Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Determining Techniques for Tracking Foodborne Illnesses and Intoxications in India

D. Kabinath[1*], K. Senthamarai Kannan[2]
[1]*Department of Statistics, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, Tamil Nadu, India*
kabi130198@gmail.com
[2]*Department of Statistics, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, Tamil Nadu, India*
senkannan2002@gmail.com

ARTICLE INFO

Received: 28 Dec 2024
Revised: 18 Feb 2025
Accepted: 26 Feb 2025

ABSTRACT

The objective of this research is to determine the percentage of foodborne infections and intoxications that can be attributed to particular sources in India. Additionally, it provides innovative outlier identification methods utilizing Machine Learning (ML) techniques in conjunction with standard statistical approaches such as ARIMA and SARIMA. The project aims to enhance the precision of identifying the main origins of foodborne infections by integrating epidemiological monitoring, genetic typing, and sophisticated machine learning-based outlier detection. The study findings indicate that more than sixty percent of foodborne diseases may be attributed to contaminated water and raw produce. Improper food handling is responsible for twenty-five percent of these infections, while infected animal products contribute to fifteen percent. These results emphasize the need for improved food safety standards and public awareness initiatives. The study emphasizes the significance of contemporary surveillance systems, which use statistical and machine learning techniques, to monitor and regulate foodborne diseases more. This, in turn, leads to enhanced public health outcomes in India.

Keywords: Foodborne Diseases, Illnesses, Intoxications, World Health Organization, ARIMA and SARIMA, Machine Learning, Outlier Detection.

INTRODUCTION

The rising incidence of food-borne diseases, which affect an estimated 600 million people each year, has elevated this public health issue to the forefront of international attention [1-2]. Both people's well-being and the economy are at risk when food-borne viruses cause illness [3]. Several diseases may be transmitted via them. Several dangerous microbes may infiltrate food throughout its production, processing, storage, and transportation phases before it is consumed [3-5]. Human pathogens in food are a leading cause of gastrointestinal diseases, which are projected to have a major detrimental impact on health and the economy [6-9]. More than 30% of deaths caused by food poisoning happen in children younger than five, according to the World Health Organization (WHO). Diarrhea, meningitis, and other infections (some of which may be deadly) can be caused by bacteria that are present in food [10-12]. Several additional symptoms can range from mild to severe. This is because food poisoning is more common in youngsters as their immune systems are still maturing. They may also be more prone to germ, virus, or parasite contamination due to putting their hands in their mouths or eating tainted food.

In many countries, poisoning from food has become a major problem in the last 20 years, impacting the economy and public health. A large number of people become sick or die from it every year. Any infectious or harmful condition traced back to eating or drinking contaminated food or water is called a "foodborne disease" [13]. To shed light on the current data gap, the World Health Organization (WHO) has initiated a new program to provide more accurate estimates of the global burden of foodborne illness. The bulk of the approximately 1.8 million deaths caused by diarrhoeal diseases in 2005 were attributed to tainted food and water [14]. Over the last 29 years (1980–2009), 3,485 individuals in India have lost their lives due to food poisoning, as per reports of illnesses and scientific investigations [15]. Approximately 47.8 million cases of food poisoning occur annually in the US, leading to 127,839 hospitalizations and 30,37 deaths [16].

The prevention and management of food-borne infections have significant public health implications for both industrialized and developing countries [17-18]. While fewer incidents of food poisoning have been reported in Japan recently, this trend is not universally true. There has been a steady decrease in the number of outbreaks and cases of food poisoning caused by Vibrio parahaemolyticus and Salmonella each year [19]. Nevertheless, outbreaks of foodborne disease caused by Campylobacter spp. and norovirus have been steadily increasing during the last few years [19]. Numerous major epidemics have occurred in Japan throughout the last few years. Raw beef products were the source of a devastating epidemic of enterohemorrhagic Escherichia coli infections in 2011. There were 205 confirmed cases, including 5 fatalities. Outdoor events served 875 instances of sushi infected with Campylobacter jejuni/coli in 2016. Finally, in 2017, 1,084 student instances of norovirus infections were linked to eating chopped seaweed (Kizami-nori) at school lunches [20].

Microorganisms Causing Food Poisoning

Microorganisms that cause illness in humans include bacteria, fungi, parasites, and viruses [21-22]. Both people and animals are susceptible to bacterial infections, which are the most common cause of food poisoning [23-24]. Some common bacteria that may be detected in food are campylobacter, clostridium, shigella, listeria, bacillus, yersinia, cholera, vibrio, and perfringens [25]. Everyone agrees that viruses are a major source of food poisoning since they are pathogens that may infect food [26]. This category includes viruses such as hepatitis and norovirus (NoV) [27]. Parasites such as Ascaris lumbricoids, Giadia lamblia, Entamoeba histolytica, and Toxoplasma gondii are among those that might compromise food safety [28-30]. The fungi that may cause food poisoning include species of Fusarium, Aspergillus flavus, Aspergillus ochraceus, Penicillium citrinum, Claviceps purpurea, and Penicillium pabulum [31-32].

Dangers of Food-Borne Infectious Agents

Bacteria use a variety of virulence factors to invade host cells during microbial pathogenesis, a process that requires both the host and the pathogen. Among the many bacterial virulence factors are phospholipids, teichoic acids, nucleic acids, polysaccharides, proteins, capsules, enzymes, poisons, spikes, and adhesins (including fibronectin, collagen, laminin, integrin, and internalin) [87-89]. Phagocytic cells, mucus, bile salts, stomach acids, antimicrobial antagonists, and other immune cells and processes are used by the host to fight infections. The outcome will be infection if the bacteria are effective. But if people's immune systems had been strong enough, we could have eliminated the illness. Bacteria, fungi, viruses, and parasites are only some of the microbes that might be present in food. These microbes may enter human beings via the mouth and go to the digestive system. Infectious disorders caused by bacteria in food may range from mild, localized infections that affect only a small area of the body to dangerous, systemic infections that affect several organs.

Food-Borne Illnesses in Machine Learning For Outliers Detections

Food poisoning data analysis requires ML outlier detection to find new dangers, reporting inaccuracies, and sickness causes. Epidemiological investigations must identify outliers because they may indicate outbreaks or contamination by deviating greatly from the predicted distribution. Time series analysis using ARIMA and SARIMA is used for foodborne disease surveillance. In complicated, multi-dimensional datasets like those used in current epidemiological monitoring, these approaches may not discover all outliers. Due to this restriction, machine learning (ML) approaches are popular for their robust outlier identification. Machine learning techniques like k-NN, Isolation Forest, and One-Class Support Vector Machine (SVM) may discover outliers by analyzing data patterns and distinguishing normal and abnormal behaviors. These methods can handle huge datasets with many variables in vast and elaborate formats, making them ideal for foodborne disease surveillance. Excellent example: training machine learning models to identify anomalous illness increases that may suggest an epidemic. Traditional ARIMA and SARIMA models with machine learning-based outlier identification may improve foodborne illness surveillance. These integrated methods may inspire public health officials to respond more quickly and correctly to reduce foodborne diseases by recognizing irregularities that may indicate new or unexpected contamination sources.

The modified Z-score (mZ-score) method was used to detect outliers in the RSS time series. When used, scale estimators operate differently according to their strengths and drawbacks. A new scale estimator with the best attributes of previous ones is the objective of scale estimator hybridization. The outlier detection performance of a Z-score method is examined by utilizing hybridization processes for Sn and MAD scale estimators. Classified hybrid scale estimators are weighted, maximum, and average. We compare the mZ-score approach to three hybrid scale estimators using three publically available and experimentally produced time-series RSS datasets. The simulations demonstrate that the weighted hybrid scale estimator detects outliers best among the three hybrid scale estimators. The mZ-score approach with the weighted hybrid scale estimator beats MOD, k-means clustering, and DBSCAN with few or no false alarms or negatives [33].

Finding outliers is a challenging but essential part of data mining. When points of data are unknown as outliers, this study investigates how they impact models trained from vast datasets. As per the suggested fuzzy c-means clustering method, outliers are individuals who have higher cluster values for membership and are located farther away from the cluster centroid. Finding the sample closest to the cluster center with a higher participation value is preferable. The weight of each sample, which indicates how representative it is of the data distribution, is computed using the AdaBoost method. This study shows how weighted loss functions work in different settings. In conclusion, our method is evaluated on twelve UCI datasets and demonstrates a 95%+ accuracy rate on the following: banknotes (99.99%), biodeg (99.01%), optdigits (99.19%), and letters (97.37%). The experimental results show the effectiveness and efficiency of the proposed method [34].

METHODS AND MATERIALS

At this time, doctors in India are required by law to notify the food safety authorities when their patients have food poisoning. Investigations into cases of food poisoning are overseen by the State Food Laboratories. Before, during, and after a study on a food-borne illness, environmental, epidemiological, and laboratory factors must all be carefully considered. India lacks such a system for investigating crimes. No comprehensive study has been carried out in India to determine the specific food or cause of this disease at this time. In collaboration with Hyderabad's renowned Ronald Ross Institute of Tropical Diseases, pilot research was conducted to evaluate food poisoning incidents to raise policymakers' knowledge of the issue.

Research area

Research on food poisoning and sickness is crucial because of the serious public health consequences of the high incidence of food poisoning in India. The goal of this project is to improve and build surveillance systems that can detect and follow food-borne diseases as they move through the food supply chain. To detect and assign foodborne infections, modern molecular and genomic methods such as whole genome sequencing and polymerase chain reaction are required. To better prepare public health organizations to respond to epidemics, this study also looks at how geographic information systems (GIS) and statistical modeling can predict and map out epidemic trends. Due to its varied sanitation and food safety requirements, as well as its varied food production and distribution networks, India requires a thorough foodborne disease monitoring system. Methods for collecting and analyzing data in real time, such as digital reporting platforms and mobile health apps, are being studied for their potential use in epidemic control and rapid reporting. In addition, it takes into account cultural and socioeconomic aspects that impact food safety to develop focused treatments to lessen contamination. This field of study aims to improve monitoring, prevention, and control of foodborne infections in India, thereby reducing their prevalence. Legislation will be impacted and food safety will be enhanced.

Dataset Source

The incidence, origins, and global distribution of food poisoning may be better understood with the use of databases documenting cases from 2012 through 2024. These databases often include comprehensive records of reported cases, associated diseases, pollution sources, and demographic information on the affected individuals. The data is often sourced from public health reporting systems, laboratory confirmations, hospital records, and surveillance systems. Molecular typing processes and digital systems for reporting are only two examples of how technological advancements over the last several decades have increased the quantity and quality of these databases. Patterns in the data show that regulatory efforts have an effect, that food safety practices have changed, and that the prevalence of specific diseases has changed. The databases are a great resource for epidemiological studies because they allow researchers to easily find novel dangers, assess food safety procedures, and develop strategies to control and prevent food poisoning on a global scale.

Case Study

Using a "case study" criteria, the cases of foodborne diseases in a community are recognized and categorized. These characteristics form the basis of systems that monitor food poisoning and intoxication. To consistently track, analyze, and compare cases of food poisoning across different regions and periods, this criterion is required. A diagnosis of food poisoning is based on the existence, severity, length, and frequency of symptoms. Some of the symptoms include a high temperature, nausea, vomiting, diarrhea, and abdominal pain. Clinical samples may be tested for toxins or pathogens using laboratory criteria that employ PCR, ELISA, or culture. Epidemiological factors are useful for tracing the origins of diseases. Some of these include a pre-symptom history of consuming food or drink that may have been contaminated. To keep track of instances, the case definition may differentiate between probable and proven occurrences. Clinical and laboratory requirements are satisfied by confirmed cases, whereas epidemiological standards are met by probable cases, but laboratory criteria are not. If you want to know how big the pandemic is and how successful the controls are, you have to make this difference. The ability to quickly react to and mitigate outbreaks of food poisoning depends on public health officials having easy access to precise case definitions.

Data Analyze and Verify

There are several important stages involved in checking and processing data on incidents of food poisoning and intoxication from 2012 to 2024 to make sure the datasets are accurate, comprehensive, and reliable. To assess trends in food-borne diseases, detect outbreaks, and direct public health measures, this approach is crucial.

Data Collection and Validation

Hospital records, digital reporting platforms, laboratory confirmations, and public health monitoring systems are among the many sources from which data is collected. To guarantee that each dataset is a true reflection of reported instances, validation is essential. This involves doing things like making sure that all pertinent clinical, laboratory, and epidemiological requirements are satisfied, verifying that case definitions are consistent, and cross-referencing with other data sources.

Data Cleaning

Data cleaning includes addressing missing or incomplete information, fixing mistakes, and eliminating duplicates. Items that do not follow the expected pattern are marked for further examination. To ensure consistency between datasets, standardized formats are used for dates, locations, and demographic information.

Data Integration

A centralized database is used to combine data from various years and areas. This ensures that factors like case definitions, detection of pathogens procedures, and reporting processes are compatible and consistent. Full geographical and temporal analysis is now possible thanks to this integration.

Data Analysis and Processing

To find patterns, correlations, and trends, the combined and cleansed datasets are run through statistical analysis. It is possible to use machine learning algorithms and other forms of advanced processing to forecast outbreaks and evaluate the efficacy of food safety measures. To visualize the distribution and pinpoint problem areas, researchers use Geographic Information Systems (GIS).

Quality Assurance

Regular audits and reviews are conducted by data scientists and epidemiologists as part of the process to ensure ongoing quality checks. The goal of establishing feedback loops is to enhance data gathering and processing systems on an ongoing basis.

Reporting and Dissemination

Lastly, public health policy and intervention methods are informed by reports, dashboards, and other outputs that are generated from processed data. To guarantee that food poisoning hazards are addressed in a timely and effective manner, findings are communicated with relevant stakeholders, such as healthcare professionals, government organizations, and the general public.

DATA ANALYSIS USING STATISTICAL METHODS

ARIMA (AutoRegressive Integrated Moving Average) and SARIMA (Seasonal AutoRegressive Integrated Moving Average) are powerful statistical methods used for time series forecasting, including in the analysis of foodborne illness data. These models are particularly effective for understanding and predicting patterns in data that are dependent on time, such as the occurrence of foodborne illnesses over months or years.

ARIMA is a model that combines three components

Auto Regressive (AR)

This component models the relationship between an observation and several lagged observations (previous time points).

Integrated (I)

This involves differencing the data to make it stationary, which means that the mean, variance, and autocorrelation structure are consistent over time.

Moving Average (MA)

This component models the relationship between an observation and a residual error from a moving average model applied to lag observations.

The ARIMA model is denoted as ARIMA (p, d, q), where:

- p is the number of lag observations in the model (AR term),
- d is the degree of differencing (I term),
- q is the size of the moving average window (MA term).

SARIMA extends ARIMA by incorporating seasonal elements into the model. It accounts for patterns that repeat over regular intervals, such as monthly or quarterly spikes in foodborne illnesses. The model is denoted as SARIMA (p, d, q)(P, D, Q)[s], where:

- P, D, and Q are the seasonal counterparts of the non-seasonal ARIMA parameters,
- s represents the seasonal period (e.g., 12 for yearly seasonality).

SARIMA is particularly useful for datasets that exhibit strong seasonal trends, allowing for more accurate predictions by considering both the regular and seasonal patterns. These models help in identifying the underlying structure of time series data, enabling better forecasting and timely intervention strategies for managing foodborne illnesses.

Two essential equations characterize the ARIMA model for food poisoning prediction: one about the AutoRegressive (AR) component and another about the Moving Average (MA) component. With p being the number of autoregressive terms, d the degree of differencing, and q the number of moving average terms, the generic ARIMA model is represented as ARIMA (p, d, q).

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t \tag{1}$$

Equation 1 illustrates, here X_t represents the differenced series of foodborne illness cases at time t and $\phi_1, \phi_2 \dots \dots . . \phi_p$ are the coefficients for the lagged values of the series. ϵ_t is the white noise error term at time t. The AR part models the relationship between the current value of the series and its previous values.

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \tag{2}$$

In this equation 2, ϵ_t is the current error term, and $\epsilon_{t-1}, \epsilon_{t-2} \dots \dots \dots \dots . . \epsilon_{t-q}$ are past error terms. The coefficientsθ_1, $\theta_2 \dots \dots . . \theta_q$ capture the influence of past errors on the current value of the series.

The SARIMA (Seasonal AutoRegressive Integrated Moving Average) model extends the ARIMA model by incorporating seasonality into the time series. It is particularly useful for data like foodborne illnesses, which may exhibit seasonal patterns (e.g., more cases in certain months due to climatic conditions or holidays). The SARIMA model is denoted as SARIMA (p, d, q) (P, D, Q) [s], where:

p, d, q are the non-seasonal ARIMA parameters,

P, D, and Q are the seasonal counterparts, and

s is the length of the seasonal cycle (e.g., s = 12 for monthly data with yearly seasonality).

Non-Seasonal ARIMA Component:

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D X_t = \theta(B)\Theta(B^s)\epsilon_t \qquad (3)$$

Where equation 3 illustrates $\phi(B) = 1 - \phi_1 B$-$\phi_2 B^2$......................$\phi_p B^p$ represents the non-seasonal autoregressive part.

$\Phi(B^s) = 1 - \phi_1 B$-$\phi_2 B^2$......................$\phi_P B^{Ps}$ represents the seasonal autoregressive part.

$(1-B)^d$ is the non-seasonal differencing.

$(1-B)^D$ is the seasonal differencing.

$\theta(B) = 1 - \theta_1 B$-$\theta_2 B^2$......................$\theta_q B^q$ represents the non-seasonal moving average part.

$\Theta(B^s) = 1 - \Theta_1 B + \Theta_2 B^{2s}$......................$\Theta_Q B^{Qs}$ represents the seasonal moving average part.

ϵ_t is the error term at time t.

SARIMA-based Seasonal AR and MA Components

Seasonal AR Component

$$\Phi(B^s)X_t = X_t - \Phi_1 X_{t-s} - \Phi_2 X_{t-2s} - \cdots - \Phi_P X_{t-Ps} \qquad (4)$$

This Equation 4, captures the seasonal autoregressive relationship where $\Phi(B^s)$ represents the influence of past seasonal values (e.g., cases from the same month in previous years).

Seasonal MA Component

$$\Theta(B^s)\epsilon_t = \epsilon_t - \Theta_1 \epsilon_{t-s} - \Theta_2 \epsilon_{t-2s} - \cdots - \Theta_Q \epsilon_{t-Qs} \qquad (5)$$

This Equation 5, captures the seasonal moving average component where $\Theta(B^s)$ accounts for the effect of past seasonal error terms on the current value.

Outlier Detection

Statistical analysis relies on outlier detection, which is finding data points that deviate significantly from the expected patterns in a dataset. These outliers need more investigation because they could indicate errors, rare events, or novel phenomena. Time series analysts often use statistical methods such as ARIMA (Autoregressive Integrated Moving Average) and SARIMA (Seasonal ARIMA) to analyze and forecast data trends. Analyzing residuals, or the differences between actual and predicted values, is a common way for these approaches to find abnormalities. When significant residuals surpass specified confidence intervals or thresholds, outliers may be found. These differences might be the result of out-of-the-ordinary events, inaccurate data entry, or modifications to basic processes. When patterns display nonlinearity or when dealing with complex, high-dimensional data, statistical methods may struggle to detect outliers. Incorporating machine learning techniques has become necessary due to the constraint. These methods have the potential to improve traditional approaches by providing more sophisticated outlier detection in diverse and complicated datasets.

Outlier Detection for Standard Deviation Method

One easy and effective way to find outliers is to use the standard deviation method. This is particularly true when tracking cases of food poisoning. This method finds outliers by measuring how much individual data points deviate from the collection's average. Epidemiological studies may use the standard deviation method to spot suspicious trends in the frequency of food poisoning over time. Possible causes of these outliers include an outbreak or inaccurate reporting shown in pseudocode for Table 1. This method requires first computing the dataset's average and standard deviation. Data points are considered outliers if they differ significantly from the mean by more than a certain number of standard deviations, often 2 or 3 standard deviations. These extreme cases can indicate a significant underreporting of diseases or a sudden surge in cases linked to contaminated food. While the standard deviation method is straightforward, it may not be the best choice when dealing with complex multivariate data or datasets with skewed distributions. More complex techniques, or a combination of traditional methods and machine learning, may be required to reliably identify outliers in certain cases.

$$X > \mu + k\sigma \text{ } or \text{ } X < \mu - k\sigma \tag{6}$$

Equation 6 illustrates, Where, X is the data point, μ is the mean of the dataset, σ is the standard deviation of the dataset, and k is a constant. That determines the sensitivity of outlier detection often set to 2 or 3. If a data point X lies outside the range of $\mu \pm k\sigma$, it is considered an outlier.

Table 1. Pseudocode for Outlier Detection using Standard Deviation Method.

Table 1: Pseudocode for Outlier Detection using Standard Deviation Method

```
# Input: Dataset (list or array of numeric values)
# Output: List of detected outliers
function detect_outliers_standard_deviation(dataset):
  # Step 1: Calculate the mean (μ) of the dataset
  mean = calculate_mean(dataset)
  # Step 2: Calculate the standard deviation (σ) of the dataset
  std_dev = calculate_standard_deviation(dataset)
  # Step 3: Define a threshold value for outlier detection
  # Typically, k is 2 or 3 standard deviations away from the mean
  k = 2  # Adjust as needed
  # Step 4: Initialize an empty list to store outliers
  outliers = []
  # Step 5: Iterate through each data point in the dataset
  for each data_point in the dataset:
    # Step 6: Check if the data point is an outlier
    if data_point > mean + k * std_dev or data_point < mean - k * std_dev:
      # Step 7: If it is an outlier, add it to the outliers list
      outliers.append(data_point)
  # Step 8: Return the list of detected outliers
  return outliers
# Function to calculate the mean
function calculate_mean(dataset):
  return sum(dataset) / length(dataset)
# Function to calculate the standard deviation
function calculate_standard_deviation(dataset):
  mean = calculate_mean(dataset)
  variance = sum((data_point - mean)^2 for each data_point in dataset) / length(dataset)
  return sqrt(variance)
```

Outlier Detection for IQR Method

The Interquartile Range (IQR) method is a useful technique for spotting outliers while monitoring occurrences of foodborne diseases. IQR is a resilient dispersion measure that analyses the distribution of data within the middle 50% to capture the central trend while being resistant against outliers. When doing monitoring on food poisoning epidemics, the IQR method might help identify anomalous patterns in data. Before using this approach, you must first identify the first and third quartiles (Q1) of the dataset. Upon deducting Q1 from Q3, the IQR is obtained. Learning the Q1 and Q3 values of the dataset is the starting point for using this method. To get the IQR, we use the difference between the third and first quarters equations 7 and 8 shown below:

$$IQR = Q3 - Q1 \tag{7}$$

Outliers are data points that do not conform to the norm if they exceed:

$$Q1 - 1.5 \times IQR \frac{Q3}{1.5 \times IQR} \tag{8}$$

This approach excels at spotting anomalous spikes in illness cases that might signal possible foodborne outbreaks because it zeroes down on the core mass of the data, which is especially useful for skewed distributions shown in the pseudocode for Table 2.

Table 2. Pseudocode for Outlier Detection using IQR Method

Table 2: Pseudocode for Outlier Detection using IQR Method

```
# Input: Dataset (list or array of numeric values)
# Output: List of detected outliers
function detect_outliers_IQR(dataset):
  # Step 1: Sort the dataset
  sorted_data = sort(dataset)
  # Step 2: Calculate Q1 (First Quartile) and Q3 (Third Quartile)
  Q1 = calculate_quartile(sorted_data, 1)
  Q3 = calculate_quartile(sorted_data, 3)
  # Step 3: Calculate the Interquartile Range (IQR)
  IQR = Q3 - Q1
  # Step 4: Define the lower and upper bounds for outliers
  lower_bound = Q1 - 1.5 * IQR
  upper_bound = Q3 + 1.5 * IQR
  # Step 5: Initialize an empty list to store outliers
  outliers = []
  # Step 6: Iterate through each data point in the dataset
  for each data_point in the dataset:
    # Step 7: Check if the data point is an outlier
    if data_point < lower_bound or data_point > upper_bound:
      # Step 8: If it is an outlier, add it to the outliers list
      outliers.append(data_point)
  # Step 9: Return the list of detected outliers
  return outliers
# Function to calculate a specific quartile (Q1 or Q3)
function calculate_quartile(sorted_data, quartile):
  n = length(sorted_data)
  if quartile == 1:
    position = 0.25 * (n + 1)
  elif quartile == 3:
    position = 0.75 * (n + 1)
  # If position is an integer, return the value at that position
  if position is an integer:
```

```
        return sorted_data[position - 1]
    # If position is not an integer, interpolate between the closest ranks
    lower_position = floor(position)
    upper_position = ceil(position)
    return (sorted_data[lower_position - 1] + sorted_data[upper_position - 1]) / 2
# Helper function to sort the dataset
function sort(dataset):
    # Implement a sorting algorithm or use a built-in sort function
    return sorted_dataset
```

Outlier Detection for Z-Score Method

A common statistical tool for detecting outliers, the Z-score approach excels in spotting irregularities in datasets like those about food poisoning. A data point's Z-score indicates the number of standard deviations it has from the dataset mean. To use the Z-score technique, the Z-score for every piece of data Xi, finds value by plugging it into the formula:

$$Z = \frac{X_i - \mu}{\sigma} \qquad (9)$$

Equation 9 illustrates, Where, X_i is the data point, μ is the mean of the dataset, σ is the standard deviation of the dataset. If the Z-score of a data point is greater than a particular threshold, it is usually |Z| > 2 or |Z| > 3.

Table 3 illustrates that in the case of food-related illnesses, this technique may spot out-of-the-ordinary increases or decreases in the prevalence of diseases, which might point to a data-reporting mistake or an epidemic. Because of its simplicity and effectiveness with normally distributed data, the Z-score approach is a great tool for epidemiological investigation.

Table 3. Pseudocode for Outlier Detection using Z-Score Method

Table 3: Pseudocode for Outlier Detection using Z-Score Method

```
# Input: Dataset (list or array of numeric values)
# Output: List of detected outliers
function detect_outliers_Z_score(dataset):
    # Step 1: Calculate the mean (μ) of the dataset
    mean = calculate_mean(dataset)
    # Step 2: Calculate the standard deviation (σ) of the dataset
    std_dev = calculate_standard_deviation(dataset)
    # Step 3: Initialize an empty list to store outliers
    outliers = []
    # Step 4: Iterate through each data point in the dataset
    for each data_point in the dataset:
        # Step 5: Calculate the Z-score for the data point
        Z = (data_point - mean) / std_dev
        # Step 6: Define a threshold for the Z-score
        # Typically, a threshold of 2 or 3 is used for outlier detection
        threshold = 3  # Adjust as needed
        # Step 7: Check if the absolute value of the Z-score is greater than the threshold
        if abs(Z) > threshold:
            # Step 8: If it is an outlier, add it to the outliers list
            outliers.append(data_point)
    # Step 9: Return the list of detected outliers
    return outliers
# Function to calculate the mean
```

```
function calculate_mean(dataset):
   return sum(dataset) / length(dataset)
# Function to calculate the standard deviation
function calculate_standard_deviation(dataset):
   mean = calculate_mean(dataset)
   variance = sum((data_point - mean)^2 for each data_point in dataset) / length(dataset)
   return sqrt(variance)
```

RESULTS

Through the use of both theoretical and practical examples, this article aims to demonstrate how some of the most frequent statistical tests are used to analyze data collected in the field of Food Science and Technology. Our goal has been to avoid using statistical jargon as much as possible and to cite more sophisticated publications where necessary. Consider the pros and cons of these tests in real-world scenarios, and then use the best techniques for ARIMA, SARIMA, and outlier identification to apply machine learning principles.

By contrast, in 2018, there was a slight decrease in the number of complaints during the spring, and the distribution of complaints throughout the year was more evenly spread compared to prior years (Fig. 1). A multinomial test revealed that the distribution of monthly complaints throughout the year was significantly dissimilar to a uniform distribution (where each month would have 1/12 of all complaints) from 2012 to 2024 (P = 0.000). However, in 2018, the distribution was not significantly different (P = 0.124), indicating that the seasonal pattern of complaints in 2018 differed from that of 2012 to 2024.

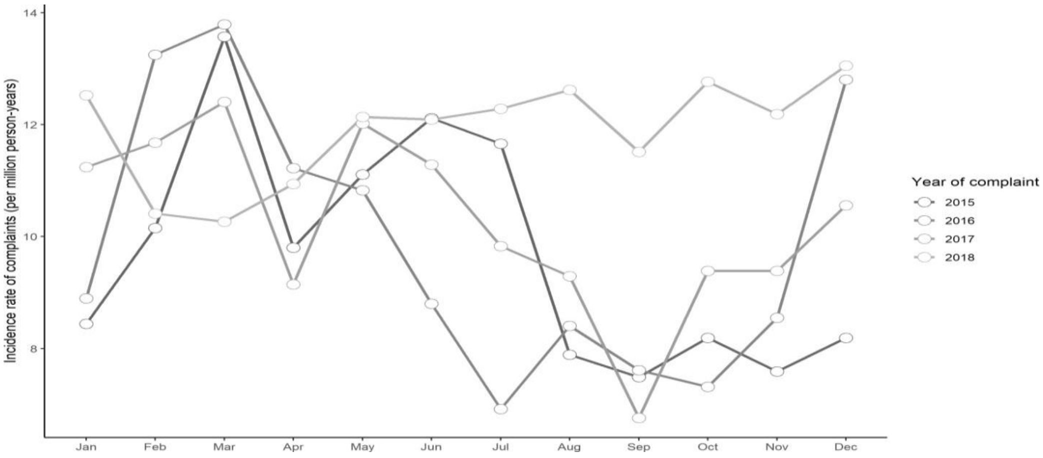

Figure 1. Month of grievance from 2015 and 2018.

Table 4. Age and Gender Compliant in 2012 to 2024

	2012	2016	2020	2024
Gender				
Male	780 (34%)	800 (34.1%)	480 (18%)	1200 (47%)
Female	670 (31%)	650 (26%)	380 (15%)	1002 (33.8%)
Missing	789 (34.4%)	900 (37%)	1200 (65.8%)	567 (39%)
Age				
Mean (SD)	40 (20.1)	41(20.1)	41(20.1)	39.4 (18.8)
Median	39.5 (284)	42.5 (148)	42.1 (247)	40.2 (205)

Complainants' age and sex breakdown

Although not required, Table 4 displays the gender and age of complainants that were most often reported. Sixty-one percent of those who filed complaints identified themselves by gender. Women were more likely to report issues (34% vs. 26%). This pattern stuck around for a long time. We were able to determine the age of 51% of the complainants. In terms of overall age, the average age of complainants remained relatively constant at 42.7 years. The age range was fairly wide, spanning from one to ninety-seven years old.

Points of contention

Figure 2 shows the data that shows how consistent the reporting is to the medical provider databases, state databases (ESSENCE, as e-mail, and FPICN), and the agency of jurisdiction. The number of complaints sent via various channels, including the online form and social media, increased significantly from 2012 to 2024. Regardless of the month, reporting to the proper agency remained the most preferred method, followed by the online form. It seems that there were no seasonal fluctuations in source use, given there was no data visualization.

Fast phone calls or face-to-face meetings with the relevant authorities were the most common means of reporting in both the 0–15 and 50–100 age groups (Fig. 3). The information we have does not make it apparent whether the stated age is of the sick individual or of someone who submitted the complaint in their stead. Since the reporting procedures for those under the age of 18 are identical to those for persons fifty and up, the second option seems to have a better chance of coming to fruition. For complaints from people in the 15–50 age bracket, the online form was the way to go. Reports on social media accounted for a small fraction of total complaints across all age categories, with the highest prevalence among those between the ages of 15 and 25.

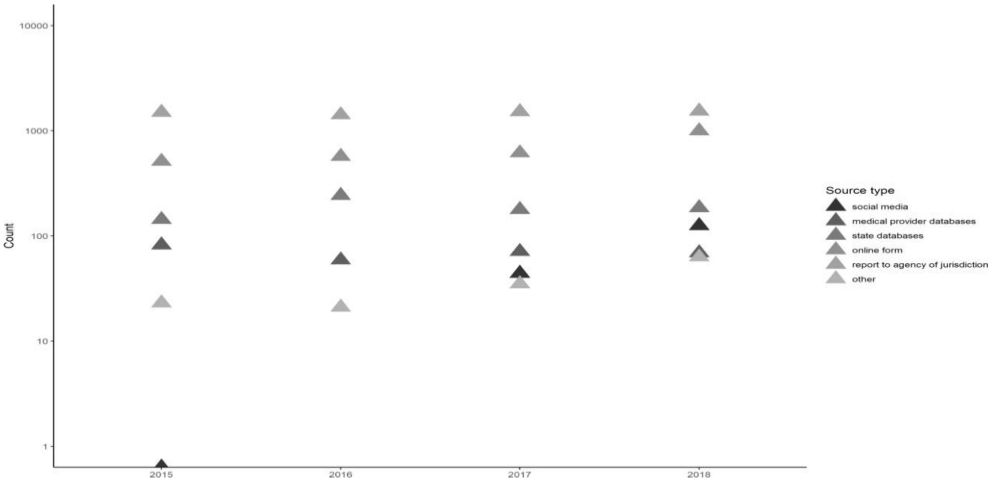

Figure 2. Source frequency on an annual basis from 2012 to 2024.

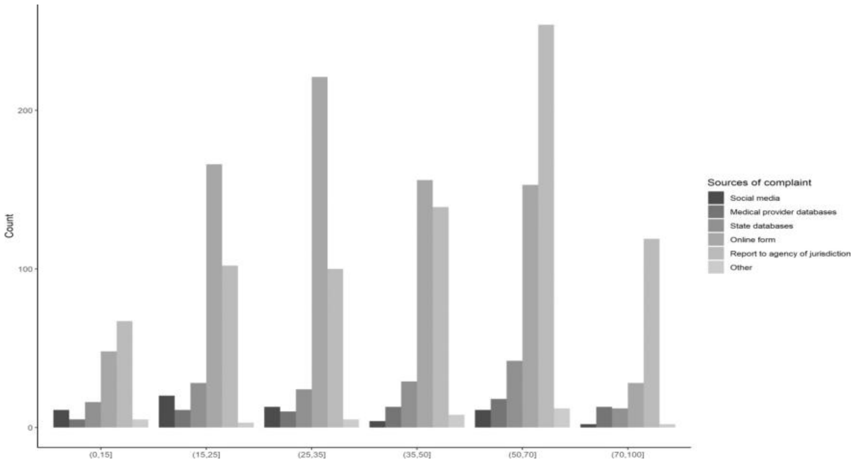

Figure 3. Age range of those who filed the complaint.

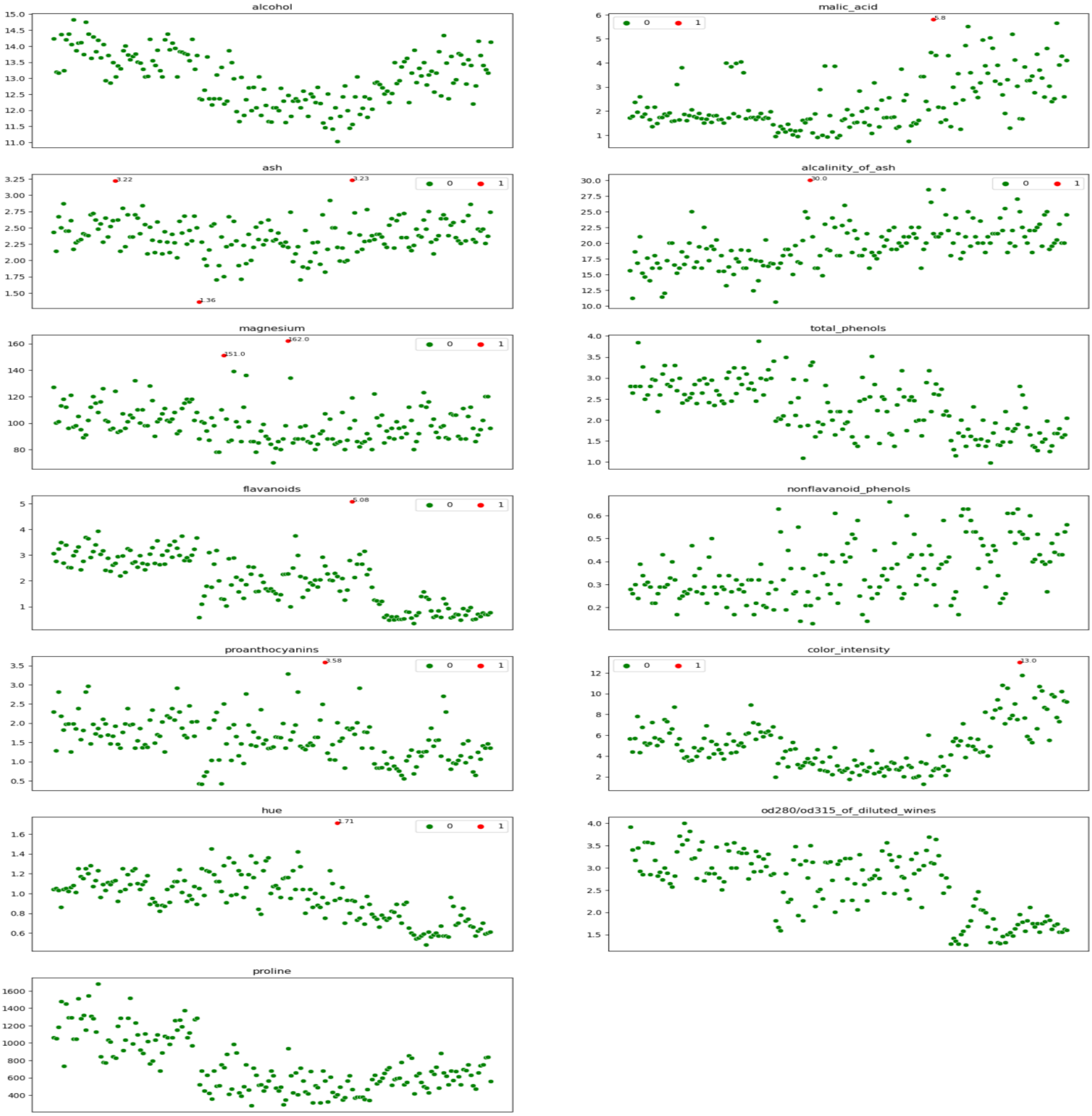

Figure 4. The complaint incidence rate broken down by county from 2012 to 2024. The rate from 2012 to 2024 is shown by dots, while the mean rate for all years is shown by open triangles.

After eliminating self-edges—complaints that come from the same county of residency and anticipated exposure the geographic network is shown in Figure 4. These county pairings were then dispersed around the plot, above and below its diagonal. The number of edges that survived in the geographical networks between 2012 and 2024 was as follows: 262 against 2,115, 289 against 2,182, 325 against 2,460, and 375 against 2,779 (after removal) are the comparisons. These figures correspond to 12%, 13%, 13%, and 14% of all complaints received within that time frame. However, the data also showed that 86% of all complaints included both the county of residence and the probable exposure during this time. Furthermore, the counties of home and presumed exposure were the same in 88% of all complaints. Geographical network demonstrates that Red County and a few other counties in India had the greatest degree of growth between 2012 and 2024.

CONCLUSION

A review of food poisoning cases in India between 2012 and 2024, which were associated with Staphylococcus aureus and Salmonella spp., highlights the critical need for improved methods of monitoring and management. Current methods of monitoring are insufficient, even though Staphylococcus aureus is a very harmful bacterium that often causes food poisoning. Improving the early detection of anomalous patterns that might indicate an impending pandemic is feasible by integrating ARIMA and SARIMA models with unique outlier detection methods such as Z-score, IQR, and Standard Deviation. A solid framework for identifying data anomalies is provided by the use of statistical and machine-learning approaches, allowing for accurate and rapid responses to any threats. Health officials must incorporate these advanced analytical methods with extensive food safety education and extension programs to improve the effectiveness of the food poisoning monitoring system. Furthermore, to provide a good picture of the true national incidence of food poisoning, government programs should include epidemiological and laboratory components. Reduced cases of food poisoning would follow from better monitoring and preventative actions made possible by this coordinated approach. An important step in modernizing public health responses to food poisoning outbreaks is integrating novel methods for detecting outliers with traditional statistical models.

REFRENCES

[1] Faour-Klingbeil D., CD Todd E. Prevention and control of foodborne diseases in Middle-East North African countries: Review of national control systems. Int. J. Environ. Res. Public Health. 2020;17:70. doi: 10.3390/ijerph17010070.

[2] Akkina R.C., Payala V., Maganti S.S. Foodborne Pathogens—Recent Advances in Control and Detection. IntechOpen; London, UK: 2022. Tools for Rapid Detection and Control of Foodborne Microbial Pathogens.

[3] Moi I.M., Ibrahim Z., Abubakar B.M., Katagum Y.M., Abdullahi A., Yiga G.A., Abdullahi B., Mustapha I., Ali J., Mahmud Z. Foodborne Pathogens. IntechOpen; London, UK: 2022. Properties of Foodborne Pathogens and Their Diseases.

[4] Martinović T., Andjelković U., Gajdošik M.Š., Rešetar D., Josić D. Foodborne pathogens and their toxins. J. Proteom. 2016;147:226–235. doi: 10.1016/j.jprot.2016.04.029.

[5] Carstens C.K., Salazar J.K., Darkoh C. Multistate outbreaks of foodborne illness in the United States associated with fresh produce from 2010 to 2017. Front. Microbiol. 2019;10:2667. doi: 10.3389/fmicb.2019.02667.

[6] Ling Y., Zhu Y., Fan H., Zha H., Yang M., Wu L., Chen H., Li W., Wu Y., Chen H. Rapid method for detection of Staphylococcus aureus in feces. J. Biomed. Nanotechnol. 2019;15:1290–1298. doi: 10.1166/jbn.2019.2781.

[7] Akter R., Rahman M., Bhattacharya T., Kaushik D., Mittal V., Parashar J., Kumar K., Kabir M., Tagde P. Novel coronavirus pathogen in humans and animals: An overview on its social impact, economic impact, and potential treatments. Environ. Sci. Pollut. Res. 2021;28:68071–68089. doi: 10.1007/s11356-021-16809-8.

[8] Prata J.C., da Costa J.P., Lopes I., Andrady A.L., Duarte A.C., Rocha-Santos T. A One Health perspective of the impacts of microplastics on animal, human and environmental health. Sci. Total Environ. 2021;777:146094. doi: 10.1016/j.scitotenv.2021.146094.

[9] Qiu H., Zhang Y., Li Z., Jiang P., Guo S., He Y., Guo Y. Donepezil ameliorates pulmonary arterial hypertension by inhibiting M2-macrophage activation. Front. Cardiovasc. Med. 2021;8:639541. doi: 10.3389/fcvm.2021.639541.

[10] Scallan Walter E.J., McLean H.Q., Griffin P.M. Hospital Discharge Data Underascertain Enteric Bacterial Infections among Children. Foodborne Pathog. Dis. 2020;17:530–532. doi: 10.1089/fpd.2019.2773.

[11] Belina D., Hailu Y., Gobena T., Hald T., Njage P.M.K. Prevalence and epidemiological distribution of selected foodborne pathogens in human and different environmental samples in Ethiopia: A systematic review and meta-analysis. One Health Outlook. 2021;3:19. doi: 10.1186/s42522-021-00048-5.

[12] Van Puyvelde L., Aissa A., Panda S.K., De Borggraeve W.M., Mukazayire M.J., Luyten W. Bioassay-guided isolation of antibacterial compounds from the leaves of Tetradenia riparia with potential bactericidal effects on food-borne pathogens. J. Ethnopharmacol. 2021;273:113956. doi: 10.1016/j.jep.2021.113956.

[13] M. R. Adams and M. O. Moss, "Significance of foodborne diseases," Food Microbiology, vol. 2, no. 163, pp. 160–164, 2003.

[14] D. G. Newell, M. Koopmans, L. Verhoef et al., "Food-borne diseases—the challenges of 20 years ago persist while new ones continue to emerge," International Journal of Food Microbiology, vol. 139, supplement 1, pp. S3–S15, 2010.

[15] R. V. Sudershan, R. N. Kumar, and K. Polasa, "Foodborne diseases in India—a review," British Food Journal, vol. 114, no. 5, pp. 661–680, 2012.

[16] CDC, "Estimates of foodborne illness in the United States," 2011, http://www.cdc.gov/foodborneburden/2011-foodborne-estimates.html.

[17] Havelaar, A. H., S. Brul, A. de Jong, R. de Jonge, M. H. Zwietering, and B. H. Ter Kuile. 2010. Future challenges to microbial food safety. Int. J. Food Microbiol. 139:S79–S94. https://doi.org/10.1016/j.ijfoodmicro.2009.10.015

[18] World Health Organization. Initiative to estimate the global burden of foodborne diseases. Available at: http://www.who.int/foodsafety/foodborne_disease/Summary_Doc.pdf?ua¼1. Accessed 15 April 2020

[19] Japan, Ministry of Health, Labour and Welfare. Foodborne disease statistics (data for foodborne disease outbreaks). Available at: https:// www.e-stat.go.jp/stat search/files?page¼1&toukei¼00450191&tstat¼000001040259. Accessed 15 April 2020

[20] Noda, M. 2017. Lessons learned from a large and diffuse outbreak of gastroenteritis caused by consumption of Kizami-nori (chopped seaweed) contaminated with norovirus. Bull. Nath. Inst. Health Sc. 135:6–12.

[21] Bari M.L., Yeasmin S. Food Safety and Preservation. Elsevier; Amsterdam, The Netherlands: 2018. Foodborne diseases and responsible agents; pp. 195–229.

[22] Aworh O.C. Food safety issues in fresh produce supply chain with particular reference to sub-Saharan Africa. Food Control. 2021;123:107737. doi: 10.1016/j.foodcont.2020.107737.

[23] Newell D.G., Koopmans M., Verhoef L., Duizer E., Aidara-Kane A., Sprong H., Opsteegh M., Langelaar M., Threfall J., Scheutz F. Food-borne diseases—The challenges of 20 years ago persist while new ones continue to emerge. Int. J. Food Microbiol. 2010;139:S3–S15. doi: 10.1016/j.ijfoodmicro.2010.01.021.

[24] Bintsis T. Foodborne pathogens. AIMS Microbiol. 2017;3:529. doi: 10.3934/microbiol.2017.3.529.

[25] Cremonesi P., Pisani L.F., Lecchi C., Ceciliani F., Martino P., Bonastre A.S., Karus A., Balzaretti C., Castiglioni B. Development of 23 individual TaqMan® real-time PCR assays for identifying common foodborne pathogens using a single set of amplification conditions. Food Microbiol. 2014;43:35–40. doi: 10.1016/j.fm.2014.04.007.

[26] Atreya C. Major foodborne illness-causing viruses and current status of vaccines against the diseases. Foodborne Pathog. Dis. 2004;1:89–96. doi: 10.1089/153531404323143602.

[27] O'Shea H., Blacklaws B.A., Collins P.J., McKillen J., Fitzgerald R. Viruses associated with foodborne infections. Ref. Modul. Life Sci. 2019 doi: 10.1016/B978-0-12-809633-8.90273-5.

[28] De Berardinis A., Paludi D., Pennisi L., Vergara A. Toxoplasma gondii, a foodborne pathogen in the swine production chain from a European perspective. Foodborne Pathog. Dis. 2017;14:637–648. doi: 10.1089/fpd.2017.2305.

[29] Ullah F., Ayaz M., Sadiq A., Ullah F., Hussain I., Shahid M., Yessimbekov Z., Adhikari-Devkota A., Devkota H.P. Potential role of plant extracts and phytochemicals against foodborne pathogens. Appl. Sci. 2020;10:4597. doi: 10.3390/app10134597.

[30] Bowman D.D. Ascaris and Toxocara as foodborne and waterborne pathogens. Res. Vet. Sci. 2021;135:1–7. doi: 10.1016/j.rvsc.2020.12.017.

[31] Bhunia A.K. Foodborne Microbial Pathogens. Springer; Berlin/Heidelberg, Germany: 2018. Molds and mycotoxins; pp. 167–174.

[32] Awuchi C.G., Ondari E.N., Ogbonna C.U., Upadhyay A.K., Baran K., Okpala C.O.R., Korzeniowska M., Guiné R.P. Mycotoxins affecting animals, foods, humans, and plants: Types, occurrence, toxicities, action mechanisms, prevention, and detoxification strategies—A revisit. Foods. 2021;10:1279. doi: 10.3390/foods10061279.

[33] Yaro, A. S., Maly, F., Prazak, P., & Malý, K. (2024). Outlier Detection Performance of a Modified Z-Score Method in Time-Series RSS Observation with Hybrid Scale Estimators. IEEE Access, 12, 12785–12796. https://doi.org/10.1109/access.2024.3356731

[34] Khan, H., Rasheed, M. T., Zhang, S., Wang, X., & Liu, H. (2024). Empirical study of outlier impact in a classification context. Expert Systems With Applications, 256, 124953. https://doi.org/10.1016/j.eswa.2024.124953