12-1-2022

# A Deep Learning-Based Framework For Food Safety, Cleanliness, And Data Privacy Assurance

Hamed Taheri Gorji

# A DEEP LEARNING-BASED FRAMEWORK FOR FOOD SAFETY, CLEANLINESS, AND DATA PRIVACY ASSURANCE

by

Hamed Taheri Gorji

Master of Science, Biomedical Engineering, Hakim Sabzevari University, 2015

Doctor of Philosophy, Behavioral Neuroscience, Sapienza University of Rome, 2019

A Dissertation

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota

December

2022

Name: Hamed Taheri Gorji

Degree: Doctor of Philosophy

This document, submitted in partial fulfillment of the requirements for the degree from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

DocuSigned by:
Dr. Kouhyar Tavakolian
A24D898B29D94B8...
Kouhyar Tavakolian

DocuSigned by:
Hassan Reza
7DBBFCC1BD6F4BE...
Hassan Reza

DocuSigned by:
Sattar Dorafshan
60308943EA7B4C7...
Sattar Dorafshan

DocuSigned by:
Bo Liang
32ADBFC8D8014D6...
Bo Liang

DocuSigned by:
Thomas Petros
C4A022288E24476...
Thomas Petros

This document is being submitted by the appointed advisory committee as having met all the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

DocuSigned by:
Chris Nelson
2E0AF088C733403...
Chris Nelson
Dean of the School of Graduate Studies

11/22/2022
Date

PERMISSION

| | |
|---|---|
| Title | A Deep Learning-Based Framework for Food Safety, Cleanliness, and Data Privacy Assurance |
| Department | Biomedical Engineering |
| Degree | Doctor of Philosophy |

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication, or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Hamed Taheri Gorji
December 1th, 2022

## Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Kouhyar Tavakolian, for his continuous guidance, assistance, support, and mentorship. During all these years, he always treated me like a friend and his younger brother. His unwavering encouragement and wise counsel have helped me grow and develop my skills and knowledge. I literally could not have completed this exciting and long journey without him.

I would like to express my appreciation to my committee members, Professor Hassan Reza, Dr. Sattar Dorafshan, Dr. Bo Liang, and Professor Thomas Petros, for accepting the invitation to serve on my committee and for their valuable time, support, and constructive feedback.

I would like to thank Dr. Fartash Vasefi, CTO of SafetySpect Inc., for his unwavering support and for giving me the opportunity to work on these fascinating projects, which were the foundation of this doctoral dissertation. I give my sincere thanks to Nicholas MacKinnon, who had a crucial role in helping me with his insightful feedback on my scholarly manuscripts.

Thanks to biomedical engineering graduate programs for providing continuous financial support during my Ph.D.

I also wish to express my appreciation to my parents and friends, who supported and encouraged me throughout my Ph.D. program.

To my lovely wife, Neda,
the reason why I am standing at this point in life and success.
Thank you for always being with me and supporting me
through this challenging journey with all its ups and downs.

## Abstract

Food safety and foodborne diseases are significant global public health concerns. Precise, reliable, and speedy contamination detection and disinfection technology while preserving the business owners' data privacy is an ongoing challenge for the food-service industry. Contamination in food-related services can cause foodborne illness, endangering customers and jeopardizing provider reputations.

This dissertation performed a cleanliness assessment and disinfection and data privacy assurance in the food services industry using fluorescence imaging, state-of-the-art deep learning algorithms, and a novel paradigm in machine learning named federated learning. In chapter 3, we combined two deep learning algorithms (EfficientNet-B0 and U-Net) and fluorescence imaging for automatic detection and precise segmentation of fecal contamination on meat carcasses to provide higher levels of food safety assurance in meat processing facilities. We achieved a 97.32% accuracy for discriminating between clean and contaminated areas on carcasses and an intersection over union (IoU) score of 89.34% for segmenting areas with fecal residue. In chapter 4, we focused on cleanliness assessment and disinfection of organic residue-based contamination in institutional kitchens and restaurants. We used new fluorescence imaging technology, applying Xception and DeepLabv3+ deep learning algorithms to identify and segment contaminated areas in images of equipment and surfaces. Deep learning models demonstrated a 98.78% accuracy for differentiation between clean and contaminated frames on various surfaces and resulted in an intersection over union (IoU) score of 95.13% for the segmentation of contamination. Further, in chapter 5, the main focus of the study was to address the concerns regarding using new technologies that can increase privacy risks

and leaks of sensitive information. Hence, we used federated learning as a new paradigm in machine learning combined with fluorescence imaging technology and two deep learning models, including MobileNetv3 and DeepLabv3+, to identify and segment the contaminated area on different equipment and surfaces. The model was trained and validated on the data of eight clients and tested on two new clients' data. The model achieved a 95.83% and 94.94% accuracy (F-scores of 96.15% and 95.61%) for classification between clean and contamination frames of the two new clients and resulted in an intersection over union (IoU) score of 91.23% and 89.45% for segmentation of the contaminated areas. Overall, the findings demonstrate that fluorescence imaging combined with state-of-the-art deep learning models not only can improve safety and cleanliness assurance but also ensure client data privacy.

# Table of Contents

## List of Tables

## List of Figures

# 1 CHAPTER 1. INTRODUCTION

## 1.1. Background

Sanitation inspection is an ongoing concern for food distributors, restaurant owners, and others within the food industry. These individuals must prevent potential contamination and infection from spreading among workers and consumers. The failure to meet legal requirements can result in damage to the institutions or restaurants' reputations, the loss of trust between the establishment and its workers and customers, and financial repercussions.

Foodborne illness outbreaks occur when people eat food contaminated with a disease-causing agent [1]. About 48 million Americans become sick each year, 128,000 are hospitalized, and 3000 die from foodborne illnesses [2, 3]. According to the Asia Pacific Society of Infection Control (APSIC) guidelines, "there are several methods for assessing environmental cleanliness: (1) a conventional program of direct and indirect observation (e.g., visual assessment, observation of performance, customer/staff satisfaction surveys); (2) an enhanced program of monitoring residual bioburden (e.g., environmental culture, adenosine triphosphate (ATP) bioluminescence); (3) and environmental marking tools (e.g., fluorescent dust marking of surfaces)" [4-6]. Current methods for identifying contamination on foods, equipment and other wide variety of surfaces in food-servies institutions are limited to simple visual examination or ATP swabbing. Visual examinationIt is difficult for human inspectors to thoroughly inspect different surfaces since some contamination is invisible or barely visible and can be missed. Swabbing methods are also not suitable for testing large surface areas, as swabs can only test a small portion of the surface, even when swiped back and forth. Hence, a more effective solution is needed to examine food processing surfaces to reduce contamination in food-service institutions, improve food safety, and boost public trust in the food processing system.

Optical fluorescence imaging can be an alternative solution as a swift, precise, and

non-destructive method for detecting organic-based residue and biofilms that can host pathogens. The visualization of fluorescence emission has great utility for food safety inspections. Various food-related biological materials have characteristic fluorescence emissions in the visible and near-infrared wavebands. However, materials fluoresce to varying degrees, so just looking for something glowing is insufficient; an observer must consider how the fluorescence pattern differs from the background. Some current research and development efforts, and some associated with the United States Department of Agriculture (USDA), offer algorithms for organic residue detection using fluorescent-based imaging [7-9]. However, there are some limitations and challenges to implementing these algorithms. Some of these suggested algorithms require predefined feature extraction based on shape or color to extract meaningful information for further analysis. Other algorithms detect the contamination regions based on thresholding methods in which determining the optimum value of the threshold level is challenging and can cause both false positive and false negative outcomes due to varying ambient light intensity and the variety of background surfaces. This is why more sophisticated and reliable methods are needed to fill the gap in current analysis methods.

Another concern that needs to be addressed is that conventional data analysis algorithms use data collected from different edge devices and brought together on a centralized server or storage. However, these centralized methods could be troublesome if the collected data contain sensitive information or the centralization is too costly. If data is not properly handled, centralized data collecting may expose people to privacy issues and corporations to legal problems. To this end, an automated, effective and reliable detection method is needed to not only be able detect contaminations precisely but also preserve the clients' data privacy. Such an approach can make the inspection process faster and more effective, drastically improving the food-service industry's safety, cleanliness assurance, and clients' data privacy.

## 1.2. Research Objectives

This research aimed to design a framework by combining state-of-the-art deep learning algorithms, multiwavelength fluorescence imaging, and federated learning to fill the gaps in current analysis methods in terms of reliability, generalization, privacy, and accuracy. By developing state-of-the-art models, not only will workplace safety for food production and food services workers and customers be improved but also food service institutions' and organizations' private data will be safe and secure.

The overall objective of this research was divided into three projects. The first project was designed to combine deep learning and fluorescence imaging to identify fecal contamination on meat carcasses automatically. The goal of the first project was to improve food safety assurance by allowing the industry to use this framework to train employees in trimming carcasses as part of their Hazard Analysis Critical Control Point zero-tolerance plan. The second project focused on using the designed framework to identify and segment the organic-based contamination on different equipment and surfaces in institutional kitchens and restaurants. The goal of the second project was to improve the level of safety and cleanliness, protecting staff and customers of companies and institutions in the food-service industry. The last project concentrated on addressing data privacy and food service providers' sensitive information leakage concerns. The aim was to employ federated learning, a decentralized privacy-preserving technology, combined with state-of-the-art deep learning algorithms and fluorescence imaging to address service providers’ data privacy issues in addition to identification and segmentation of contamination.

## 1.3. Dissertation Outline

This dissertation is organized as follows. Chapter 2 presents background on fluorescence imaging in the food industry and the technology used for data collection for all projects, as well as explanations about the data analysis methodology, including deep

learning, semantic segmentation, and federated learning. Chapter 3 discusses developing a model using deep learning and fluorescence imaging to identify fecal contamination on meat carcasses automatically. Chapter 4 explores the efficiency of deep learning algorithms and fluorescence imaging for detecting and segmenting organic-based residue and biofilms to improve cleanliness assurance in the food services industry, including institutional kitchens and restaurants. Chapter 5 describes using federated learning as a new privacy-preserving approach to training deep learning algorithms for contamination identification and segmentation using data from different clients (institutions or organizations) without direct access to their data. Finally, chapter 6 describes the concluding remarks and the potential future work.

## 1.4. Publications

This section contains the journal and conference papers published while working on the Ph.D. There are also more published journal papers that were not part of the Ph.D. projects and are not included in this dissertation.

### Peer-Reviewed Journal Papers

**J.1.** **Gorji, H. T.,** Shahabi, S. M., Sharma, A., Tande, L. Q., Husarik, K., Qin, J., ... & Tavakolian, K. (2022). Combining deep learning and fluorescence imaging to automatically identify fecal contamination on meat carcasses. Scientific Reports, 12(1), 1-11.

**J.2.** **Gorji, H. T.**, et al. "Deep Learning and Multiwavelength Fluorescence Imaging for Cleanliness Assessment and Disinfection in Food Services." Frontiers in Sensors: 25.

**J.3.** **Gorji, H. T.**, et al. " Federated Learning for Clients' Data Privacy Assurance in Food Service Industry." [*Under Preparation*]

**J.4.** Sueker, M., Stromsodt, K., **Gorji, H. T**., Vasefi, F., Khan, N., Schmit, T., ... &

Tavakolian, K. (2021). Handheld multispectral fluorescence imaging system to detect and disinfect surface contamination. Sensors, 21(21), 7222.

## Peer-Reviewed Conference Papers

**C.1.** **Gorji, H. T**., Shahabi, S. M., Tande, L. Q., Sharma, A., Qin, J., Chan, D. E., ... & Tavakolian, K. (2022, May). Food safety assurance and training of meat inspectors using handheld fluorescence imaging with deep learning detection algorithm. In Sensing for Agriculture and Food Quality and Safety XIV (p. PC121200I). SPIE.

**C.2.** Husarik, K., **Gorji, H. T**., Qin, J., Chan, D. E., Baek, I., Kim, M. S., ... & Tavakolian, K. (2022, May). Handheld dual-wavelength fluorescence imaging system for improving food safety: case study in restaurants and institutional kitchens. In Sensing for Agriculture and Food Quality and Safety XIV (p. PC121200F). SPIE.

**C.3.** **Gorji, H. T**, Mahdi Saeedi, Hossein Kashani Zadeh, Kaylee Husarik, …& Kouhyar Tavakoliana. Federated Learning for Contamination Detection and Data Privacy in Food Service Industry.

**C.4.** Kaylee Husarik, **Gorji, H. T**, Jianwei Qin, …& Kouhyar Tavakolian. Cleanliness Assessment in Long-term Care Facilities Using Deep Learning and Multiwavelength Fluorescence Imaging.

**C.5.** Connor Propp, Mitch Sueker, Kaylee Husarik a,c , **Gorji, H. T**, Luke Woods , Jianwei Qin b ,… & Kouhyar Tavakolian. Dual-Excitation Fluorescence Imaging System for Contamination Detection in Food Facilities.

**C.6.** Luke Woods, Connor Propp, Mitch Sueker, Kaylee Husarik, **Gorji, H. T**, Jianwei Qin,… & Kouhyar Tavakolian. Sanitization Efficacy in Healthcare

Using Multiwavelength Fluorescence Imaging and Deep Learning Fluorescence Imaging.

**C.7.** Tellinghuisen, M., Carriere, C., Husarik, K., Elderini, T., **Gorji, H. T**., Qin, J., ... & Tavakolian, K. (2022, May). Autonomous robot with fluorescence imaging system for invisible contamination detection and pathogen deactivation. In Sensing for Agriculture and Food Quality and Safety XIV (p. PC121200J). SPIE.

## 1.5. Dissertation contribution

This dissertation demonstrated the potential of deep learning-based algorithms for contamination detection, segmentation, and clients' data privacy assurance in the food service industry. The key contributions of the dissertation are summarized below.

- Collected, labeled, and annotated data from three meat processing facilities (*chapter 3*). This contribution refers to J.1. and C.1.
- Developed two deep learning-based algorithms, including EfficientNet-B0 and U-Net, to automatically detect and precisely segment areas of fecal matter contamination on meat carcasses (*chapter 3*). This contribution refers to J.1. and C.1.
- Collected, labeled, and annotated data from two restaurants and six institutional kitchens (*chapter 4*). This contribution refers to J.2., C.2., C.4., and C.6.
- Developed two deep learning-based algorithms (Xception, DeepLabv3+) to identify and segment contaminated areas in images of equipment and a wide variety of surfaces (*chapter 4*). This contribution refers to J.2., C.2., C.4., and C.6.
- Collected, labeled, and annotated data from two more institutional kitchens (*chapter 5*). This contribution refers to J.3., C.3., C.4., and C.6.

- Developed two federated learning-based deep learning algorithms, including MobileNetv3 and DeepLabv3+, to not only identify and segment the contaminated area on different equipment and surfaces but also preserve client data privacy (*chapter 5*). This contribution refers to J.3., C.3., C.4., and C.6.

# 2 CHAPTER 2. METHODOLOGY

## 2.1. Fluorescence imaging in the food industry

Food-related biological residues have been shown to have characteristic fluorescence emission spectra in visible (VIS) and near-infrared (NIR) wavelengths. Dairy cow feces show red fluorescence emissions peaking at 680 nm when excited by ultraviolet (UV) radiation (360 nm) [10, 11]. Chlorophyll (Chl) in green plants has unique fluorescence emissions in the red and far-red regions, peaking at 685 nm and 730 nm [12, 13]. Additionally, a number of plant constituents have been reported to have a UV emission at 340 nm and blue and green emissions peaking near 450 nm and 530 nm [12-14]. Meat products have been shown to have fluorescence emission in UV, blue, and green wavelengths. Proteins are known to emit UV fluorescence, and a variety of aromatic compounds emit fluorescence in blue and green wavelengths [14-18].

Multiple imaging inspection techniques and systems have been used for food safety inspection. The online inspection of poultry carcasses for fecal contamination has been developed using a multispectral imaging system to visualize reflectance spectra features of visible wavelength regions [19, 20]. A hyperspectral imaging system to detect fecal contamination on apples was developed, and this system can measure both reflectance and fluorescence in the visible to near-infrared [21, 22]. A portable hyperspectral imaging system has also been developed to monitor sanitation procedures in food processing facilities. It showed the ability to detect minute quantities of juice from produce on food processing equipment [23, 24]. An imaging device that is portable and capable of fluorescence-based contaminant detection on food products and food processing equipment can easily be integrated with workflows and sanitation audits in food handling facilities.

Some efforts have been made to commercialize line scan spectral imaging systems without disinfection capability and with some documentation capabilities. Headwall

Photonics, Inc. [25] has commercialized a line scan system after licensing a patent. P&P Optica [26], Inc. has commercialized a similar line scan spectral imaging system that they claim is “able to detect, identify, and remove many types of foreign objects on production lines” as well as using “artificial intelligence to provide insights about shelf life, product composition, flavor, fat content, quality variation and much more.” VERITIDE Ltd. [27] is commercializing a fluorescence-based point measurement system to detect fecal contamination on meat carcasses, as well as a fluorescence-based production line imaging system for meat carcasses.

## 2.2. CSI-D Technology

In this dissertation, we used a fast, convenient, and easy-to-use handheld system developed by SafetySpect Inc. for “contamination, sanitization inspection, and disinfection (CSI-D)” that enables the rapid detection of organic residues and biofilms that are present in kitchens, dining areas, and food processing facilities a well as saliva and respiratory droplets. The system provides immediate disinfection and documentation of contaminants on surfaces that may cause disease spread. CSI-D can wirelessly communicate the inspection process, which allows remotely located personnel to immediately provide oversight and respond to inspection issues. The CSI-D system is not intended to be a primary disinfection or cleaning tool; instead, it acts as a post-cleaning audit solution complementary to other post-cleaning auditing tools (ATP, FT-IR, etc.), as well as providing documentation of cleanliness. The system’s disinfection capability is intended to provide spot disinfection only during audits or incident responses and is not employed as a large-area disinfection method (e.g., fogging).

The key innovations of this device encompass the visualization of contamination using fluorescence imaging, the disinfection of the contamination using UVC illumination, and the documentation of cleanliness. The combination of detection, disinfection, documentation, and verification is the core innovation from an operator’s point of view.

Specific technical innovations include the ability to capture fluorescence images under bright ambient light situations in food processing facilities, institutional kitchens, and dining facilities. Previous systems (described above) had difficulties with bright ambient light and, often, could only function in darker rooms or under shrouding. Other innovations related to the UVC germicidal LEDs include the integration of safety systems based on sensors and software (LIDAR, gyroscope, motion detector, etc.) that help protect the operator and other personnel from accidental UVC light exposure. These sensors are also used to monitor motion and distance during the image capture process to ensure images are free from motion artifacts, such as image blur, and provide software-based guidance for operators when they are moving the camera too quickly or too far away or too close. Finally, the image database and records of contamination for each location at each facility, combined with the local hazards and disease prevalence, will enable the future delivery of intelligent dynamic risk assessment associated with each surface to guide cleaning and inspection processes.

### 2.2.1. CSI-D System Description

As shown in Figure 2.1, the CSI-D system consists of a handheld device that incorporates illumination, imaging, battery power, display, and processing units in a single system. The illumination module includes the 405 nm and 275 nm LED arrays, heat management, and driver circuits. The 275 nm LEDs were chosen because they were very close to the 282 nm excitation maximum wavelength of salivary amylase, were commercially available, and had high optical power. This wavelength is also a very effective germicidal wavelength. The 405 nm LEDs were selected because we previously used them for the detection of other organic residues, such as food residues containing fluorophores like collage, flavins, bacterial porphyrins, and chlorophyll. They are not used for the detection of saliva and respiratory droplets.

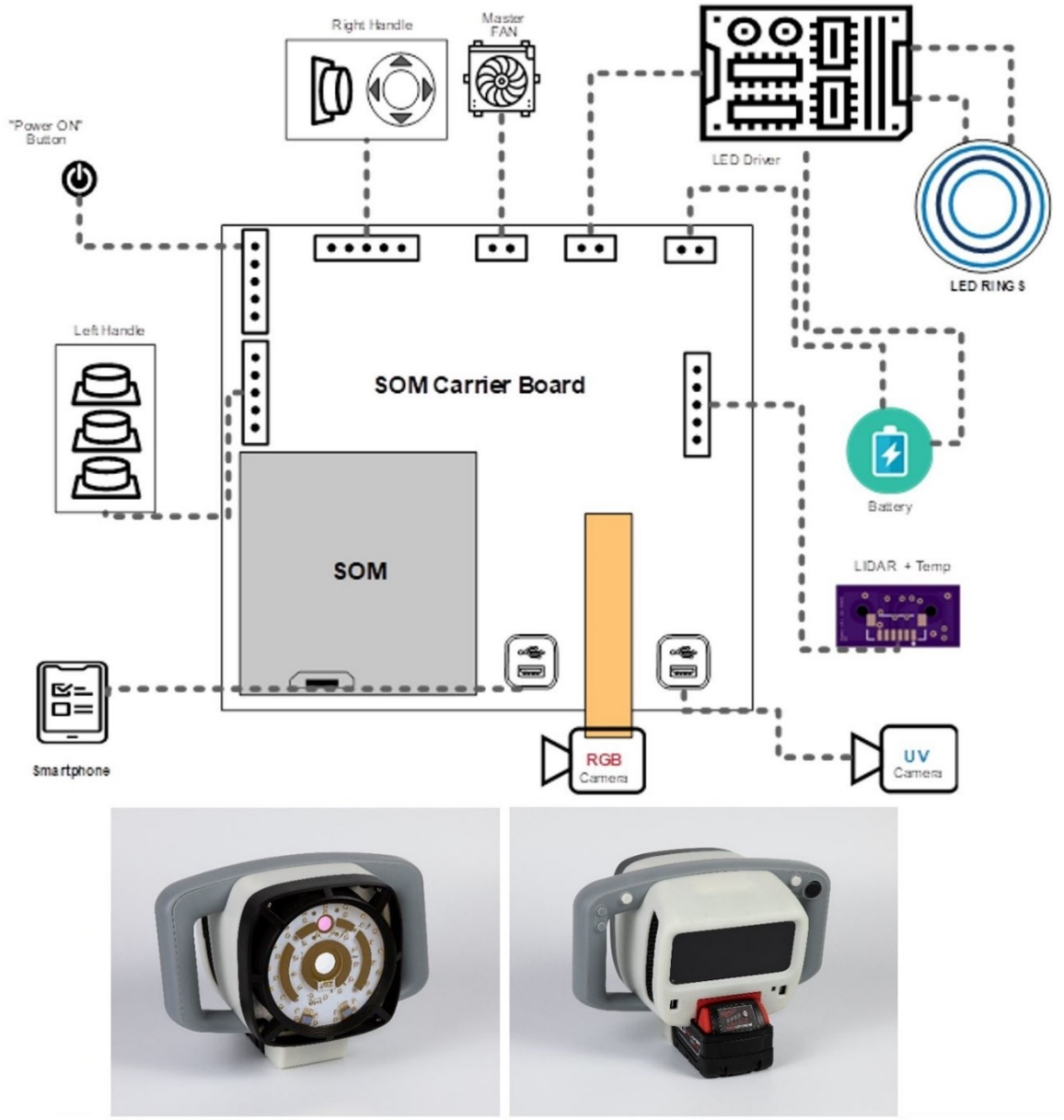


**Figure 2.1.** CSI-D system (top: system block diagram, bottom: CSI-D picture).

During the fluorescence imaging mode, the 405 nm or 275 nm LED arrays are turned on and off sequentially via electronic signals. During the disinfection mode, the 275 nm LEDs are turned on continuously for 2–5 s. The imaging system includes an RGB camera and a UV camera that communicate with the processing unit, which triggers the image acquisition and storage of fluorescence images of organic residues (RGB camera), or saliva and respiratory droplets, and certain organic residues (UV camera) during

fluorescence imaging. The RGB camera is also used in “ViewFinder” mode, whereby an operator can locate the area of interest to be scanned. The camera systems include lenses and spectral bandpass filters that select wavelengths specific to the contamination emission wavelength ranges.

The processing unit includes a system-on-module (SOM) board that controls the illumination and imaging modules to capture the fluorescence and background images under the appropriate illumination. The SOM processes images to provide meaningful information to the operator and for inspection records. The CSI-D system also includes a LIDAR module that communicates with the SOM module, which initializes and controls the LIDAR module and receives distance information from the rangefinder and temperature information from its temperature sensor. CSI-D uses an Android device as a smart display to provide an operator interface. The CSI-D system is designed to communicate with a dedicated cloud server in which all task lists are assembled, and inspection reports and video data are stored and managed.

The operator can select a disinfection mode using the hand controls and user interface. The system calculates how long the UVC illumination should be activated by calculating the surface distance using the LIDAR module.

## 2.3. Deep Learning

Deep learning is a subset of machine learning inspired by the human brain's structure in which neurons in one layer take data as input, analyze it, and transmit the results to the next layer. Deep learning employs multi-layer neural networks, which may include thousands of interconnected nodes (neurons) in different layers so that each node receives input data from several other nodes from the previous layer and delivers processed output data to several other nodes in the next layer [28]. Deep learning models are representation-learning techniques that use multiple levels of representation. They are created by combining straightforward but non-linear modules that each convert the

representation at one level (beginning with the raw input) into a representation at a higher, marginally more abstract level. By composing a large number of such transformations, very complicated functions can be learned. The most important characteristic of deep learning is that these feature layers are learned from data using a general-purpose learning technique rather than being created by human engineers [29].

The field of deep learning has made significant progress in recent years in overcoming complex problems and automating tasks in different fields, including computer vision, natural language processing (NLP), medical diagnosis, biology, the food industry, self-driving vehicles, recommendation systems, playing games, robotics, etc. [30-38]. One of the most important reasons for the tremendous growth in popularity and usefulness of deep learning is its ability to automatic feature extraction [39]. Learning what and how to measure is an essential factor in the success of any data-driven operation. This is why selecting and designing features are crucial steps in the machine learning process. Feature extraction and selection often include in-depth domain knowledge, statistical data analysis, and repeated experimentation when developing models with different feature sets. Deep learning may be far superior in feature creation compared to conventional machine learning methods since traditional machine learning models often require extensive human intervention in feature design. In contrast to traditional approaches for feature extraction, deep learning automatically learns the features from the raw data. Given sufficiently big datasets, deep learning models have shown to be so efficient at learning meaningful features and are currently more accurate for a wide variety of applications than many other machine learning models that employ handcrafted features [39, 40].

Among various types of deep neural networks, convolutional neural networks (CNNs) have received the most attention and are primarily used to solve complex image-driven pattern recognition tasks with their precise yet straightforward architecture [41, 42].

### 2.3.1. Convolutional Neural Network

Convolutional neural network (CNN) is a well-known deep learning architecture that Yann LeCun and colleagues invented in 1998 [43]. CNN architecture was inspired by a work from Nobel prize winners Hubel & Wiesel in 1959 [44], in which they investigated the animal visual cortex and discovered links between certain sections of the visual field and the corresponding brain regions. In some situations, they could even identify the specific neurons responsible for a certain visual field region. As a result of these findings, the idea of the receptive field was developed to characterize the relationship between certain regions of the visual field and the neurons responsible for processing the information. The concept of a receptive field was the main building block of the CNN architecture which refers to the region's size in the input space responsible for generating the feature [45, 46].

The CNNs were developed to handle data sets consisting of many arrays, such as a color picture consisting of three 2D arrays storing the intensities of pixels in each of the three color channels. 1D for signals and sequences, including language; 2D for pictures or audio spectrograms; 3D for video or volumetric images are all common data array formats [29].

In recent years, a variety of CNN architectures have been proposed. However, their fundamental elements are quite similar, including three types of layers: convolutional, pooling, and fully connected (dense) layers. As can be seen in Figure 2.2, a typical CNN architecture consists of a sequence of layers. The convolutional layers are made up of several convolution kernels (filters) that are used to calculate and generate feature maps for recognizing the spatial patterns in input, such as an image. Filters are simply random vectors of weights and biases generated by the network, and it is possible to design a wide variety of filters, each one able to extract a unique feature of the input. In other words, convolutional layers take the input, convolve it, then apply an element-wise nonlinear activation function and pass the result (feature map) to the next layer.

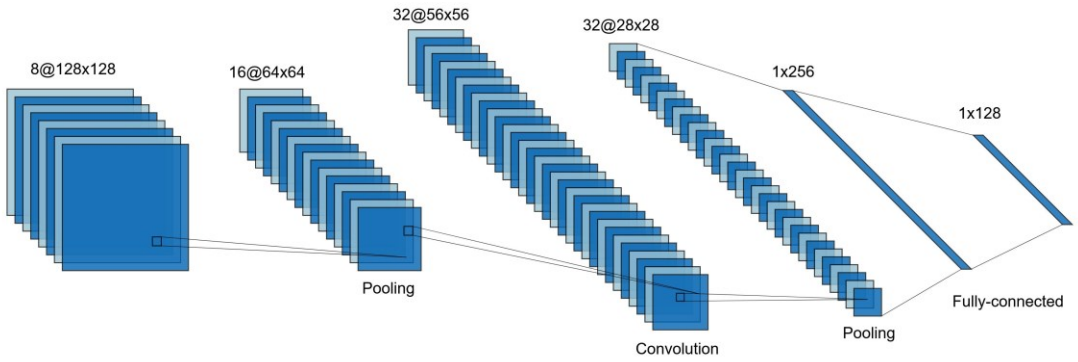


**Figure 2.2.** A typical CNN architecture.

This is similar to how a neuron in the visual cortex reacts to a single stimulus. Each convolutional neuron processes information exclusively for the receptive field to which it is dedicated.

Given that most real-world problems that need to be solved by CNN models are not linear, the primary role of activation functions is to introduce non-linearity into the model [28]. There are different types of activation functions [47], sigmoid ($\sigma(x) = \frac{1}{1+e^{-x}}$), softmax ($\sigma(x_i) = \frac{e^{x_i}}{\sum_{i=0}^{n} e^{x_i}}$), and rectified linear unit ReLU ($\sigma(x) = max\ (0, x)$) are the most popular ones [28].

As explained earlier, a convolutional layer is a stack of feature maps; one feature map corresponds to each filter. Adding more filters increases the dimension of the feature maps and consequently increases the network's number of parameters and computations. The pooling layer in CNN reduces the feature map dimensionality by discarding irrelevant or redundant information without sacrificing any useful information. CNNs use a pooling layer to provide resilience against clutter, compactness of representation, and invariance to changes in location and illumination [48]. In a nutshell, the pooling layer summarizes the results from groups of neurons within the same kernel map [49]. By pooling over a local neighborhood on the feature maps from the preceding layer, the resolution of the feature maps is decreased, which improves the invariance to distortions in the inputs. There are

two well-known pooling methods, average pooling (calculates the average of the values inside a window on feature maps) and max-pooling (takes the maximum value inside a window on feature maps). Max-pooling with a window size of 2×2 and stride of 2 (the amount of steps the window is shifted along each dimension) is the most common method used in CNNs [41]. Stacking multiple convolutional and pooling layers enables the CNN model to progressively extract more complex and higher-level feature representations [42, 48].

Usually, a stack of convolutional and pooling layers is followed by one or more fully-connected layers [50] in order to interpret the features extracted by the previous convolutional and pooling layers. Fully-connected layer connects all the neurons from the previous layer to all neurons in the next layer to generate global semantic information [51].

In CNNs, the last layer is the output layer that depending on the task (binary or multiclass), can consist of one or multiple nodes. Sigmoid and softmax are common activation functions used in the output layer to convert the raw values into probabilities. The former is used for binary tasks, and the latter is used for multi-class problems. Typically, in neural networks, the learning process is done using backpropagation of error to calculate the gradients of the loss function with respect to the model parameters (weights and biases). After that, an optimization algorithm will make use of the gradient in order to update the model weights and biases to minimize the loss function. Currently, a wide variety of loss functions (like Mean Squared Error, Binary Cross Entropy, Categorical Cross Entropy, etc.) and optimization algorithms (such as Stochastic Gradient Decent (SGD), Root Mean Squared Propagation (RMSprop), Adaptive Moment Estimation (Adam), etc.) can be chosen to train a neural network architecture based on the task and problem that need to be solved.

### 2.3.2. Semantic Segmentation

Semantic segmentation is the process of identifying and clustering image pixels that all represent the same object class [52]. Semantic segmentation is also known as pixel-

level classification since it classifies each image pixel into a specific class, and the algorithm should determine which pixels in a new image are semantically related to one another. Clustering is the foundation of most conventional image segmentation techniques, with extra information provided by contours and edges [53, 54]. However, emerging deep learning has shifted the focus to the use of deep neural network architectures, typically Convolutional Neural Networks (CNNs), to solve semantic segmentation problems [55-59]. These models vastly outperform previous methods in terms of accuracy and, in some cases, efficiency [60]. Deep learning techniques for semantic segmentation are advantageous because they can automatically learn appropriate feature representations from labeled pixels compared to using hand-crafted features such as SIFT [61] and HOG [62] that are not specifically developed to perform image segmentation and require domain expertise, effort, and often too much fine-tuning [53, 60]. There are several deep neural networks, such as AlexNet [49], GoogLeNet [50], ResNet [63], VGG-16 [64], etc. [65], which are well-known and now generally recognized as the main building block of the semantic segmentation algorithms.

Currently, fully convolutional network (FCN) [66] is the most common approach for semantic segmentation [60]. FCN strategy's insight was to use pre-existing CNNs as robust visual models with the ability to learn feature hierarchies. They swapped out the fully connected layers with convolutional ones, allowing the original well-known classification models to provide spatial maps for each class label as output rather than classification scores. FCN also uses fractionally strided convolutions (deconvolutions [67]) to upsample the spatial maps to provide dense per-pixel labeled outputs. Figure 2.3 shows one example of the FCN architecture.

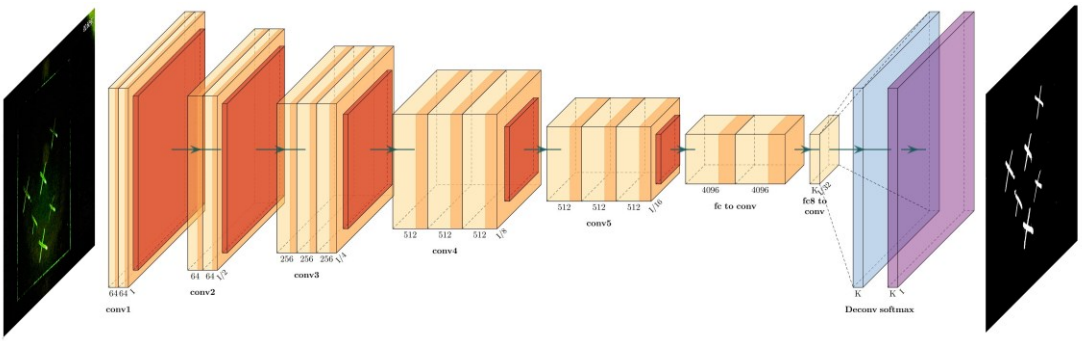


**Figure 2.3.** Example of an FCN architecture.

Another popular approach in semantic segmentation is the encoder-decoder deep neural network architecture [68]. The encoder network, similar to FCN, consists of several convolutions and max-pooling layers to extract the features from the input. In this network, the deeper layers are responsible for extracting information with higher-level semantic meanings. However, the deeper the layers go, the low-resolution image representations are produced [69]. To address this issue, the decoder network with a symmetric structure to the encoder network is used, which is made up of upsampling and convolution layers, and a softmax classifier to predict pixel-wise labels for an output with the same resolution as the input. In the decoder stage, each upsampling layer corresponds to a max-pooling layer in the encoding phase. Then the upsampled feature maps are convolved using a group of trainable filters to generate dense feature maps. Once the feature maps are returned to their original resolution, they are fed to a softmax classifier to generate the final segmentation [70]. An example architecture of an encoder-decoder semantic segmentation model is shown in Figure 2.4.

Although the CNN-based semantic segmentation methods achieve promising results in terms of accuracy and efficiency, one of the main drawbacks of such methods is the need for a massive amount of pixel-level annotated images for the training process [71]. As is known, image annotation is a tedious, time-consuming, and expensive effort. That is why new research fields named weakly-supervised semantic segmentation and semi-supervised semantic segmentation have been introduced [72, 73]. In weakly-supervised

semantic segmentation, data annotation is performed using four primary labeling methods, including image-level, bounding boxes, scribbling, and point annotation, which make the annotation much simpler than pixel-level annotations [74]. Semi-supervised semantic segmentation differs from weakly-supervised in that it presumes a smaller number of fully annotated training data instead of larger data labeled using weakly-supervised approaches [75].

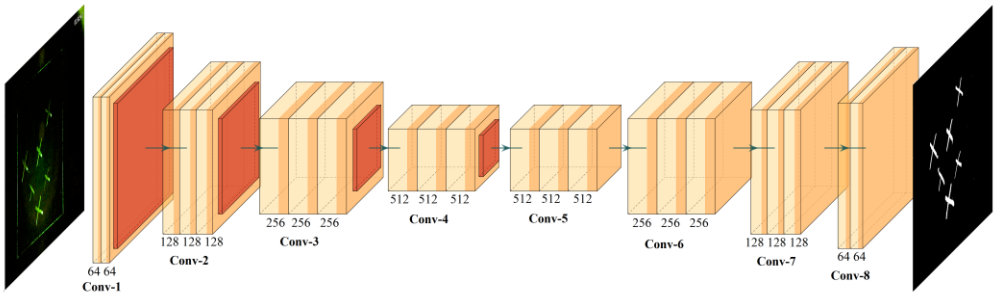


**Figure 2.4.** Example of an encoder-decoder architecture.

## 2.4. Federated Learning

With the advent of smartphones, wearable devices, IoT gadgets, autonomous vehicles, etc., every day, an unprecedented volume of data is generated worldwide. This huge amount of data raises some challenges and concerns. The former is that collecting and storing such data from a wide variety of distributed sources in centralized storage is time-consuming and costly. The second is data privacy and user data confidentiality, which are also major concerns since users' data often include sensitive personal information such as financial data, medical information, location-based services, facial images, etc. [76, 77]. Leaking sensitive information in an untrusted environment can be harmful and damaging. Hence, preserving personal information and privacy is a priority not only for individuals but also for institutions, organizations, and other groups in society [78]. General Data Protection Regulations (GDPR) [79] is an effort by the European Union that went into effect on May 25, 2018, to ensure the privacy and security of users' information. Similarly, the US and china enforce low to protect personal information

privacy and security [80, 81]. The other challenge may appear when centralized data processing is not feasible due to legal or regulatory constraints, high transmission costs, or some other form of technological limits.

For conventional centralized machine learning models to work, data from distributed devices must be transferred to a centralized data repository. Many machine learning (ML) methods, specifically deep learning models, are data-hungry. However, data typically exists in the form of data islands, and because of the competitive nature of the industry, privacy concerns, and complicated administrative procedures even within the same organization, data integration across departments is very challenging [78, 82]. In addition to the challenges and concerns mentioned above, due to infrastructure limitations such as low communication bandwidth, sporadic network access, and strict latency requirements, traditional centralized learning methods cannot handle ML's rapid evolution [83].

Federated learning (FL) is an emerging learning paradigm and a promising solution to tackle the challenges and concerns mentioned earlier. For the first time, the concept of FL as a distributed machine learning framework was proposed by Google researchers McMahan et al. in 2016 [84]. FL is a machine learning paradigm in which several entities (clients) collaboratively train an ML model under the orchestration of a central server or service provider while clients' data is kept decentralized and never shared with either other clients or the central server [85]. More specifically, a central server shares a global model with clients. Then each client trains the global model using local data, and the trained parameters (e.g., weights in neural networks) of each client's model are sent back to the server for aggregation. Finally, the server updates the global model and distributes it between the clients to continue the training process. This process will be repeated until the model achieves the learning goal and a satisfactory result.

### 2.4.1. Categorizations of federated learning

Generally speaking, federated learning based on the distribution of sample space and

feature space can be categorized into three types: horizontal federated learning, vertical federated learning, and federated transfer learning [82].

Horizontal FL can be used when clients' datasets have the same feature space but a different sample space. The horizontal FL framework can be beneficial when several clients try to improve their model performance on the same task. Since the feature space is the same, the horizontal FL can expand the sample size, and consequently, the model will be trained on more data. For instance, four institutional kitchens, each in a different country, collect fluorescence imaging data in the same feature space for contamination detection. Given that each institutional kitchen trains the model based on its own data, using the horizontal FL, these four institutional kitchens can develop the training model collaboratively, increasing sample size and, as a result, the model performance and reliability.

Vertical FL, also named feature-based FL, is applicable when clients' datasets have the same sample space but are different in feature space. Vertical FL often uses entity alignment methods to acquire overlapping samples [86, 87], and later these samples and their features can be used for training the machine learning model. Vertical FL can increase the feature space dimension by combining different features. For example, in a city, there are three different hospitals or institutions that, each of them perform one of the cognitive tests, MRI, or blood gene expression on AD patients. Since these three tests are the most common AD testing, many patients will probably visit these institutions. So there will be a significant intersection of patients, and vertical FL can aggregate these different tests as features for the same patients to increase the feature space for more robust and reliable training of ML models.

Federated transfer learning can be used in a case where both feature space and sample space of the clients' datasets have a relatively very small overlap. Following the above example for diagnosis of Alzheimer’s Disease, consider three institutions or hospitals in three different countries, each of which does one of the cognitive tests, MRI,

or blood gene expression on the AD patients. Because of the geographical distance, the sample space might have no overlap or barely overlap of information. Also, each institution or hospital does a different test on AD patients, resulting in no overlap in feature space. In such scenarios, transfer learning can be used to create an efficient and reliable model while transferring and adapting knowledge from different but related tasks.

### 2.4.2. Federated learning algorithm

Conventional optimization approaches, such as distributed SGD, are often inappropriate in FL and may result in high communication costs. Many federated optimization approaches address this issue by using local client updates, in which clients update their models locally many times before communicating with the server. This can significantly reduce communication costs when training a model [88].

Federated averaging (FedAvg) is the first and the most commonly used algorithm for federated learning [84]. In order to update the model, it performs a predefined number of steps of stochastic gradient descent (SGD) in parallel on a small selection of devices, and then model updates are averaged on a central server. FedAvg does more computations locally and less communication than SGD and its derivatives [89]. In recent years, many variants of FedAvg (FedProx [90], FedOpt [88], FedPAQ [91], FedBN [92], FedNova [93], etc. ) have been proposed to tackle possible issues such as convergence and heterogeneity in federated networks.

### 2.4.3. Federated learning frameworks

In recent years, several FL open-source frameworks and software have been developed, and five popular ones are discussed in the following.

FedML [94] is an open-source research library and benchmark designed to make developing and fairly performance comparison of FL algorithms easier with respect to the wide range of computing paradigms and configurations. FedML is compatible with several computing paradigms, including on-device training for edge devices, distributed computing, and single-machine simulation. FedML also supports a variety of algorithmic

studies by providing a generic API architecture and extensive reference baseline implementations, including optimizers, models, and datasets. FedML library is comprised of two main components: FedML-API, which represents high-level API, and FedML-core, which stands for low-level API. FedML's training engine and distributed communication infrastructures are built into the FedML-core. On top of this, the FedML-API includes training models, datasets, and FL algorithms.

FATE (Federated AI Technology Enabler) [95] is an industrial-level FL framework that allows businesses and academic institutions to work together on data while maintaining privacy and security. Currently, FATE can be used to train a wide variety of machine learning models in a horizontal or vertical federated setup. It protects communications through homomorphic encryption and secure multi-party computation. Given that FATE provides algorithm-level interfaces, developers need to modify the source code to implement their FL system.

TFF (TensorFlow Federated) [96] is a lightweight framework developed by Google that allows researchers to create and implement novel TensorFlow-based FL algorithms. The TFF interface is divided into two main layers: Federated Learning (FL) API and Federated Core (FC) API. Using the FL API, developers can easily integrate the implemented and provided federated training and evaluation into their existing TensorFlow models. FC API is the foundation of federated learning in TFF and combines TensorFlow with distributed communication operators, allowing for the succinct expression of unique federated algorithms.

PySyft [97, 98] is an open-source library that enables private and secure machine learning by transparently encapsulating and extending popular deep learning frameworks like PyTorch [99]. Several privacy methods are available as add-ons in PySyft, such as secure multi-party computation and differential privacy. PySyft also uses the WebSocket API, allowing client-to-client communication regardless of whether the service is hosted on a single or distributed set of computers [100].

FLOWER [101] is an FL framework that facilitates experimenting with both algorithmic and systemic challenges in FL. Compared to other platforms, FLOWER stands out because of its unique capabilities for conducting large-scale FL experiments and considering diverse situations involving heterogeneous FL device configurations. Flower offers higher-level abstractions and makes it easier for researchers to try out novel approaches while still operating on a solid foundation. Flower facilitates using the existing ML algorithms and pipelines in an FL configuration to assess the convergence and training times of such pipelines in a distributed environment. It is worth noting that FLOWER enables the use of FL implementations on mobile and wireless clients with heterogeneous computing power, memory, and network resources.

# 3 CHAPTER 3.

## COMBINING DEEP LEARNING AND FLUORESCENCE IMAGING TO AUTOMATICALLY IDENTIFY FECAL CONTAMINATION ON MEAT CARCASSES

## 3.1. Summary

**Background:** Food safety and foodborne diseases are significant global public health concerns. Meat and poultry carcasses can be contaminated by pathogens like E. coli and salmonella by contact with animal fecal matter and ingesta during slaughter and processing. Since fecal matter and ingesta can host these pathogens, detection and excision of contaminated regions on meat surfaces are crucial. Fluorescence imaging has proven its potential for the detection of fecal residue but requires expertise to interpret. In order to be used by meat cutters without special training, automated detection is needed. **Methods:** This study used fluorescence imaging and deep learning algorithms to automatically detect and segment areas of fecal matter in carcass images using EfficientNet-B0 to determine which meat surface images showed fecal contamination and then U-Net to precisely segment the areas of contamination. **Results:** The EfficientNet-B0 model achieved a 97.32% accuracy (precision 97.66%, recall 97.06%, specificity 97.59%, F-score 97.35%) for discriminating clean and contaminated areas on carcasses. U-Net segmented areas with fecal residue with an intersection over union (IoU) score of 89.34% (precision 92.95%, recall 95.84%, specificity 99.79%, F-score 94.37%, and AUC 99.54%). **Conclusion:** These results demonstrate that the combination of deep learning and fluorescence imaging techniques can improve food safety assurance by allowing the industry to use CSI-D fluorescence imaging to train employees in trimming carcasses as part of their Hazard Analysis Critical Control Point zero-tolerance plan.

## 3.2. Background

Unsafe food poses a threat to people worldwide, causing many illnesses and deaths every year [102]. Foodborne diseases are a public health challenge and major contributors to morbidity and mortality. About 600 million cases of foodborne illness and 420 thousand of deaths globally are caused by unsafe food each year [103].

According to the US Centers for Disease Control and Prevention (CDC), foodborne illnesses affect millions in the US, causing thousands of deaths and billions of dollars in economic losses annually [104]. Each year about 48 million Americans become sick (1 in 6), 128,000 are hospitalized, and 3000 die from foodborne illnesses [2, 3, 105]. The economic burden of foodborne illness was over $15.5 billion in 2013 [106], and the U.S. Department of Agriculture (USDA) stated that this amount was increased by about 13% to $17.6 billion in 2018 [107].

In the United States, the foods that are most likely to become contaminated and, therefore unsafe, are raw foods of animal origin [108]. Currently, the meat and poultry industry is the largest segment of agriculture. For instance, in 2017, 52 billion pounds of meat and 48 billion pounds of poultry were produced in the US [109]. With the high consumption of meat and poultry comes a high number of human foodborne illnesses.

There are numerous foodborne pathogens associated with meat, including Salmonella spp., E. coli, Campylobacter jejuni, Yersinia, and Listeria monocytogenes. These are the most often detected pathogens that can cause significant public health problems [3, 110]. Becoming infected with any of these pathogens can lead to severe diarrhea, vomiting, abdominal cramps, and even death in some cases [3]. Outbreaks of foodborne illness resulting from contamination by these pathogens tend to occur by exposure of the animal carcass to feces, ingesta, and soiled hides during meat processing. According to USDA policy, the regulatory standard for US Food Safety and Inspection Service (FSIS) is “zero tolerance” [111]. “Zero tolerance” is a visual standard by which all surfaces of processed

meat and poultry carcasses are required to be free of visible fecal contamination before they can enter the carcass chiller.

Current methods for identifying fecal contamination on carcasses are limited to simple visual examination during meat processing. It is difficult for human inspectors to thoroughly inspect carcasses since some contamination is invisible or barely visible and can be missed. Having a “zero tolerance” requirement for visible fecal contamination on the carcasses creates a problem. The visual inspection method becomes insufficient because of human errors that happen during processing, especially “painting.” In painting, the slaughter knife is not disinfected and cleaned properly of feces or ingesta and is used later to skin the carcass. This error dilutes the fecal contamination, making it difficult to visually detect on the carcass with the naked eye. Another source of error is when fecal contamination is not detected on the lower or upper parts of the carcass, which are more difficult for the inspector to reach and see.

To combat meat and poultry contamination, a solution is needed to more effectively inspect meat surfaces during meat processing to overcome potential human error, increase food safety, and increase public confidence in the meat processing system while allowing a fast enough production speed. Although there is no way to ensure that our food is completely secure, new developments in optical fluorescence imaging can provide improved confidence in food processing operations. Fluorescence imaging can play a crucial role in food safety as a swift, precise, and non-destructive technique that is able to detect chlorophyll and its metabolites as well as other fluorophores within fecal matter and ingesta.

There are two problems to overcome when deploying fluorescence imaging technology in meat processing facilities. One of these is the problem of doing fluorescence imaging in a bright ambient light environment. Another is the need for the interpretation and analysis of fluorescence images by untrained operators. New technologies in Light-emitting diode (LED) illumination and image sensors have been developed to overcome

the bright ambient light problem, and modern machine learning algorithms can be applied to automate image analysis and provide immediate feedback to inspectors and meat cutters.

In this study, we employ a fluorescence imaging device developed by SafetySpect and two state-of-the-art deep learning algorithms, including EfficientNet and U-Net, to first discriminate the video frames with fecal residue on meat surfaces and then accurately segment and identify the corresponding areas of fecal contamination in the image. The handheld automated imaging inspection device (CSI-D), created by SafetySpect [112], offers mobility and flexibility for fluorescence-based contaminant detection on both food products (e.g., carcasses) and equipment surfaces in meat processing facilities under bright ambient light. Visualization by fluorescence emission has excellent potential for food safety procedures. However, fluorescence imaging, like many other imaging techniques, delivers video or images that need to be interpreted to be used for live visual inspection. These images need automated methods to identify contamination and make decision-making easier for human inspectors and meat cutters excising the contaminated areas from meat. This is the next level of safety that needs to be developed and which we address in this research.

Research in applications of image processing and machine learning algorithms over the last few years has rapidly increased in the area of food safety and food security, especially in the meat processing industry [113, 114]. Researchers evaluated the quality of 16 types of pork and poultry using image processing techniques, including segmentation and histogram-based analysis [115]. Another study used 4-bit monochromatic images of meat cuts to extract features using image processing techniques from fatty regions of meat images and then used neural network and multiple regression analysis to grade the meat quality [116]. To assure meat safety, several studies used threshold-based algorithms on fluorescence images to detect fecal contamination on beef and poultry meat surfaces [7, 117, 118]. However, most of these techniques are

prone to error and are not reliable enough to be applicable in the real world for the following reasons.

Many of these experiments were conducted in a laboratory environment with low ambient light intensity and on surfaces that were dark as opposed to the stainless steel and bright plastic surfaces typically found in meat processing plants. The meats being measured were mostly fixed in position as opposed to the constant motion of carcasses in meat processing plants. Additionally, many of the techniques discussed required predefined feature segmentation based on feature shape or color. These identified features are then used to extract meaningful information for further analysis for classification. Several studies used thresholding-based algorithms to detect the contamination regions in which determining the optimum value of the threshold level is very challenging and can cause both false positive and false negative outcomes due to varying ambient light intensity and the variety of background surfaces found in meat processing plants. This is why more sophisticated and reliable algorithms are needed to fill the gap in current analysis methods.

During the past few years, deep neural network (DNN) algorithms, and more specifically, convolutional neural network (CNN) algorithms, have grabbed much attention in many fields as state-of-the-art machine learning methods [119, 120], and certainly in food safety-related technologies [38, 121]. One of the main advantages of CNN models is that they can automatically extract important and meaningful information from images and videos and learn from them without image feature extraction by human-derived techniques. The learned weights of a trained CNN model can also be saved and reused with a new dataset without re-training the model [122]. DNN algorithms have a wide variety of applications in the food domain, including food recognition and classification, food calorie estimation, food supply chain monitoring, food quality detection, food contamination detection, etc. [38, 123]. Although applications of DNN algorithms have been studied in several food-related domains, to our knowledge, there is no study where

the main focus is on the detection and segmentation of fecal contamination on meat surfaces.

## 3.3. Methods

### 3.3.1. Fecal contamination measurement technology

In this study, to collect fecal contamination information, we used a handheld contamination, sanitization inspection, and disinfection device (CSI-D). One of its capabilities is the detection of organic residues in meat processing facilities. The CSI-D device integrates illumination, imaging, data processing, and display into a single portable device. Its illumination module is composed of 270 nm and 405 nm light-emitting diode (LED) arrays, a suitable heatsink for the dissipation of heat from the LEDs and driver circuits. In addition, the device incorporates a Wi-Fi transmitter for streaming captured video to supplementary monitoring devices like tablets and cell phones.

During imaging mode, the LED arrays are automatically pulsed—switched off and on in rapid sequence through electrical signals to enable background image subtraction. For fluorescence imaging, there are two cameras: an RGB camera and a UV camera, which are controlled by a processing unit that initiates the capture of fluorescence images. The RGB camera captures images of organic residues, and the UV camera captures images of saliva, respiratory droplets, and other organic residues. CSI-D is also equipped with lenses and spectral bandpass filters that pass specific wavelengths emitted by contamination fluorescence. Figure 3.1 shows the CSI-D device.

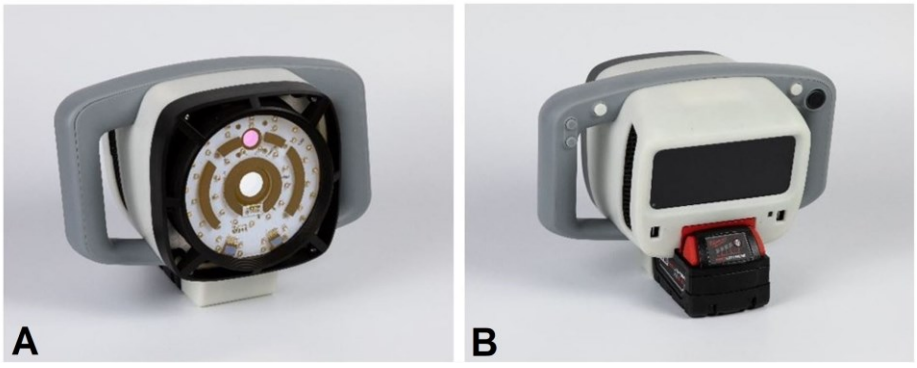


**Figure 3.1.** CSI-D device. (A) Front view. (B) Rear view.

### 3.3.2. Data collection

We collected data at three meat processing facilities (two cattle and one sheep) in North Dakota, USA. Before starting data acquisition, there was a detailed discussion with the owners and local meat inspectors to better understand the current methods used in their facilities, including any perceived needs or shortcomings. SafetySpect's CSI-D device was used to record videos from processed raw meat carcasses, initially under the direction of a retired federal meat inspector and later without direction. If we found anything suspicious during the scan, we asked the inspector to review it. We scanned fourteen beef and six sheep carcasses on the left and right sides of the carcasses after skinning and before they were placed in the chiller.

In addition to the video recording of the scans, we also created a video record of the procedure with the meat inspector using a GoPro camera. We also secured samples of meat, fat, and sheep feces for further measurements in the lab at the University of North Dakota. All measurement scans were captured with a resolution of 1024 × 768 and at 24 frames per second (FPS). We also measured the meat processing environment light intensity which varied between 50 to 90 Foot-candle (FC). In total, 1 h and 15.2 min of video scans were used for further analysis. Figure 3.2 shows examples of image data collected from clean meat surfaces, and Figure. 3.3 shows examples of image data from meat surfaces contaminated by fecal matter.

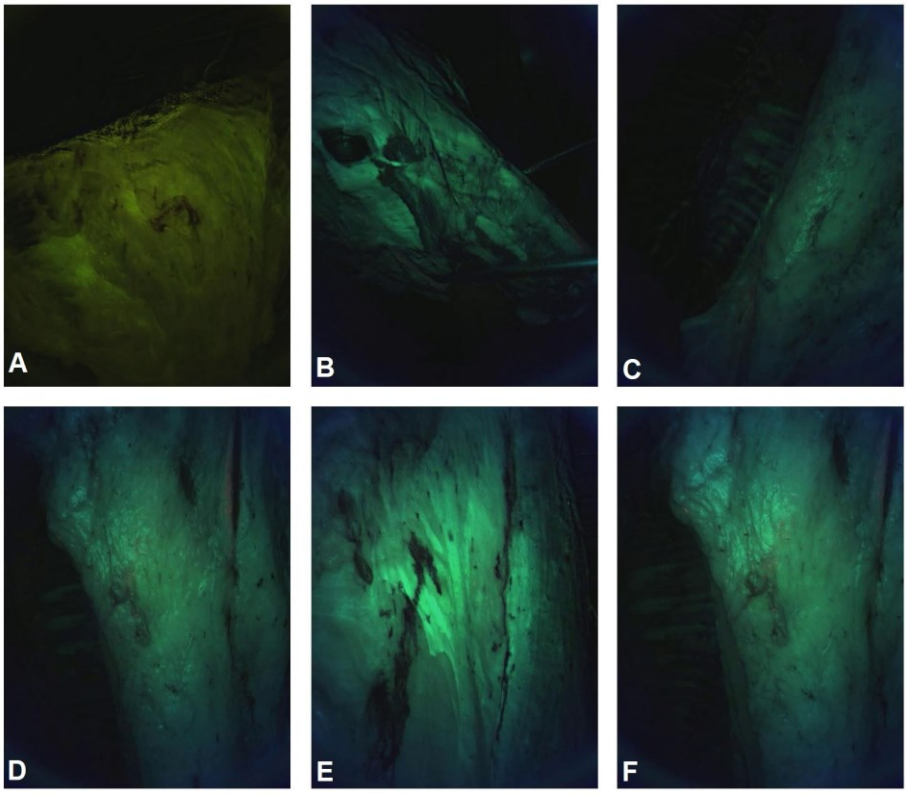


**Figure 3.2.** Six CSI-D fluorescence images of clean meat surfaces (A–F).

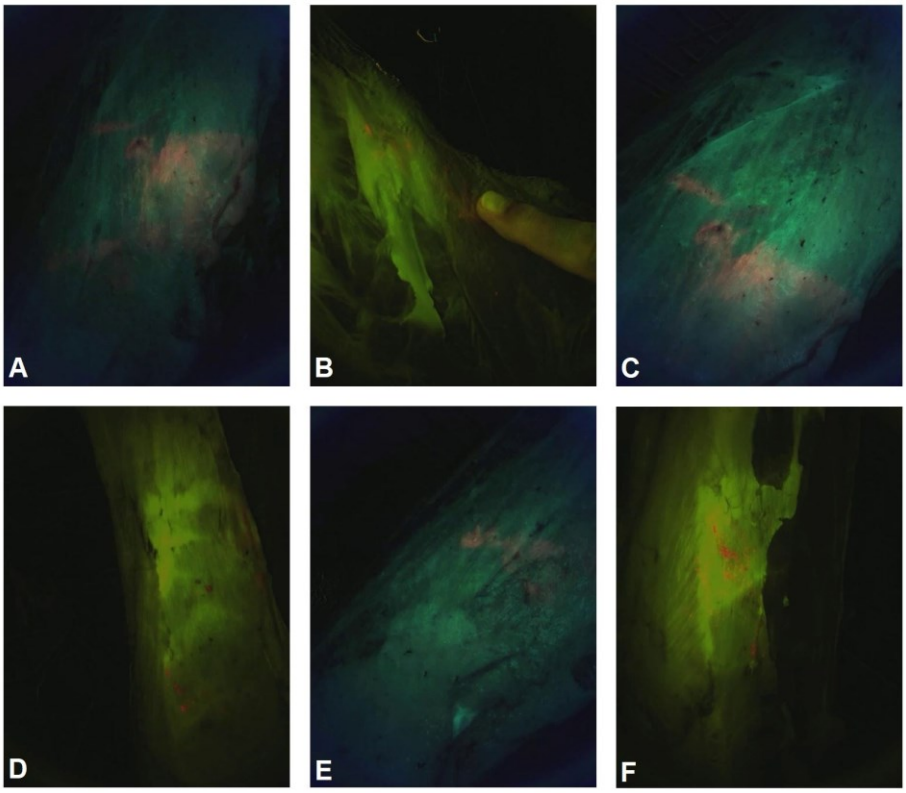


**Figure 3.3.** Six CSI-D fluorescence images of meat surfaces with fecal contamination (A–F).

### 3.3.3. Methodology

For this study, we had two aims. The first aim was to differentiate the frames showing fecal contamination on meat surfaces (contamination) versus the frames with no contamination on meat surfaces (clean) using a DNN model. The second aim was to identify the specific areas of fecal contamination using a semantic segmentation algorithm on the frames already classified as contamination.

### 3.3.4. Deep learning model architecture

In this study, a CNN model is used to classify the captured video frames from the CSI-D device into two classes of *clean* and *contamination*. To do so, first, all the captured videos were converted to frames, and then each frame that showed contamination was labeled as contamination, and each frame without any contamination was labeled as clean. We then employed a state-of-the-art CNN model (EfficientNet) for the classification between *clean* and *contamination* frames.

The EfficientNet model was proposed by Google researchers [124], and their main idea was that balanced scaling up of a CNN in terms of depth, width, and input image resolution could lead to better accuracy and efficiency. Their empirical investigation revealed that balancing such dimensions resulted in a more accurate outcome. They used a neural architecture search [125] and designed a baseline network (EfficientNet-B0), and proposed a compound scaling method in which all network dimensions (width, depth, and image resolution) can be uniformly scaled using a set of predefined scaling factors to generate a family of EfficientNets (B1–B7). We used EfficientNet-B0 and Figure. 3.4 depicts a condensed schematic representation of this model.

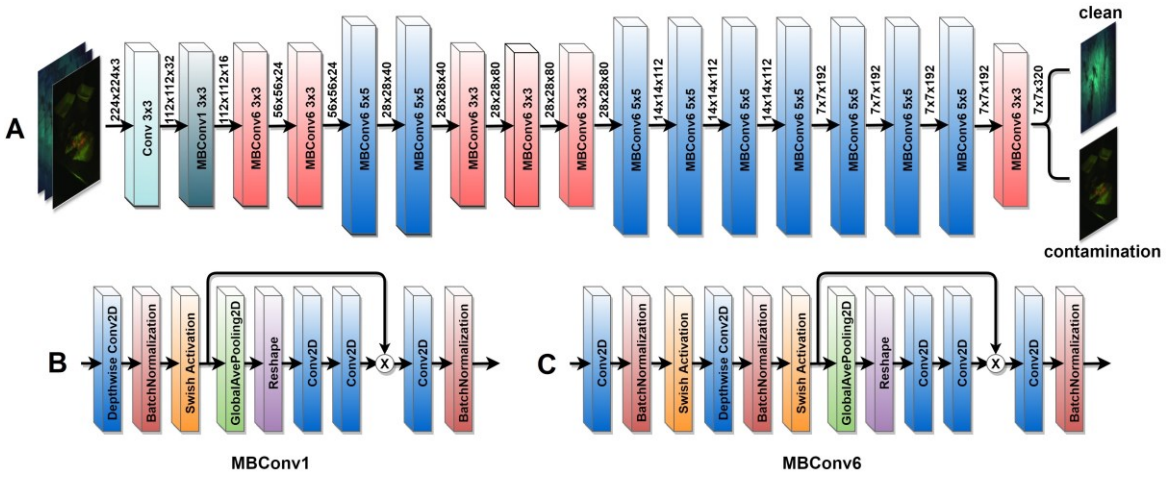


**Figure 3.4.** (A) A concise representation of the EfficientNet-B0 model. (B) The building blocks of MBConv1. (C) The building blocks of MBConv6.

As can be seen, the main building block of EfficientNet-B0 is a mobile inverted bottleneck convolution (MBConv) [125, 126], which is slightly modified by adding squeeze-and-excitation optimization blocks [127]. Each MBConv block relies on depthwise convolutions [128] and shortcut connections between the blocks. The squeeze-and-excitation enhances the network's representative power by explicitly modeling interdependencies among channels, resulting in dynamic channel-wise feature recalibration. Unlike typical two-dimensional (2D) convolutions, depthwise convolutions can apply a single convolutional filter to each input channel and not only decrease the number of parameters and computational costs but improve the network's representational efficiency.

The input image size for EfficientNet-B0 should be (224, 224, 3), and the model expects the inputs to be floating pixel tensors with values in the range of 0–255. The input images pass through a 2D convolutional layer with 32 filters and a kernel size of $3 \times 3$. Then the images flow through sixteen other MBConvs, including one MBConv1 and fifteen MBConv6 with different filter and kernel sizes. It is worth mentioning that each convolutional layer is followed by a batch normalization [129] along with a swish activation function [130]. The batch normalization by smoothing the optimization landscape can

make the gradients more stable and predictable, making optimization faster and more effective, and can speed up the training process [131].

The last block of the EfficientNet-B0 is composed of another convolution layer followed by another batch normalization and the swish activation function. More specifically, a 1 × 1 convolution layer [132] is used in the final block, which enables the network for channel-wise pooling or cross-channel downsampling. By doing so, the 1 × 1 convolution layer functions as a projection layer that aggregates information across channels and reduces the dimensionality by decreasing the number of filters while introducing non-linearity and keeping essential, feature-related information.

Given that the first aim of our experiment was binary classification between clean and contamination frames, we employed a sigmoid activation function as the output activation layer. The sigmoid function is a nonlinear activation function that transforms the values between 0 and 1, and in our case, its output represents the probabilities that an input frame belongs to the *clean* and *contamination* classes.

Next, the EfficientNet-B0 model needs to be trained to recognize and classify the input frames correctly, and the model error for the prediction of the correct classes needs to be minimized. A loss function is needed to compute the error and quantify the performance of the model. Since our task is a binary classification, we used Binary Cross-entropy as our loss function. Binary Cross-entropy calculates the cross-entropy loss between the predicted labels and true labels, which can be 0 or 1. By diverging the predicted labels from true labels, the cross-entropy loss will be increased. The Binary Cross-entropy Eq. (3.1) is as follows:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i \times \log\big(p(y_i) + (1 - y_i) \times \log\big(1 - p(y_i)\big)\big) \tag{3.1}$$

where N is the number of training samples, $y_i$ denotes the true label, $p(y_i)$ represents the predicted probability of class 1 (in our case, contamination) and $1 - p(y_i)$ is the predicted probability of class 0 (*clean*).

To minimize the loss function and improve the model performance, an optimization algorithm is needed. In this study, we used the Adam optimizer, a gradient-based optimization of stochastic objective functions, which adaptively estimates the first and second-order moments [133]. In other words, the Adam optimizer individually adjusts the learning rate of the parameters using the estimation of the first and second moments of the gradients.

### 3.3.5. Contamination segmentation

The second aim of this research was to segment the regions with fecal contamination after finding and classifying the video frames as *contamination* using the model described above. The accurate segmentation of areas contaminated by fecal matter can be very beneficial for the following reasons. The first and foremost reason is that finding *contamination* frames does not always lead to finding a contaminated area on the meat surface. In many cases, the fecal matter on the meat surface is very small, and finding areas of contamination even after detecting the corresponding frame could be challenging during a live inspection process.

Since carcass inspection must comply with the "zero tolerance" standard, segmenting and pseudo-coloring these areas can make it more straightforward for inspectors to recognize them and not miss any contamination. Another reason for pseudo-coloring is that the segmentation can improve inspector discrimination during the monitoring process. During the inspection, there might be several contaminated regions of different sizes or colors at the same time on meat surfaces, and without accurate segmentation, the inspector might miss some contaminated areas. Developing a model for accurate segmentation of the fecal contaminated areas on meat surfaces may be crucial. Further, the segmentation of the fecal contaminated area can help the training process of new FSIS inspectors in recognition of the possible fecal contamination areas and in trimming carcasses as part of their Hazard Analysis Critical Control Point (HACCP) zero-tolerance plan.

In the past, fecal contamination segmentation on the surface of different types of meats and vegetables has been tackled mainly using threshold-based algorithms. However, such algorithms are prone to error because they are sensitive to ambient light intensity, and determining an appropriate thresholding level can be challenging. In real-world situations, the presence of other objects with similar pixel intensity in the background can also lead to false positives [7-9, 118].

In this study, we used the semantic segmentation method to precisely segment fecal contaminated regions on meat surfaces. In semantic segmentation, a pixel-level annotation assigns a class to each pixel of an image. Then a deep CNN is first trained to learn to classify pixels according to their corresponding classes (in our case, *clean* and *contamination*), and then the trained model is used to predict the class of each pixel for the unseen data (test set). To build the semantic segmentation training and testing datasets, we employed MATLAB Video Labeler for pixel-wise labeling of each video frame. The Video Labeler provides a straightforward way to label ground truth data in a video or image sequence by marking rectangular or polyline region of interest (ROI) labels, pixel ROI labels, and scene labels.

Since the aim was to segment the fecal contamination regions on meat surfaces, the ground truth labels consisted of two classes of *contamination* and *background*, and all pixels were labeled accordingly. We labeled every pixel of 55,114 frames that had already been classified as *contamination* in the previous DNN classification described above. It is worth mentioning that the ground truth labeling process was conducted by two labelers and reviewed by two supervisors who were present during the data collection.

To accomplish the semantic segmentation task, we used U-Net, a CNN model initially developed for biomedical image segmentation [134]. In the U-Net architecture (Figure. 3.5), the input images pass through three sections, including a contraction path (left side), bottleneck, and expansion path (right side). Before feeding the images to the contraction

path, we first resized the video frame to 512 × 512 × 3 pixels and then rescaled the range of each pixel intensity from [0, 255] to [0, 1] to speed up the model convergence learning process.

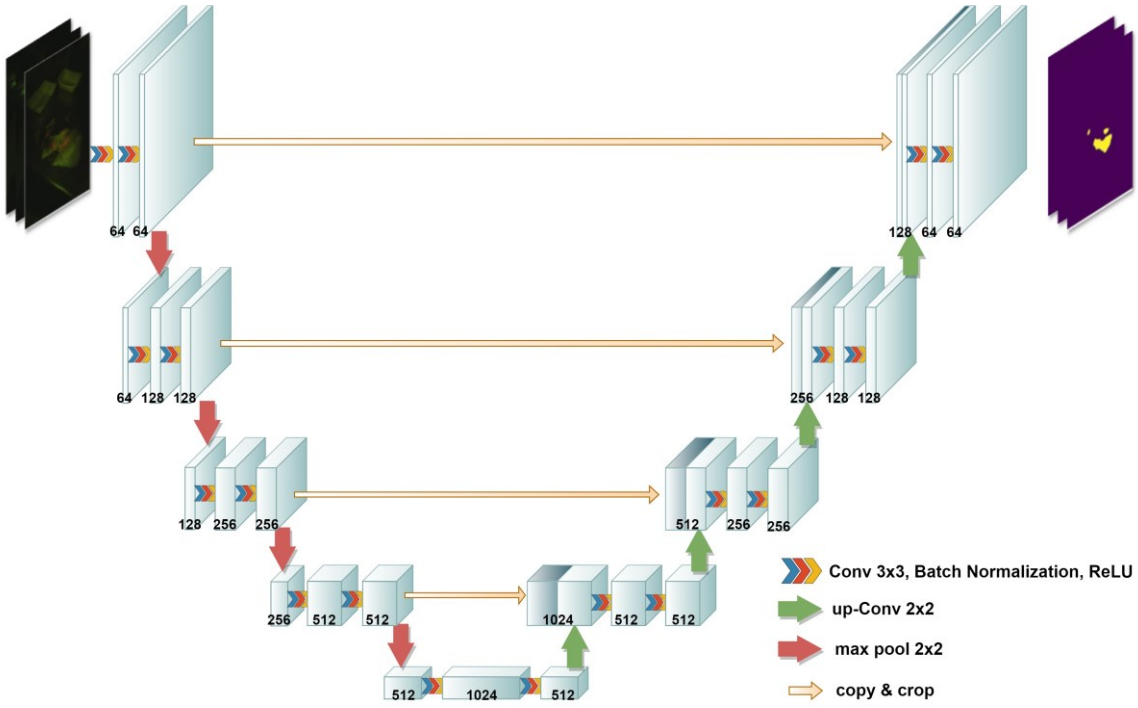


**Figure 3.5.** U-Net architecture.

The contraction path also called the downsampling path, comprises two 3 × 3 convolutions that are repeated several times. These two convolution layers are followed by a rectified linear unit (ReLU) and a downsampling layer composed of a 2 × 2 max-pooling process with a stride of 2. The number of feature channels (filters) at each downsampling step was doubled, allowing the architecture to learn more complex structures.

The data then passes through the bottleneck, a nonlinear dimensionality reduction layer, which is the nethermost section of the U-Net architecture and operates as a bridge between the contraction and expansion paths. The data then flows through the expansion path, each step of which is similar to the contraction path. It comprises repeated application of two 3 × 3 convolutions, followed by a 2 × 2 up convolution layer that reduces

the number of feature channels by half to maintain the architecture's symmetry. It is worth mentioning that the feature maps of each of the two-convolution layers in the contraction path are concatenated with the feature maps of the same layer in the expansion path. By doing so, the high-resolution feature maps from the contraction path are concatenated with the upsampled features of the expansion path, which can improve the model's ability to learn and localize the representations. It is also worth noting that after each convolution layer, we used batch normalization to solve the problem of internal covariate shift [129].

Finally, since our task is a binary classification at the pixel level (contamination vs. background), a sigmoid function is used for the output layer. Also, binary cross-entropy was employed as the model loss function, and the Adam optimizer was employed to update model weights and learning rate to minimize the loss.

### 3.3.6. Experimental setup

When evaluating the models for classification between clean and contamination frames, and also when evaluating the image segmentation of the fecal matter on meat surfaces, all codes were implemented using the Keras framework (open-source library for solving machine learning problems). We used the Tensorflow-GPU v2.6.0 backend on a GPU-enabled workstation with an NVIDIA GeForce GTX 1080 FTW 8 GB GDDR5. The experiments were conducted with Windows10 as the operating system.

## 3.4. Results

### 3.4.1. Model performance for clean vs. contamination classification

The classification between clean and contamination was carried out on a total of 108,296 frames, including 53,182 clean frames and 55,114 contamination frames. We chose 70% of frames at random as our training set, 20% for validation, and 10% of the frames as the test set to evaluate the performance of our classification model. Since deep learning models result in higher performance in the presence of rich and sufficient data, we used some data augmentation methods such as random flip and rotation to improve our model's accuracy and generalization.

We trained and validated our model over 150 epochs by choosing binary cross-entropy as the loss function and Adam optimizer with a learning rate of 0.001. We also set the batch size to 32. Batch size is the number of propagated frames through the network in one iteration during the model training. It is worth noting that in addition to Adam, four other optimization algorithms, including SGD, RMSprop, Adamax, and Adagrad, were used to train the model, and the highest performance was achieved using the Adam optimizer.

Our model performance was evaluated using the six metrics of accuracy, precision, recall, specificity, F-score, and area under the curve (AUC). The first five metrics are defined in Eqs. (3.2) and (3.3):

$$\text{Accuracy} = \frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})}, \text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}, \text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{3.2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}}, \quad \text{F}_{\text{score}} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}} \tag{3.3}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. In addition, we used the area under the curve (AUC) of the receiver operating characteristic (ROC), as an indicator of the model's ability to separate the two classes.

As shown in Table 3.1, the model could successfully discriminate between the two classes of clean and contamination with an accuracy of 97.32%, precision of 97.66%, recall of 97.06%, specificity of 97.59%, F-score of 97.35%, and AUC of 99.54%. Figure 3.6 shows the accuracy and loss of the model during training and validation. Figure 3.7 illustrates the confusion matrix of the model on the test set, which is used as a visual evaluation tool in classification. The rows of the confusion matrix show the actual class for clean and contamination, and the columns represent the predicted label of each class.

**Table 3.1.** Performance of the EfficientNet-B0 for discrimination between *clean* and *contamination* frames.

| Evaluation metrics | Accuracy (%) | Precision (%) | Recall (%) | Specificity (%) | F-score (%) | AUC (%) |
|---|---|---|---|---|---|---|
| | 97.32 | 97.66 | 97.06 | 97.59 | 97.35 | 99.54 |

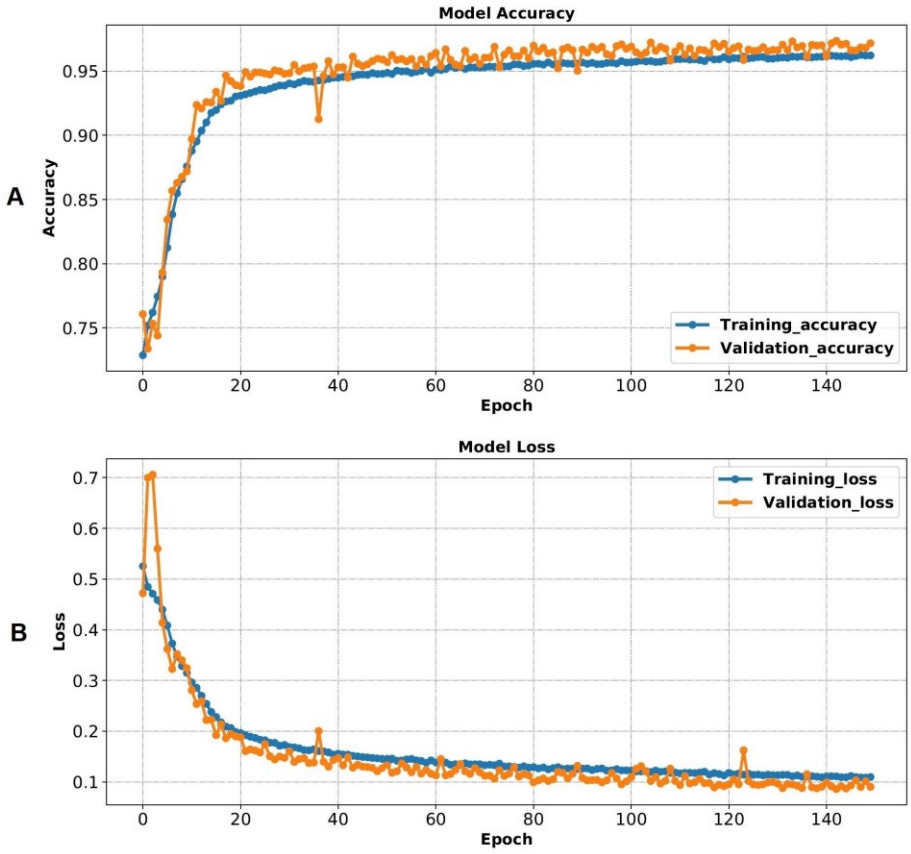


**Figure 3.6. (A)** The model accuracy during training and validation. **(B)** The model loss during training and validation.

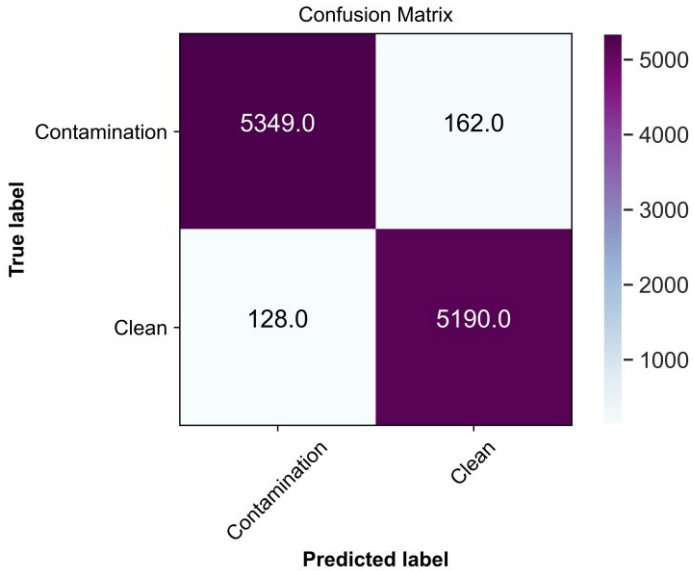


**Figure 3.7.** The confusion matrix of the model when applied to the test set.

### 3.4.2. Model performance on fecal contamination image segmentation

The segmentation of areas of fecal matter in meat surface images is used to determine the presence and the exact location of contamination. The model was trained, evaluated, and tested on the 55,114 frames already labeled as *contamination* in the classification section. Like the previous section, 70% of the data was chosen for training, 20% for validation, and 10% as the test set.

Our semantic segmentation model was trained and validated over 100 epochs. Like our classifier, we chose binary cross-entropy as the loss function, the Adam optimizer with a learning rate of 0.001, and a batch size of 32.

The model performance was evaluated using the same metrics as described above. However, instead of accuracy, we used another metric, intersection over union (IoU), to assess the model segmentation performance. The rationale for not using accuracy is that our semantic segmentation model is a pixel-wise binary classifier (contamination vs. background). Since the fecal matter usually covered just a tiny portion of the frames, the accuracy would be consistently over 99%, making it not useful for evaluating the model performance.

IoU, also known as Jaccard Index, is a well-known metric for assessing how accurately a segmentation method can segment the ROI compared to ground truth segmentation. The IoU metric quantifies the percentage of overlap between the ground truth and the model prediction.

Equation (3.4) shows the IoU definition.

$$\mathrm{IoU} = \frac{\mathrm{TP}}{(\mathrm{TP+FP+FN})} \tag{3.4}$$

Our semantic segmentation model segmented the fecal matter on meat surfaces with an IoU of 89.34%, precision of 92.95%, recall of 95.84%, specificity of 99.79%, F-score of 94.37%, and AUC of 99.89% (Table 3.2).

**Table 3.2.** Performance of the U-Net for segmentation of fecal matter in meat surface images

| Evaluation metrics | IoU (%) | Precision (%) | Recall (%) | Specificity (%) | F-score (%) | AUC (%) |
|---|---|---|---|---|---|---|
| | 89.34 | 92.95 | 95.84 | 99.79 | 94.37 | 99.89 |

For better visualization to show how accurately our model can segment fecal matter areas in meat surface images, we show six sample images that were randomly chosen from the test set (Figure. 3.8). The first row of Figure. 8 shows the input frames to be analyzed using the semantic segmentation model. The second row is the segmented image output by the model, and the third row shows the ground truth segmented image where a human expert has labeled each pixel.

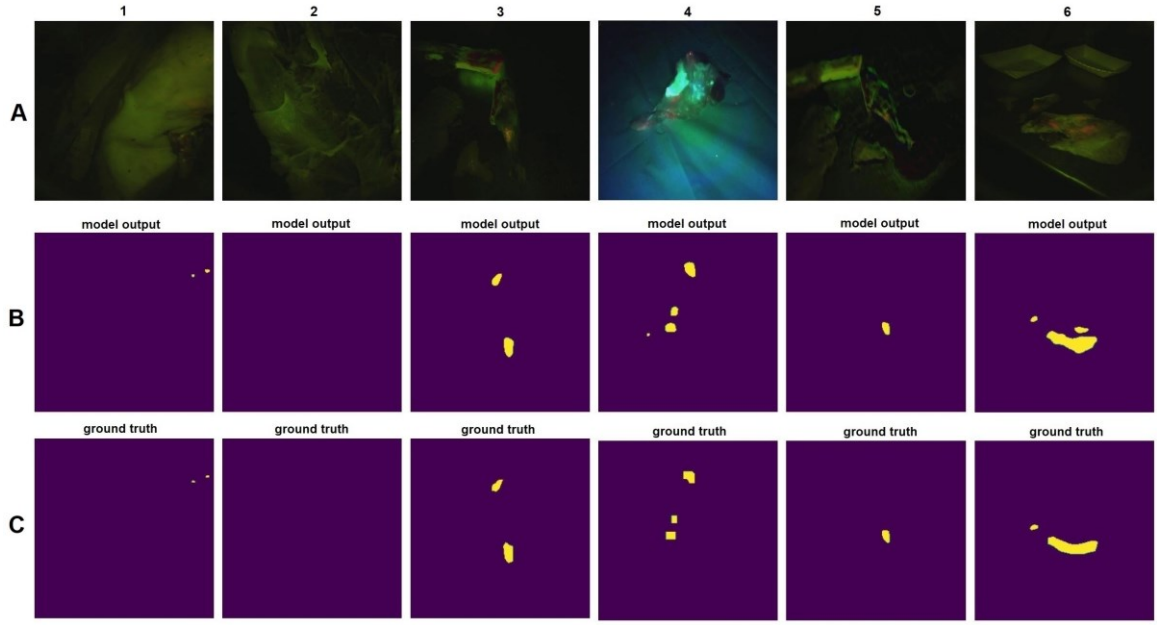


**Figure 3.8.** Performance of the semantic segmentation method on some randomly selected test frames. **(A)** The input frames to the semantic segmentation model. **(B)** Segmented image output by the model. **(C)** The ground truth segmented image by human experts.

Comparing the model output with ground truth shows that our model can segment fecal matter on meat surfaces with high accuracy. Interestingly, in columns four and six, there are two small areas that the model segmented as fecal matter but were not labeled in the ground truth by our human experts. A closer look at the input frames reveals that those areas were chosen correctly and precisely as fecal contamination by the model, and

the three human experts, when reviewing the images, agreed that they had missed these tiny spots during the labeling of the images.

## 3.5. Discussion

The goal of this research is to develop a system and software that can be used to make processed meat safer for consumers. Detecting contamination can be extremely challenging in production environments that have a lot going on at once. We need to overcome fluorescence imaging issues caused by bright ambient light, continually changing shadows as carcasses move about the processing area, and often high-speed production requirements. It is difficult for a human to recognize these kinds of contamination without some form of automated analysis. In this paper, we have demonstrated the ability of state-of-the-art machine learning models to quickly analyze the presence of contamination with a portable handheld carcass scanning system. This system has been well received by the meat processing personnel working on the floors of the meat processing plants where we have been testing.

The detection of contaminated regions using CSI-D is beneficial because action can be taken immediately to cut away meat with fecal contamination from the carcass and help the facility meet the "zero tolerance" standard for fecal contamination that is required by the USDA FSIS. CSI-D can also help train new processing plant staff and FSIS inspectors in locating contamination that might be difficult to see during visual inspections. Although fluorescence-based imaging can be beneficial, the detection of fecal matter on meat surfaces can be challenging because the fecal contamination may be very small, there may be multiple tiny spots on a carcass, and the contamination may be a very faint smear, making detection difficult during the live inspection. The contaminated regions that fluoresce can be of different sizes and of varying shades or hues depending on the color of the meat or fat in the background and the mix of fluorophores being detected as well as their concentration. This can increase the risk of missing some contaminated areas by

inspectors. That is why developing a model for accurate detection of the contaminated areas on meat could be crucial. The inspection process will be more straightforward for inspectors to recognize fecal residue on meat surfaces if we have an accurate and reliable model that can highlight contamination.

The model we have developed provides a substantive improvement in the ability to detect even faint indications of fluorescence of fecal contamination.

## 3.6. Conclusion

These results demonstrate that applying deep learning algorithms in the food inspection domain can provide higher levels of food safety assurance. State-of-the-art deep learning algorithms can be used with new fluorescence imaging systems for detecting fecal residues on meat surfaces to keep our meat and poultry safer than is possible with conventional visual inspection alone.

We have presented a contamination detection technique based on deep learning that automatically identifies unclean meat surfaces contaminated with fecal matter in a fluorescence video image. We have also shown how video image processing and new LED illumination and imaging sensor can overcome the problem of detecting contamination using fluorescence under the bright ambient light found in meat processing facilities. Our methods can also segment and highlight areas of fecal residue in images of contaminated meat surfaces for easy identification and trimming from carcasses in real-time during meat processing.

We employed a state-of-the-art deep convolutional neural network architecture named EfficientNet-B0 on 108,296 images extracted from videos and labeled as "clean" and "contamination," using 70% of the images for training and 20% of the images for validating the model. When tested on the remaining 10% of the images, classification results yielded a 97.32% accuracy (97.66% precision, 97.06% recall, 97.59% specificity, 97.35% F-score, and 99.54% AUC). We segmented areas of fecal residue in images classified as

"contamination" using the U-Net semantic segmentation algorithm. Segmentation results from 55,114 "contamination" frames yielded an intersection over union (IoU) score of 89.34% (precision of 92.95%, recall of 95.84%, specificity of 99.79%, F-score of 94.37%, and AUC of 99.89%).

# 4 CHAPTER 4.

# DEEP LEARNING AND MULTIWAVELENGTH FLUORESCENCE IMAGING FOR CLEANLINESS ASSESSMENT AND DISINFECTION IN FOOD SERVICES

## 4.1. Summary

**Background:** Precise, reliable, and speedy contamination detection and disinfection is an ongoing challenge for the food-service industry. Contamination in food-related services can cause foodborne illness, endangering customers and jeopardizing provider reputations. Fluorescence imaging has been shown to be capable of identifying organic residues and biofilms that can host pathogens. **Methods:** We use new fluorescence imaging technology, applying state-of-the-art deep learning algorithms, namely Xception and Deeplabv3+, to identify and segment contaminated areas in images of equipment and surfaces. **Results:** Deep learning models demonstrated a 98.78% accuracy for differentiation between clean and contaminated frames on various surfaces and resulted in an intersection over union (IoU) score of 95.13% for the segmentation of contamination. The portable imaging system's intrinsic disinfection capability was evaluated on S. enterica, E. coli, and L. monocytogenes, resulting in up to 8-log reductions in under 5 seconds. **Conclusions**: Results showed that fluorescence imaging with deep learning identification algorithms can help assure safety and cleanliness in the food-service industry.

## 4.2. Background

Eating at restaurants and other institutional kitchens is an integral part of our lives worldwide. In 2016, the U.S. restaurant industry had sales of $799 billion [135]. The National Restaurant Association indicated that 47% of every food dollar in 2016 was spent in a restaurant, and in 2015 the average American ate out five times per week [135]. Restaurants handle a wide variety of raw foods, which presents a risk of cross-contamination, leading to foodborne illness. Cross-contamination can happen at any stage of food handling, from food processing to food serving or in domestic kitchens, when there is an unintentional transfer of bacteria or other microorganisms from contaminated surfaces to another contact surface [136]. In restaurants, cross-contamination can occur during the food preparation stages due to poor cleaning and sanitization procedures, improper hand hygiene, or improper separation of ready-to-eat foods (e.g., salad and fresh fruits) from foods requiring cooking (e.g., meat), [137, 138].

Foodborne illness outbreaks occur when people eat contaminated food with a significant dose of a disease-causing agent [1]. Since restaurants provide a way for people to gather for shared food experiences, they also provide opportunities for outbreaks to happen. Restaurants are more often linked to food illness outbreaks than other food preparation sites. In 2017, 841 foodborne disease outbreaks were reported in the U.S., with restaurants accounting for two-thirds of these outbreaks [139]. The impact of these outbreaks is significant in both the cost on human health and mortality and in the economic implications for commercial establishments. The economic impact comes from loss of revenue and damage to brand strength. Loss of income is due to the cost of a facility shutdown while investigating and remediating the problem, lost customers, and lawsuits [135].

In 2015 the share value of a reputable restaurant chain in the U.S. decreased by 18% in the first two months of an E. coli outbreak, and their brand strength index fell by 27%

[140]. A 2013 study in the International Journal of Hospitality Management reported that the negative impact on investment value expressed as cumulative abnormal return (CAR) for multiple publicly traded firms was significant in magnitude [141]. A 2018 study conducted by Brand Finance estimating the impact on business equity showed that "Businesses in the Restaurant & Food Services industry face a risk of losing up to 19% of their Enterprise Value as a result of a food safety incident." Using the example of Darden Restaurants Group, they calculated the business value at risk for Darden to be $2.4 billion and for the industry as a whole the business risk to be $104 billion [140].

During meal preparation, residues left on food processing surfaces or equipment can create a favorable environment for microorganisms to grow, often forming biofilms5. Foodborne pathogens associated with food poisoning include Salmonella enterica, Listeria monocytogenes, and Escherichia coli O157:H7, which form robust biofilms [136]. Biofilms consist of a consortium of sessile microbial species formed on biotic or abiotic surfaces and a self-produced extracellular polymeric substance (EPS) matrix. EPS comprises exopolysaccharides, extracellular proteins, eDNA, and lipids, and provides structural rigidity as well as protection against antimicrobials, antibiotics, starvation, and dehydration [142-144]. Bacterial cells can be easily released from a biofilm and quickly spread to other surfaces in the food preparation environment [145]. The release of pathogenic bacteria from biofilms leads to cross-contamination from preparation surfaces to the prepared food. Effective sanitization methods and food handling procedures must be implemented to reduce the spread of foodborne pathogens.

Currently, retail food and food-service businesses in the U.S. are regulated by more than 3,000 state, local, and tribal entities [146]. These agencies are responsible for maintaining oversight and inspection of over one million food establishments, including restaurants, cafeterias, and dining centers in healthcare facilities, schools, and correctional facilities [146]. The most common testing method for microorganism contamination in the food industry is swabbing of food contact surfaces and equipment for

Adenosine Triphosphate (ATP) testing or laboratory culturing [147]. These methods can be time-consuming and may take hours or days to complete. They do not provide real-time detection of potentially harmful microorganisms on the food preparation surfaces. The lack of real-time inspection allows for the additional spread of microorganisms through the food preparation facility. Swabbing methods cannot test large surface areas since swabs can only collect residue from small areas, even when swiped back and forth (10cm x 10cm). To combat cross-contamination in restaurants and institutional kitchens, more effective solutions are needed to inspect food processing surfaces, to increase food safety and public confidence in the food processing system, and increase public confidence in the cleanliness of restaurants and institutional kitchens. Optical fluorescence imaging can provide this improved confidence. Fluorescence imaging can play an essential role in food safety and provide a fast and effective method to detect biofilms and organic residues that are less visible or invisible to the human eye. Real-time analysis to identify invisible biofilms and organic residues and immediately disinfect them could reduce foodborne illnesses in restaurants and institutional kitchens.

Applications of image processing and machine learning algorithms have rapidly increased in food safety and food security, especially in meat processing [114, 119]. Several studies have used threshold-based algorithms to detect contaminated areas in the processing environment. Determining optimum threshold values is challenging, and inappropriate thresholds can contribute to false-positive and false-negative results. More sophisticated and reliable algorithms are needed to fill the gap in current analysis methods. Deep neural network (DNN) algorithms, especially convolutional neural network (CNN) algorithms, have become popular machine learning methods in many fields [148-151], including food safety-related technologies [38, 121]. An advantage of CNN models is that they can automatically extract information from images and videos and learn from them without human-derived image feature extraction techniques. Learned weights of a trained CNN model can be saved and reused with a new dataset without re-training the model

[122]. DNN algorithms have many food safety applications, including food recognition and classification, calorie estimation, supply chain monitoring, food quality assessment, and contamination detection [123, 152-154]. Food-related applications of DNN algorithms have been studied, but we know of no study where the main focus is on detection and segmentation of organic residue or biofilm contamination in institutional kitchens and restaurants.

In this research, we use a new fluorescence imaging system developed by SafetySpect Inc (Grand Forks, ND) and two state-of-the-art deep learning algorithms, Xception, to identify video frames with contamination and then DeepLabv3+, to precisely segment and label the associated regions of contamination in the frame identified as contaminated. A portable fluorescence imaging system for "contamination, sanitization inspection, and disinfection" (CSI-D) has been developed by SafetySpect [112]. It provides mobility and flexibility for contamination identification and disinfection on surfaces in institutional kitchens and restaurants, even under bright ambient light. However, to tap the enormous potential of this technology for use in the food and food service industries, the fluorescence video or image frames produced by the system must be appropriately processed and interpreted. An automated method for analyzing these video frames and detecting contamination will make work easier for human inspectors. When the CSI-D detects contamination missed during cleaning, its UVC germicidal illumination can be used to disinfect the missed area. CSI-D is not intended to provide primary sanitization but can disinfect small areas missed during cleaning. We used several bacterial strains to evaluate the efficacy of the CSI-D system's disinfection performance. We characterized the killing percentage based on a range of UVC dosages applied to multiple strains of *S. enterica*, *E. coli*, and *L. monocytogenes*. We designed this part of the study to evaluate the effectiveness of this additional safety barrier during food processing.

## 4.3. Methods

### 4.3.1. Contamination detection and disinfection technology

Much of the fluorescence in biofilms and food residues come from a limited number of organic compounds characterized by conjugated double bonds (alternating single and double bonds) that can fluoresce. Polyaromatic hydrocarbons, heterocyclic compounds, and several highly unsaturated aliphatic molecules comprise intrinsic fluorescent components [155, 156]. The CSI-D is a fast, convenient, and easy-to-use handheld automated imaging inspection device that offers mobility and flexibility for fluorescence-based detection of intrinsic fluorophore residues on various surfaces [112]. CSI-D provides detection, disinfection, and documentation of contaminants on surfaces that might spread pathogens and cause disease. Novel aspects of the CSI-D solution include the combination of contamination identification and immediate remediation of the potential threat (bacteria, virus) using UVC light disinfection and image- or video-based documentation of this process to provide traceable evidence of disinfection. Its lighting module consists of 275-nm and 405-nm light-emitting diode (LED) arrays and a heatsink for heat dissipation from the LEDs and driver circuits. SafetySpect designed the CSI-D illumination system to ensure adequate illumination uniformity and intensity at the desired working distance (between 5 to 8 inches) for fluorescence detection and effective disinfection using the 275-nm LEDs for treatments of 2 to 5 seconds. Wireless communication enables remotely placed staff to supervise inspections in real-time and instantly respond to inspection concerns. Figure 4.1 shows the CSI-D device.

**Figure 4.1.** CSI-D device, Front view, and Rear view.

### 4.3.2. Contamination detection data collection

We collected data at six Edgewood long-term care facility (LTCF) institution kitchens in North Dakota and two restaurants in Los Angles, California, USA. Before data acquisition, we had detailed discussions with each facility manager to better understand their high-touch and high-risk areas, the cleaning methods used, and any perceived sanitization concerns. The CSI-D system was used to record videos from multiple high-risk areas. We scanned all identified high-risk areas, such as trash bins, fridge door handles, cutting boards, and preparation tables. We captured the videos at 1024 × 768 resolution and 24 frames per second (FPS). One hour and 35 minutes of video scans were used for further analysis.

Figure 4.2A,B show examples of clean and contaminated surfaces. Since many materials fluoresce naturally, clean surfaces can show both green and red fluorescence. It is insufficient to simply look for different colors in the dual-band fluorescence images. What is important is to look for fluorescence differences in the image that are different from the background objects. Figure 4.2A, columns 4 and 5 show some fluorescence differences that follow the background object's shape and are not *contamination*. In Figure 4.2B, we see fluorescence differences in the form of spray and splash patterns that are

clearly different from the background object fluorescence and are likely *contamination*. Depending on the excitation wavelength, fluorescing food components can include aromatic amino acids, vitamins A, E, B2, and B6, NADH coenzymes, phenolic compounds, chlorophylls, and porphyrins. Process-derived elements that also fluoresce consist of Maillard reaction products, food additives, and contaminants, including antibiotics, pesticide residues, mycotoxins and aflatoxins, and fecal contamination [155].

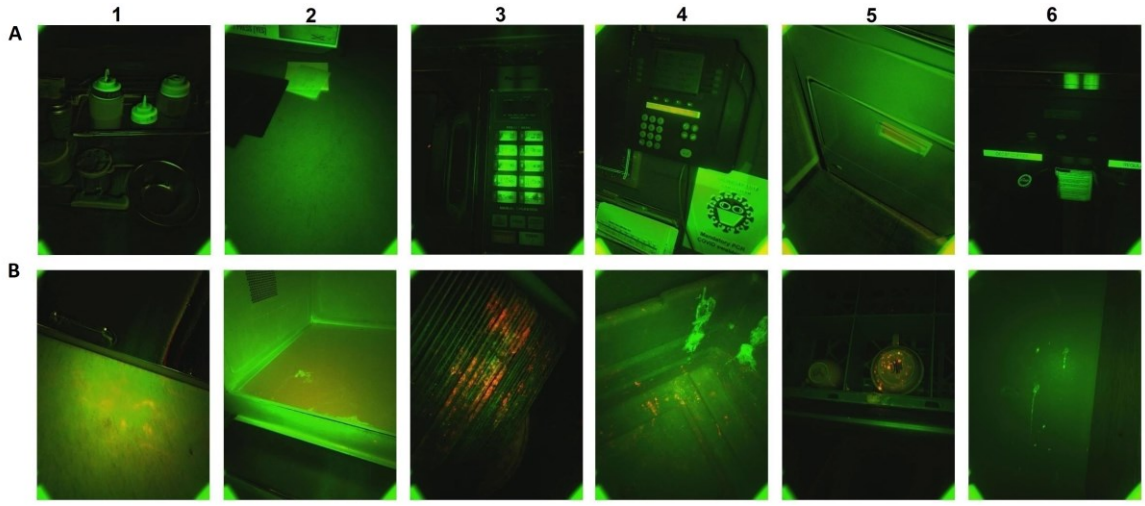


**Figure 4.2.** Fluorescence images captured by CSI-D. **(A)** CSI-D fluorescence images of surfaces without indications of contamination. **(B)** CSI-D fluorescence images of surfaces with indications of contamination.

### 4.3.3. Contamination detection and segmentation

We set out to accomplish two aims with our investigation. Using a DNN model, we initially sought to distinguish between images showing contamination on different surfaces (*contamination*) and those showing no contamination (*clean*). The precise locations of the *contamination* in the images we previously classified as contamination images were subsequently detected using a semantic segmentation approach.

#### 4.3.3.1. Contamination detection model architecture

For the image classification aim, we used a state-of-the-art CNN model named Xception to classify the video frames recorded using the CSI-D device into two categories:

clean and contamination. Before feeding the data to the model, we extracted frames from videos. Selected frames showing contamination were labeled as contamination, while selected frames showing no contamination were labeled clean.

The Xception model was developed by Google [157] by modifying modules of Inception [158] with depthwise separable convolutions (DSC). Xception's high-performance architecture depends on two key components: DSC and residual connections between convolution blocks. Using DSC increases representational efficiency while generating fewer parameters than standard CNNs, reducing computational costs and memory requirements. Using residual connections [63] addresses the vanishing gradient problem [159] common in very deep networks, which causes architecture performance to saturate or degrade.

The backbone of Xception's feature extraction is its 36 convolutional layers. These layers are divided into 14 modules connected by linear residual connections, except for the first and last modules. The network performs channel-wise pooling or cross-channel downsampling with each residual 1×1 convolution layer block [132]. The 1×1 convolution layer acts as a projection layer, aggregating information across channels and reducing dimensionality. This layer reduces the number of filters while adding non-linearity and retaining crucial, feature-related information. In the Xception architecture, the 1×1 convolution layer has a stride of 2×2, reducing the feature map size in the residual path to match it with the feature map size of the max pooling layer.

Figure 4.3 shows the Xception model's architecture, where the data enters the input flow, then passes to the middle flow for eight iterations, and finally, the exit flow. All convolution and SeparableConvolution layers are followed by batch normalization and a Rectified Linear Unit (ReLU) activation function [129, 160]. The batch normalization can stabilize the learning process and significantly reduce the number of training epochs needed by reducing the internal covariate shift when training a deep learning model [129]. The default input image size for Xception is (299, 299, 3), and the images will pass through

the entry flow.

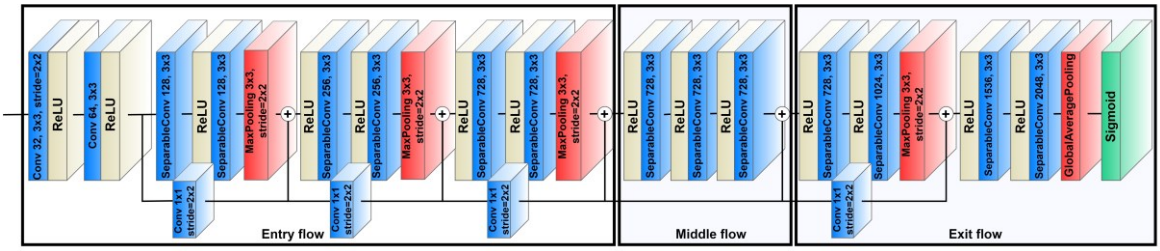


**Figure 4.3.** The architectures of Xception model (Conv stands for convolutional layer).

The entry flow comprises four modules, including one module of convolutional layers and three modules of SeparableConvolution layers. The first module comprises a layer of 32 filters with a kernel size of 3×3 and a stride of 2×2, and a layer of 64 filters with a kernel size of 3×3. For the other three entry flow modules, each has two SeparableConvolution layers and one max-pooling layer. The filter sizes in these three modules are 128, 256, and 728, used in that order. The kernel sizes are 3×3, and the strides are 2×2. The input image dimensions change from 299×299×3 on entry to 19×19×728 on exit from the entry flow.

The feature maps then pass to the middle flow, consisting of eight repeated modules, each of which includes three SeparableConvolution layers with a filter size of 728 and a kernel size of 3×3. The middle flow mainly learns the correlations and optimizes the features after the feature spatial dimensions are reduced in the entry flow.

In the final stage, the feature maps proceed to the exit flow consisting of two modules, a residual block and a non-residual block. The residual block consists of two SeparableConvolution layers, with filter sizes of 728 and 1024 used in that order, followed by a max-pooling layer to reduce the dimensions of the feature maps. The non-residual block also comprises two SeparableConvolution layers, with filter sizes of 1536 and 2048 used in that order. However, instead of using a max-pooling layer, the two SeparableConvolution layers are followed by a global average pooling layer that can minimize model overfitting by reducing the final number of parameters in the model. The

model will use these final 2048 features for classification. Since the first aim of this study is to classify frames as *Clean* or *Contamination*, a binary classification task, we used a sigmoid function as the activation function for the output layer of the model. The sigmoid function is a nonlinear activation function that squashes all values to a range between 0 and 1, providing the probability of belonging to either the *Clean* or *Contamination* classes. We selected binary cross-entropy as the model loss function due to our task's binary nature. The cross-entropy loss between the predicted and actual labels is calculated using the binary cross-entropy method, and the results can be either 0 or 1. The cross-entropy loss will increase if the predicted labels deviate from the actual ones. The binary cross-entropy loss is defined as follows:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i \times \log\big(p(y_i) + (1 - y_i) \times \log\big(1 - p(y_i)\big)\big) \quad (4.1)$$

Where N denotes the training samples size, $y_i$ is the actual label, $p(y_i)$ represents the predicted probability of class 1 (*Contamination*) and $1 - p(y_i)$ denotes the predicted probability of class 0 (*Clean*).

After defining the loss function, the model needs an optimization algorithm to enhance the model performance by changing the model weights and learning rate while minimizing overall loss. We employed the Adam optimizer [133], a stochastic gradient descent approach that estimates the first- and second-order moments to calculate individual adaptive learning rates for different parameters. Adam is computationally efficient, has low memory demand, is invariant to diagonal rescaling of gradients, and is useful for problems with big data and parameters.

#### 4.3.3.2. Contamination segmentation model architecture

Accurate segmentation of contaminated areas on a range of different surfaces is very important because just identifying contamination frames using the method described above does not necessarily result in finding all the contamination. Contaminated areas can be spread sporadically across the surface, and it can be challenging to identify every

single area of contamination in the field of view. Organic residue and biofilms may also contaminate minute areas on a surface, making it difficult to detect even when contamination is identified in a video frame during an inspection procedure. This difficulty is why we focused on the segmentation of contaminated regions in the video frames already identified as contamination as the second aim of this study.

Segmenting and pseudo-coloring contaminated regions can make it easier for inspectors to identify and not overlook any contamination. The many different types of objects and surfaces in restaurants and kitchens can produce different background fluorescence and reflections. There may be many varying kinds of contamination present at the same time, so the combination of these can make it likely that an inspector could miss some area of contamination during an inspection. Pseudo-coloring the contaminations—in our case, green for green fluorescence and red for red fluorescence—can increase the inspector's discrimination of contamination during monitoring. An effective model that accurately segments contaminated regions for a range of surfaces in real-time needs to be developed. The segmentation of the contamination can also be helpful for both restaurant cleaning crews and new health inspectors learning to recognize potential contamination areas and understand how to address potential sources of contamination and avoid cross-contamination during their work. This knowledge can be beneficial for assessing and improving cleaning processes and sanitation standard operating procedures (SSOPs).

In this research, instead of using traditional threshold-based algorithms to segment the contamination, we used a semantic segmentation method to conduct pixel-level classification for each frame to precisely segment contaminated regions on various surfaces. The first step requires pixel-level annotation to assign each pixel of a frame to a specific class. A deep CNN is then trained on the annotated data to classify clean, green fluorescent, and red fluorescent pixels. Later the trained model is tested on an unseen data test set to predict the pixel class. Since we are dealing with three classes in this

research, the segmentation task can be considered a multi-class pixel-wise classification. We used MATLAB image labeler to label each video frame pixel to create the semantic segmentation training and testing datasets. The MATLAB image labeler provides a swift and easy way to annotate images by drawing shapes that can be assigned region of interest (ROI) labels. Our study only segmented the contamination regions on the surfaces, and the remaining pixels were labeled as background. The ground-truth-labeled pixels were categorized as background, green fluorescent, and red fluorescent. A total of 12,000 frames were labeled at pixel level by four image labelers under the supervision of two experts present throughout the data collection.

To carry out the semantic segmentation task, we employed DeepLabv3+ [161], a state-of-the-art semantic segmentation algorithm designed by Google. Figure 4.4 shows how DeepLabv3+ is composed of an encoder and decoder architecture. The encoder can encode multi-scale contextual information, and the decoder module provides for a more precise and accurate recovery of object boundaries. The encoder section consisted of three key components: the ResNet architecture, Atrous convolutions, and Atrous Spatial Pyramid Pooling. The ResNet is used as the network backbone to extract the features. In this study, we used ResNet50. The Atrous convolution, also called dilated convolution, is a powerful approach in dense prediction applications. When using a deep CNN, Atrous convolution enables the model to fine-tune the resolution at which feature responses are generated. It also enables the model to efficiently expand the field of view of the filters to include more context without increasing the number of parameters and computing requirements. DeepLabv3+ also employs Atrous Spatial Pyramid Pooling (ASPP) applied on top of the extracted features from the backbone network. Parallel ASPP layers, with an increasing dilation rate, are used to aggregate multi-scale context, which can robustly segment objects while considering picture context at multiple scales. Since the contamination can be either a tiny or extensive area in this study, we used the default DeepLabv3+ dilation rate of 6. In order to acquire the final segmented mask for the input

image, the output of the ASPP layers is first concatenated and then fed through a 1×1 convolution layer with 256 filters that generate rich semantic information.

For the decoder section, the features generated by the encoder are bilinearly upsampled by a factor of four before being concatenated with the equivalent low-level features from the network backbone that have the same spatial resolution. A 1×1 convolution layer is applied to the low-level features extracted by the backbone network to reduce the channel numbers and avoid outweighing the importance of the encoder features and making the training process more difficult. Finally, after concatenating the low-level features with the rich features from the encoder, a few 3×3 convolution layers are applied to achieve sharper segmentation, and then upsampling by a factor of 4 is performed to generate the final segmented image output.

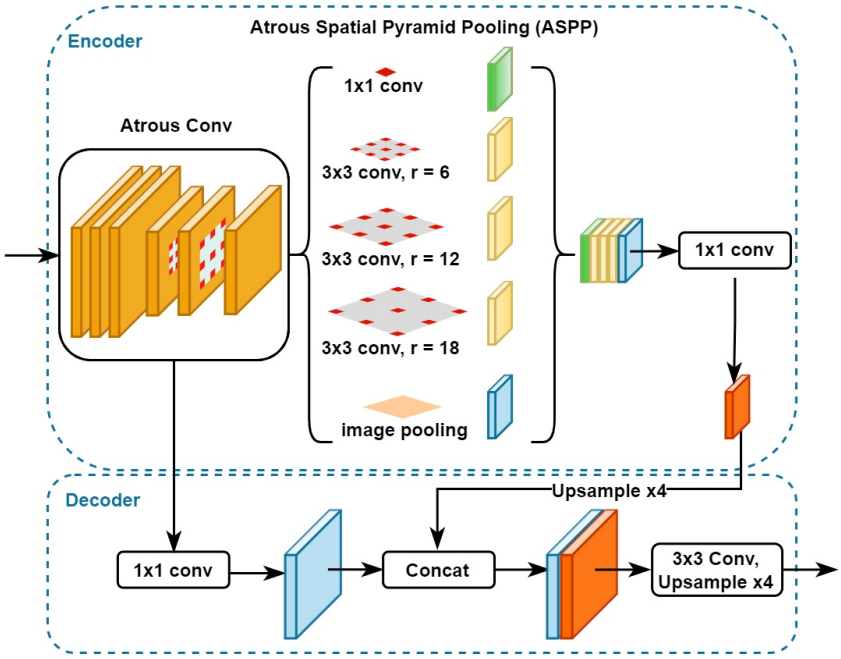


**Figure 4.4.** The architectures of DeepLabv3+ model.

## 4.3.4. Bacteriological methods

### 4.3.4.1. Strains

Four strains each of *Escherichia coli*, *Salmonella enterica* subspecies *enterica*, and *Listeria monocytogenes* were selected for the study. *E. coli* strains included one non-

pathogenic strain (ATCC 25922 (American Type Culture Collection, Manassas, VA)), one antimicrobial-susceptible strain with few virulence factors, and two pathogenic strains (one O157:H7 strain and one multi-drug resistant strain). *Salmonella* strains representing four serotypes were selected: Enteritidis, Infantis, Typhimurium LT2, and Heidelberg. To represent *L. monocytogenes*, strains from serotypes 1/2a and 1/2b were selected, along with two hypervirulent serotype 4b strains.

#### 4.3.4.2. Culture Methods

Bacterial strains were streaked from frozen biomass onto selective agar and confirmed by transferring isolate biomass onto a panel of selective agars. *E. coli* was streaked on MacConkey agar (BD, Franklin Lakes, NJ) and confirmed on MacConkey agar, MacConkey agar with sorbitol, Simmons Citrate agar, and L-agar (Lennox LB Broth base with 1.5% agar; Neogen, Lansing, MI); plates were incubated at 37°C for 18-24 h. Salmonella strains were streaked onto XLT4 agar (XLT4 agar base and supplement, Neogen), incubated at 37°C, and scored at 24 h and 48 h. Isolates were confirmed on XLT4, Brilliant Green, and Lennox LB agars (Neogen), which were incubated for 18-24 h at 37°C. *L. monocytogenes* strains were streaked on Modified Oxford agar ("MOX", Neogen) and incubated for 48 h at 37°C, with plates scored at 24 h and 48 h. Isolates were confirmed on MOX, Palcam (BD), an *L. monocytogenes* chromogenic agar (R&F products, Downers Grove, IL), and tryptic soy agar with yeast extract ("TSA-YE"; tryptic soy broth with 1.5% agar and 0.6% yeast extract; Neogen), and incubated for 24 h at 37°C.

For *E. coli* and *Salmonella*, biomass from L-agar (< 7 days old) was transferred to L-broth (10 mL) and incubated at 37°C overnight (18-24h). For *L. monocytogenes*, tryptic soy broth (10 mL) was inoculated with biomass from TSA-YE and incubated at 37°C for 48 h, with the tube caps loosened.

#### 4.3.4.3. Disinfection

Within 1-1.5 hours prior to UVC exposure, broth cultures were serially diluted in 1X Phosphate Buffered Saline, pH 7.4 (Gibco Laboratories, Gaithersburg, MD), and 100 µL of two consecutive dilutions were manually spread plated onto agar plates. For E. coli and *Salmonella*, 1:105 and 1:106 dilutions were plated onto L-agar. For *L. monocytogenes*, 1:104 and 1:105 dilutions were plated onto TSA-YE. Four replicates per set of experimental conditions were plated, along with four control plates per strain. Initial S. enterica concentrations were $1.7 \times 10^8 – 1.0 \times 10^9$, initial E. coli concentrations were $4.6 \times 10^8 – 8.1 \times 10^8$, and initial L. monocytogenes concentrations were $2.2 \times 10^7 – 5.3 \times 10^7$.

Inoculated agar plates were exposed to UVC at two intensities (10 mW/cm2 or 5 mW/cm2), for durations of 1s, 3s, or 5s. The distance of the light from the agar was 12 cm for the high-intensity samples and 20 cm for low-intensity samples. To ensure colonies would not be too large or crowded to count the next day, all plates were incubated first at room temperature for 2-3 hours, and then at 37°C for 18-20 h for *E. coli* and *S. enterica*, and at 37°C for 48 h for *L. monocytogenes*, scoring at 24 h and 48 h. Colony counts were performed manually. Experiments were conducted on three separate days, each focusing on one species.

### 4.3.5. Contamination detection system setup

To implement the models for classification and segmentation, we used the Keras framework as a high-level API for neural network development. We used a Tensorflow-GPU v2.7.0 backend on a GPU-enabled workstation with an NVIDIA GeForce RTXTM 3090 with 24GB of G6X memory. We used the Windows 10 operating system.

## 4.4. Results

### 4.4.1. Model performance for identification between clean and contamination frames

A total of 72,381 frames were used for clean and contamination classification, comprising 35,858 clean frames and 36,523 contamination frames, and used two

approaches for evaluation. First, we randomly chose 70% of the frames for training and 20% to validate the model for each epoch, holding out the remaining 10% of the frames to test the model performance. Secondly, we used leave-one-out cross-validation (LOOCV), but applied it to the facility level rather than the image frame level. From eight institutional kitchens and restaurants, we trained and validated our model on seven facilities, tested it on one facility, and iterated this eight times. Each facility was part of training, validation, and testing, so we could evaluate how well the model could generalize when facing a new facility and environment.

The model was trained and validated over 100 epochs, choosing binary cross-entropy as the loss function and Adam as the model optimizer, with a learning rate of 0.0001. The model batch size, or the number of frames utilized in one iteration during the training process, was set to 32.

We evaluated the model performance using six metrics: accuracy, precision, recall, F-score, and area under the curve (AUC). The first five metrics are described in equations (4.2) and (4.3) below:

$$\text{Accuracy} = \frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})},\ \text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}},\ \text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{4.2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}},\quad \text{F}_{\text{score}} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}} \tag{4.3}$$

Here TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively. The accuracy shows how accurate the model is at differentiating clean and contamination frames. The precision shows what proportion of the positive samples (in our case, contamination) were identified correctly. The recall or sensitivity is the number of the true predicted positives divided by the total number of true positives and false-negative predictions and indicates the proportion of actual positive observations (contamination class) that were detected correctly. F-score is a harmonic mean of precision and recall used to evaluate model performance. We used the area under the curve (AUC) of the receiver operating characteristic (ROC) to assess the

model's ability to differentiate between clean and contamination classes.

We trained and validated the model over 100 epochs and monitored accuracy and loss, as shown in Figure 4.5A,B. We then applied the model to the test set. The Xception model achieved an accuracy of 98.78%, precision of 99.12%, recall of 98.47%, specificity of 98.44%, F-score of 98.80%, and AUC of 99.91%. Figure 5C shows the confusion matrix representing the model's classification performance.

Table 4.1 shows the results of the LOOCV approach, where each time, we removed one facility from the training process, trained the model on the seven remaining facilities, and then tested it on the removed facility. As expected, the lowest performances resulted when we removed either of the two facilities with the most frames (Facility No. 6 or 7). By withholding facility no. 6, 20,526 frames (28.35% of the total frames) were removed from training and validation, and similarly, by withholding facility no. 7, 17,527 frames (24.21% of the total frames) were removed from training and validation. The drop in performance from removing these facilities with relatively large numbers of frames shows that they contained meaningful information. These could include the facility architecture, ambient light, variety of the background, and a greater variety of contamination patterns, which can help the model learn more robustly.

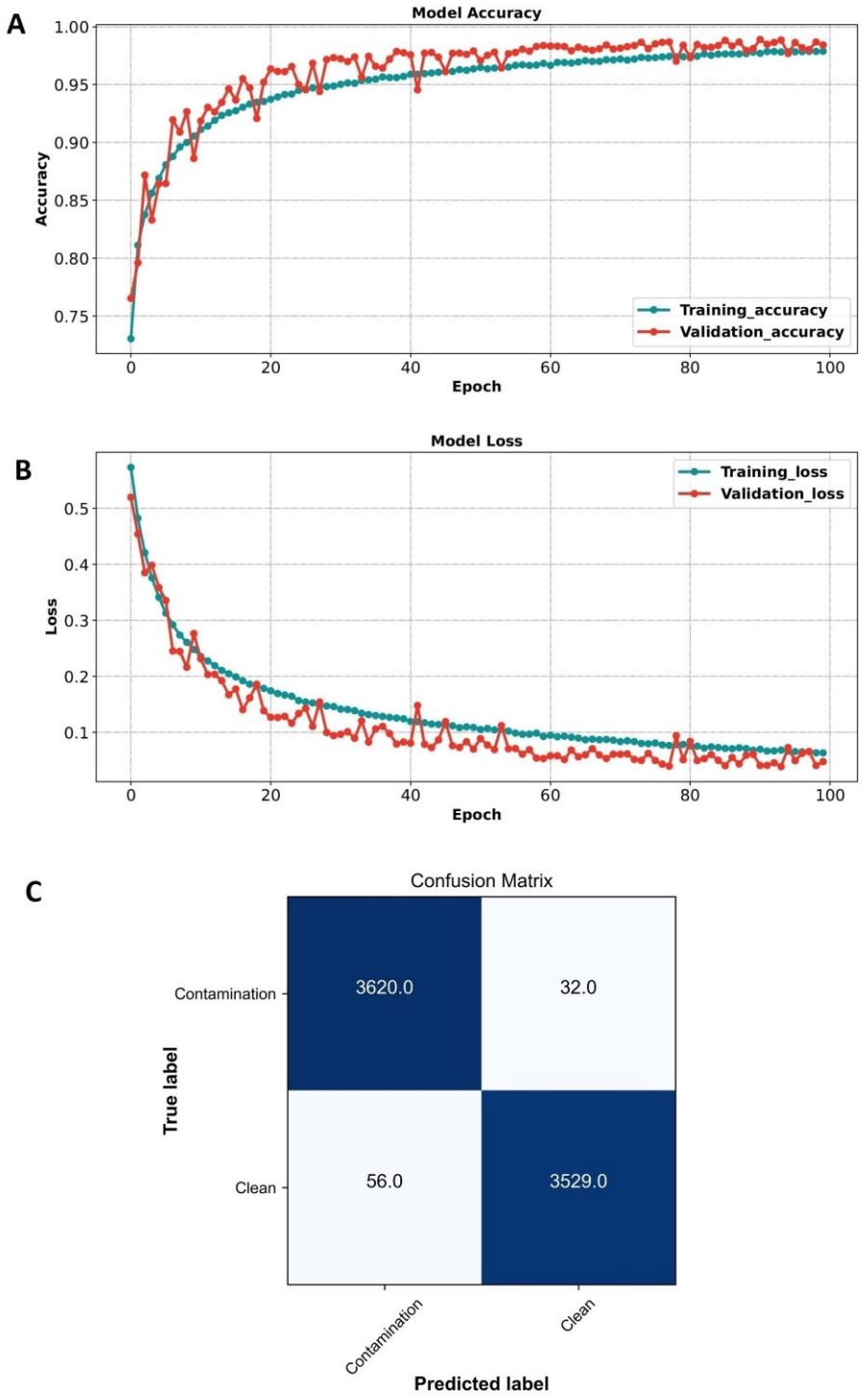


**Figure 4.5**. Performance of the Xception model (clean vs. contamination frames). **(A)** Xception model accuracy during training and validation. **(B)** Xception model loss during training and validation. **(C)** Xception model confusion matrix when applied to the test set.

**Table 4.1.** The Xception model performance for differentiation between clean and contamination frames using the LOOCV approach.

| Facility No. | No. of "Clean" Frames | No. of "Contamination" Frames | Accuracy (%) | Precision (%) | Recall (%) | Specificity (%) | F-score (%) | AUC |
|---|---|---|---|---|---|---|---|---|
| 1 | 708 | 3,242 | 93.30 | 97.71 | 94.13 | 89.27 | 95.89 | 0.9802 |
| 2 | 1,585 | 7,207 | 92.88 | 97.92 | 93.30 | 90.98 | 95.55 | 0.9766 |
| 3 | 2,740 | 1,815 | 91.17 | 93.75 | 83.42 | 96.31 | 88.28 | 0.9339 |
| 4 | 3,893 | 2,574 | 92.14 | 88.15 | 92.74 | 91.75 | 90.38 | 0.9617 |
| 5 | 2,679 | 2,320 | 95.87 | 95.36 | 95.77 | 95.97 | 95.57 | 0.9924 |
| 6 | 11,897 | 8,629 | 86.51 | 86.64 | 80.29 | 91.02 | 83.34 | 0.9092 |
| 7 | 11,041 | 6,486 | 87.13 | 78.65 | 89.70 | 85.62 | 83.76 | 0.9129 |
| 8 | 1,315 | 4,250 | 90.64 | 92.95 | 94.94 | 76.73 | 92.97 | 0.9535 |
| | **35,858** | **36,523** | **91.20** | **91.39** | **90.53** | **89.70** | **90.71** | **0.9525** |

### 4.4.2. Model performance for image segmentation of contamination

We used the DeepLabv3+ semantic segmentation model to identify precise areas of contamination on a variety of surfaces in 12,000 image frames that had been annotated for contamination by four experts. We used 70% of the data for training, 20% for validation, and 10% for testing, over 100 epochs. Since the task is a multi-class pixel-wise classification (background, green fluorescence, and red fluorescence), we chose categorical cross-entropy for the model loss function. We used Adam (learning rate of 0.0001) for the model optimizer and a batch size of 32.

To evaluate the DeepLabv3+ performance, we used the same metrics as for classification evaluation. However, we employed a different metric to evaluate the segmentation performance, replacing accuracy with intersection over union (IoU). Since contamination is often a tiny part of the image, accuracy is commonly greater than 99%, making it less suitable for evaluating performance. IoU, also known as the Jaccard index, quantifies the percentage overlap between the regions annotated by human experts (ground truth) and the model prediction. The following equation defines IoU:

$$\mathrm{IoU} = \frac{\mathrm{TP}}{(\mathrm{TP+FP+FN})} \tag{4.4}$$

Equation (4) shows how we also used false positive rate (FPR) and false-negative rate (FNR) to evaluate the segmentation model performance.

$$\mathrm{FPR} = \frac{\mathrm{FP}}{(\mathrm{FP+TN})}, \qquad \mathrm{FNR} = \frac{\mathrm{FN}}{(\mathrm{FN+TP})} \tag{4.5}$$

The model was able to segment contamination on multiple surfaces into background, green-fluorescence, and red-fluorescence classes, with a mean IoU of 95.13%. Background IoU was 99.65%, green-fluorescence IoU 92.57%, and red-fluorescence IoU 93.18%. The model segmented contamination with 98.99% precision, 99.89% recall, 99.95% specificity, 99.89% F-score, 0.05% FPR, and 0.11% FNR.

A visual representation of how accurately our model could segment the red and green fluorescence contaminations on different surfaces shows example images from our test set selected randomly from six types of surfaces (Figure 4.6). CSI-D fluorescence image frames are shown in the first row, segmentation by the model in the second row, and segmentation by our human experts (ground truth) in the last row.

Comparing ground truth with model output shows that the model accurately segments green and red fluorescence from contamination on various surfaces, including a cabinet door, cup rack, kitchen range, microwave oven, counter-top, and toilet seat. The images show that the model can address contamination in the form of tiny droplets, sprays, splashes, and larger areas like spills or smears. They also show that the model can detect green fluorescence, red fluorescence, or red-green combinations.

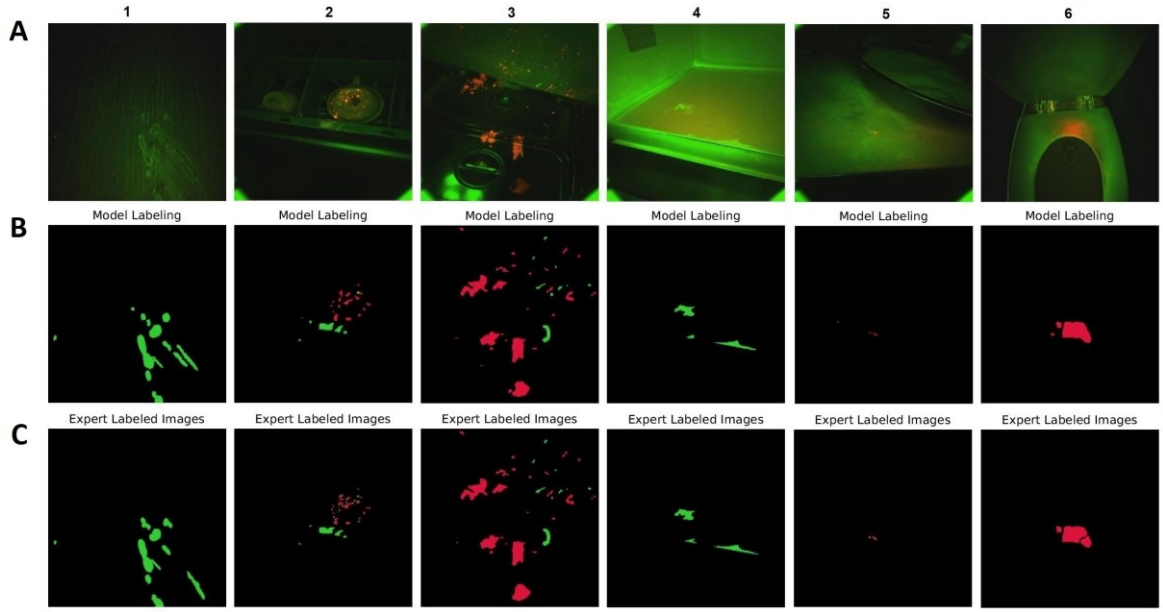


**Figure 4.6.** DeepLabv3+ model performance on selected test set video frames. **(A)** CSI-D fluorescence image frames input to DeepLabv3+. **(B)** DeepLabv3+ segmented image output. **(C)** Same images segmented and labeled by human experts.

### 4.4.3. Bacterial disinfection

When plated cultures of the four strains of *E. coli* were subjected to high-intensity (10 mW/cm2) UVC for 1, 3, or 5s, complete growth inactivation was observed. Similarly, complete inactivation was observed for all strains exposed to lower intensity (5 mW/cm2) UVC for 3s and 5s. Some growth was observed for two strains exposed to the low-intensity UVC for 1s; however, this treatment reduced the final colony count by 90-99%.

Similar results were observed when *S. enterica* was exposed to UVC. Growth was completely inactivated when the Enteriditis strain was subjected to UVC under all conditions. As with the *E. coli*, the growth of the other three S. enterica serovars was completely inactivated under all intensity/time combinations except when exposed to low-intensity UVC for 1s.

Complete growth inactivation was also observed for *L. monocytogenes* strains subjected to high-intensity UVC at all exposure durations except for strain LM3, where growth was reduced by 2.3 Logs (>99%) after 1s. As with the *S. enterica* and *E. coli* strains, at the low UVC intensity, *L. monocytogenes* growth was completely inactivated

when exposure was 3 or 5s, but some cells grew after the 1s exposure. Additional strain details and treatment results can be found in the following tables.

**Table 4.2.** Concentrations and Log reductions of E. coli after cultures were exposed to UVC illumination (5 or 10 mW/cm2) for 1, 3, or 5s (NTS = non-type-specific; ST = Strain Type; MDR = multi-drug resistant).

| Strain ID | Strain Notes | Intensity, mW/cm$^2$ | Exposure Duration (s) | Mean CFU/mL | Log Reduction |
|---|---|---|---|---|---|
| **EC1** | ATCC 92522 | 0 | 0 | 5.50E+08 | 0.00 |
| **EC1** | ATCC 92522 | 10 | 1 | 0.00E+00 | 8.00 |
| **EC1** | ATCC 92522 | 10 | 3 | 0.00E+00 | 8.00 |
| **EC1** | ATCC 92522 | 10 | 5 | 0.00E+00 | 8.00 |
| **EC1** | ATCC 92522 | 5 | 1 | 0.00E+00 | 8.00 |
| **EC1** | ATCC 92522 | 5 | 3 | 0.00E+00 | 8.00 |
| **EC1** | ATCC 92522 | 5 | 5 | 0.00E+00 | 8.00 |
| **EC2** | O157:H7 | 0 | 0 | 4.55E+08 | 0.00 |
| **EC2** | O157:H7 | 10 | 1 | 0.00E+00 | 8.00 |
| **EC2** | O157:H7 | 10 | 3 | 0.00E+00 | 8.00 |
| **EC2** | O157:H7 | 10 | 5 | 0.00E+00 | 8.00 |
| **EC2** | O157:H7 | 5 | 1 | 0.00E+00 | 8.00 |
| **EC2** | O157:H7 | 5 | 3 | 0.00E+00 | 8.00 |
| **EC2** | O157:H7 | 5 | 5 | 0.00E+00 | 8.00 |
| **EC3** | NTS, Susceptible | 0 | 0 | 8.05E+08 | 0.00 |
| **EC3** | NTS, Susceptible | 10 | 1 | 0.00E+00 | 8.00 |
| **EC3** | NTS, Susceptible | 10 | 3 | 0.00E+00 | 8.00 |
| **EC3** | NTS, Susceptible | 10 | 5 | 0.00E+00 | 8.00 |
| **EC3** | NTS, Susceptible | 5 | 1 | 1.25E+07 | 1.81 |
| **EC3** | NTS, Susceptible | 5 | 3 | 0.00E+00 | 8.00 |
| **EC3** | NTS, Susceptible | 5 | 5 | 0.00E+00 | 8.00 |
| **EC4** | ST457, MDR | 0 | 0 | 5.78E+08 | 0.00 |
| **EC4** | ST457, MDR | 10 | 1 | 0.00E+00 | 8.00 |
| **EC4** | ST457, MDR | 10 | 3 | 0.00E+00 | 8.00 |
| **EC4** | ST457, MDR | 10 | 5 | 0.00E+00 | 8.00 |
| **EC4** | ST457, MDR | 5 | 1 | 1.00E+07 | 1.76 |
| **EC4** | ST457, MDR | 5 | 3 | 0.00E+00 | 8.00 |
| **EC4** | ST457, MDR | 5 | 5 | 0.00E+00 | 8.00 |

**Table 4.3.** Concentrations and Log reductions of S. enterica subspecies enterica after cultures were exposed to UVC illumination (5 or 10 mW/cm2) for 1, 3, or 5s.

| Strain ID | Serotype | Intensity, mW/cm$^2$ | Exposure Duration (s) | Mean CFU/mL | Log Reduction |
|---|---|---|---|---|---|
| **S1** | Enteritidis | 0 | 0 | 1.70E+08 | 0.00 |
| **S1** | Enteritidis | 10 | 1 | 0.00E+00 | 8.00 |
| **S1** | Enteritidis | 10 | 3 | 0.00E+00 | 8.00 |
| **S1** | Enteritidis | 10 | 5 | 0.00E+00 | 8.00 |
| **S1** | Enteritidis | 5 | 1 | 0.00E+00 | 8.00 |
| **S1** | Enteritidis | 5 | 3 | 0.00E+00 | 8.00 |
| **S1** | Enteritidis | 5 | 5 | 0.00E+00 | 8.00 |
| **S2** | Infantis | 0 | 0 | 7.88E+08 | 0.00 |
| **S2** | Infantis | 10 | 1 | 0.00E+00 | 8.00 |
| **S2** | Infantis | 10 | 3 | 0.00E+00 | 8.00 |
| **S2** | Infantis | 10 | 5 | 0.00E+00 | 8.00 |
| **S2** | Infantis | 5 | 1 | 2.53E+07 | 1.49 |
| **S2** | Infantis | 5 | 3 | 0.00E+00 | 8.00 |
| **S2** | Infantis | 5 | 5 | 0.00E+00 | 8.00 |
| **S3** | Typhimurium LT2 | 0 | 0 | 6.53E+08 | 0.00 |
| **S3** | Typhimurium LT2 | 10 | 1 | 0.00E+00 | 8.00 |
| **S3** | Typhimurium LT2 | 10 | 3 | 0.00E+00 | 8.00 |
| **S3** | Typhimurium LT2 | 10 | 5 | 0.00E+00 | 8.00 |
| **S3** | Typhimurium LT2 | 5 | 1 | 5.00E+06 | 2.12 |
| **S3** | Typhimurium LT2 | 5 | 3 | 0.00E+00 | 8.00 |
| **S3** | Typhimurium LT2 | 5 | 5 | 0.00E+00 | 8.00 |
| **S4** | Heidelberg | 0 | 0 | 1.03E+09 | 0.00 |
| **S4** | Heidelberg | 10 | 1 | 0.00E+00 | 9.00 |
| **S4** | Heidelberg | 10 | 3 | 0.00E+00 | 9.00 |
| **S4** | Heidelberg | 10 | 5 | 0.00E+00 | 9.00 |
| **S4** | Heidelberg | 5 | 1 | 1.25E+07 | 1.91 |
| **S4** | Heidelberg | 5 | 3 | 0.00E+00 | 9.00 |
| **S4** | Heidelberg | 5 | 5 | 0.00E+00 | 9.00 |

**Table 4.4.** Concentrations and Log reductions of L. monocytogenes after cultures were exposed to UVC illumination (5 or 10 mW/cm2) for 1, 3, or 5s (CC = clonal complex).

| Strain ID | Serotype | Intensity, mW/cm$^2$ | Exposure Duration (s) | Mean CFU/mL | Log Reduction |
|---|---|---|---|---|---|
| **LM1** | 1/2a | 0 | 0 | 2.20E+07 | 0.00 |
| **LM1** | 1/2a | 10 | 1 | 0.00E+00 | 7.00 |
| **LM1** | 1/2a | 10 | 3 | 0.00E+00 | 7.00 |
| **LM1** | 1/2a | 10 | 5 | 0.00E+00 | 7.00 |
| **LM1** | 1/2a | 5 | 1 | 2.00E+06 | 1.04 |
| **LM1** | 1/2a | 5 | 3 | 0.00E+00 | 7.00 |
| **LM1** | 1/2a | 5 | 5 | 0.00E+00 | 7.00 |

| | | | | | |
|---|---|---|---|---|---|
| **LM2** | 4b, CC4 | 0 | 0 | 5.33E+07 | 7.00 |
| **LM2** | 4b, CC4 | 10 | 1 | 0.00E+00 | 7.00 |
| **LM2** | 4b, CC4 | 10 | 3 | 0.00E+00 | 7.00 |
| **LM2** | 4b, CC4 | 10 | 5 | 0.00E+00 | 7.00 |
| **LM2** | 4b, CC4 | 5 | 1 | 5.75E+06 | 0.97 |
| **LM2** | 4b, CC4 | 5 | 3 | 0.00E+00 | 7.00 |
| **LM2** | 4b, CC4 | 5 | 5 | 0.00E+00 | 7.00 |
| **LM3** | 1/2b | 0 | 0 | 5.03E+07 | 0.00 |
| **LM3** | 1/2b | 10 | 1 | 2.50E+05 | 2.30 |
| **LM3** | 1/2b | 10 | 3 | 0.00E+00 | 7.00 |
| **LM3** | 1/2b | 10 | 5 | 0.00E+00 | 7.00 |
| **LM3** | 1/2b | 5 | 1 | 5.75E+06 | 0.94 |
| **LM3** | 1/2b | 5 | 3 | 0.00E+00 | 7.00 |
| **LM3** | 1/2b | 5 | 5 | 0.00E+00 | 7.00 |
| **LM4** | 4b, CC6 | 0 | 0 | 3.63E+07 | 0.00 |
| **LM4** | 4b, CC6 | 10 | 1 | 0.00E+00 | 7.00 |
| **LM4** | 4b, CC6 | 10 | 3 | 0.00E+00 | 7.00 |
| **LM4** | 4b, CC6 | 10 | 5 | 0.00E+00 | 7.00 |
| **LM4** | 4b, CC6 | 5 | 1 | 6.00E+06 | 0.78 |
| **LM4** | 4b, CC6 | 5 | 3 | 0.00E+00 | 7.00 |
| **LM4** | 4b, CC6 | 5 | 5 | 0.00E+00 | 7.00 |

## 4.5. Discussion

Inspection of institutional kitchens and restaurants plays a critical role in preventing foodborne disease outbreaks. Visual inspection is the most common method for cleanliness assurance, although inspection results vary depending on the person performing the inspection's food safety knowledge, hygiene, and diligence39. Because contamination may not be perceptible or due to location, low concentration, or interference by surface reflections, the inspecting person misses some areas of contamination (false negative) or identifies some areas as contaminated that are not (false positive). This study aims to improve cleanliness assurance in institutional kitchens and restaurants by combining new imaging technology and state-of-the-art deep learning algorithms. For accurate, reliable, and speedy inspections, automated analysis of potential contamination is needed, and we have shown that deep learning-based models can swiftly identify contamination using a portable handheld scanning technology.

Using CSI-D to identify contamination in institutional kitchens and restaurants allows fast detection and immediate response to clean contaminated areas. In addition to visual inspection, swab-based methods such as adenosine triphosphate (ATP) testing are the most frequent scientific methods used to ensure environmental cleanliness in institutional kitchens and restaurants40. CSI-D can complement swab-based methods by providing guidance on where to swab, thereby increasing swab analysis efficiency and reducing costs and production downtime waiting for cleanliness assessment. Since CSI-D is a handheld device with automated contamination detection, it can substantially decrease person-hours and costs required for sanitization inspection.

CSI-D technology can also be used to train inexperienced health inspectors (and managers) to identify and locate contamination that is less visible to the eye during visual inspections. It is extremely important for managers to know these patterns as well so that cleaning and sanitation are done in the most efficient and effective manner. Identification of contamination on surfaces with a variety of colors or background textures can be tricky, even with fluorescence-based imaging. Sometimes the contaminated area can be a tiny area on a big surface or many contaminated spots across multiple views, which can increase risk. Materials fluoresce to varying degrees, so just looking for something glowing is insufficient; an observer must consider how the fluorescence pattern differs from the background. With deep learning and an effective and reliable detection model, CSI-D provides a precise and consistent method for detecting contamination, making the inspection process faster and more effective.

Protection from contamination for inspectors and staff while addressing sanitization issues is the goal of the CSI-D's disinfection capability. E. coli, S. enterica, and L. monocytogenes are diverse species, and the strains tested in this study may not represent the range of within-species responses to UVC exposure. However, under laboratory-based conditions in this study, the UVC system is highly effective for inactivating these common foodborne pathogens. More work is needed to explore environments (matrices,

densities, mixed populations, etc.) where food production and preparation systems encounter bacteria, and the appropriate UVC intensity/time combinations required to inactivate them.

## 4.6. Conclusion

This study presents a fluorescence-imaging technology combined with deep learning algorithms to capture images, identify video frames with contamination, segment the contamination in identified video frames, and disinfect organic-based residue and biofilms. We used CSI-D, a handheld fluorescence imaging device (SafetySpect Inc), to collect data from eight institutional kitchens and restaurants. To classify "clean" and "contamination" frames, we used a state-of-the-art deep convolutional neural network architecture, Xception. Classification results for 72,381 "clean" and "contamination" frames yielded a 98.78% accuracy. To precisely segment contamination, we used a semantic segmentation algorithm, DeepLabv3+, on 12,000 "contamination" frames achieving an IoU score of 89.34%. We verified the CSI-D's UVC disinfection ability on three foodborne illness-associated pathogens, including *S. enterica*, *E. coli*, and *L. monocytogenes*. All were deactivated in less than 5 seconds after being exposed to UVC illumination from the CSI-D, achieving log reductions of up to 8.0 for active bacteria.

We have demonstrated that fluorescence-imaging technology combined with deep learning algorithms can improve the level of safety and cleanliness, protecting staff and customers of companies and institutions in the food-service industry.

# CHAPTER 5.

## FEDERATED LEARNING FOR CLIENTS' DATA PRIVACY ASSURANCE IN FOOD INDUSTRY

## 5.1. Summary

**Background:** The food service industry faces the challenge of ensuring that service facilities are clean and free of foodborne pathogens hosted by organic residues and biofilms. Such contamination can lead to foodborne diseases, putting customers at risk and compromising the reputations of service providers. New fluorescence technology empowered by state-of-the-art artificial intelligence algorithms may address this issue. However, using such advanced technologies raises concerns about data privacy and possible leakage of service providers' sensitive information. **Methods:** In this study, we employed federated learning, a decentralized privacy-preserving technology, to address client data privacy issues. By using federated learning, there is no need for data sharing across clients or data centralization on a server. We used a new fluorescence imaging technology with two deep learning models, MobileNetv3 and DeepLabv3+, to identify and segment the contaminated areas on different equipment and surfaces. We used FedML as our federated learning framework and Fedavg as the aggregation algorithm. The model was trained and validated on data from eight clients and tested on two new clients' data. **Results:** The model achieved 95.83% and 94.94% accuracies (F-scores of 96.15% and 95.61%) for classification between clean and contamination frames for the data from the two new clients and resulted in intersection over union (IoU) scores of 91.23% and 89.45% for segmentation of the contaminated areas. **Conclusions**: The results demonstrated that using federated learning combined with fluorescence imaging and deep learning algorithms can improve safety and cleanliness assurance while assuring client data privacy.

## 5.2. Background

Foodborne illness contributes significantly to morbidity and mortality as a public health issue. In 2010, contaminated food caused around 600 million cases of foodborne disease and 420 thousand fatalities worldwide [103]. According to the US Centers for Disease Control and Prevention (CDC), approximately 48 million Americans are sickened each year, 128,000 are hospitalized, and 3000 people are killed by foodborne illnesses [2, 3]. According to the CDC, restaurants and institutional kitchens are the places with the most foodborne disease-associated outbreaks [162]. Since restaurants usually handle a broad range of raw foods, the risk of cross-contamination leading to outbreaks of foodborne disease grows significantly. Cross-contamination in restaurants during the food preparation phases can be caused by inadequate hygiene and sanitization methods and contaminated equipment and surfaces.

Bacteria can adhere to food-contact surfaces, facilitated by the formation of protective film coatings from organic components and nutrients that can impede the sanitization of these surfaces [163, 164]. Contamination detection in restaurants and institutional kitchens is primarily limited to visual examination during the inspection. Another way of testing is swabbing food contact surfaces for Adenosine Triphosphate (ATP) testing or for laboratory culture [147]. In many cases, the visual inspection method becomes insufficient because the contaminated areas are invisible to the naked eye. When it comes to testing large surfaces, swabs are not the best option. They can be time-consuming and costly since they can only contact a tiny fraction of the surface, even if swabbed back and forth.

Newer developments in imaging technology and machine learning (ML), especially deep learning algorithms, have been proposed to improve food safety and surface cleanliness. The authors of [165] proposed a system for the early detection of apple bruises using hyperspectral imaging (HSI) with shortwave infrared (SWIR) illumination and a line-scan camera combined with a partial least squares discriminant analysis (PLS-DA)

classifier. In [166], real-time detection of parasites in shell-off cooked clams using transillumination imaging and binary decision trees was proposed. A method for detecting the neurotoxin acrylamide in potato chips was suggested by [167], using an SVM classifier on spatial domain statistical features extracted from images. The authors of [168] proposed foreign object detection (FOD) in meat products using a sequential deep-learning framework. In [169], a three-level classification of cleanliness in restrooms as dirty, average, and clean using deep convolutional neural networks (CNN) was proposed. In a previous research project, our group combined two deep learning algorithms (EfficientNet-B0 and U-Net) with fluorescence imaging to automatically identify and segment fecal contamination on meat carcasses [152]. In other work, we used Xception and DeepLabV3+ deep learning algorithms with multiwavelength fluorescence imaging to identify and segment contaminated areas in images of equipment and surfaces in the food-service industry [170].

ML models, generally, and deep learning algorithms specifically, need vast and diversified sets of training data to achieve accurate and reliable performance. These datasets are usually collected from multiple devices, such as cameras, sensors, cell phones, etc., at the interface or edge between the real world and the abstract world of data. The data from these "edge" devices are then brought together on a central server or storage system. The ML models use this data to train themselves and eventually predict the outcomes for new data. However, these centralized methods can be troublesome if the collected data contain sensitive information or the centralization is too costly. Centralized data collection may create privacy issues for people and liability problems for corporations if data is not handled correctly. Since people are increasingly concerned about data privacy and security, governments and organizations have begun enforcing legislation to safeguard personal information [171-173].

Since the most frequent sites of foodborne illness outbreaks are restaurants and institutional kitchens, sanitation and infection prevention in these facilities is vital. New

technologies like the SafetySpect sanitation inspection devices [112] can help restaurants and institutional kitchens keep their environment safer and improve their level of cleanliness. However, companies and institutions are concerned that using new technologies can increase privacy risks, and leaks of sensitive information (e.g., the presence of contamination in the kitchen or dining area) could jeopardize their reputation. There is a need to convince such organizations and facilities that their sensitive information and data will be safe and secure when participating in shared-data projects.

New avenues for ML research have opened up with the advent of federated learning (FL) [84]. FL allows ML algorithms to learn from data without transferring data to a centralized server. FL lets the training data remain distributed on multiple client's devices while the ML models are downloaded and the models trained on the decentralized local data. The updated model parameters are then returned to the centralized server without transferring any of the client's raw data. FL allows several clients to work together to train an ML model without sharing any private data or sensitive information. In order to train the FL model, it is necessary to have a coordinating agent that is responsible for handling the information exchange [174]. FL spreads the computational power requirements across all clients while preserving client privacy and eliminating the cost of data transfer and storage needed with centralized processing. Each client's computational power is used to analyze their own data.

The application of FL has recently gained attention in various "edge" device applications, such as in mobile devices for text prediction on virtual keyboards [175], improving the quality of search suggestions [176], and emoji suggestion based on the text typed on a keyboard [177]. Another notable application of FL is in the medical field, where data typically includes private and sensitive patient information that cannot be shared outside of an organization, limiting its use for public research. For example, in [178], the authors proposed a framework to investigate changes in the subcortical brain caused by neurological disorders. In [179], a heterogeneous FL method was proposed to train ML

models for electroencephalography data classification, and the effectiveness of FL for the identification of COVID-19 using chest X-ray images was investigated in [180]. FL has also been investigated in many other areas [181], but to our knowledge, there are no publications focusing on using FL in food safety and cleanliness assessment in the food service industry. In this study, we use a fluorescence-imaging technology combined with an FL-based deep learning model for the detection of contamination and image segmentation of the contamination in fluorescence images of a variety of food handling surfaces. Combining FL-based deep learning models with fluorescence imaging for contamination detection could improve the level of safety and cleanliness in institutional kitchens and restaurants while assuring the safety and privacy of client data.

## 5.3. Methods

### 5.3.1. Data Collection Technology

A portable automated imaging inspection system for "contamination, sanitization inspection, and disinfection" (CSI-D) has been developed (SafetySpect Inc, Grand Forks, ND). This handheld system provides mobility and flexibility for fluorescence-based detection of organic residue, bacterial biofilms, saliva, and respiratory droplets on a variety of surfaces [112]. CSI-D is able to detect, disinfect, and document the presence of contamination on food preparation surfaces that might potentially harbor pathogens or disease organisms. The illumination of the CSI-D includes 270 nm and 405 nm light-emitting diode (LED) arrays, programmed to rapidly turn on and off to allow image capture and removal of ambient light reflectance from fluorescence images of the surface. CSI-D uses two cameras for fluorescence imaging; an RGB camera to capture images of various organic residues and a UV camera to capture saliva and respiratory droplets as well as aromatic amino acids such as tryptophan and other residues.

### 5.3.2. Data Collection Technology

We collected data from seven Edgewood LTCF kitchens in North Dakota, a kitchen facility that prepares meals and snacks for multiple public schools in Grand Forks, ND, and two restaurants in Los Angeles, CA. We discussed cleaning procedures, high-touch and high-risk locations, and any perceived sanitization problems with each facility manager prior to data collection. We used CSI-D to record videos from a variety of high-risk regions, including doorknobs, garbage cans, oven and refrigerator door handles, chopping boards, and preparation tables. All videos were recorded at 1024 × 768 resolution and 24 frames per second (FPS). We analyzed one hour and 58 minutes of video.

In Figures 5.1 and 5.2, we have examples of kitchen equipment and other high-touch areas showing clean surfaces (Figure 5.1) and surfaces with contamination (Figure 5.2). Many materials both fluoresce and reflect ambient light that can be detected by the CSI-D scanner. It is important to consider both the fluorescence and the differences between irregular fluorescence on a surface and regular fluorescence on that surface. The irregular patterns of fluorescence are more consistently associated with contamination. By looking for differences between patterns on a surface and patterns of a surface, we can identify the presence of contamination.


**Figure 5.1**. Six CSI-D fluorescence images of clean surfaces.

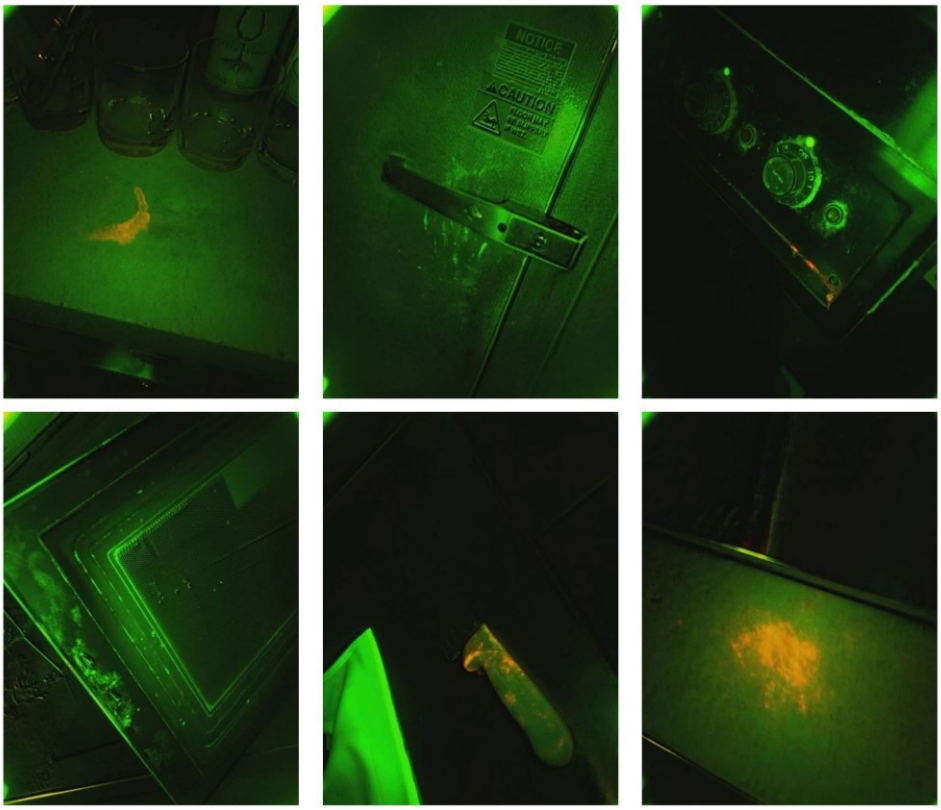


**Figure 5.2.** Six CSI-D fluorescence images of contamination on different surfaces.

### 5.3.3. Federated Learning Architecture

A standard FL system comprises a central server and multiple individual clients with their own raw datasets that are stored locally. The central server is responsible for coordinating the training by selecting a random set of clients, randomly initializing the model's global weights, broadcasting it to the clients, collecting and aggregating the updated weights received from the clients after training the model (local weights), and finally, sending the updated global model back to the clients for the next round of training. This process will be repeated until the model meets the criteria defined by the system owner. From the client viewpoint, they receive the global model broadcast by the server, and train this model for several epochs using locally stored data, then send the updated model back to the server for aggregation. Clients then wait to receive the updated global model for the next round of training. It is worth emphasizing that the clients only share the updated model weights with the server and that the server has no access to the client's private data. Figure 5.3 shows a concise illustration of an FL system. To implement an FL system, several algorithms and open-source frameworks have been proposed [94, 95, 98, 101, 182-184]. In this study, we use the Federated Averaging (FedAvg) algorithm and FedML framework to implement our FL system. MobileNetV3 is used for the classification of the clean and contamination frames, and DeepLabv3+ for the precise segmentation of contaminated areas in each video frame.

### 5.3.4. Federated Averaging (FedAvg)

FedAvg [84] is the most common and well-known method used in the implementation of FL to aggregate the local weights coming from clients and update the global model on the server. Generally, FedAvg comprises four steps. In the first step, the global model is sent synchronously from the server to a randomly chosen subset of clients. Then each client, in parallel, performs the gradient descent steps during training on its local data and then updates the model with a determined learning rate, epoch number, and batch

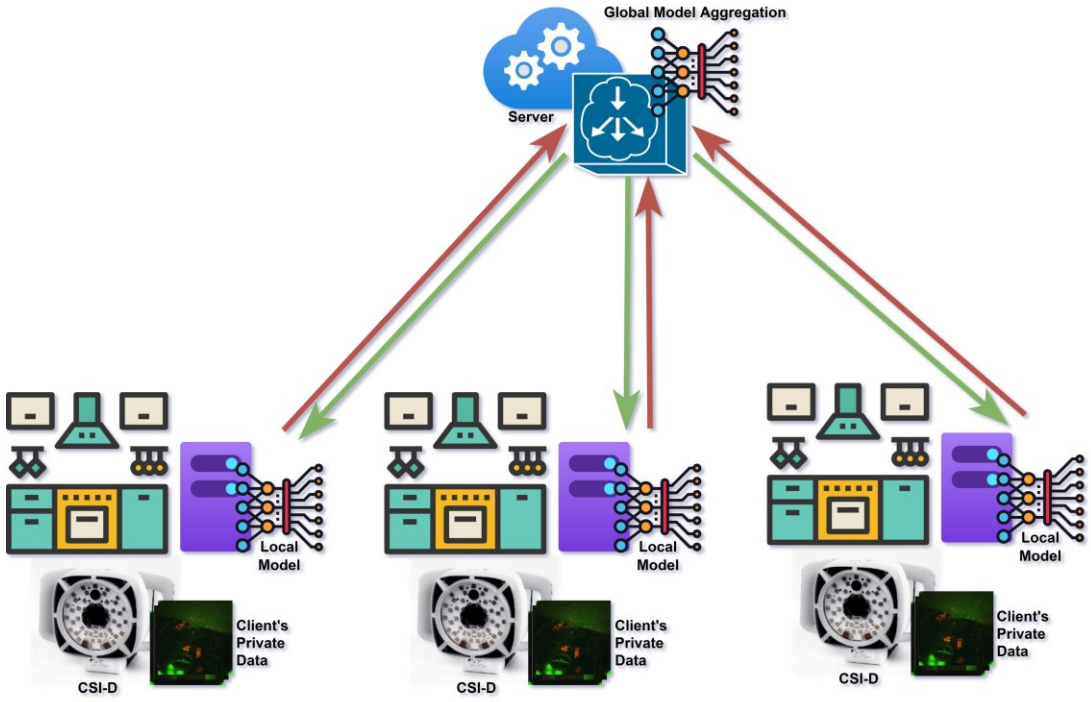


**Figure 5.3.** A concise illustration of an FL system

size, as shown in (5.1). After that, each client returns the updated model weights to the server. Finally, the server aggregates the client model weights (5.2) and updates the global model, sending it back to the clients for the next round. These four steps make one round of communication and are repeated several times until the global model converges. In each communication round, the FedAvg takes a weighted average of the weights from the local models based on the size of each client's local dataset. This means clients with more training data contribute more to the global model update.

$$w \leftarrow w - \eta \nabla \ell(w; b) \tag{5.1}$$

$$w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k \tag{5.2}$$

In (1), η is the local learning rate, w is the model weights, b is the local batch size, and $\ell()$ is the local loss function. In (5.2), K is the total number of selected clients, and k shows each client, n denotes the total number of data samples and $n_k$ is the sample size of each client. Despite the simplicity of FedAvg, this algorithm has proven to be not only robust to unbalanced data that is not “independent and identically distributed” (non-IID) but also

able to decrease the frequency and the amount of data transferred to that server that is normally required for training a deep neural network model [88].

### 5.3.5. FedML

Any problem being framed through a federated machine learning scenario will have several aspects in common. This characteristic allows developers to adopt an open-source model to expand the capabilities of machine learning. One of the frameworks that enable deployments and research using FL is FedML [94]. FedML aims to address challenges in the space of algorithmic development by offering APIs for secure aggregation, communication, and benchmarking tools.

FedML is designed with two layers of high-level and low-level APIs. Low-level APIs handle the establishment of secure communication between servers and clients. High-level APIs enable the developer to handle model type and data manipulations. FedML's main goal is to provide an adequate platform that supports algorithm and development and has standard benchmarking tools. These benchmarking tools can serve as models for setting up benchmarking with additional model types. The FedML platform currently provides FedAvg, FedOpt, FedNova, and FedNAS as FL algorithms. However, the authors have built-in mechanisms to allow outside developers to add their own optimization algorithm and run standard tests on it.

A federated learning environment has two aspects: infrastructure and training/aggregation. The infrastructure establishes secure communication across the nodes and handles the initiation of the training. The training and aggregation aspect is the part that usually requires more fine-tuning compared to the infrastructure. This is what motivated the development of FedML, where the design principle allows the developers to standardize benchmarking. A key feature of FedML is its support for a diverse set of topologies dictating how nodes can communicate with one another. The current version of FedML supports standalone simulation, distributed computation, topology customization, flexible and customizable message passing, and custom algorithm

implementation. In this study, we deployed the standalone simulation capabilities of FedML to study the training capabilities on fluorescence images of 10 clients.

### 5.3.6. Contamination Classification

To classify CSI-D recorded video frames into clean and contamination categories, we first selected video excerpts and converted them to frames that were labeled as clean and *contamination*. We then fed the labeled frames to a CNN model, MobileNetV3 [185], developed by Google, which is the next generation of the MobileNet family [125, 126, 128]. It uses a novel architecture and a mix of complementary search methods to provide high accuracy and a computationally efficient model for mobile computer vision tasks. There are two models of MobileNetV3, MobileNetV3-Small, and MobileNetV3-Large, respectively, designed for low or high resource usage needs. The architecture of these two models differs in terms of the number of blocks, expansion sizes, activation functions, etc.

#### 5.3.6.1. Classification Model Architecture

MobileNetV3 depends on AutoML to identify the best feasible architecture in a search space for the given tasks. MnasNet [125] and NetAdapt [186] are used sequentially to exploit the search space more efficiently and discover and optimize the network architecture. MnasNet first uses reinforcement learning to identify the optimum global network structure by optimizing each network block, and then the NetAdapt algorithm tries to fine-tune the architecture by adjusting the number of filters per layer. Combining these two techniques can determine the best model for a specific hardware platform. MobileNetV3 also incorporates a squeeze-and-excitation block [127] into its core architecture to improve the network representational ability (emphasizing informative features and suppressing less helpful features) via adaptive channel-wise feature recalibration by explicitly modeling between the channels' interdependencies. MobileNetV3 uses an activation function called hard-swish, a modified version of the swish [130] activation function. Swish nonlinearity uses a sigmoid function, which is not

computationally efficient. Hard-swish replaces the sigmoid with its piece-wise linear hard analog, as shown in (5.3):

$$hard-swish[x] = x\frac{ReLU6(x+3)}{6} \tag{5.3}$$

This replacement not only eliminates any numerical precision loss caused by various implementations of the approximate sigmoid but can also significantly minimize latency cost by reducing the number of memory accesses. In addition to the modifications mentioned above, some computationally expensive layers at the beginning and end of MobileNetV3 were redesigned to reduce the cost of feature generation while maintaining accuracy.

To train the MobileNetV3 for classification between clean and contamination frames, we needed to choose a suitable loss function and optimizer. Since we are dealing with a binary classification task, binary cross-entropy (BCE) was chosen as the model loss function. Comparing predicted probability to actual class labels using BCE yields a result that is either 0 or 1. It then creates a score that penalizes the probability depending on the difference between the predicted and actual values. If the predicted value is far from the actual value, the BCE loss will be increased. Equation (5.4) shows the definition of BCE.

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i \times \log\big(p(y_i) + (1-y_i) \times \log\big(1-p(y_i)\big)\big) \tag{5.4}$$

Where N shows the training sample size, $y_i$ is the assigned label, $p(y_i)$ shows the predicted probability of class 1 (*Contamination*), and $1\text{-}p(y_i)$ represents the predicted probability of class 0 (*Clean*).

After the loss function is defined, the model requires an optimization algorithm to minimize the loss by changing the model weights and learning rate. We used the Adam (Adaptive Moment Estimation) optimizer [133] as a straightforward and computationally efficient approach for first-order gradient-based stochastic optimization. Adam estimates the first- and second-order moments to calculate individual adaptive learning rates for different parameters. Adam's mathematical expressions include:

$$v_t \leftarrow \beta_1 v_{t-1} + (1-\beta_1) g_t \quad (5.5)$$

$$s_t \leftarrow \beta_2 s_{t-1} + (1-\beta_2) g_t^2 \quad (5.6)$$

$$g_t \leftarrow \nabla_\theta f_t(\theta_{t-1}) \quad (5.7)$$

$$\hat{v}_t \leftarrow \frac{v_t}{1-\beta_1^t} \ , \quad \hat{s}_t \leftarrow \frac{s_t}{1-\beta_2^t} \quad (5.8)$$

$$\theta_t \leftarrow \theta_{t-1} - \frac{\eta \hat{v}_t}{\sqrt{\hat{s}_t + \epsilon}} \quad (5.9)$$

Where $g_t$ is the gradient and step t, $v_t$ and s_t denote the exponential moving average of $g_t$ and $g_t^2$, respectively. $\beta_1$ and $\beta_2$ are smoothing parameters, f() is the loss function to minimize, θ represents the parameter (weights), η shows the learning rate, and ϵ is a small number.

## 5.3.7. Contamination Segmentation

The precise segmentation of contaminated regions on a variety of surfaces is critical since identifying contamination video frames using a classification model does not necessarily lead to detecting all contaminated areas. There might be many tiny, contaminated spots strewn across a surface, making it difficult to recognize them all during a live inspection. Restaurants and kitchens usually contain many surfaces and objects that may create background fluorescence or reflection artifacts, making it more likely that inspectors overlook some regions of contamination during an inspection. Segmenting and pseudo-coloring contaminated regions can make it easier for inspectors to identify and not miss any contamination. This is why we focused on the segmentation of contaminated regions in video frames already classified as a contamination for the second goal of this study.

### 5.3.7.1. Semantic Segmentation and Pixel-Level Annotation

Instead of relying on threshold-based techniques, which are prone to error, we employed a semantic segmentation approach to perform accurate pixel-level classification in each frame classified as contamination. To categorize every pixel of a frame into a specific class (in our case, green fluorescence, red fluorescence, and background), pixel-

level annotation is required. The annotated data is used to train a deep CNN to classify pixels, and the trained model can then be used to predict the probable class of pixels in unseen video frames (test set).

MATLAB image labeler was used for annotating when building the semantic segmentation training and testing datasets. Images can be annotated quickly and easily by sketching shapes that can be assigned region of interest (ROI) labels. In MATLAB image labeler, a rectangle, line, polygon, and projected cuboid can be used to construct a ground truth annotation for a single image or a series of images. As mentioned above, we have three different classes, and all pixels need to be annotated accordingly. Four image labelers annotated a total of 17859 frames, supervised by two experts present throughout the data collection and another expert who provided review and training remotely.

#### 5.3.7.2. Semantic Segmentation Model Architecture

To accomplish the semantic segmentation task, we used DeepLabv3+, a state-of-the-art semantic segmentation algorithm developed by Google researchers [161]. The DeepLabv3+ architecture comprises an encoder and a decoder. Multi-scale contextual information from the image is encoded, and object boundaries are precisely and accurately recovered by the decoder module. The encoder comprises three essential components: ResNet, atrous convolution, and atrous spatial pyramid pooling. In this research, we employed ResNet50 as the network backbone to extract features. Atrous convolution is a helpful technique that enables the model to directly modify the resolution of features generated by deep convolutional neural networks and alter the filter's field of view to collect multi-scale information. The mathematical expression of atrous convolution is as follows:

$$y[i] = \sum_k x[i + r.k]w[k] \tag{5.10}$$

Where r is the atrous rate (by changing it, the field of view of the filters can be modified adaptively), w depicts the convolution filter, and i and k show the pixel locations.

Atrous spatial pyramid pooling (ASPP), the other main module of DeepLabv3+, is used

to resample the features extracted from the model backbone at several rates before convolution. This is equivalent to scanning the original image with several filters, each with a complimentary effective field of view, in order to capture objects and valuable visual context at different scales. We chose the default DeepLabv3+ dilation rate of 6 since contamination might occur in a very small or very large region. The outputs of the ASPP are concatenated and pass through a 1×1 convolution layer with 256 filters that can generate rich semantic information.

A bilinear upsampling factor of four is then applied to the encoder's features before passing them to the decoder section and concatenating them with low-level features from the backbone. The low-level features from the backbone network are subjected to a 1×1 convolution layer that limits the number of channels to prevent outweighing the encoder features importance and complicating the training process. Once the low-level features and the encoder's rich features are combined, a few 3x3 convolution layers are used to improve the generated features, and finally, a bilinear upsampling by a factor of four is applied to generate the segmentation output.

### 5.3.8. Experiment Setting

To implement federated learning for the classification of clean and contamination frames, we used PyTorch v1.11.0. We resized all the images to (300, 300) for training and testing purposes. To implement FL for semantic segmentation, we used Tensorflow v2.2.0. Since semantic segmentation is computationally more expensive than classification, we resized all the images to (256, 256). The FL framework was trained and tested using eight NVIDIA Tesla V100 GPUs with 32GB RAM on a Red Hat Enterprise Linux Server 7.9 operating system.

## 5.4. Results

### 5.4.1. Federated Learning Model Classification Performance

Performance testing of classification between the clean and contamination frames

using FL was carried out on data from ten clients (facilities), including 44,185 clean frames and 46,882 contamination frames. The Dataset description is shown in Table 5.1. We used data from eight initial clients (35,858 clean frames and 36,523 contamination frames) for training and validation of the model and data from two new clients (8,327 clean frames and 10,359 contamination frames) for the final testing. Data from the eight clients were randomly assigned to training (70%) and validation (30%) sets. The FedAvg algorithm requires that in each round, a subset of clients be selected for training the model and then evaluated on all clients before starting the next round. In this study, four out of eight clients were randomly chosen for the training subset for each of the 150 rounds.

**Table 5.1.** Description of datasets.

| | **No. of clients** | **No. of "Clean" frames** | **No. of "Contamination" frames** | **Total No. of frames** |
|---|---|---|---|---|
| Training/ Validation | 1 | 708 | 3242 | 3950 |
| | 2 | 1585 | 7207 | 8792 |
| | 3 | 2740 | 1815 | 4555 |
| | 4 | 3893 | 2574 | 6467 |
| | 5 | 2679 | 2320 | 4999 |
| | 6 | 11897 | 8629 | 20526 |
| | 7 | 11041 | 6486 | 17527 |
| | 8 | 1315 | 4250 | 5565 |
| External Testing | 9 | 6354 | 7606 | 13960 |
| | 10 | 1973 | 2753 | 4726 |

In each round, clients download the global model and train it on their local dataset over 100 epochs, with a local batch size of 32, using binary cross-entropy as the model loss function. We used Adam as the model optimizer with a learning rate of $1e^{-5}$ and weight decay of $1e^{-6}$. Each client then returns the locally trained model to the server, the local models are aggregated using the FedAvg algorithm, and an updated global model is sent back to all clients. The updated global model is validated using each client's validation dataset before starting the next training round.

The FL classification model was evaluated using the six metrics of accuracy, precision, recall, specificity, F-score, and area under the curve (AUC). The first five metrics are

defined as follows:

$$\text{Accuracy} = \frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{TN}+\text{FP}+\text{FN})},\ \text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}},\ \text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{5.11}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}},\quad \text{F}_{\text{score}} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}} \tag{5.12}$$

Where TP, TN, FP, and FN show true positive, true negative, false positive, and false negative, respectively. In our case, TP shows correctly identified contamination frames, and TN shows clean frames. FP denotes the clean frame wrongly classified as contamination, and FN shows the contamination frame misclassified as clean.

We trained and validated the FL classification model for 150 rounds, and the updated global model was validated using the clients' validation dataset. Figure 5.4 shows the model accuracy and loss for each round of communication. The accuracy and loss of each round are the average accuracies and average losses of all eight clients.

After the model was trained and validated over 150 rounds, we tested it on two new clients. The FL model could identify clean and contamination frames with accuracies of 95.84% (precision of 96.88%, recall of 95.44%, specificity of 96.32%, F-score of 96.15%), and 94.92% (precision of 96.11%, recall of 95.13%, specificity of 96.63%, F-score of 95.62%) for clients 9 and 10, respectively. Figure 5.5 shows the FL model confusion matrix on the two new clients' datasets. The rows show the true label of clean and contamination frames, and the columns show the predicted labels.

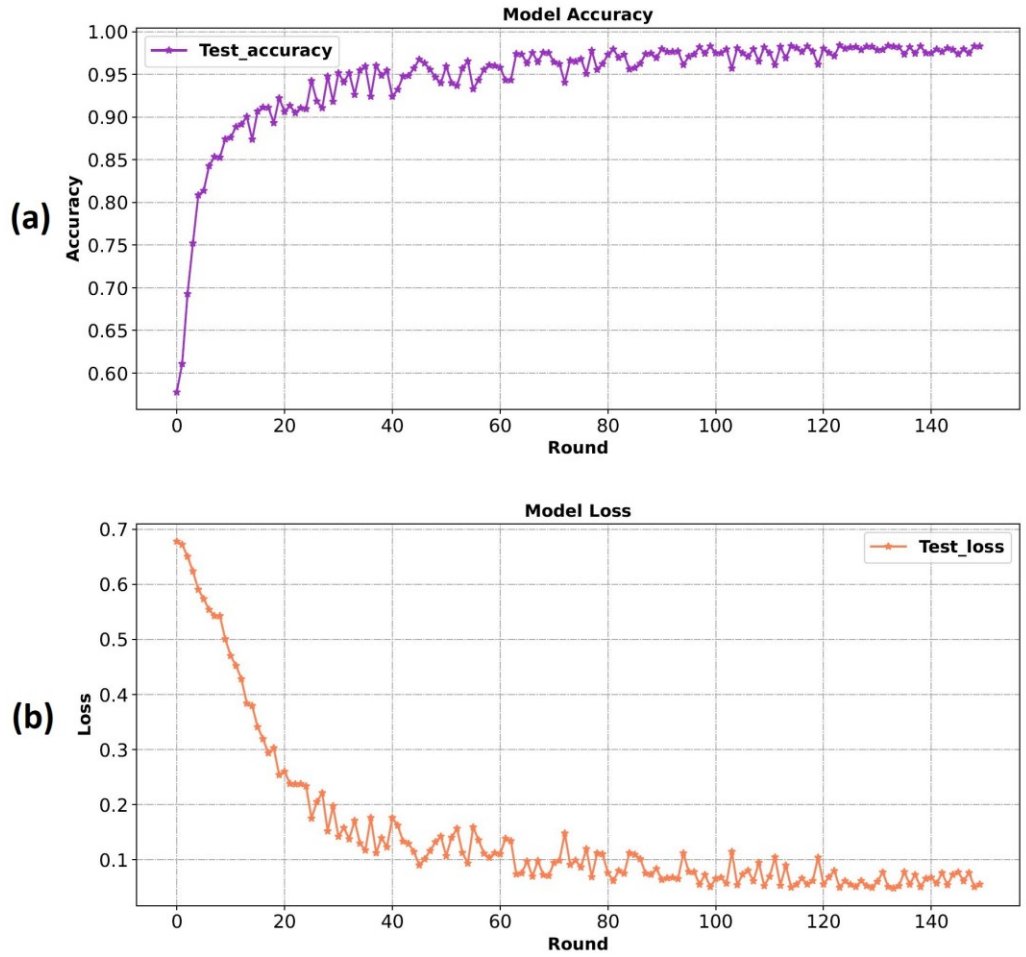


**Figure 5.4. (A)** FL model accuracy during training and validation. **(B)** FL model loss during training and validation.

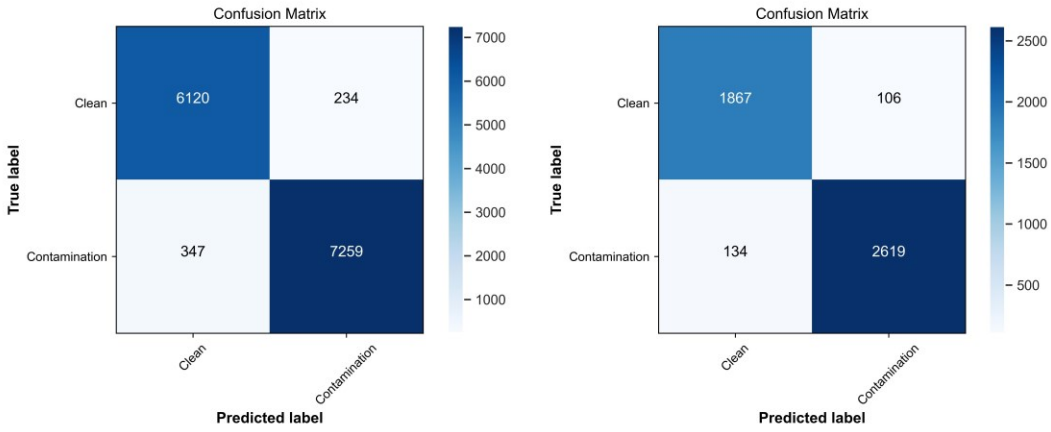


**Figure 5.5.** FL model confusion matrix when applied to two new clients' data

### 5.4.2. Federated Learning Model Segmentation Performance

We employed DeepLabv3+ to perform FL semantic segmentation to precisely identify contamination areas on various surfaces. We used 12,000 annotated contamination frames from eight clients to train and validate the model and 5859 (3770 and 2089) annotated frames from two clients for testing to evaluate how effectively the FL model could generalize when confronted with a new client's data. We randomly split the data from the eight clients to train (70%) and validate (30%) the FL model. Similar to the frame classification, we trained the model on four randomly chosen clients in each round and validated it on all eight clients after model aggregation and before moving on to the next round.

We trained and validated the FL semantic segmentation model over 150 rounds. For each round, after downloading the global model, each client trains it on its local data for over 100 epochs with a local batch size of 16. For each client, we used categorical cross-entropy as the model loss function and Adam as the model optimizer with a learning rate of $1e^{-5}$. The locally trained models are sent to the server, aggregated using the FedAvg method, and a new global model is sent back to all clients to validate the model on the validation set and then start the next round.

To evaluate the FL semantic segmentation model performance, we used five metrics: intersection over union (IoU), precision, recall, specificity, and F-score. IoU is a common metric in semantic segmentation problems that measures the overlap between the ground truth (regions annotated by a human expert) and the model prediction. The following equation shows the IoU definition.

$$\mathrm{IoU} = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FP}+\mathrm{FN})} \tag{5.13}$$

The reason for not using accuracy as a semantic segmentation evaluation metric is that contamination usually affects a tiny fraction of the frames, and accuracy would always be more than 99%, making it inappropriate to evaluate the model's performance.

The FL semantic segmentation model achieved a mean IoU of 91.23% and 89.45% for segmenting frames into the background, green-fluorescence, and red-fluorescence classes for clients 9 and 10, respectively.

In Figure 5.6, we show six image frames from a range of kitchen surfaces in the two new client datasets and the corresponding ground-truth and model outputs to show how accurately the FL semantic segmentation algorithm was able to segment the red and green fluorescence contamination. The first row is the raw image captured by CSI-D, the second row depicts the FL model segmentation results, and the last row shows the human expert annotation (ground-truth). When comparing FL model output to ground truth, it can be shown that the model successfully distinguishes and segments green and red fluorescence on various surfaces, including (Figure 5.6, left to right) under-sink plumbing, kitchen wall, kitchen countertop, a microwave door inner side, toaster, and refrigerator door handle. This comparison shows the model's ability to recognize either red fluorescence or green fluorescence contamination, or a mixture of these two types of contamination, that takes the shape of minute drops, sprays, splashes, and larger regions like spills or stains.

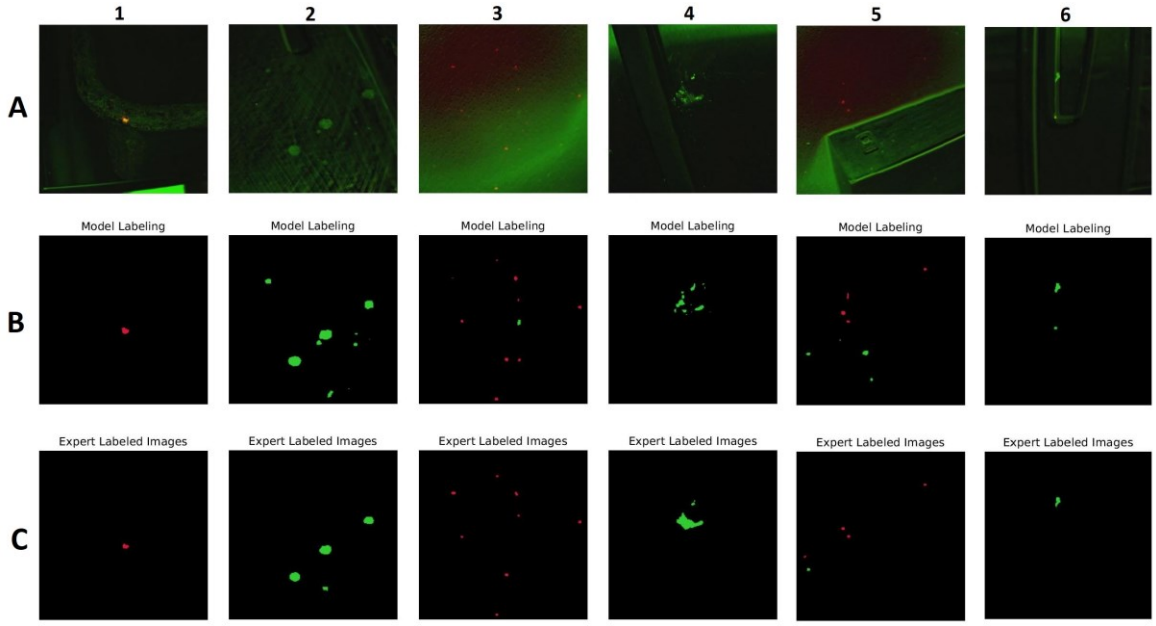


**Figure 5.6.** FL semantic segmentation model performance on new client's dataset. **(A)** Raw frames captured by CIS-D. **(B)** Segmented frames by FL model. **(C)** Annotated frames by human experts.

## 5.5. Conclusion

This study presents a federated learning model for identifying and segmenting organic-based contamination and biofilms on various surfaces. We captured video frames using a fluorescence-imaging technology developed by SafetySpect Inc and two state-of-the-art deep learning algorithms, including MobileNetV3 and DeepLabv3+, for classification and semantic segmentation. FL allows training a model without transferring client datasets to a central server, helping to address some client data privacy and confidentiality concerns. In this study, we used FedML with some modifications as our FL framework and FedAvg as the aggregation algorithm. We collected data from 10 institutional kitchens and restaurants. We trained and validated the model on eight clients, including 35,858 clean and 36,523 contamination frames, and tested it on two clients (clean: 8,327, contamination: 10,359). For differentiation between clean and contamination frames, the model achieved respective accuracies of 95.84% and 94.92% for clients 9 and 10. The FL-based semantic segmentation model was trained and validated on 12,000 annotated contamination frames from eight clients and tested on 5859 (3770 and 2089) annotated frames of two new clients. The model resulted in a mean IoU score of 91.23% for client 9 and 89.45% for client 10.

The results demonstrated that using new fluorescence imaging technology combined with FL-based deep learning models not only can improve the cleanliness and safety level of food preparation facilities but also can improve data privacy assurance for clients in the food-service industry.

# Chapter 6.

# Conclusions and Future Directions

## 6.1. Conclusions

Assurance of cleanliness and sanitation are ongoing concerns in the food industry, including producers, distributors, restaurant owners, and others. Foodborne illness outbreaks can be brought on by the spread of potential contamination and infection among workers and customers. About 48 million Americans get sick from foodborne illnesses each year, 128,000 are hospitalized, and 3000 people pass away. The US Department of Agriculture (USDA) reported that the financial burden of foodborne illness increased by about 13% to $17.6 billion in 2018 from over $15.5 billion in 2013.

There are currently only two main ways to determine whether food, equipment, or other surfaces in food-services establishments are contaminated: a quick visual inspection and adenosine triphosphate (ATP) swabbing. Visual examination is challenging because some contamination is invisible or barely visible and can be missed, and swabs can only test a small area of a surface, even when wiped back and forth.

Optical fluorescence imaging is a fast, accurate, and non-destructive way to look for organic residues and biofilms that can host pathogens. It can be used as an alternative to other methods. But different materials glow in different ways, so it's not enough to just look for something that glows. Some recent research has come up with algorithms for finding organic residues by using fluorescent imaging and extracting features based on their shape or color. Some other algorithms use thresholding, which makes it hard to find the best value for the threshold level and can lead to both false positives and false negatives because of the amount of light in the room and the different surfaces in the background change. Thus, the gap in current analysis methods necessitates more complex and trustworthy algorithms. Moreover, companies and institutions are concerned that using

new technologies can increase privacy risks, and leaks of sensitive information (e.g., the presence of contamination in the kitchen or dining area) could jeopardize their reputation. Also, there is a need to convince such organizations and facilities that their sensitive information and data will be safe and secure.

Therefore, this dissertation aims to design a framework by combining state-of-the-art deep learning algorithms, multiwavelength fluorescence imaging, and federated learning not only to improve the cleanliness assurance in food production and food services environment but also to keep the food service institutions and organizations' private data safe and secure.

In chapter 3, we have introduced a deep learning-based method for detecting contamination, which can instantly spot contaminated surfaces on meat in a fluorescence video image. We have also demonstrated that the challenge of using fluorescence to detect contamination in the bright ambient light of meat processing facilities can be overcome with the help of video image processing, new LED illumination, and imaging sensor technology. Our techniques can also be used to segment and highlight fecal residue in images of contaminated meat surfaces, making it possible to remove this contamination from carcasses in real-time. Using 108,296 images extracted from videos and labeled as "clean" and "contamination," we trained a state-of-the-art deep convolutional neural network architecture called EfficientNet-B0 on 70% of the images and validated it with 20%. Classification results were 97.32% accurate when tested on the remaining 10% of images (97.06% precise, 97.06% recall, 97.59% specificity, 97.35% F-score, and 99.54% AUC). Utilizing the U-Net semantic segmentation algorithm, we were able to segment regions of feces in images labeled as "contamination." The segmentation results from 55,114 "contamination" frames had an IoU score of 89.34% (precision of 92.95%, recall of 95.84%, specificity of 99.79%, F-score of 94.37%, and AUC of 99.89%).

In chapter 4, in order to eliminate organic residue and biofilms, this research introduces a fluorescence-imaging technology coupled with deep learning algorithms for image capture, video frame contamination detection, segmentation, and disinfection. To do this, we utilized a portable fluorescence imaging device called CSI-D (SafetySpect Inc) in a total of eight commercial kitchens and eating establishments. We employed a deep convolutional neural network architecture, Xception, to classify images into "*clean*" and "*contamination*" categories. For a total of 72,381 frames, both "*clean*" and "*contamination*," the accuracy rate was 98.78%. We used the DeepLabv3+ semantic segmentation algorithm on 12,000 "*contamination*" frames, achieving an IoU score of 89.34%, to accurately segment contamination. We tested the CSI-D's capacity for UVC disinfection using three pathogens linked to food foodborne illness: Salmonella enterica, Escherichia coli, and Listeria monocytogenes. Less than 5 seconds of exposure to UVC illumination from the CSI-D was enough to inactivate all of them, with log reductions of up to 8.0 for active bacteria. We have shown that integrating fluorescence-imaging technology with deep learning algorithms can increase safety and cleanliness in the food-service industry, safeguarding both employees and patrons.

In chapter 5, to identify and classify organic-based contamination and biofilms on different surfaces while preserving the client's data privacy, we introduced a federated learning model. In order to classify and semantically segment the contamination, we used a fluorescence-imaging technology created by SafetySpect Inc and two cutting-edge deep learning algorithms, namely MobileNetV3 and DeepLabv3+. Since federated learning enables model training without transferring client datasets to a central server, it can be used to alleviate some concerns over the security and privacy of client data. For this research, we made some adjustments to the FedML framework and used FedAvg as our aggregation algorithm. We collected data from ten different institutional kitchens and restaurants. Eight clients, including 35,858 clean and 36,523 contamination frames, were

used for training and model validation, and two clients were used for testing (clean: 8,327, contamination: 10,359). For clients 9 and 10, the model's accuracy for differentiating between clean and contaminated frames was 95.84% and 94.92%, respectively. The FL-based semantic segmentation model was tested on 5859 (3770 and 2089) annotated frames from two new clients after being trained and validated on 12,000 annotated contamination frames from eight clients. For clients 9 and 10, the model produced mean IoU scores of 91.23% and 89.45%, respectively. The findings showed that combining new fluorescence imaging technology with FL-based deep learning models can enhance data privacy assurance for customers in the food-service industry while increasing the cleanliness and safety level of food preparation facilities.

## 6.2. Future Direction

### 6.2.1. Database Expansion

While this dissertation resulted in accurate identification and segmentation of contamination on a wide variety of surfaces, there is a need for collecting data from more facilities to improve the models' generalization and reliability both in centralized and federated architectures. By collecting more data under different ambient light intensities and scanning different surfaces and objects, we will allow the models to learn new contamination patterns. So we aim to visit more meat processing facilities, institutional kitchens, and restaurants to increase of database for future training and testing of the models.

### 6.2.2. Solving Data Annotation Bottleneck

Supervised learning is the backbone of the current projects. Although supervised learning is a well-known method that can lead to building high-performance models, it needs massive amounts of manually labeled data which slows down model construction, is costly, and is prone to errors. Self-supervised learning (SSL) is an emerging method to

address the data labeling challenge. The SSL relies on the underlying structure of the data to provide supervisory information and train itself instead of learning from labeled data. Since we record videos during the inspection of the different facilities and with respect to the fact that we aim to expand our database, a massive amount of data will be generated, making the data annotation cumbersome and expensive. Hence, we aim to use SSL as our future step to tackle the labeling bottleneck.

## References


[1] C. W. Hedberg, "Explaining the risk of foodborne illness associated with restaurants: The environmental health specialists network (ehs-net)," *J. Food Prot,* vol. 76, no. 12, p. 2124, 2013.

[2] R. L. Scharff, J. Besser, D. J. Sharp, T. F. Jones, G.-S. Peter, and C. W. Hedberg, "An economic evaluation of PulseNet: a network for foodborne disease surveillance," *American journal of preventive medicine,* vol. 50, no. 5, pp. S66-S73, 2016.

[3] E. Scallan *et al.*, "Foodborne illness acquired in the United States—major pathogens," *Emerging infectious diseases,* vol. 17, no. 1, p. 7, 2011.

[4] S. Sutton, "The Contamination Control Plan in Facility Validation," *Journal of Validation technology,* vol. 18, no. 2, 2012.

[5] M. L. Ling, A. Apisarnthanarak, L. T. A. Thu, V. Villanueva, C. Pandjaitan, and M. Y. Yusof, "APSIC Guidelines for environmental cleaning and decontamination," *Antimicrobial resistance and infection control,* vol. 4, no. 1, pp. 1-9, 2015.

[6] B. G. Mitchell, F. Wilson, S. J. Dancer, and A. McGregor, "Methods to evaluate environmental cleanliness in healthcare facilities," *Healthcare infection,* vol. 18, no. 1, pp. 23-30, 2013.

[7] M. Oh, H. Lee, H. Cho, S.-H. Moon, E.-K. Kim, and M. S. Kim, "Detection of fecal contamination on beef meat surfaces using handheld fluorescence imaging device (HFID)," in *Sensing for Agriculture and Food Quality and Safety VIII*, 2016, vol. 9864: International Society for Optics and Photonics, p. 986411.

[8] B. Park, K. C. Lawrence, W. R. Windham, and D. P. Smith, "Performance of hyperspectral imaging system for poultry surface fecal contaminant detection," *Journal of Food Engineering,* vol. 75, no. 3, pp. 340-348, 2006.

[9] H. Cho, M. S. Kim, S. Kim, H. Lee, M. Oh, and S. H. Chung, "Hyperspectral determination of fluorescence wavebands for multispectral imaging detection of multiple animal fecal species contaminations on romaine lettuce," *Food and bioprocess technology,* vol. 11, no. 4, pp. 774-784, 2018.

[10] M. S. Kim, A. M. Lefcourt, and Y.-R. Chen, "Optimal fluorescence excitation and emission bands for detection of fecal contamination," *Journal of food protection,* vol. 66, no. 7, pp. 1198-1207, 2003.

[11] M. S. Kim, A. M. Lefcourt, and Y.-R. Chen, "Multispectral laser-induced fluorescence imaging system for large biological samples," *Applied optics,* vol. 42, no. 19, pp. 3927-3934, 2003.

[12] E. W. Chappelle, F. M. Wood, J. E. McMurtrey, and W. W. Newcomb, "Laser-induced fluorescence of green plants. 1: A technique for the remote detection of plant stress and species differentiation," *Applied Optics,* vol. 23, no. 1, pp. 134-138, 1984.

[13] M. S. Kim, J. E. McMurtrey, C. L. Mulchi, C. S. Daughtry, E. W. Chappelle, and Y.-R. Chen, "Steady-state multispectral fluorescence imaging system for plant leaves," *Applied Optics,* vol. 40, no. 1, pp. 157-166, 2001.

[14] L. A. Corp, J. E. McMurtrey, E. M. Middleton, C. L. Mulchi, E. W. Chappelle, and C. S. Daughtry, "Fluorescence sensing systems: In vivo detection of biophysical variations in field corn due to nitrogen supply," *Remote sensing of environment,* vol. 86, no. 4, pp. 470-479, 2003.

[15] M. Lang, F. Stober, and H. Lichtenthaler, "Fluorescence emission spectra of plant leaves and plant constituents," *Radiation and environmental biophysics,* vol. 30, no. 4, pp. 333-347, 1991.

[16] H. Swatland and S. Barbut, "Fluorimetry via a quartz-glass rod for predicting the skin content and processing characteristics of poultry meat slurry," *International journal of food science & technology,* vol. 26, no. 4, pp. 373-380, 1991.

[17] J. Wold and K. Kvaal, "Mapping lipid oxidation in chicken meat by multispectral imaging of autofluorescence," *Applied spectroscopy,* vol. 54, no. 6, pp. 900-909, 2000.
[18] J. Wold, F. Lundby, and B. Egelandsdal, "Quantification of connective tissue (hydroxyproline) in ground beef by autofluorescence spectroscopy," *Journal of Food Science,* vol. 64, no. 3, pp. 377-383, 1999.
[19] B. Park, K. C. Lawrence, W. R. Windham, and D. P. Smith, "Detection of cecal contaminants in visceral cavity of broiler carcasses using hyperspectral imaging," *Applied Engineering in Agriculture,* vol. 21, no. 4, pp. 627-635, 2005.
[20] B. Park, K. C. Lawrence, W. R. Windham, and R. J. Buhr, "Hyperspectral imaging for detecting fecal and ingesta contamination on poultry carcasses," in *2001 ASAE Annual Meeting*, 1998: American Society of Agricultural and Biological Engineers, p. 1.
[21] M. Kim and A. Lefcourt, "Online hyperspectral line-scan fluorescence imaging for safety inspection of apples," in *XXVII International Horticultural Congress-IHC2006: International Symposium on The Role of Postharvest Technology in the 768*, 2006, pp. 385-390.
[22] B.-K. Cho, Y.-R. Chen, and M. S. Kim, "Multispectral detection of organic residues on poultry processing plant equipment based on hyperspectral reflectance imaging technique," *Computers and Electronics in Agriculture,* vol. 57, no. 2, pp. 177-189, 2007.
[23] A. M. Lefcourt, M. S. Wiederoder, N. T. Liu, M. S. Kim, and Y. M. Lo, "Development of a portable hyperspectral imaging system for monitoring the efficacy of sanitation procedures in food processing facilities," *Journal of Food Engineering,* vol. 117, no. 1, pp. 59-66, 2013.
[24] M. S. Wiederoder, N. T. Liu, A. M. Lefcourt, M. S. Kim, and Y. M. Lo, "Use of a portable hyperspectral imaging system for monitoring the efficacy of sanitation procedures in produce processing plants," *Journal of Food Engineering,* vol. 117, no. 2, pp. 217-226, 2013.
[25] "Headwall Photonics." https://www.headwallphotonics.com/ (accessed 5 July, 2022).
[26] "P&P Optica." https://ppo.ca/ (accessed 5 July, 2022).
[27] "Vertide." https://www.veritide.com/ (accessed 5 July, 2022).
[28] H. Saleh, *The The Deep Learning with PyTorch Workshop: Build deep neural networks and artificial intelligence applications with PyTorch*. Packt Publishing Ltd, 2020.
[29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature,* vol. 521, no. 7553, pp. 436-444, 2015.
[30] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience,* vol. 2018, 2018.
[31] D. W. Otter, J. R. Medina, and J. K. Kalita, "A survey of the usages of deep learning for natural language processing," *IEEE transactions on neural networks and learning systems,* vol. 32, no. 2, pp. 604-624, 2020.
[32] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technologies and Interaction,* vol. 2, no. 3, p. 47, 2018.
[33] T. Ching *et al.*, "Opportunities and obstacles for deep learning in biology and medicine," *Journal of The Royal Society Interface,* vol. 15, no. 141, p. 20170387, 2018.
[34] Q. Rao and J. Frtunikj, "Deep learning for self-driving cars: Chances and challenges," in *Proceedings of the 1st International Workshop on Software Engineering for AI in Autonomous Systems*, 2018, pp. 35-38.
[35] A. Da'u and N. Salim, "Recommendation system based on deep learning methods: a systematic review and new directions," *Artificial Intelligence Review,* vol. 53, no. 4,

pp. 2709-2748, 2020.
[36] H. A. Pierson and M. S. Gashler, "Deep learning in robotics: a review of recent research," *Advanced Robotics,* vol. 31, no. 16, pp. 821-835, 2017.
[37] N. Justesen, P. Bontrager, J. Togelius, and S. Risi, "Deep learning for video game playing," *IEEE Transactions on Games,* vol. 12, no. 1, pp. 1-20, 2019.
[38] L. Zhou, C. Zhang, F. Liu, Z. Qiu, and Y. He, "Application of deep learning in food: a review," *Comprehensive reviews in food science and food safety,* vol. 18, no. 6, pp. 1793-1811, 2019.
[39] J. D. Kelleher, *Deep learning*. MIT press, 2019.
[40] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electronic Markets,* vol. 31, no. 3, pp. 685-695, 2021.
[41] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458,* 2015.
[42] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern recognition,* vol. 77, pp. 354-377, 2018.
[43] Y. LeCun *et al.*, "Handwritten digit recognition with a back-propagation network," *Advances in neural information processing systems,* vol. 2, 1989.
[44] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *The Journal of physiology,* vol. 195, no. 1, pp. 215-243, 1968.
[45] A. Araujo, W. Norris, and J. Sim, "Computing receptive fields of convolutional neural networks," *Distill,* vol. 4, no. 11, p. e21, 2019.
[46] S. Skansi, *Introduction to Deep Learning: from logical calculus to artificial intelligence*. Springer, 2018.
[47] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *towards data science,* vol. 6, no. 12, pp. 310-316, 2017.
[48] D. Yu, H. Wang, P. Chen, and Z. Wei, "Mixed pooling for convolutional neural networks," in *International conference on rough sets and knowledge technology*, 2014: Springer, pp. 364-375.
[49] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems,* vol. 25, 2012.
[50] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
[51] M. Sewak, M. R. Karim, and P. Pujari, *Practical convolutional neural networks: implement advanced deep learning models using Python*. Packt Publishing Ltd, 2018.
[52] M. Thoma, "A survey of semantic segmentation," *arXiv preprint arXiv:1602.06541,* 2016.
[53] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," *International journal of multimedia information retrieval,* vol. 7, no. 2, pp. 87-93, 2018.
[54] D. Weinland, R. Ronfard, and E. Boyer, "A survey of vision-based methods for action representation, segmentation and recognition," *Computer vision and image understanding,* vol. 115, no. 2, pp. 224-241, 2011.
[55] M. Rezaei *et al.*, "A conditional adversarial network for semantic segmentation of brain tumor," in *International MICCAI Brainlesion Workshop*, 2017: Springer, pp. 241-252.
[56] D. Feng *et al.*, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Transactions on Intelligent Transportation Systems,* vol. 22, no. 3, pp. 1341-1360, 2020.
[57] M. Wurm, T. Stark, X. X. Zhu, M. Weigand, and H. Taubenböck, "Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks," *ISPRS journal of photogrammetry and remote sensing,* vol. 150, pp. 59-69, 2019.

[58] A. Ouahabi and A. Taleb-Ahmed, "Deep learning for real-time semantic segmentation: Application in ultrasound imaging," *Pattern Recognition Letters,* vol. 144, pp. 27-34, 2021.
[59] A. Milioto, P. Lottes, and C. Stachniss, "Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs," in *2018 IEEE international conference on robotics and automation (ICRA)*, 2018: IEEE, pp. 2229-2235.
[60] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," *arXiv preprint arXiv:1704.06857,* 2017.
[61] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision,* vol. 60, no. 2, pp. 91-110, 2004.
[62] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, 2005, vol. 1: Ieee, pp. 886-893.
[63] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
[64] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.
[65] F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing,* vol. 338, pp. 321-348, 2019.
[66] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.
[67] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*, 2014: Springer, pp. 818-833.
[68] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence,* vol. 39, no. 12, pp. 2481-2495, 2017.
[69] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Systems with Applications,* vol. 169, p. 114417, 2021.
[70] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," *Applied Soft Computing,* vol. 70, pp. 41-65, 2018.
[71] Z.-Q. Zhao, H. Glotin, Z. Xie, J. Gao, and X. Wu, "Cooperative sparse representation in two opposite directions for semi-supervised image annotation," *IEEE Transactions on Image Processing,* vol. 21, no. 9, pp. 4218-4231, 2012.
[72] Y. Wei *et al.*, "Stc: A simple to complex framework for weakly-supervised semantic segmentation," *IEEE transactions on pattern analysis and machine intelligence,* vol. 39, no. 11, pp. 2314-2320, 2016.
[73] D. Zhang, H. Zhang, J. Tang, X.-S. Hua, and Q. Sun, "Causal intervention for weakly-supervised semantic segmentation," *Advances in Neural Information Processing Systems,* vol. 33, pp. 655-666, 2020.
[74] Y. Mo, Y. Wu, X. Yang, F. Liu, and Y. Liao, "Review the state-of-the-art technologies of semantic segmentation based on deep learning," *Neurocomputing,* vol. 493, pp. 626-646, 2022.
[75] S. Hao, Y. Zhou, and Y. Guo, "A brief survey on semantic segmentation with deep learning," *Neurocomputing,* vol. 406, pp. 302-321, 2020.
[76] M. Abadi *et al.*, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308-318.

[77] L. Lyu, H. Yu, and Q. Yang, "Threats to federated learning: A survey," *arXiv preprint arXiv:2003.02133,* 2020.
[78] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowledge-Based Systems,* vol. 216, p. 106775, 2021.
[79] J. P. Albrecht, "How the GDPR will change the world," *Eur. Data Prot. L. Rev.,* vol. 2, p. 287, 2016.
[80] E. L. Harding, J. J. Vanto, R. Clark, L. Hannah Ji, and S. C. Ainsworth, "Understanding the scope and impact of the California Consumer Privacy Act of 2018," *Journal of Data Protection & Privacy,* vol. 2, no. 3, pp. 234-253, 2019.
[81] M. Parasol, "The impact of China's 2016 Cyber Security Law on foreign technology firms, and on China's big data and Smart City dreams," *Computer law & security review,* vol. 34, no. 1, pp. 67-98, 2018.
[82] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST),* vol. 10, no. 2, pp. 1-19, 2019.
[83] H. Li, K. Ota, and M. Dong, "Learning IoT in edge: Deep learning for the Internet of Things with edge computing," *IEEE network,* vol. 32, no. 1, pp. 96-101, 2018.
[84] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*, 2017: PMLR, pp. 1273-1282.
[85] P. Kairouz *et al.*, "Advances and open problems in federated learning," *Foundations and Trends® in Machine Learning,* vol. 14, no. 1–2, pp. 1-210, 2021.
[86] Y. Zhuang, G. Li, and J. Feng, "A survey on entity alignment of knowledge base," *Journal of Computer Research and Development,* vol. 53, no. 1, pp. 165-192, 2016.
[87] P. Christen, "The data matching process," in *Data matching*: Springer, 2012, pp. 23-35.
[88] S. Reddi *et al.*, "Adaptive federated optimization," *arXiv preprint arXiv:2003.00295,* 2020.
[89] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of fedavg on non-iid data," *arXiv preprint arXiv:1907.02189,* 2019.
[90] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine Learning and Systems,* vol. 2, pp. 429-450, 2020.
[91] A. Reisizadeh, A. Mokhtari, H. Hassani, A. Jadbabaie, and R. Pedarsani, "Fedpaq: A communication-efficient federated learning method with periodic averaging and quantization," in *International Conference on Artificial Intelligence and Statistics*, 2020: PMLR, pp. 2021-2031.
[92] X. Li, M. Jiang, X. Zhang, M. Kamp, and Q. Dou, "Fedbn: Federated learning on non-iid features via local batch normalization," *arXiv preprint arXiv:2102.07623,* 2021.
[93] J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, "Tackling the objective inconsistency problem in heterogeneous federated optimization," *Advances in neural information processing systems,* vol. 33, pp. 7611-7623, 2020.
[94] C. He *et al.*, "Fedml: A research library and benchmark for federated machine learning," *arXiv preprint arXiv:2007.13518,* 2020.
[95] Y. Liu, T. Fan, T. Chen, Q. Xu, and Q. Yang, "FATE: An Industrial Grade Platform for Collaborative Learning With Data Protection," *J. Mach. Learn. Res.,* vol. 22, no. 226, pp. 1-6, 2021.
[96] Google. "TensorFlow Federated: Machine Learning on Decentralized Data " https://www.tensorflow.org/federated (accessed July 14, 2022).
[97] T. Ryffel *et al.*, "A generic framework for privacy preserving deep learning," *arXiv preprint arXiv:1811.04017,* 2018.
[98] A. Ziller *et al.*, "Pysyft: A library for easy federated learning," in *Federated Learning Systems*: Springer, 2021, pp. 111-139.

[99] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems,* vol. 32, 2019.
[100] D. Skvorc, M. Horvat, and S. Srbljic, "Performance evaluation of Websocket protocol for implementation of full-duplex web streams," in *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2014: IEEE, pp. 1003-1008.
[101] D. J. Beutel *et al.*, "Flower: A friendly federated learning research framework," *arXiv preprint arXiv:2007.14390,* 2020.
[102] F. Fung, H.-S. Wang, and S. Menon, "Food safety in the 21st century," *Biomedical journal,* vol. 41, no. 2, pp. 88-95, 2018.
[103] W. H. Organization, *WHO estimates of the global burden of foodborne diseases: foodborne disease burden epidemiology reference group 2007-2015*. World Health Organization, 2015.
[104] R. L. Scharff, "Economic burden from health losses due to foodborne illness in the United States," *Journal of food protection,* vol. 75, no. 1, pp. 123-131, 2012.
[105] E. Scallan, P. M. Griffin, F. J. Angulo, R. V. Tauxe, and R. M. Hoekstra, "Foodborne illness acquired in the United States—unspecified agents," *Emerging infectious diseases,* vol. 17, no. 1, p. 16, 2011.
[106] S. A. Hoffmann, B. Maculloch, and M. Batz, "Economic burden of major foodborne illnesses acquired in the United States," ed: United States Department of Agriculture, Economic Research Service, 2015.
[107] USDA. "Cost Estimates of Foodborne Illnesses." https://www.ers.usda.gov/data-products/cost-estimates-of-foodborne-illnesses/ (accessed July 14, 2022).
[108] C. f. D. C. a. P. (CDC). "Foods That Can Cause Food Poisoning." https://www.cdc.gov/foodsafety/foods-linked-illness.html (accessed July 14, 2022).
[109] NAMI. "The United States Meat Industry at a Glance." https://www.meatinstitute.org/index.php?ht=d/sp/i/47465/pid/47465 (accessed July 14, 2022).
[110] M. Doyle, *Foodborne bacterial pathogens*. CRC Press, 1989.
[111] J. Rasekh, A. Thaler, D. Englejohn, and N. Pihkala, "Food safety and inspection service policy for control of poultry contaminated by digestive tract contents: a review," *Journal of applied poultry research,* vol. 14, no. 3, pp. 603-611, 2005.
[112] M. Sueker *et al.*, "Handheld multispectral fluorescence imaging system to detect and disinfect surface contamination," *Sensors,* vol. 21, no. 21, p. 7222, 2021.
[113] T.-C. Chen and S.-Y. Yu, "The review of food safety inspection system based on artificial intelligence, image processing, and robotic," *Food Science and Technology,* vol. 42, 2021.
[114] Y. Shi *et al.*, "A review on meat quality evaluation methods based on non-destructive computer vision and artificial intelligence technologies," *Food science of animal resources,* vol. 41, no. 4, p. 563, 2021.
[115] P. Zapotoczny, P. M. Szczypiński, and T. Daszkiewicz, "Evaluation of the quality of cold meats by computer-assisted image analysis," *LWT-Food Science and Technology,* vol. 67, pp. 37-49, 2016.
[116] K. Shiranita, K. Hayashi, A. Otsubo, T. Miyajima, and R. Takiyama, "Grading meat quality by image processing," *Pattern Recognition,* vol. 33, no. 1, pp. 97-104, 2000.
[117] B. Park, S.-C. Yoon, W. R. Windham, K. C. Lawrence, M. S. Kim, and K. Chao, "Line-scan hyperspectral imaging for real-time in-line poultry fecal detection," *Sensing and instrumentation for food quality and safety,* vol. 5, no. 1, pp. 25-32, 2011.
[118] Y. Seo *et al.*, "Multispectral fluorescence imaging technique for on-line inspection of fecal residues on poultry carcasses," *Sensors,* vol. 19, no. 16, p. 3483, 2019.
[119] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems,* 2021.

[120] Y. Li, Z. Hao, and H. Lei, "Survey of convolutional neural network," *Journal of Computer Applications,* vol. 36, no. 9, p. 2508, 2016.
[121] L. Zhu, P. Spachos, E. Pensini, and K. N. Plataniotis, "Deep learning and machine vision for food processing: A survey," *Current Research in Food Science,* vol. 4, pp. 233-249, 2021.
[122] R. Ribani and M. Marengoni, "A survey of transfer learning for convolutional neural networks," in *2019 32nd SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T)*, 2019: IEEE, pp. 47-57.
[123] J. Nayak, K. Vakula, P. Dinesh, B. Naik, and D. Pelusi, "Intelligent food processing: Journey from artificial neural network to deep learning," *Computer Science Review,* vol. 38, p. 100297, 2020.
[124] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, 2019: PMLR, pp. 6105-6114.
[125] M. Tan *et al.*, "Mnasnet: Platform-aware neural architecture search for mobile," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2820-2828.
[126] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510-4520.
[127] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132-7141.
[128] A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861,* 2017.
[129] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, 2015: PMLR, pp. 448-456.
[130] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *arXiv preprint arXiv:1710.05941,* 2017.
[131] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?," *Advances in neural information processing systems,* vol. 31, 2018.
[132] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400,* 2013.
[133] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.
[134] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015: Springer, pp. 234-241.
[135] S. M. Bartsch, L. Asti, S. Nyathi, M. L. Spiker, and B. Y. Lee, "Estimated cost to a restaurant of a foodborne illness outbreak," *Public Health Reports,* vol. 133, no. 3, pp. 274-286, 2018.
[136] E. Giaouris *et al.*, "Attachment and biofilm formation by foodborne bacteria in meat processing environments: causes, implications, role of bacterial interactions and control by alternative novel methods," *Meat science,* vol. 97, no. 3, pp. 298-309, 2014.
[137] L. A. Jackson *et al.*, "Where's the Beef?: The Role of Cross-contamination in 4 Chain Restaurant–Associated Outbreaks of Escherichia coli O157: H7 in the Pacific Northwest," *Archives of Internal Medicine,* vol. 160, no. 15, pp. 2380-2385, 2000.
[138] J. Weinstein, "The clean restaurant. II: Employee hygiene," *Restaurants & institutions,* vol. 101, no. 13, pp. 138-9, 142, 144 passim, 1991.
[139] M. J. Firestone, S. Rajamani, and C. W. Hedberg, "A Public Health Informatics Solution to Improving Food Safety in Restaurants: Putting the Missing Piece in the Puzzle," *Online Journal of Public Health Informatics,* vol. 13, no. 1, 2021.

[140] "Brand Finance Food Safety 2018." https://brandirectory.com/reports/food-safety-2018 (accessed July 19, 2022).
[141] S. Seo, S. S. Jang, L. Miao, B. Almanza, and C. Behnke, "The impact of food safety events on the value of food-related firms: An event study approach," *International Journal of Hospitality Management,* vol. 33, pp. 153-165, 2013.
[142] L. Hobley, C. Harkins, C. E. MacPhee, and N. R. Stanley-Wall, "Giving structure to the biofilm matrix: an overview of individual strategies and emerging common themes," *FEMS microbiology reviews,* vol. 39, no. 5, pp. 649-669, 2015.
[143] D. Lahiri *et al.*, "Amylases: biofilm inducer or biofilm inhibitor?," *Frontiers in Cellular and Infection Microbiology,* vol. 11, p. 660048, 2021.
[144] K. P. Rumbaugh and K. Sauer, "Biofilm dispersion," *Nature Reviews Microbiology,* vol. 18, no. 10, pp. 571-586, 2020.
[145] N. Trachoo, "Biofilms and the food industry," *Biofilms,* vol. 25, no. 6, p. 808, 2003.
[146] U. S. F. D. Administration. "Retail Food Protection." https://www.fda.gov/food/guidance-regulation-food-and-dietary-supplements/retail-food-protection (accessed July 19, 2022).
[147] J. Verran, J. Redfern, L. Smith, and K. Whitehead, "A critical evaluation of sampling methods used for assessing microorganisms on surfaces," *Food and Bioproducts processing,* vol. 88, no. 4, pp. 335-340, 2010.
[148] K. Rezaee, S. Savarkar, X. Yu, and J. Zhang, "A hybrid deep transfer learning-based approach for Parkinson's disease classification in surface electromyography signals," *Biomedical Signal Processing and Control,* vol. 71, p. 103161, 2022.
[149] H. Taheri Gorji and N. Kaabouch, "A deep learning approach for diagnosis of mild cognitive impairment based on MRI images," *Brain sciences,* vol. 9, no. 9, p. 217, 2019.
[150] R. Zuo, Y. Xiong, J. Wang, and E. J. M. Carranza, "Deep learning and its application in geochemical mapping," *Earth-science reviews,* vol. 192, pp. 1-14, 2019.
[151] Z. Gao, Z. Luo, W. Zhang, Z. Lv, and Y. Xu, "Deep learning application in plant stress imaging: a review," *AgriEngineering,* vol. 2, no. 3, p. 29, 2020.
[152] H. T. Gorji *et al.*, "Combining deep learning and fluorescence imaging to automatically identify fecal contamination on meat carcasses," *Scientific Reports,* vol. 12, no. 1, pp. 1-11, 2022.
[153] P. Pouladzadeh, P. Kuhad, S. V. B. Peddi, A. Yassine, and S. Shirmohammadi, "Food calorie measurement using deep learning neural network," in *2016 IEEE international instrumentation and measurement technology conference proceedings*, 2016: IEEE, pp. 1-6.
[154] H. Chen, Z. Chen, F. Lin, and P. Zhuang, "Effective management for blockchain-based agri-food supply chains using deep reinforcement learning," *IEeE Access,* vol. 9, pp. 36008-36018, 2021.
[155] E. Sikorska, I. Khmelinskii, and M. Sikorski, "Fluorescence spectroscopy and imaging instruments for food quality evaluation," in *Evaluation technologies for food quality*: Elsevier, 2019, pp. 491-533.
[156] B. Valeur and M. N. Berberan-Santos, *Molecular fluorescence: principles and applications*. John Wiley & Sons, 2012.
[157] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251-1258.
[158] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818-2826.
[159] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* vol. 6, no. 02, pp. 107-116, 1998.

[160] A. F. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint arXiv:1803.08375,* 2018.
[161] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801-818.
[162] D. Dewey-Mattia, K. Manikonda, A. J. Hall, M. E. Wise, and S. J. Crowe, "Surveillance for foodborne disease outbreaks—United States, 2009–2015," *MMWR Surveillance Summaries,* vol. 67, no. 10, p. 1, 2018.
[163] S. Abban, M. Jakobsen, and L. Jespersen, "Attachment behaviour of Escherichia coli K12 and Salmonella Typhimurium P6 on food contact surfaces for food transportation," *Food microbiology,* vol. 31, no. 2, pp. 139-147, 2012.
[164] Y. Quan, H.-Y. Kim, and I.-S. Shin, "Bactericidal activity of strong acidic hypochlorous water against Escherichia coli O157: H7 and Listeria monocytogenes in biofilms attached to stainless steel," *Food science and biotechnology,* vol. 26, no. 3, pp. 841-846, 2017.
[165] J. C. Keresztes, M. Goodarzi, and W. Saeys, "Real-time pixel based early apple bruise detection using short wave infrared hyperspectral imaging in combination with calibration and glare correction techniques," *Food Control,* vol. 66, pp. 215-226, 2016.
[166] P. A. Coelho *et al.*, "A machine vision system for automatic detection of parasites Edotea magellanica in shell-off cooked clam Mulinia edulis," *Journal of Food Engineering,* vol. 181, pp. 84-91, 2016.
[167] M. K. Dutta, A. Singh, and S. Ghosal, "A computer vision based technique for identification of acrylamide in potato chips," *Computers and Electronics in Agriculture,* vol. 119, pp. 40-50, 2015.
[168] M. Al-Sarayreh, M. M. Reis, W. Q. Yan, and R. Klette, "A sequential CNN approach for foreign object detection in hyperspectral images," in *International Conference on Computer Analysis of Images and Patterns*, 2019: Springer, pp. 271-283.
[169] L. Jayasinghe, N. Wijerathne, and C. Yuen, "A deep learning approach for classification of cleanliness in restrooms," in *2018 International Conference on Intelligent and Advanced System (ICIAS)*, 2018: IEEE, pp. 1-6.
[170] C. J. Harrison *et al.*, "Deep learning and multiwavelength fluorescence imaging for cleanliness assessment and disinfection in Food Services," 2022.
[171] P. Voigt and A. Von dem Bussche, "The eu general data protection regulation (gdpr)," *A Practical Guide, 1st Ed., Cham: Springer International Publishing,* vol. 10, no. 3152676, pp. 10-5555, 2017.
[172] B. M. Gaff, H. E. Sussman, and J. Geetter, "Privacy and big data," *Computer,* vol. 47, no. 6, pp. 7-9, 2014.
[173] S. L. Pardau, "The California consumer privacy act: Towards a European-style privacy regime in the United States," *J. Tech. L. & Pol'y,* vol. 23, p. 68, 2018.
[174] N. Rodríguez-Barroso, D. J. López, M. V. Luzón, F. Herrera, and E. Martínez-Cámara, "Survey on federated learning threats: concepts, taxonomy on attacks and defences, experimental study and challenges," *Information Fusion,* 2022.
[175] A. Hard *et al.*, "Federated learning for mobile keyboard prediction," *arXiv preprint arXiv:1811.03604,* 2018.
[176] T. Yang *et al.*, "Applied federated learning: Improving google keyboard query suggestions," *arXiv preprint arXiv:1812.02903,* 2018.
[177] S. Ramaswamy, R. Mathews, K. Rao, and F. Beaufays, "Federated learning for emoji prediction in a mobile keyboard," *arXiv preprint arXiv:1906.04329,* 2019.
[178] S. Silva, B. A. Gutman, E. Romero, P. M. Thompson, A. Altmann, and M. Lorenzi, "Federated learning in distributed medical databases: Meta-analysis of large-scale subcortical brain data," in *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, 2019: IEEE, pp. 270-274.
[179] D. Gao, C. Ju, X. Wei, Y. Liu, T. Chen, and Q. Yang, "Hhhfl: Hierarchical

heterogeneous horizontal federated learning for electroencephalography," *arXiv preprint arXiv:1909.05784,* 2019.
[180] B. Liu, B. Yan, Y. Zhou, Y. Yang, and Y. Zhang, "Experiments of federated learning for covid-19 chest x-ray images," *arXiv preprint arXiv:2007.05592,* 2020.
[181] L. Li, Y. Fan, M. Tse, and K.-Y. Lin, "A review of applications in federated learning," *Computers & Industrial Engineering,* vol. 149, p. 106854, 2020.
[182] Q. Li *et al.*, "A survey on federated learning systems: vision, hype and reality for data privacy and protection," *IEEE Transactions on Knowledge and Data Engineering,* 2021.
[183] W. L. Yuhong Wen, Holger Roth and Prerna Dogra. "Federated Learning powered by NVIDIA Clara." https://developer.nvidia.com/blog/federated-learning-clara/ (accessed.
[184] S. Caldas *et al.*, "Leaf: A benchmark for federated settings," *arXiv preprint arXiv:1812.01097,* 2018.
[185] A. Howard *et al.*, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1314-1324.
[186] T.-J. Yang *et al.*, "Netadapt: Platform-aware neural network adaptation for mobile applications," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 285-300.