 0.78-0.90). In the age- and sex-adjusted models, patients with food allergies were significantly more likely to have one of the following mental illnesses: anxiety (CI, 1.63-2.19), depression (OR, 1.17; 95% CI, 1.02-1.35), attention deficit/hyperactivity disorder (OR, 1.81; 95% CI, 1.66-1.96), and autism (OR, 1.89; 95%). [11].

Figure 3. Food and food products associated with food allergy.

Disorders are handled by the computer-aided allergy diagnosis method developed by researchers. The technique was developed using datasets collected from allergy testing facilities in South India. After the results of 878 patients' intradermal skin tests were recorded, it was discovered that there weren't many samples for coexisting conditions in the data. Modified data sampling techniques were applied to rectify this data imbalance and boost the learning algorithms' efficacy. The algorithms were cross-validated in order to determine which trained model was optimal for multi-label classification. The system test was to confirm the performance of a trained random forest model using test data. The training and validation accuracy rates of the decision tree, random forest, and support vector machine are 81.62, 81.04, and 83.07, respectively. During the evaluation process, random forest showed an overall accuracy rate of 86.39, with a sensitivity rate of 75% for the comorbid Rhinitis-Urticaria class. Every feature of the framework was designed to function on mobile devices. The average performance of the physicians was 77.21% before and 81.80% after using the decision assistance system, respectively. The diagnostics system's mobile application-integrated knowledge source allows junior clinicians to verify their diagnostic theories [14].

The survey comprised 70 valid questionnaires in total was conducted in Lebanon. Males were less than females (81.4% vs. 19.6%). Furthermore, 58.6% of participants were residents of Beirut's capital, 72.9% had a university degree, 30% had a background in health-related education, and 62.9% of participants were under 45. As far as their medical history goes, 70.7% said that a blood test (an IgE specific to an allergen or a total serum IgE) was used to identify them. Furthermore, it was found that 21.4% of participants had members of their immediate family with a diagnosis of food allergies, 54.4% had had their first allergic response before turning 21, and 54.3% had been diagnosed earlier [16].

Fish, wheat, and shellfish were the top 3. When it comes to seafood, shrimp was the most frequently reported, followed by tuna, eggplant, almonds, strawberries, and pepper under trees and nuts, respectively. The most prevalent symptoms of a food allergy were itchy skin and rash (71.4% and 68.6%, respectively). Only one (1.4%) respondent reported visiting the emergency room, while 52 (74.3%) reported having no food allergic reactions in the year before to the survey's closing date. 67 (95.7%) reported not missing any workdays due to a food allergy. Conversely, 48 people (68.6%) made at least one ER visit in their [16].

Food intolerances, both immune-mediated and non-immune-mediated, are common unpleasant reactions that arise from eating specific foods. Since there is now no known cure for these responses, adopting a rigorous diet that eliminates items that cause allergies or intolerances is a preventive step [17], [19]. Dairy, eggs, fish, shellfish, tree nuts, peanuts, soybeans, and wheat-based goods are among the common food allergies. Foods heavy in gluten, biogenic amines, and specific dietary additives can be harmful to people who are lactose intolerant. Uncertainty surrounds the specific mechanisms underlying these dietary allergies and intolerances [18], [17].

The food industry and regulatory bodies are developing risk-based strategies for controlling allergens to assist allergy patients in avoiding inadvertent exposure. High-quality clinical data to determine safe food quantities and analytical techniques to quantify allergenic proteins are needed to implement such initiatives [21], [22]. By creating a prototype reference method that uses harmonized quantitative mass spectrometry to identify different food allergens in standardized matrices such broth powder and chocolate, the Thrall project seeks to assist these strategies [24]. Hen's eggs, peanuts, soybeans, hazelnuts, and almonds are all found in these matrices. In order to determine minimal eliciting doses and thresholds, the project's second purpose is gathering and organizing data on oral food challenges. The established protocols for gathering data on allergic foods will close current gaps [25].

Uncovering the molecular pathways underlying food allergies has progressed significantly in the last few decades. This improved knowledge has made it possible to build cutting-edge treatment approaches, prevention measures, and diagnostic instruments. These scientific advancements represent a hopeful evolution in the field and give promise for improved outcomes and quality of life for those affected by food allergies. They are also crucial in helping us recognize, monitor, and potentially treat these disorders. Figure 4 illustrates the urgent need for advancements in diagnostics, prophylactic measures, and treatments given the increased incidence of food allergies [9].

"Adverse reactions to food" is the broad phrase used to describe abnormal reactions associated with food consumption. The academy of allergens in Europe classifies them as either toxic or non-toxic based on the response mechanism. All people experience the same adverse effects from toxic responses, which are thought to be dose dependent. In certain circumstances, poisoning occurs right after eating. Non-toxic reactions are categorized as immunological (food allergy) and non-immunological (food intolerance) and are attributed to an individual's predisposition. They are also typically not dose dependent.

Recent years have seen a progressive rise in the prevalence of food allergies, making them a serious global public health concern. Food allergies can produce severe symptoms that significantly impair a person's quality of life. New foods may include unidentified allergen components, and cross-contamination during food preparation can result in hidden allergens that represent a substantial risk to allergy sufferers. Therefore, in order to detect or examine allergic components on-site and determine the likelihood of an allergy in a timely manner, quick and multiplex detection techniques are needed.

Figure 4. Food Allergy Interventions

In recent decades, significant progress has been made in understanding the molecular mechanisms of food allergy, resulting in innovative diagnostics, prevention methods, and therapies

Concern over food safety has grown as it poses a significant hazard to public health worldwide. In order to guarantee food safety, effective methods for detecting food dangers are necessary. Numerous techniques have been developed to identify food hazards in order to address issues pertaining to food safety. However, the development of effective and trustworthy detection systems has proven to be difficult work because of the trace quantities of viruses, poisons, and dietary pollutants. The development of food safety tests has accelerated due to the introduction of new and speedy sensors, which are fast, portable, and sensitive.

Fundamentally, a model functions as a transformative instrument, revealing the complex relationships among food allergy data with a unique emphasis on enhancing patient outcomes. In addition to identifying foods that trigger allergies, the technology uses cutting-edge machine learning techniques to move medicine closer to a new understanding of and approach to treating food allergies.

3. Methodology

The model's applicability goes beyond simple identification; by anticipating the presence of allergens in food, it suggests a proactive strategy. This feature is very helpful to medical professionals and dieticians when creating customized meal menus where non-allergenic foods are kept at the top. By doing this, the model turns into a vital ally for patients balancing the tightrope of food limitations, reducing the likelihood of nutritional deficiencies.

In order to prevent food allergy and save lives, machine learning (ML) approaches have demonstrated encouraging outcomes when applied to medical data. In this work, we present a health care system based on machine learning techniques that can monitor patients in real time for food allergies. The six machine learning techniques used in the food allergy detection model are support vector machine, naive bayes, logistic regression, random forest, decision tree and k nearest neighbors. To maximize the classification of food allergy products and ingredients, the system highly accurately determines the appropriate classification technique. According to the experimental data, the Decision Tree in the suggested system has the maximum accuracy of 98.4%.

The two primary components of ML algorithms are prediction and classification. The initial step is to recognize, understand, and categorize ideas and objects into predefined categories, or "sub-populations." Furthermore, a number of classification algorithms, like k nearest neighbor and naïve bayes, utilize the pertinent datasets to understand the problem and identify possible features and labels. These aspects stand in for the traits or qualities that influence the label's outcomes. Learning and evaluation are the two stages of the classification process. The classifier uses a provided dataset to train its model in the first phase, and its performance is evaluated in the second. Recall, accuracy, and precision are three of the metrics used to evaluate the classifier's efficacy. The prediction process, which can be employed by any decision maker, is the subject of the second section. It leverages the features of the dataset to forecast a new value based on the information at hand. Numerous prediction algorithms exist, including SVM.

3.1. The Proposed System

The key elements of the system are patients, physicians, and medical dataset management using machine learning techniques. The four key phases of the system are described below. See figure 5.

3.1.1 Setup Phase

The setup phase consist of two steps: data gathering and data pre-processing.

3.1.1.1. Data gathering

The dataset used in our work is obtained the publicly accessible kaggle, an open source website. There are 400 records in the dataset, each of which represents a distinct food item and the allergies connected to it. A detailed list of all the allergens present in the food products is included in the dataset. A broad variety of ingredients are included in this list of allergens, including dairy, wheat, nuts (almonds, peanuts, pine nuts), seafood (anchovies, fish, shellfish), grains (oats, rice), plant-based ingredients (celery, mustard, soybeans), animal-based ingredients (chicken, pork), and common ingredients (cocoa, eggs). The collection also includes instances in which no specific allergens are mentioned. The dataset consist of attributes: food product, main ingredient, sweetener, fat/oil, seasoning, allergens and prediction.

3.1.1.2. Data pre-processing

Since medical data cannot have a zero value and must be normalized, this approach deals with dataset variations to produce more accurate and exact results. Duplicate data must be removed and missing values must be filled in by calculating the dataset's mean for each attribute.

3.1.2. Classification Phase

The development of the food allergy prediction model is the core component of the suggested system. We used a variety of machine learning methods, including K nearest neighbors (KNN), Naive Bayes (NB), Logistic Regression (LR), Random Forest (RF), Decision Tree (DT) and Support Vector Machine (SVM). The training stage and the evaluation stage are two other phases that are part of this phase.

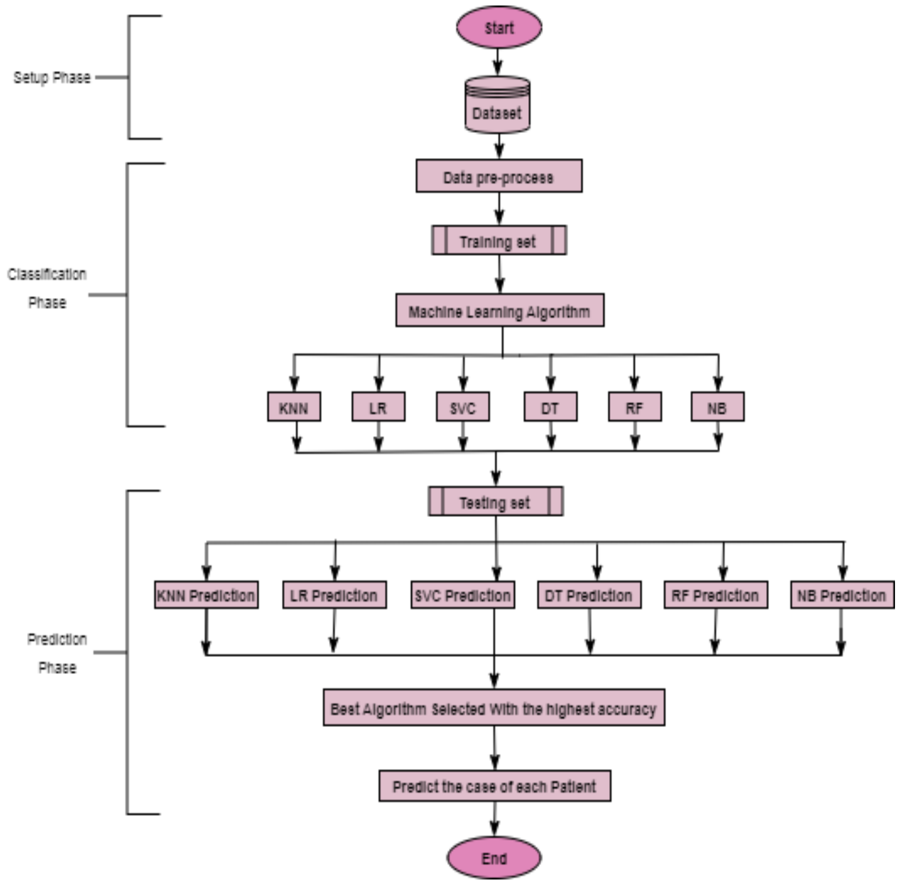

Figure 5. Work Flow of prediction model

3.1.2.1. Building the dataset

A training set, a validation set, and a testing set comprise the three datasets that are currently created from the data. Through the use of the training set, the model is initially trained and instructed in information processing. Model correctness is evaluated and parameter adjustments are made using the validation set. Model performance and accuracy are evaluated using the testing set.

3.1.2.2. Training stage

The algorithm learns how to extract characteristics for classification from the training data. Using the validation dataset, the model is improved after training in order to get the best accuracy by modifying variables and settings. In order to extract features depending on algorithm-specific parameters, this step is essential.

3.1.2.3. Evaluation stage

Using precise characteristics and a testing set, this step validates the model. Feedback may lead to changes, such as deploying the model or going back to training for more accuracy. The primary objectives of the algorithm during this stage are listed below, and Figure 6 shows the workflow of the classification process.

3.1.3. Prediction Phase

During this stage, we classified each food product as either non-allergic or allergic using the machine learning method that had the highest accuracy on the new dataset. For each record, the outcome is either 0 or 1. The prediction stage operates in this manner:

- Establishing a dataset for the purpose of training and feature extraction.
- Prediction error is decreased by the fitting phase, which determines the mapping between inputs and output labels by applying the training data to the machine learning algorithm. The training dataset is

estimated or forecasted using the model throughout the prediction phase. The model is then evaluated by generating a categorization report and comparing the calculation to the expected label.

- Apply a fresh data entry one row and several columns to the machine learning algorithm to make use of the modified dataset. Create a new prediction with a single output based on this input. Then, repeat the fitting and prediction steps again in accordance with the learning stage. The new data row and the matching expected class label (0 or 1) should be added to the database.

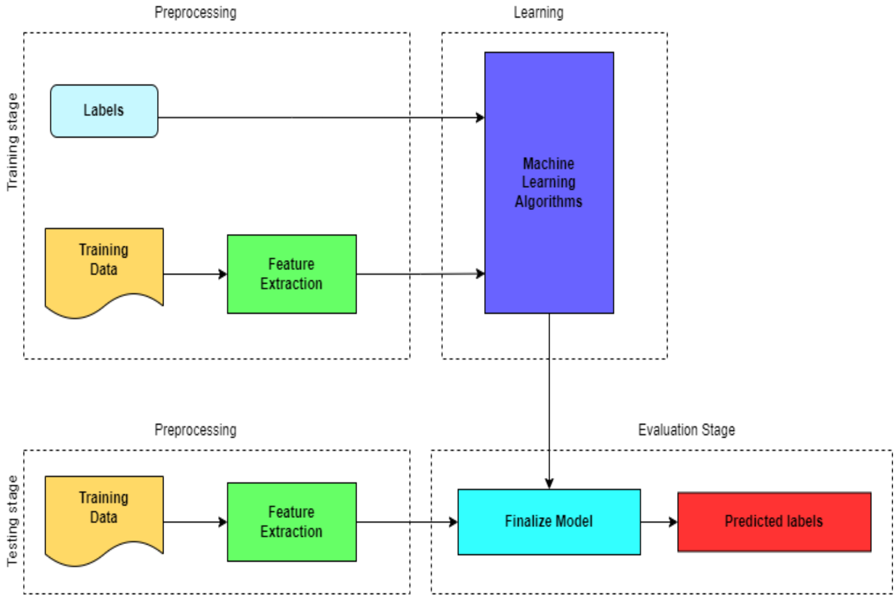

Figure 6. Classification Phase Workflow

3.1.4. Health Care Phase

Nutritionist consultants, doctors, and certified restaurants can effortlessly utilize this system for detecting food allergy triggers. The dataset makes it possible to identify and analyses possible allergens in various foods and goods by providing allergy information for a variety of food constituents. Researchers, food manufacturers, medical professionals, and people with food allergies can all benefit from it. The mechanism of the present phase of our system is explained in the steps that follow shown in figure 7.

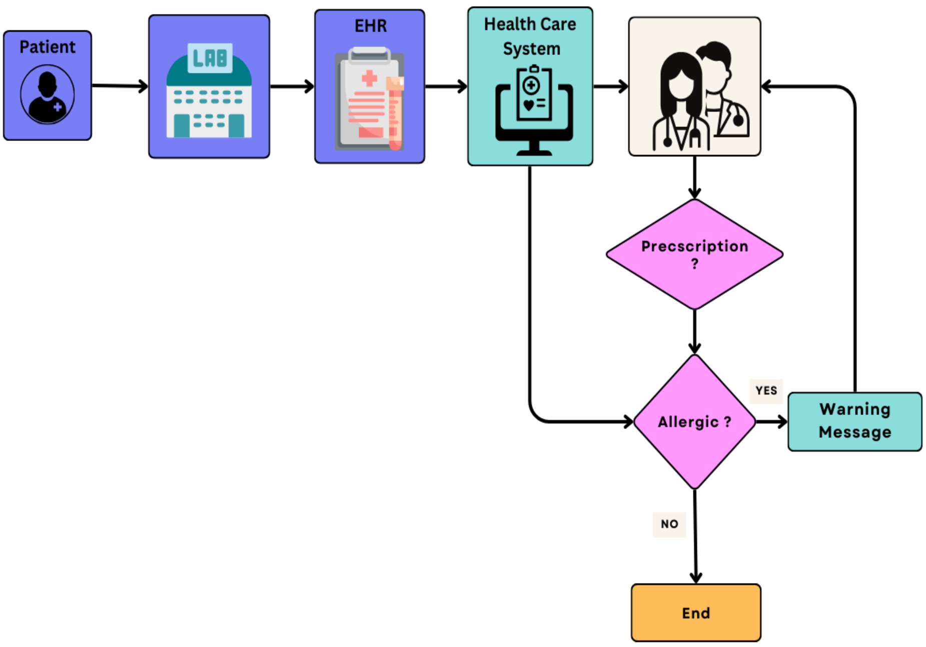

Figure 7. Demonstrate the health care scenario

Step 1: New patient is added by setting up their electronic health record (EHR), which includes their biographical data, test results from lab work, and information about any allergies to foods group.

Step 2: From among the suggested systems, the optimal algorithm is chosen during the prediction phase.

Step 3: Based on the following, the report is forwarded to the physician for ultimate care.

- The doctor prepares the prescription after reviewing the patient report.
- Our system verifies the list of food ingredients in conflict with the food allergy.
- If it is, the system notifies the physician to alter the prescription.
- The doctor provides a list of food products tailored to an individual's food allergies, suggesting alternative ingredients to replace those causing allergies.

4. Results

This research explores a profound understanding of the nature of food allergies, going beyond its immediate uses. An in-depth examination of allergy prevalence provides insight into the complex interactions between various allergens in a wide range of food products.

A food allergen with a dairy allergy cannot consume products containing dairy ingredients. To address this, the system filters out products that include dairy, ensuring they are safe for individuals with a dairy allergy. Generate a list that excludes any items with dairy ingredients.

This system can be utilized for establishing a restaurant that caters to individuals with food allergies. It can provide a list of menu items with chicken as the primary ingredient in case a person with a beef allergy goes to a restaurant and only wants chicken-based selections. This guarantees that the individual is served appropriate meals that meet their dietary requirements, resulting in a safer and more pleasurable dining experience.

Patients with celiac disease must follow a gluten-free diet precisely to prevent side effects. In people with celiac disease, gluten—which is frequently present in wheat and similar grains—causes inflammation and minor intestinal damage. They should therefore choose gluten-free options instead, like goods produced with rice flour, almond flour, or other non-gluten flours. This dietary change is essential for both choosing packaged foods and cooking at home, with the significance of closely reading ingredient labels and trying out different gluten-free recipes emphasized. It's also crucial to know the hidden sources of gluten in processed foods to ensure that the patient adheres to a strict gluten-free diet and manages their sickness appropriately.

In order to analyze the experimental results based on the basic measurements mentioned, we compare the different machine learning algorithms that we used in our study to evaluate the diagnosis of food allergy goods. Table 1 demonstrates that Decision Tree was the most accurate categorization method, with an accuracy rate of 98.4%.

Table 1. Comparative analysis of model's accuracy (%)

	Model Name	Score
1	Logistic Regression	0.580645
2	Naive Bayes	0.661290
3	Random Forest	0.967742
4	Decision Tree	0.983871
5	K Nearest Neighbors	0.951613
6	Support Vector Classifier	0.935484

Finally, the results showed that Decision Tree (DT) is considered the most accurate approach to satisfy the requirements of our proposed system when compared to other categorization strategies. To assess the efficacy of the model, a rigorous scoring system evaluates how well food descriptors perform across various machine learning techniques. While the random forest, support vector machine, and k-nearest neighbors models all remarkably demonstrate strong training and validation accuracies, the decision tree model emerges see figure 8.

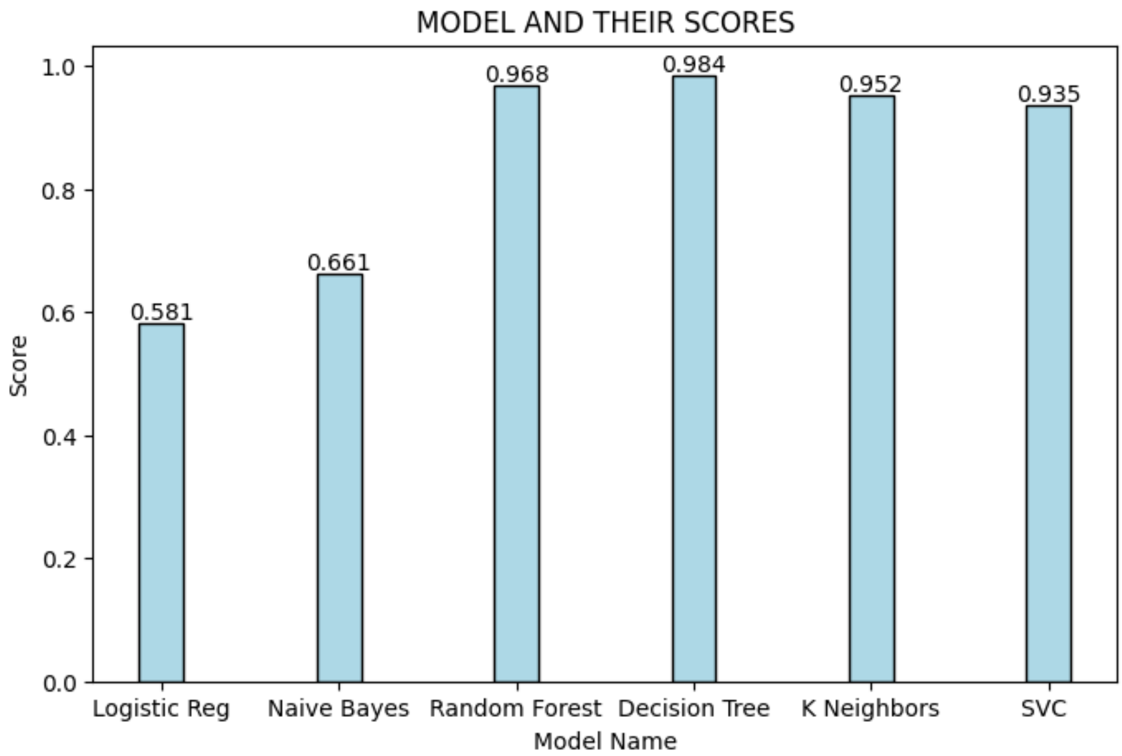

Figure 8. Accuracy comparison of different algorithms

5. Conclusion

A vast amount of data, including studies on food allergies, has been produced as a result of amazing developments in the healthcare and IT sectors. Technological improvements have highlighted the importance of conducting thorough investigations in food allergy research, which offer vital insights into our understanding of food-related allergic illnesses. Health information technology, which gives doctors access to accurate patient records and a complete understanding of their food allergy history, is essential to improving the quality of care in medical clinics. Using machine learning algorithms is very important when looking for patterns associated with food allergies, as these algorithms have demonstrated significant success in analyzing a variety of patient data. This work evaluates six machine learning algorithms using different metrics with the goal of creating a system to analyze and predict food allergies. Decision tree method with the best accuracy of 98.4%, is used to predict a patient's likelihood of developing a food allergy. The suggested system is a useful resource for medical facilities, patients, and healthcare providers since it reduces the need to prescribe allergy-causing foods.

The future implications of this approach could reach the field of individualized medicine. The system lays the groundwork for personalized recommender systems by imagining a time when people with dietary limitations or allergies may make safe and educated food selections. This suggests a radical change in the way we think about nutritional well-being, with technology acting as a reliable guide through the complexity of food options. This journey delves into the intersection of machine learning, food safety, and healthcare.

References

1. Kamphorst, K., Lopez-Rincon, A., Vlieger, A. M., Garssen, J., van't Riet, E., & van Elburg, R. M. (2023). Predictive factors for allergy at 4–6 years of age based on machine learning: A pilot study. PharmaNutrition, 23, 100326.
2. Man Liu, B. H., Rayner, A., Mendelsohn, A. R., Shneyderman, A., Chen, M., & Pun, F. W. (2023). Applying Artificial Intelligence to Identify Common Targets for Treatment of Asthma, Eczema, and Food Allergy. bioRxiv, 2023-06.
3. Zhang, J., Lee, D., Jungles, K., Shaltis, D., Najarian, K., Ravikumar, R., ... & Gryak, J. (2023). Prediction of oral food challenge outcomes via ensemble learning. Informatics in Medicine Unlocked, 36, 101142.
4. Ktona, A., Mitre, A., Shehu, D., & Xhaja, D. (2022). Support Allergic Patients, using Models Found by Machine Learning Algorithms, to Improve their Quality of Life. International Journal of Intelligent Systems and Applications in Engineering, 10(4), 512-517.
5. Metwally, A. A., Yu, P. S., Reiman, D., Dai, Y., Finn, P. W., & Perkins, D. L. (2019). Utilizing longitudinal microbiome taxonomic profiles to predict food allergy via long short-term memory networks. PLoS computational biology, 15(2), e1006693.
6. Yang, H., Qu, Y., Gao, Y., Sun, S., Ding, R., Cang, W., ... & Wu, J. (2022). Role of the dietary components in food allergy: A comprehensive review. Food Chemistry, 386, 132762.
7. Turner, P. J., Arasi, S., Ballmer-Weber, B., Baseggio Conrado, A., Deschildre, A., Gerdts, J., .& Global Allergy, Asthma European Network (GA2LEN) Food Allergy Guideline Group. (2022). Risk factors for severe reactions in food allergy: rapid evidence review with meta-analysis. Allergy, 77(9), 2634-2652.
8. Wang, L., Niu, D., Zhao, X., Wang, X., Hao, M., & Che, H. (2021). A comparative analysis of novel deep learning and ensemble learning models to predict the allergenicity of food proteins. Foods, 10(4), 809.
9. Sindher, S. B., Long, A., Chin, A. R., Hy, A., Sampath, V., Nadeau, K. C., & Chinthrajah, R. S. (2022). Food allergy, mechanisms, diagnosis and treatment: Innovation through a multi-targeted approach. Allergy, 77(10), 2937-2948.
10. Eiwegger, T., Hung, L., San Diego, K. E., O'Mahony, L., & Upton, J. (2019). Recent developments and highlights in food allergy. Allergy, 74(12), 2355-2367.
11. Singer, A. G., Kosowan, L., Soller, L., Chan, E. S., Nankissoor, N. N., Phung, R. R., & Abrams, E. M. (2021). Prevalence of physician-reported food allergy in Canadian children. The Journal of Allergy and Clinical Immunology: In Practice, 9(1), 193-199.
12. Nedyalkova, M., Vasighi, M., Azmoon, A., Naneva, L., & Simeonov, V. (2023). Sequence-Based Prediction of Plant Allergenic Proteins: Machine Learning Classification Approach. ACS omega, 8(4), 3698-3704.
13. Sookrah, R., Dhowtal, J. D., & Nagowah, S. D. (2019, July). A DASH diet recommendation system for hypertensive patients using machine learning. In 2019 7th international conference on information and communication technology (ICoICT) (pp. 1-6). IEEE.
14. Kavya, R., Christopher, J., Panda, S., & Lazarus, Y. B. (2021). Machine learning and XAI approaches for allergy diagnosis. Biomedical Signal Processing and Control, 69, 102681.
15. Carter, C. A., Pistiner, M., Wang, J., & Sharma, H. P. (2020). Food allergy in restaurants work group report. The Journal of Allergy and Clinical Immunology: In Practice, 8(1), 70-74.
16. Dimassi, H., Haidar, S., Issa, S., & Hassan, H. F. (2020). Food allergies and allergens: Characterization and perceptions among diagnosed food allergic individuals in Lebanon. World Allergy Organization Journal, 13(11), 100481.
17. Jan, L., & Naig, A. (2021). Food allergen labeling of commonly used ingredients in Chinese restaurants in the United States. Food Control, 130, 108381.
18. Cao, S., Borro, M., Alonzi, S., Sindher, S., Nadeau, K., & Chinthrajah, R. S. (2021). Improvement in health-related quality of life in food-allergic patients: a meta-analysis. The Journal of Allergy and Clinical Immunology: In Practice, 9(10), 3705-3714.
19. Devika Varshini, G., & Chinnaiyan, R. (2020). Optimized machine learning classification approaches for prediction of autism spectrum disorder. Ann Autism Dev Disord. 2020; 1 (1), 1001.
20. Noori, N. A., & Yassin, A. A. (2021). Towards for Designing Intelligent Health Care System Based on Machine Learning. Iraqi Journal for Electrical & Electronic Engineering, 17(2).
21. Makridis, G., Mavrepis, P., Kyriazis, D., Polychronou, I., & Kaloudis, S. (2020). Enhanced food safety through deep learning for food recalls prediction. In Discovery Science: 23rd International Conference, DS 2020, Thessaloniki, Greece, October 19–21, 2020, Proceedings 23 (pp. 566-580). Springer International Publishing.
22. Sena-Torralba, A., Pallás-Tamarit, Y., Morais, S., & Maquieira, Á. (2020). Recent advances and challenges in food-borne allergen detection. TrAC Trends in Analytical Chemistry, 132, 116050.

23. Muthukumar, J., Selvasekaran, P., Lokanadham, M., & Chidambaram, R. (2020). Food and food products associated with food allergy and food intolerance–An overview. Food Research International, 138, 109780.
24. Schroer, B., Groetch, M., Mack, D. P., & Venter, C. (2021). Practical challenges and considerations for early introduction of potential food allergens for prevention of food allergy. The Journal of Allergy and Clinical Immunology: In Practice, 9(1), 44-56.
25. Mills, E. C., Adel-Patient, K., Bernard, H., De Loose, M., Gillard, N., Huet, A. C., ... & Monaci, L. (2019). Detection and quantification of allergens in foods and minimum eliciting doses in food-Allergic individuals (ThRAll). Journal of AOAC International, 102(5), 1346-1353.
26. Noori, N. A., & Yassin, A. A. (2021). Towards for Designing Intelligent Health Care System Based on Machine Learning. Iraqi Journal for Electrical & Electronic Engineering, 17(2).
27. Renz, H., Allen, K. J., Sicherer, S. H., Sampson, H. A., Lack, G., Beyer, K., & Oettgen, H. C. (2018). Food allergy. Nature reviews Disease primers, 4(1), 1-20.
28. Alag, A. (2019). Machine learning approach yields epigenetic biomarkers of food allergy: A novel 13-gene signature to diagnose clinical reactivity. PLoS One, 14(6), e0218253.
29. Foong, R. X., & Santos, A. F. (2021). Biomarkers of diagnosis and resolution of food allergy. Pediatric Allergy and Immunology, 32(2), 223-233.
30. van Breugel, M., Fehrmann, R. S., Bügel, M., Rezwan, F. I., Holloway, J. W., Nawijn, M. C., ... & Koppelman, G. H. (2023). Current state and prospects of artificial intelligence in allergy. Allergy, 78(10), 2623-2643.
31. Foers, A. D., Shoukat, M. S., Welsh, O. E., Donovan, K., Petry, R., Evans, S. C., ... & Soilleux, E. J. (2021). Classification of intestinal T-cell receptor repertoires using machine learning methods can identify patients with coeliac disease regardless of dietary gluten status. The Journal of Pathology, 253(3), 279-291.