ORIGINAL ARTICLE

# Detection of aflatoxin $B_1$ in chili powder using attenuated total reflectance–Fourier transform infrared spectroscopy

Thin Thin Sein [a,b,1], Urairat Mongmonsin [b,1], Patutong Chatchawal [c], Molin Wongwattanakul [c,e], Oranee Srichaiyapol [c], Rungrueang Pattanakul [d], Patcharaporn Tippayawat [b,e,*]

[a] Medical Technology Program, Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, 40002, Thailand
[b] Center for Research and Development of Medical Diagnostic Laboratory (CMDL), Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, 40002, Thailand
[c] Center for Innovation and Standard for Medical Technology and Physical Therapy (CISMaP), Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, 40002, Thailand
[d] Synchrotron Light Research Institute (SLRI), 111 University Avenue, Muang District, Nakhon Ratchasima, 30000, Thailand
[e] Department of Medical Technology, Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, 40002, Thailand

**Abstract**

**Aflatoxin $B_1$, a major global food safety concern, is produced by toxigenic fungi during crop growing, drying, and storage, and shows increasing annual prevalence. This study aimed to detect aflatoxin $B_1$ in chili samples using ATR–FTIR coupled with machine learning algorithms. We found that 83.6% of the chili powder samples were contaminated with *Aspergillus* and *Penicillium* species, with aflatoxin $B_1$ levels ranging from 7.63 to 44.32 μg/kg. ATR–FTIR spectroscopy in the fingerprint region (1800–400 $cm^{-1}$) showed peak intensity variation in the bands at 1587, 1393, and 1038 $cm^{-1}$, which are mostly related to aflatoxin $B_1$ structure. The PCA plots from samples with different trace amounts of aflatoxin $B_1$ could not be separated. Vibrational spectroscopy combined with machine learning was applied to address this issue. The logistic regression model had the best *F1 score* with the highest %accuracy (73%), %sensitivity (73%), and %specificity (71%), followed by random forest and support vector machine models. Although the logistic regression model contributed significant findings, this study represents a laboratory research project. Because of the peculiarities of the ATR–FTIR spectral measurements, the spectra measured for several batches may differ, necessitating running the model on multiple spectral ranges and using increased sample sizes in subsequent applications. This proposed method has the potential to provide rapid and accurate results and may be valuable in future applications regarding toxin detection in foods when simple onsite testing is required.**

*Keywords:* Aflatoxin $B_1$, Attenuated total reflection–Fourier transform infrared spectroscopy (ATR–FTIR), Chili powder, Machine learning, Principal component analysis

## 1. Introduction

The chili pepper (*Capsicum* spp.) is the most widely used spice and condiment in the world for its pungency and for adding special flavor to many cuisines. Chilis are consumed as fresh unripe fruits, ripened (red or other colors), and dried [1]. In Thailand, the average consumption of Thai chilis was 5 g/d, as surveyed by the Ministry of Public Health [2]. A considerable concern regarding chili consumption is the reported high prevalence of aflatoxin $B_1$ ($AFB_1$) in chili powder, which was estimated to range from 96.7% to 100% from 2014 to 2016 [3,4].

Received 21 December 2023; accepted 5 February 2024.
Available online 15 June 2024

* Corresponding author at: 123 Mittraphap Rd., Department of Medical Technology, the Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, Thailand.
E-mail address: patchatip@kku.ac.th (P. Tippayawat).
[1] These authors contributed equally to this work.

https://doi.org/10.38212/2224-6614.3497

Aflatoxins are a group of highly mutagenic, teratogenic, and carcinogenic mycotoxins produced by certain species of fungi (including the genera *Aspergillus*, *Penicillium*, and *Fusarium*) under favorable conditions before and after harvest and during storage [5]. The presence of mycotoxins in food and feed is a major global food safety issue, as they contribute to potential health hazards to humans and animals. Exposure to aflatoxins is an important risk factor for the development of hepatocellular carcinoma [6]. The toxic effects of aflatoxins mainly depend on the exposure levels and amount of food consumed and may extend from acute infirmity or death to chronic problems [7]. The four types of aflatoxins produced in nature belong to a group of difuranocoumarins and include $AFB_1$, aflatoxin $B_2$ ($AFB_2$), aflatoxin $G_1$ ($AFG_1$), and aflatoxin $G_2$ ($AFG_2$) [8,9]. $AFB_1$ is the most dominant toxin form in agricultural products and is classified as a Group I compound (carcinogenic to humans) by the International Agency for Research on Cancer [10,11]. Currently, aflatoxins induce approximately 4.6% of the total annual hepatocellular carcinoma cases and 28.2% of all cases of hepatocellular carcinoma globally, and more than 55 billion people worldwide have been exposed to uncontrolled toxin contamination [12].

Several technologies have been explored for the detection of $AFB_1$ in foods, including thin-layer chromatography (TLC), high-performance liquid chromatography (HPLC), and enzyme-linked immunosorbent assay (ELISA) [13,14]. Despite yielding accurate results, these procedures are generally time-consuming, sample-destructive, costly, and require expert staff. These conditions are largely unattainable within the massive scope of non-destructive screening and real-time and on-site analysis [15]. Rapid and accurate techniques to detect $AFB_1$ in foods are thus needed urgently. Infrared (IR) spectroscopy comprises rapid and non-destructive techniques that require minimal technical training and sample preparation, are not labor-intensive, and can be performed using relatively minute quantities of chemicals [16]. Attenuated total reflection–Fourier transform IR (ATR–FTIR) spectroscopy is a frequently used vibrational spectroscopic technique that has been experimentally applied for identification of multiple samples. Recently, machine learning has gained popularity and has been employed in several fields, such as pattern recognition, object detection, text interpretation, and different research areas [17–19]. Previous studies have applied vibrational spectroscopy combined with machine learning algorithms to discriminate various samples, including sera from breast cancer patients against healthy sera [20], pharmaceutical polymorphs [21], and biofluids [22]. Thus, vibrational spectroscopy combined with machine learning algorithms is a promising alternative technique to detect $AFB_1$ in foods.

In this study, we used ATR–FTIR spectroscopy to investigate contamination of chili samples with $AFB_1$ and aimed to develop a model to discriminate $AFB_1$ spectra using chemometrics. The unsupervised and machine learning algorithms, principal component analysis (PCA), support vector machine (SVM), logistic regression, k-nearest neighbors, and random forest (RF) were established and evaluated by calculating *F1 scores*, %accuracy, %sensitivity, and %specificity.

## 2. Materials and methods

### 2.1. Fungal strains

Aflatoxin-positive *Aspergillus parasiticus* (ATCC® 26691™) was purchased from Biomedia, Thailand. The strain was inoculated onto Czapek Dox medium according to the ATCC product sheet instructions.

### 2.2. Instruments and chemicals

An FTIR spectrometer (TENSOR II; Bruker Optics, Ettlingen, Germany) equipped with a platinum ATR accessory was used. The MaxSignal Aflatoxin $B_1$ ELISA kit was purchased from PerkinElmer Inc. (Texas, USA). The $AFB_1$ standard was purchased from Merck (USA). A 365-nm UV lamp was purchased from China. Peptone (bacteriological), Dichloran Glycerol Medium Base, and Czapek Dox agar were purchased from Himedia (India). Commercial coconut milk was purchased from Thailand.

### 2.3. Provision of chili powder samples

From September 2020 to April 2021, 73 chili powder samples were randomly collected from four merchants and vendors in Khon Kaen province, Thailand. The majority (35 samples) were obtained from the Bang-Lampu market (market 1), 20 from the Non-Maung market (market 2), 11 from the Kumhai market (market 3), and 7 from the Ban Non-Than market (market 4). The samples were packed in sterile plastic bags and kept at 2–8 °C until analysis.

#### 2.3.1. Fungal isolation from chili powder samples

Fungal isolation followed the ISO 21527-2 standard protocol. Briefly, 10 g of each chili powder sample was suspended in 90 mL of 0.1% (w/v)

peptone water and mixed vigorously. The initial suspension was serially diluted to a final concentration of 0.01 g/mL using 0.1% (w/v) peptone water. Each dilution of chili powder suspension was spread on Dichloran 18% glycerol (DG18) agar plates and incubated at 25 °C for 5 d, before counting the fungal colonies. The total fungal count of each contaminated sample was expressed as colony-forming units per sample g (CFU/g).

#### 2.3.2. Screening of aflatoxin-producing fungal species using UV fluorescence

Screening of aflatoxin-producing fungi in the chili powder samples was based on the fluorescence of positive colonies after exposure to 365-nm UV radiation. The fungi were incubated in coconut agar medium using the modified formula described by Lin and Dianese [23]. The isolated *Aspergillus* strains were placed in the center of coconut agar medium and cultured for 3 d at 25 °C. After incubation, the presence or absence of blue or blue-green fluorescence in the agar surrounding the colony was evaluated as indicative of aflatoxin-producing fungi, using the *A. parasiticus* strain ATCC® 26691™ as a positive reference.

#### 2.3.3. Identification of fungal species

*Aspergillus* and *Penicillium* species were isolated on DG18 plates and identified using lactophenol cotton blue stain. The taxonomic classification was done following the Manual of Identification, as detailed in Description of Medical Fungi (https://www.adelaide.edu.au/mycology/ua/media/1596/fungus3-book.pdf). Isolated colonies of identified *Aspergillus* species were subcultured on Czapek Dox agar and incubated at 25 °C for 3 d to study their morphological characteristics. Yellow-green colonies growing on Czapek Dox agar were identified as *Aspergillus flavus* and dark green colonies as *A. parasiticus*. The incidence of each fungal species was calculated as follows:

$$\text{Species occurrence (\%)} = (\text{number of species isolates /total number of isolates}) \times 100$$

### 2.4. Quantification of $AFB_1$ content using competitive ELISA

$AFB_1$ content was assessed using competitive ELISA following the manufacturer's instructions. Briefly, 5 g of chili powder sample was subjected to extraction with 25 mL of 70% v/v methanol in water. The chili powder extraction was centrifuged at 4000×*g* for 10 min to collect the supernatant that contained $AFB_1$. The levels of $AFB_1$ in the supernatant were measured in duplicate using the competitive ELISA kit. The absorbance was set at 450 nm for the primary filter and at 630 nm for an additional differential filter to decrease background measurements. The optimum dilution of the antibody to ensure maximum sensitivity was determined by the displacement values of $B/B_0$, where B is the extinction coefficient of the sample containing toxins and $B_0$ of the sample without toxins, derived from the slope of the calibration curves. Samples containing $AFB_1$ in concentrations below the limit of <20 μg/kg were considered uncontaminated [14].

### 2.5. ATR–FTIR spectroscopy for $AFB_1$ analysis

The methanolic extract of the chili powder was centrifuged at 1509×*g* for 15 min. Six microliters of the supernatant were analyzed using ATR–FTIR spectroscopy over the surface of an ATR–FTIR sampling window (diameter 3 mm) and allowed to dry for 3 min in air at room temperature. After complete drying, the transparent film was examined to determine absorbance at a spectral resolution of 8 $cm^{-1}$ and 64 scans in the 4000–400 $cm^{-1}$ spectral range. Spectral data were collected from five replicates for each sample.

### 2.6. ATR–FTIR spectral preprocessing and unsupervised analysis: PCA

PCA allows dimensionality reduction and simplification of datasets with multiple variables to increase interpretability and minimize information loss [24]. The spectral data, raw (with 17 smoothing points using the Gaussian filter algorithm and Rubberband baseline correction) and preprocessed (with 2nd derivative with 17 smoothing points using the Savitzky–Golay algorithm and unit vector normalization), were used for PCA in the following six spectral ranges: (i) 4000–400 $cm^{-1}$, (ii) 2950–2800 $cm^{-1}$, (iii) 1700–900 $cm^{-1}$, (iv) 1450–900 $cm^{-1}$, (v) 2950–2800 + 1700–900 $cm^{-1}$, and (vi) 2950–2800 + 1450–900 $cm^{-1}$ using the Orange software (version 3.35.0, University of Ljubljana, Slovenia).

### 2.7. Establishing advanced machine learning models

Different machine learning algorithms were employed to evaluate classification performance with the parameters listed in Table 1. The calibration set was validated using the leave-one-out method. Each sample was sequentially omitted from the calibration set and the machine learning model

*Table 1. Parameters for advanced machine learning analysis.*

| Machine learning models | Parameter | |
|---|---|---|
| Support vector machine | Radial basis function (RBF) | 1.00 |
| | Cost | 0.10 |
| | Regression loss epsilon | 0.0010 |
| | Numerical tolerance | 100 |
| | Iteration limit | |
| Logistic regression | Regularization type | Ridge |
| | Cost strength | 1 |
| K-nearest neighbor | Number of neighbors | 5 |
| | Metric | Euclidean |
| | Weight | Uniform |
| Random forest | Number of trees | 10 |

(determined based on the reduced dataset) was used to predict contamination with $AFB_1$ in the omitted sample. The predictive model with an *F1 score* value approaching 1 was considered acceptable [25]. To avoid overly optimistic modeling, no technical replicates from the same sample were included in the training or test sets, i.e., the technical replicate trap. The *F1 score* was calculated as follows:

$$F1\ score = \left( \frac{TP}{TP + \frac{1}{2}(FP + FN)} \right)$$

Where TP is the true positive value, FN is the false negative value, and FP is the false positive value.

### 2.8. Method evaluation and calculation

The predictive results of each model are presented in Table 2 for comparison of the actual $AFB_1$ levels in the chili powder samples and the index test results. Percent accuracy, sensitivity, and specificity were calculated using the following formulas:

$$\%\ Accuracy = \left( \frac{TP + TN}{TP + FP + FN + TN} \right) \times 100$$

$$\%\ Sensitivity = \left( \frac{TP}{TP + FN} \right) \times 100$$

*Table 2. Prediction performance between reference and index tests.*

| Index test (Predicted value) | Actual value | |
|---|---|---|
| | $AFB_1$ contaminated | $AFB_1$ uncontaminated |
| $AFB_1$ contaminated | TP | FP |
| $AFB_1$ uncontaminated | FN | TN |

Chili powder samples with aflatoxin $B_1$ levels <20 μg/kg were considered uncontaminated.
Definitions: TP, true positive; FP, false positive; FN, false negative; TN, true negative.

$$\%\ Specificity = \left( \frac{TN}{FP + TN} \right) \times 100$$

Where TP is the true positive value, FP is the false positive value, FN is the false negative value, and TN is the true negative. The evaluation results of % accuracy, %sensitivity, and %specificity were expressed as the average value over classes ($AFB_1$ contaminated and uncontaminated).

## 3. Results

### 3.1. Incidence of $AFB_1$ in the chili powder samples

The occurrence of toxigenic fungal contamination in chili powder gathered in Khon Kaen province is displayed as a heatmap (Fig. 1), with the levels of total fungal contamination ranging from $1.0 \times 10^2$ to $3.5 \times 10^4$ CFU/g (shift from blue to red in the heatmap). The results showed that 83.6% (61 of 73) of the chili powder samples contained toxigenic fungi. The most frequently contaminating fungi were *Aspergillus* species (93.4%; 57 of 61 samples), with more than one species per sample occasionally detected. We identified *A. flavus* in 68.2% (22 of 61), *A. parasiticus* in 55.8% (43 of 61), and *Aspergillus niger* in 87.0% (46 of 61) of the samples. *Penicillium* species were present in 6.6% of the samples (4 of 61; Table 3). Only 16.4% (12 of 73) of the chili samples were found to be uncontaminated.

### 3.2. Detection of $AFB_1$ with ELISA

Aflatoxin production was evaluated using the competitive ELISA method. The levels of $AFB_1$ were calculated following formula $Y = -0.158\ln(x) + 0.2281$ ($R^2 = 0.99$) according to the standard curve (Fig. S1). The $AFB_1$ levels in the chili powder extracts were in the range of 7.63–44.32 μg/kg, with levels exceeding the limit of 20 μg/kg being considered contaminated by the US Food and Drug Administration (FDA). Therefore, the assessment of $AFB_1$ levels led to division into contaminated (57.5%) and uncontaminated (42.5%) samples, as presented in Table 4.

### 3.3. Analysis of $AFB_1$ using ATR–FTIR spectroscopy

The presence of $AFB_1$ in the chili powder extracts was assessed using ATR–FTIR spectroscopy in the spectral range 4000–400 $cm^{-1}$. Variations in terms of absorbance peaks were apparent in the fingerprint region (1800–400 $cm^{-1}$), as shown in Fig. 2. Higher absorbance peak was observed in the contaminated (red line) than in the uncontaminated

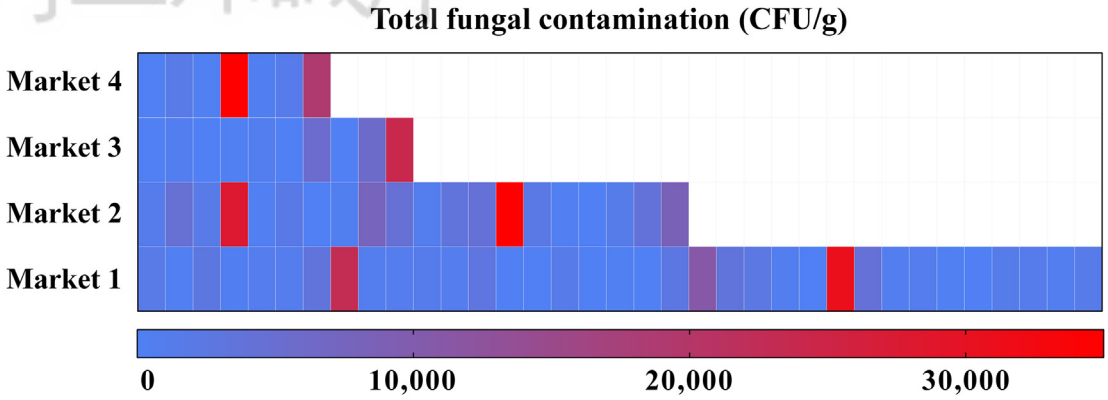



*Fig. 1. Heatmap of total fungal contamination using the ISO 21527-2 method. A blue-to-red shift corresponds to increasing fungal concentration (colony-forming units per g [CFU/g]). Samples with <10 CFU/g were considered uncontaminated.*

*Table 3. Evaluation of fungal contamination using the ISO 21527-2 protocol.*

| Interpretation using the ISO 21527-2 protocol[a] | No. of contaminated or uncontaminated samples[b] | Species occurrence (%)[c] |
|---|---|---|
| **Contaminated $AFB_1$ samples** | 83.6% (61/73) | *Aspergillus* spp. (93.4%)<br>*A. flavus* (68.2%)<br>*A. parasiticus* (55.8%)<br>*A. niger* (87.0%)<br>*Penicillium* spp. (6.6%) |
| **Uncontaminated $AFB_1$ samples** | 16.4% (12/73) | – |

-: not detected.
[a] The maximum acceptable value for chili powder samples contaminated with fungal was <10 CFU/g.
[b] Number of uncontaminated samples or contaminated samples/number of tested samples.
[c] (Number of species isolates)/(total number of isolates) × 100.

*Table 4. Assessment of contamination of chili powder samples with aflatoxin $B_1$ using competitive ELISA.*

| | Uncontaminated samples | Contaminated samples |
|---|---|---|
| Concentration of aflatoxin $B_1$ (μg/kg) | 7.63–19.89 | 20.22–44.32 |
| % frequency distribution | 42.5% (31 of 73) | 57.5% (42 of 73) |

samples (black line). The fingerprint region shows variation in the contaminated and uncontaminated samples in the bands at 1587, 1393, and 1038 $cm^{-1}$. The ATR–FTIR spectra showed a band at 2925–2853 $cm^{-1}$ for aromatic (=CH), =C–H, C=C, and phenyl structures; 1706 $cm^{-1}$ for C=O; 1587 $cm^{-1}$ for aromatic ring (C–C) and C=C stretching; 1393 $cm^{-1}$ for –C–H bending; and 1038 $cm^{-1}$ for =C–O–C or symmetric bending of phenyls; these are mostly related to the $AFB_1$ structure.

### 3.4. $AFB_1$ spectral discrimination using unsupervised analysis: PCA

The spectra acquired from the 31 uncontaminated and 42 contaminated chili powder samples were preprocessed and used to investigate sample clustering to help determine sample similarity and dissimilarity. The PCA plots from the spectral regions 1450–900 $cm^{-1}$ exhibited the highest explained variance (>99%) among other spectral regions (Fig. 3). However, the discriminatory power of PC1 (*x*-axis) versus other components (*y*-axis) was unclear among the contaminated (red dots) and uncontaminated (green squares) $AFB_1$ samples in the spectral range 1450–900 $cm^{-1}$.

### 3.5. Establishing advanced machine learning models

Vibrational spectroscopy combined with machine learning algorithms was used to classify $AFB_1$ levels in chili powder samples using predictive models. As presented in Table 5, based on the *F1 scores* determined using machine learning algorithms in six spectral ranges, the range 1700–900 $cm^{-1}$ contributed to better *F1 score* values than the combined spectra 2950–2800 + 1700–900 $cm^{-1}$ and other ranges. The raw spectral data obtained by preprocessing 17 smoothing points with Rubberband baseline correction and vector normalization were suitable for use with the predictive models. The logistic regression model based on these raw spectral data resulted in the highest *F1 score* (0.725), with 73% accuracy, 73% sensitivity, and 71% specificity, followed by RF (*F1 score* 0.657, 66% accuracy, 66%

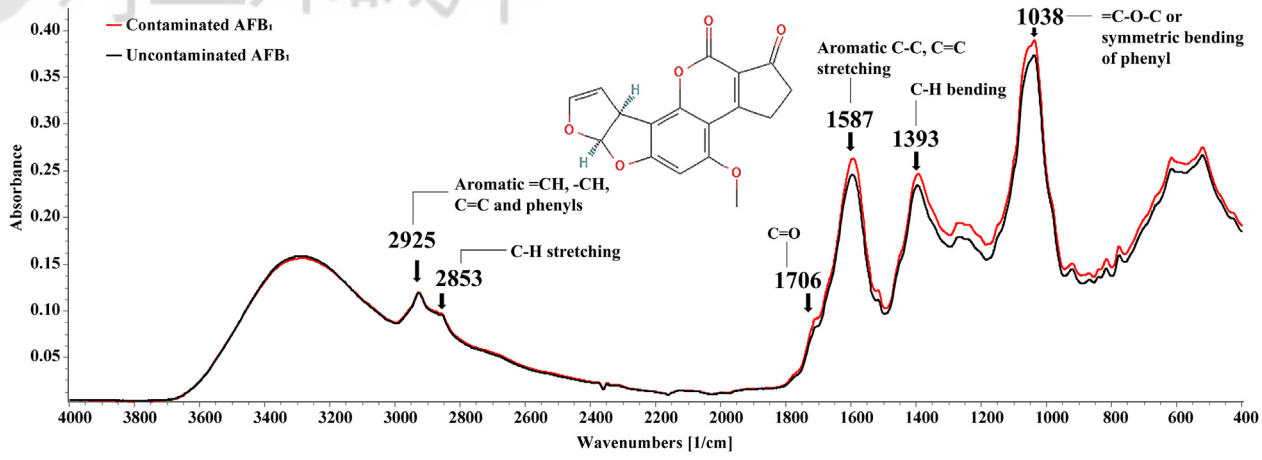



*Fig. 2. Averaged ATR–FTIR spectra profiles from chili powder samples containing various aflatoxin $B_1$ ($AFB_1$) levels. The contaminated (red) and uncontaminated (black) samples were collected in the spectral range of 4000–400 $cm^{-1}$.*

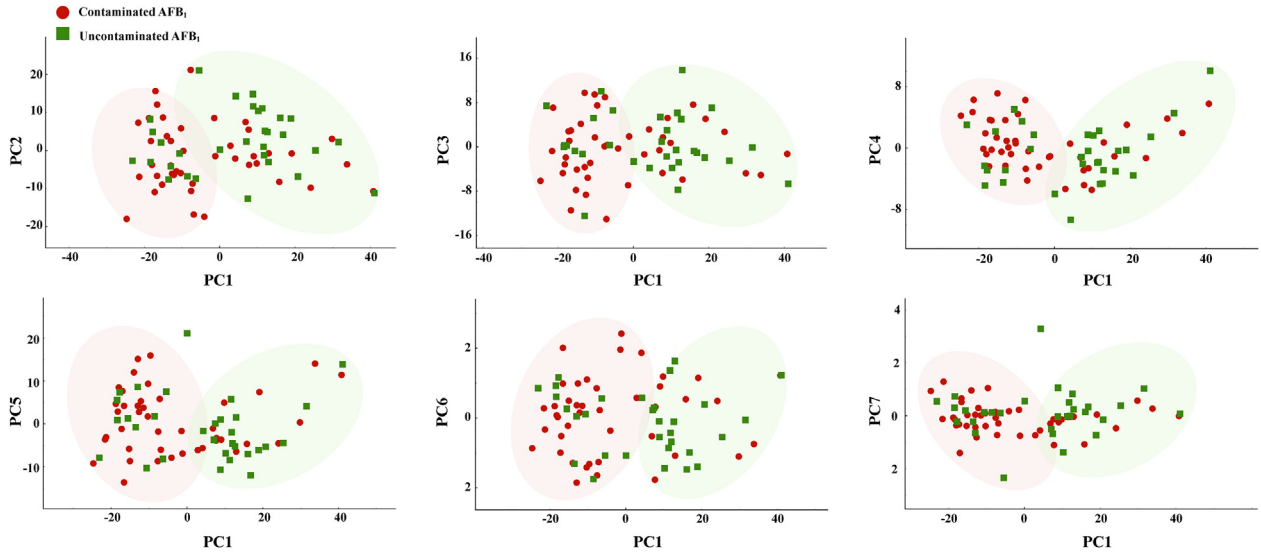



*Fig. 3. Principal component score plot representing clusters of $AFB_1$ levels contained in chili powder samples at the wavenumber range of 1450–900 $cm^{-1}$ with 99% of explained variance.*

sensitivity, and 64% specificity) and SVM (*F1 score* 0.645, 66% accuracy, 66% sensitivity, and 60% specificity) in the spectral range 1700–900 $cm^{-1}$ (Table S1 and Table S2). These results indicate that logistic regression was suitable for detecting $AFB_1$ in the chili powder samples.

## 4. Discussion

Toxigenic fungi can attack and colonize chilies, grains, peanuts, and dairy products during crop production, drying, and storage. Warm temperatures and high humidity are crucial factors that influence fungal growth and mycotoxin production. Aflatoxin contamination can be divided into two phases: during early crop development and after development, where contamination increases until consumption [26]. Our findings revealed contamination with toxigenic fungus in 83.6% of the chili powder samples collected in Khon Kaen province. The contaminating fungi were most commonly of the *Aspergillus* species (93.4%), including *A. flavus*, *A. parasiticus*, and *A. niger*, followed by *Penicillium* species (6.6%). These two fungal species are typically present in spices collected from various locations [27]. It has been reported that a significant percentage of chili powder samples (86.7%) and dried chili pods (96.7%) are contaminated with fungi. The most common strains of fungi found in

*Table 5. Predictive models for evaluation of aflatoxin $B_1$ levels using advanced machine learning.*

| Spectral preprocessing | PCA (% explained variance) | Spectral region ($cm^{-1}$) | *F1 score* SVM | LR | kNN | RF |
|---|---|---|---|---|---|---|
| 17 smoothing points + Rubberband baseline correction | 92% | 4000–400 | 0.529 | 0.489 | 0.532 | 0.508 |
| | 99% | 2950–2800 | 0.531 | 0.531 | 0.671 | 0.581 |
| | 99% | 1700–900 | 0.497 | 0.618 | 0.609 | 0.548 |
| | 99% | 1450–900 | 0.569 | 0.587 | 0.543 | 0.541 |
| | 98% | 2950–2800 + 1700–900 | 0.525 | 0.658 | 0.667 | 0.583 |
| | 98% | 2950–2800 + 1450–900 | 0.612 | 0.605 | 0.544 | 0.513 |
| 17 smoothing points + Rubberband baseline correction + Vector normalization | 89% | 4000–400 | 0.531 | 0.489 | 0.569 | 0.562 |
| | 99% | 2950–2800 | 0.621 | 0.605 | 0.494 | 0.544 |
| | 98% | 1700–900 | 0.645 | **0.725** | 0.577 | 0.657 |
| | 99% | 1450–900 | 0.602 | 0.609 | 0.583 | 0.521 |
| | 97% | 2950–2800 + 1700–900 | 0.571 | **0.670** | 0.637 | 0.584 |
| | 98% | 2950–2800 + 1450–900 | 0.571 | 0.590 | 0.544 | 0.541 |
| 2nd derivatives + 17 smoothing points + Rubberband baseline | 50% | 4000–400 | 0.571 | 0.574 | 0.576 | 0.612 |
| | 95% | 2950–2800 | 0.600 | 0.534 | 0.549 | 0.497 |
| | 86% | 1700–900 | 0.569 | 0.541 | 0.576 | 0.537 |
| | 88% | 1450–900 | 0.629 | 0.640 | 0.576 | 0.501 |
| | 84% | 2950–2800 + 1700–900 | 0.576 | 0.588 | 0.565 | 0.449 |
| | 83% | 2950–2800 + 1450–900 | 0.600 | 0.621 | 0.501 | 0.525 |
| 2nd derivatives + 17 smoothing points + Rubberband baseline + Vector normalization | 50% | 4000–400 | 0.571 | 0.565 | 0.547 | 0.457 |
| | 92% | 2950–2800 | 0.617 | 0.587 | 0.632 | 0.655 |
| | 83% | 1700–900 | 0.576 | 0.548 | 0.587 | 0.640 |
| | 86% | 1450–900 | 0.600 | 0.565 | 0.547 | 0.543 |
| | 87% | 2950–2800 + 1700–900 | 0.605 | 0.553 | 0.569 | 0.492 |
| | 86% | 2950–2800 + 1450–900 | 0.514 | 0.593 | 0.553 | 0.469 |

Definitions: PCA, principal component analysis; SVM, support vector machine; LR, logistic regression; kNN, k-nearest neighbors algorithm; RF, random forest.

these samples include *Aspergillus* section *Flavi*, *Aspergillus* section *Nigri*, *Penicillium* spp., *Aspergillus* section *Aspergillus*, *Aspergillus* section *Circumdati*, and *Rhizopus* [28]. Aflatoxins are naturally occurring toxins that can be found in certain types of food, such as peanuts, corn, and other grains. Among the different types of naturally occurring aflatoxins, B aflatoxins ($AFB_1$ and $AFB_2$) are more prevalent in chilis and peanut products than G aflatoxins [9,29].

Here, we used ATR–FTIR spectroscopy as a rapid and accurate strategy for the detection of $AFB_1$. The ATR–FTIR spectroscopy results revealed variations in the fingerprint region (1800–400 $cm^{-1}$), with higher absorbance peaks for the contaminated than the uncontaminated samples (Fig. 2). These differences are attributed to the interaction between the biochemical components in the samples and their amounts. Our findings show absorbance peak variation for the bands at 1587, 1393, and 1038 $cm^{-1}$, which are mostly related to the $AFB_1$ structure. The band at 1062–1000 $cm^{-1}$ has the strongest association between absorbance responses and aflatoxin levels [30]. The contamination of aflatoxins includes B1, B2, G1, and G2 subtypes exhibits the FTIR characteristic bands at ~3000–2900 $cm^{-1}$ and ~1800–900 $cm^{-1}$, therefore the spectral rationing exposed minor differences along each pure subtype standards [31]. Since overlapping spectral patterns were observed in the major peaks, spectral discrimination of aflatoxin subtypes cannot be conducted. Hence chosen spectral range cannot rule out the contamination of other subtypes so the limitation should be concerned. However, the qualitative and quantitative differentiation between aflatoxin subgroups can be identified by using chemometric model of each standard subtype [31]. ATR–FTIR spectroscopy and chemometrics were combined to categorize the uncontaminated and contaminated chili powder samples. The PCA score plots for $AFB_1$ contamination at different trace amounts could not be separated (Fig. 3) likely because the absorption peaks for samples with trace amounts of aflatoxin subtypes were barely discernible. Thus, PCA failed to discriminate the levels of $AFB_1$ in the chili samples.

Vibrational spectroscopy combined with machine learning was applied to address these issues and to classify samples that contained various $AFB_1$ levels. The logistic regression model resulted in the best *F1 score* with the highest %accuracy, %sensitivity, and %specificity (Table 5) followed by RF and SVM. A previous study demonstrated that the RF model

achieved high specificity (>80%) and sensitivity (76.2%) for $AFB_1$ detection in maize samples [32]. However, this preliminary study demonstrated the choice of spectral preprocessing procedures and machine learning algorithms to identify $AFB_1$ contamination. Improving the test accuracy by increasing the sample size and performing algorithm tuning may be eligible for a routine approach [33]. Moreover, a simple sample handling procedure also needed for routine application. The study of Sulistiawan A. et al. evidenced the high performance of the PLS model from peanut paste spectra [34].

In conclusion, our study demonstrated the application of a robustness tool to identify the contamination of aflatoxin in chili powder samples. Additionally, the spectral preprocessing procedures (17 smoothing points, rubber band baseline correction, vector normalization at 1700–900 $cm^{-1}$) and integration of machine learning for pattern-based modeling executed the impact of the screening method. The combination of logistic regression algorithm for a spectral analysis accounted 73% accuracy, 73% sensitivity, and 71% specificity. The limitations of our study included 1) low sample size for machine learning, 2) fine-tuning for the best parameters, and 3) application of aflatoxin subtype standards and chemometric model for subtype identification. Therefore, test performance improvement could be conducted by increasing the sample size and performing algorithm tuning which may be eligible for a routine approach. The establishment of chemometric models for identifying of each subtype of aflatoxin (B1, B2, G1, and G2) could be conducted using the standard of individual subtypes.

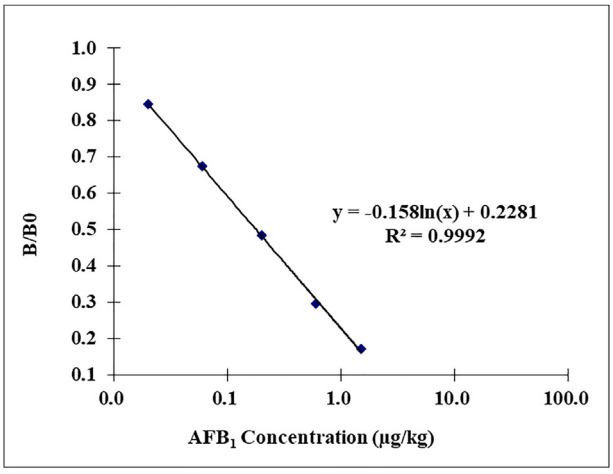
## Funding/support

This study was supported by a research grant from the Center for Research and Development of Medical Diagnostic Laboratory (CMDL), Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, Thailand and the machine and program were supported by the Center for Innovation and Standard for Medical Technology and Physical Therapy (CISMaP), Faculty of Associated Medical Sciences, Khon Kaen University, Thailand.

## Conflicts of interest

The authors declare that there is no conflict of interest related to this study.

## Appendix




*Supplementary Figure S1. Competitive ELISA standard curve for aflatoxin $B_1$ ($AFB_1$). The data represent the mean of duplicates. The graph was generated by plotting the percentage of binding ($B/B_0$) against toxin concentration (µg/kg).*

*Supplementary Table 1. Evaluation results using a logistic regression model for the spectral region 1700–900 $cm^{-1}$.*

| Predictive value | Actual value | | |
|---|---|---|---|
| | | Contaminated $AFB_1$ | Uncontaminated $AFB_1$ |
| | **Contaminated $AFB_1$** | 33 | 11 |
| | **Uncontaminated $AFB_1$** | 9 | 20 |

*Supplementary Table 2. Performance of predictive models in different spectral regions for aflatoxin $B_1$ levels.*

| Preprocessing methods | algorithms | Spectral region ($cm^{-1}$) | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 4000–400 | | | 2950–2800 | | | 1700–900 | | | 1450–900 | | | 2950–2800 + 1700 –900 | | | 2950–2800 + 1450 –900 | | |
| | | %Acc | %Sen | %Spec | %Acc | %Sen | %Spec | %Acc | %Sen | %Spec | %Acc | %Sen | %Spec | %Acc | %Sen | %Spec | %Acc | %Sen | %Spec |
| 17 smoothing points + RBB | SVM | 56 | 56 | 48 | 55 | 55 | 49 | 52 | 52 | 45 | 60 | 60 | 52 | 53 | 53 | 49 | 63 | 63 | 57 |
| | LR | 49 | 49 | 46 | 55 | 55 | 49 | 62 | 62 | 61 | 59 | 59 | 56 | **66** | **66** | **65** | 60 | 60 | 61 |
| | kNN | 53 | 53 | 50 | **67** | **67** | **66** | 62 | 62 | 57 | 56 | 56 | 50 | **67** | **67** | **64** | 55 | 55 | 51 |
| | RF | 52 | 52 | 47 | 59 | 59 | 55 | 56 | 56 | 51 | 55 | 55 | 51 | 60 | 60 | 54 | 52 | 52 | 48 |
| 17 smoothing points + RBB + VN | SVM | 58 | 58 | 48 | 63 | 63 | 58 | 66 | 66 | 60 | 64 | 64 | 55 | 59 | 59 | 53 | 59 | 59 | 53 |
| | LR | 49 | 49 | 46 | 62 | 62 | 57 | **73** | **73** | **71** | 62 | 62 | 57 | **67** | **67** | **65** | 59 | 59 | 58 |
| | kNN | 58 | 58 | 53 | 49 | 49 | 47 | 60 | 60 | 53 | 60 | 60 | 54 | 64 | 64 | 60 | 55 | 55 | 51 |
| | RF | 56 | 56 | 54 | 55 | 55 | 51 | 66 | 66 | 64 | 52 | 52 | 50 | 59 | 59 | 55 | 55 | 55 | 51 |
| SD + 17 smoothing points + RBB | SVM | 59 | 59 | 53 | 62 | 62 | 56 | 60 | 60 | 52 | 66 | 66 | 58 | 59 | 59 | 54 | 62 | 62 | 56 |
| | LR | 58 | 58 | 55 | 53 | 53 | 51 | 55 | 55 | 51 | 64 | 64 | 61 | 60 | 60 | 55 | 63 | 63 | 58 |
| | kNN | 58 | 58 | 56 | 55 | 55 | 53 | 59 | 59 | 54 | 59 | 59 | 54 | 58 | 58 | 53 | 51 | 51 | 47 |
| | RF | 62 | 62 | 58 | 51 | 51 | 46 | 55 | 55 | 50 | 51 | 51 | 47 | 45 | 45 | 42 | 53 | 53 | 49 |
| SD + 17 smoothing points + RBB + VN | SVM | 59 | 59 | 53 | 63 | 63 | 58 | 59 | 59 | 54 | 62 | 62 | 56 | 62 | 62 | 57 | 53 | 53 | 47 |
| | LR | 58 | 58 | 53 | 59 | 59 | 56 | 56 | 56 | 51 | 58 | 58 | 53 | 56 | 56 | 52 | 60 | 60 | 56 |
| | kNN | 55 | 55 | 52 | 63 | 63 | 64 | 59 | 59 | 56 | 55 | 55 | 52 | 58 | 58 | 53 | 56 | 56 | 52 |
| | RF | 48 | 48 | 41 | 66 | 66 | 63 | 64 | 64 | 61 | 56 | 56 | 50 | 49 | 49 | 47 | 48 | 48 | 43 |

Definitions: %Acc, %accuracy; %Sen, %sensitivity; %Spec, %specificity; SVM, support vector machine; LR, logistic regression; kNN, k-nearest neighbors algorithm; RF, random forest; SD, 2nd derivative spectra; RBB; Rubberband baseline correction; VN, vector normalization. The evaluation results of %accuracy, %sensitivity, and %specificity were expressed as the average value over classes (contaminated $AFB_1$ and uncontaminated $AFB_1$ samples).

## References


[1] Saleh BK, Omer A, Teweldemedhin BJMFPT. Medicinal uses and health benefits of chili pepper (*Capsicum* spp.): a review. MOJ Food Process Technol 2018;6:325–8. https://doi.org/10.15406/mojfpt.2018.06.00183.

[2] Ooraikul S, Siriwong W, Siripattanakul S. Dietary intake of chili for local people living in chili farm area, Ubonratchathani Province, Thailand. Bus Off 2011;11:10–1.

[3] Tansakul N, Limsuwan S, Böhm J, Hollmann M, Razzazi-Fazeli E. Aflatoxins in selected Thai commodities. Food Addit Contam Part B 2013;6:254–9. https://doi.org/10.1080/19393210.2013.812148.

[4] Thanaboripat D, Petcharawan O, Sukonthamut S. Aflatoxin contamination of chili powder in canteens of king mongkut's institute of Technology Ladkrabang, Bangkok, Thailand. Curr Appl Sci Technol 2016;16:1–17.

[5] Bhat R, Rai RV, Karim AA. Mycotoxins in food and feed: present status and future concerns. Compr Rev Food Sci Food Saf 2010;9:57–81. https://doi.org/10.1111/j.1541-4337.2009.00094.x.

[6] Chitapanarux T, Phornphutkul K. Risk factors for the development of hepatocellular carcinoma in Thailand. J Clin Transl Hepatol 2015;3:182. https://doi.org/10.14218/JCTH.2015.00025.

[7] Zain ME. Impact of mycotoxins on humans and animals. J Saudi Chem Soc 2011;15:129–44. https://doi.org/10.1016/j.jscs.2010.06.006.

[8] Pei SC, Zhang YY, Eremin SA, Lee WJ. Detection of aflatoxin $M_1$ in milk products from China by ELISA using monoclonal antibodies. Food Control 2009;20:1080–5.

[9] Golge O, Hepsag F, Kabak B. Incidence and level of aflatoxin contamination in chilli commercialised in Turkey. Food Control 2013;33:514–20. https://doi.org/10.1016/j.foodcont.2013.03.048.

[10] Ostry V, Malir F, Toman J, Grosse Y. Mycotoxins as human carcinogens—the IARC Monographs classification. Mycotoxin Res 2017;33:65–73. https://doi.org/10.1007/s12550-016-0265-7.

[11] Ozbey F, Kabak B. Natural co-occurrence of aflatoxins and ochratoxin A in spices. Food Control 2012;28:354–61. https://doi.org/10.1016/j.foodcont.2012.05.039.

[12] Liu Y, Wu F. Global burden of aflatoxin-induced hepatocellular carcinoma: a risk assessment. Environ Health Perspect 2010;118:818–24. https://doi.org/10.1289/ehp.0901388.

[13] Martínez J, Hernández-Rodríguez M, Méndez-Albores A, Téllez-Isaías G, Mera Jiménez E, Nicolás-Vázquez MI, et al. Computational studies of aflatoxin $B_1$ ($AFB_1$): a review. Toxins 2023;15:135. https://doi.org/10.3390/toxins15020135.

[14] Song H, Li F, Guang P, Yang X, Pan H, Huang F. Detection of aflatoxin $B_1$ in peanut oil using attenuated total reflection Fourier transform infrared spectroscopy combined with partial least squares discriminant analysis and support vector machine models. J Food Prot 2021;84:1315–20. https://doi.org/10.4315/JFP-20-447.

[15] Qiang Z, Fuguo J, Chenghai L, Jingkun S, Xianzhe Z. Rapid detection of aflatoxin $B_1$ in paddy rice as analytical quality assessment by near infrared spectroscopy. Int J Agric Biol Eng 2014;7:127–33. https://doi.org/10.3965/j.ijabe.20140704.014.

[16] McMullin D, Mizaikoff B, Krska R. Advancements in IR spectroscopic approaches for the determination of fungal derived contaminations in food crops. Anal Bioanal Chem 2015;407:653–60. https://doi.org/10.1007/s00216-014-8145-5.

[17] Dhall D, Kaur R, Juneja M. Machine Learning: A review of the algorithms and its applications. In: Singh P, Kar A, Singh Y, Kolekar M, Tanwar S. (eds) Proceedings of ICRIC 2019. Lect Notes Electr Eng 2020; 597. Springer, Cham. https://doi.org/10.1007/978-3-030-29407-6_5.

[18] Alzubi J, Nayyar A, Kumar A. Machine learning from theory to algorithms: an overview. In: J phys conf ser. vol. 1142. IOP Publishing; 2018. p. 12012. https://doi.org/10.1088/1742-6596/1142/1/012012.

[19] Ray S. A quick review of machine learning algorithms. In: 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon). IEEE; 2019. p. 35–9. https://doi.org/10.1109/COMITCon.2019.8862451.

[20] Chatchawal P, Wongwattanakul M, Tippayawat P, Kochan K, Jearanaikoon N, Wood BR, et al. Detection of human cholangiocarcinoma markers in serum using infrared spectroscopy. Cancers 2021;13:5109. https://doi.org/10.3390/cancers13205109.

[21] Salari A, Young RE. Application of attenuated total reflectance FTIR spectroscopy to the analysis of mixtures of pharmaceutical polymorphs. Int J Pharm 1998;163:157–66. https://doi.org/10.1016/S0378-5173(97)00378-5.

[22] Giamougiannis P, Morais CL, Rodriguez B, Wood NJ, Martin-Hirsch PL, Martin FL. Detection of ovarian cancer (±neoadjuvant chemotherapy effects) via ATR-FTIR spectroscopy: comparative analysis of blood and urine biofluids in a large patient cohort. Anal Bioanal Chem 2021;413:5095–107. https://doi.org/10.1007/s00216-021-03472-8.

[23] Lin MT, Dianese JC. A coconut-agar medium for rapid detection of aflatoxin production by *Aspergillus* spp. Phytopathology 1976;66:1466–9.

[24] Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans Royal oc 2016;374:20150202.

[25] Phillips B, Gamess E, Krishnaprasad S. An evaluation of machine learning-based anomaly detection in a SCADA system using the modbus protocol. In: Proceedings of the 2020 ACM southeast conference; 2020. p. 188–96. https://doi.org/10.1145/3374135.3385282.

[26] Cotty PJ, Jaime-Garcia R. Influences of climate on aflatoxin producing fungi and aflatoxin contamination. Int J Food Microbiol 2007;119:109–15. https://doi.org/10.1016/j.ijfoodmicro.2007.07.060.

[27] El-Kady IA, El-Maraghy SM, Eman Mostafa M. Contribution of the mesophilic fungi of different spices in Egypt. Mycopathologia 1992;120:93–101. https://doi.org/10.1007/BF00578294.

[28] Chuaysrinule C, Maneeboon T, Roopkham C, Mahakarnchanakul W. Occurrence of aflatoxin-and ochratoxin A-producing Aspergillus species in Thai dried chili. J Agric Food Res 2020;2:100054. https://doi.org/10.1016/j.jafr.2020.100054.

[29] Chen YC, Liao CD, Lin HY, Chiueh LC, Shih DYC. Survey of Aflatoxin contamination in peanut products in Taiwan from 1997 to 2011. J Food Drug Anal 2013;21. https://doi.org/10.1016/j.jfda.2013.07.001.

[30] Salisu B, Anua SM, Rosli WIW, Mazlan N. An improved Fourier-Transform Infrared Spectroscopy combined with partial least squares regression for rapid quantification of total aflatoxins in commercial chicken feeds and food grains. J Adv Vet Anim Res 2022;9:546. https://doi.org/10.5455/javar.2022.i624.

[31] Mirghani ME, Che Man YB, Jinap S, Baharin BS, Bakar J. A new method for determining aflatoxins in groundnut and groundnut cake using Fourier transform infrared spectroscopy with attenuated total reflectance. J Am Oil Chem Soc 2001;78:985–92. https://doi.org/10.1007/s11746-001-0376-y.

[32] Temba BA, Darnell RE, Gichangi A, Lwezaura D, Pardey PG, Harvey JJ, et al. The influence of weather on the occurrence of aflatoxin $B_1$ in harvested maize from Kenya and Tanzania. Foods 2021;10:216. https://doi.org/10.3390/foods10020216.

[33] Lalor JP, Wu H, Yu H. CIFT: crowd-informed fine-tuning to improve machine learning ability vol. 1702; 2017. p. 8563. arXiv preprint arXiv.

[34] Sulistiawan A, Setyaningsih W, Rohman A. A new FTIR method combined with multivariate data analysis for determining aflatoxins in peanuts (*Arachis hypogaea*). J Appl Pharm Sci 2022;12:199–206. https://doi.org/10.7324/JAPS.2022.120720.