TED UNIVERSITY

Faculty of Engineering

Department of Computer Engineering

Project Specifications Report

CMPE 491 – Senior Design Project I

by

Berk Kaya - 56296039538

İlhan Ün - 23825061872

İrem Ayça Uçankale - 12293337022

Alperen Aktaş - 45359027308

Onur Turan - 12172236576

Project Specifications Report

1. Introduction

1.1. Description

The main aim of this project is to detect Aflatoxin contamination, a toxic and carcinogenic compound, in dried figs, one of Türkiye's important export products, using a non-contact and high-speed machine learning–driven system. In current export processes, fig batches are frequently rejected at customs due to the detection of aflatoxin, causing economic and international reputation losses.

In the traditional quality control chain, Aflatoxin detection proceeds in three stages:

1. **Preliminary Control:** Inspection of the outer surface of the figs by workers under UV light.
2. **Advanced Pre-Inspection:** Performing deeper UV detection through needle-based tissue penetration.
3. **Definitive Analysis:** The sample is minced and chemically analyzed in the laboratory.

The system to be developed will analyze the surface properties of each fig using image processing and classification techniques at the preliminary control stage.The data obtained, it aims to detect and separate figs at risk of aflatoxin in real time.

1.2. Constraints

Dataset Preparation Constraints: A key limitation of the project is the lack of a large, commercially available labeled dataset demonstrating the relationship between aflatoxin concentration and surface fluorescence patterns of figs under UV light. This will require time and expense for product sourcing (supply of healthy and aflatoxin containing dried figs) and imaging them to generate data in the initial phase of the project.

Speed and Production Line Constraints: It is assumed that approximately 175 kg of figs are processed per hour on existing manual lines. This speed is determined as the minimum processing capacity that the AI system must meet. The number of figs to be scanned per second will be calculated based on the average fig weight and the latency determined by the AI model for each fig must adapt to this production rate. The ultimate goal of the project is to significantly exceed this speed achieved with current systems.

Performance Metrics Constraints: Because this project directly impacts public health and commercial reliability, the following metrics, not just the overall accuracy of the model, are considered particularly critical constraints:

Recall (Sensitivity): How many figs containing aflatoxin were correctly identified. (A high Recall is vital to ensuring no at-risk figs are missed.)
F1 Score: The balanced average of Recall and Precision.
ROC AUC: The model's overall classification ability.

Budget Constraints:

- Transportation costs are incurred for field visits.
- Logistics and supply costs are incurred for accessing sample figs.
- UV LED creates a budget constraint.
- Curtains and other products to be used to provide a dark environment
- The computational resources required for model training are costly.
- It is difficult to find suitable funding in this field in Turkey and worldwide.

Time Constraints:

- The collection of sufficient numbers of dried figs may make the data set creation time longer than planned.
- The timeline for completing the project is tight.

Technical Constraints: Due to the lack of a standardized environment, there are camera and lighting systems deficiencies. The project team's lack of experience with image processing and hardware may impact the pace of progress in the initial stages of the process. To address these shortcomings, the plan is to utilize existing resources efficiently, utilize open source software, and increase technical proficiency through consultant support.

Professional Support Constraints: There are few experts available for consultation on this subject.

Environmental Constraints: Local interest and ecosystem support for this field in Turkey is limited; collaboration and finding resources may be difficult.

Sustainability Constraints:
Cameras and optics: The repair/replacement time for hardware used in data collection in the event of a malfunction affects the project.

Training machine (GPU): The energy cost resulting from continuous use makes execution difficult; an alternative such as repair or new hardware procurement is required in the event of a malfunction.

Repair and maintenance: Failure to secure backups for essential components such as LED drivers, power adapters, and cables causes the process to halt in the event of a malfunction. It is important that simple maintenance can be performed by the team.

Supply and lifespan: The light intensity of UV LEDs decreases over time; periodic replacement and lack of local supply limit sustainability.

1.3. Professional and Ethical Issues

Public Health Responsibility: The project's fundamental ethical responsibility is to protect the health of consumers. Generating a high rate of false negative results leads to the introduction of toxic products into the market and is an unacceptable ethical risk. Therefore, the model's Recall metric must be carefully considered.

Commercial and Financial Responsibility: The model's high rate of false positive results (labeling a healthy fig as "aflatoxin-containing") leads to unnecessary product loss, destruction, and financial losses for the exporter. Professionally, the commercial cost of the model must be minimized.

Data Transparency: The industry-standard chemical HPLC analysis is inherently destructive; the sample is ground for analysis. This makes it methodologically impossible to simultaneously image a fig that has been fragmented for chemical analysis under UV light.

Due to this methodological challenge and practical access constraints, this project will use a proxy method to create the dataset: visual inspection by experts under standard UV light. Therefore, the ethical validity of the project relies on the transparent documentation and consistent application of these visual criteria.

2. Requirements

Functional Requirements

Scan Start / Stop: The scanning process should be started and stopped with a single button on the system interface. The color or icon of the button should clearly indicate to the user whether the system is active or paused.

Live Counters and Status Indicator:
The following information should be displayed in real time on the application screen:
• Total number of scanned figs
• Number of figs containing aflatoxin
• Number of healthy figs
• Percentage of figs containing aflatoxin (%)

Report Generation and Export: All screening results must be exportable in CSV format based on date and batch. Report columns: Date, Time, Band_ID, Fig_ID, Decision

Image Recording and Archiving: The UV image of each scanned fig should be automatically recorded along with the decision tag.

Session Management and Counter Reset: When a user starts a new session, counters should be reset, and the system should create a new CSV file and image folder.

Session Log: The system should display a list of the last figs scanned in the current session (since the scan was started) and their results ("Aflatoxin" / "Healthy") in a simple table/list format on the interface.

Hardware Status Notification: The system should check the camera and UV light source (if controllable) connections at startup. If the camera connection is missing, the system should not crash; instead, it should clearly warn the user, "Camera not found. Please check the USB connection."

Real-Time Visual Feedback: The system should display the live image (or the last scanned frame) from the camera on the interface. When the model makes a decision ("Healty" or "Aflatoxin"), that decision should be instantly overlaid on the image (with a green/red frame or text). This allows the user to monitor the system in real time.

Non-Functional Requirements

Classification Performance: The system's key success metric is to ensure that figs containing aflatoxin are not missed (Public Health Obligation 1.3.1). Therefore, in the test dataset: Recall (for Aflatoxin Class): Must be at least 95%. (That is, at least 95 out of 100 figs containing aflatoxin must be correctly labeled as "Aflatoxin"). Precision (for Aflatoxin Class): Must be as high as possible to avoid financial loss (False Positive), aiming for a minimum of 85%. (That is, at least 85 out of 100 figs labeled "Aflatoxin" must actually be aflatoxin).

Environmental Robustness: The system should be designed to operate within a closed test box (prototype conveyor belt), isolated from external ambient light and under a constant UV light source.

Usability: The interface should be simple and understandable, and the user should be able to learn the scanning process with a short (under 30 minutes) training.

Maintainability: The software's code structure must be modular; the model, camera driver, and user interface components must be able to be updated independently.

Portability: The developed application must be able to run on a standard computer or a server (with an external USB camera and UV light source).

Latency: The time it takes to capture an image of a fig, make a decision for the model, and display the result in the interface should not exceed 1 second. The system should match the speed of a manual processing pipeline (approximately 175 kg/hour). This performance should be provided on a standard laptop CPU (or integrated GPU), without Google Colab or an external GPU, for inference, not training.

Data Integrity: The system must consistently save an image of a scanned fig (e.g., fig_105.jpg) and its CSV entry (fig_105, Contains Aflatoxin). There must be no duplicates or missing labels (image present, CSV missing).

Reliability: The system must be able to operate without errors during long-term operations (e.g., 2 hours of uninterrupted scanning).

Dataset Requirements

Data Structure:
Each fig record must contain the following information:

- Fig_ID
- Batch_ID
- Image (photograph taken under UV light)
- Aflatoxin label (Aflatoxin/Healthy)
- Scan time (date-time)

Data Quality: The dataset must be free of identical image records (exact duplicates) or visually indistinguishable near-duplicates. All images must meet a minimum resolution, be in sharp focus, and use a consistent, high-quality format (e.g., .PNG or high-quality .JPEG). Images must be checked for uniform exposure, correct color depth (24-bit RGB), and accurate labeling consistency ('Stained' or 'Healthy') according to the established strategy.

The following strategies will be followed for labeling:

• Samples with "ground truth" confirmation of aflatoxin presence through HPLC or similar laboratory analysis are used.

• Labeling is based on visual confirmation by an expert.

• Labeling is divided into two categories based on fluorescence observed under UV light: 'Stained' and 'Healthy.' The 'Stained' group includes any specimen exhibiting even the slightest fluorescent spot or trace, while the 'Healthy' group is reserved exclusively for those specimens showing absolutely zero signs of fluorescence.

Dataset Extensibility: To overcome the limited data situation and increase the small number of examples in the "Aflatoxin" class, data augmentation (translation, rotation, brightness/contrast adjustment) techniques will be applied during the training phase.

Target Data Set Size: For the prototype model to undergo robust and effective training, a total of 10,000 unique fig images should be collected, ideally consisting of at least 5,000 'Stained' (Aflatoxin) and 5,000 'Healthy' images.

3. References

1. Kuru incirlerde AFLATOKSİN ve okratoksin a bulaşisinin önlenmesi. (n.d.). https://www.tarimorman.gov.tr/GKGM/Belgeler/Uretici_Bilgi_Kosesi/Egitim/Hijyen_Kilavuz/kuru_incir_aflatoksin_okratoksin_a_onleme_azaltma.pdf

2. Ömer Barış Özlüoymak. (2014). Development of an UV-Based Imaging System for Real-Time Aflatoxin Contaminated Dried Fig Detection and Separation. Tarım Bilimleri Dergisi/Ankara Üniversitesi Ziraat Fakültesi Tarım Bilimleri Dergisi, 20(3), 302–302. https://doi.org/10.15832/tbd.87873

3. Kılıç, C., Özer, H., & İnner, B. (2024). Real-time detection of aflatoxin-contaminated dried figs using lights of different wavelengths by feature extraction with deep learning. Food Control, 156, 110150. https://doi.org/10.1016/j.foodcont.2023.11015