Classification of Salmonella Serotypes with Hyperspectral Microscope Imagery

Bosoon Park¹*, Youngwook Seo², Matthew Eady¹, Seung-Chul Yoon¹, Arthur Hinton¹, Jr., Kurt Lawrence¹, and Gary Gamble¹

¹United States Department of Agriculture, Agricultural Research Service, U.S. National Poultry Research Center, USA
²Choongnam National University, Republic of Korea

Abstract

Among serious foodborne outbreaks, Salmonella has the most infections and incidence cases. Because Salmonella is a leading cause of foodborne illness and a zoonotic agent capable of causing gastroenteritis and septicemia, Salmonella detection and identification has become an important subject of research for the poultry industry. Based on the numerous culture protocols to characterize Salmonella spp., traditional culture-based methods are still the most reliable and accurate “gold standard” techniques for presumptive-positive pathogen detection. However, they are laborious and time consuming processes. Therefore, rapid detection and identification of pathogenic microorganisms naturally occurring during food processing are important in developing intervention and verification strategies. Since current detection methods for Salmonella are limited for a practical use, a more sensitive, accurate and rapid pathogen detection method is needed to prevent foodborne outbreaks. Non-destructive advanced optical methods, such as hyperspectral imaging for evaluation of foodborne pathogens could enhance the presumptive-positive screening method by reducing labor and increasing detection speed. Among the several different hyperspectral imaging platforms, acousto-optic tunable filter (AOTF)-based hyperspectral imaging method was developed for microscopic imaging of live bacterial cells from microcolony on agar plates. Thus, the objective of this research is to develop a hyperspectral microscopic imaging method to classify Salmonella serotypes with their spectral signatures from the cells. Five Salmonella serotypes including Enteritidis (SE), Typhimurium (ST), Kentucky (SK), Heidelberg (SH) and Infantis (SI) and five different machine learning algorithms including Mahalanobis distance (MD), k-nearest neighbor (k-NN), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and support vector machine (SVM) were used for classification method development. The SVM algorithm performed better than other algorithms with average classification accuracy of 93.6% (SE), 97.6% (ST), 90.7% (SK), 93.0% (SH), and 94.2% (SI).

ABBREVIATIONS


INTRODUCTION

Salmonella bacteria are commonly found living in the entrails of poultry, which often acquire the bug through their parents as well as environment for living conditions. Most types of Salmonella don’t make the birds ill. In humans, however, many of those same types of Salmonella can result in health problems from a minor gastrointestinal illness to a life-threatening infection in the bloodstream.

The Centers for Disease Control and Prevention (CDC) estimates that approximately 48 million people in the U.S. become ill each year from a foodborne pathogen infection. Of these, an estimated 128,000 are hospitalized and 3,000 die with over 95 percent of these caused by only fifteen pathogens including...
Salmonella, Campylobacter, Listeria, and Shiga toxin-producing E. coli (STEC) [1,2]. More than one million people are sickened by Salmonella in the United States each year with approximately 200,000 cases from poultry alone [3] and the average national cost of foodborne illness was estimated from $55.5 billion up to $93.2 billion [4]. Thus there is a need to reduce foodborne illnesses, especially in poultry. To reduce the risk, real-time and deployable microbial detection and source identification has become increasingly important. Conventional culture methods for the detection and identification of foodborne pathogens usually require sample pre-treatment, colony isolation, and confirmation. Molecular methods, such as polymerase chain reaction (PCR), labeled oligonucleotide probes and DNA microarray, are effective for microbiological detection. Moreover, extensive research into both conventional and molecular methods is worldwide with goals of increasing sensitivity and specificity, while trying to reduce time for analysis. Yet there are still needs to develop faster, more reliable, and more cost-effective methods to quantify and identify pathogenic bacteria in poultry products. One methodology that shows significant promise in rapidly identifying and quantifying pathogenic bacteria is hyperspectral microscope imaging (HMI). This technology combines the resolving power of a microscope with the spectral discriminating ability of a spectrometer. With HMI, there is the potential of identifying single bacteria cells by combining cell morphology with their spectral profile into an automated method for counting and classifying pathogenic bacteria. Since a microscope can image single cells, the challenge is how cellular images are analyzed for identification with other microbiota background.

The primary Hazard Analysis and Critical Control Point (HACCP) concern in poultry industry is typically microbiological, due to the widespread illness that can be associated with a foodborne disease outbreak. For HACCP plans, there is an emphasis placed on process validation and monitoring ready-to-ship product for contaminates. Validating these processes can take days with the standard detection methods with nutrient enriched growth media with polymerase chain reaction (PCR) confirmation [5]. While these methods are widely accepted for microbial detection, they are time-consuming for growing media [6] and require additional plating on selective agar or serological testing [5]. PCR can be completed in as little as a few hours, but has a high recurring cost for target specific cell lysing reagent kits [7,8].

Optical methods, such as laser-induced breakdown spectroscopy (LIBS) [9], Fourier transform infrared (FT-IR) spectroscopy [10], and surface enhanced Raman spectroscopy (SERS) [11,12], have previously been used for food safety applications. Hyperspectral imaging (HSI) has been used to detect defects in food product [13], viable bacteria on raw chicken breasts [14], waterborne bacterial species [15-18], presumptive detection of foodborne pathogenic bacteria colonies of Campylobacter [19] and shiga toxin-producing E. coli (STEC) [20,21] with colonies formed in agar plates.

Typically, spectra obtained from food products [14,22,23], bacterial colonies [19,21,24,25], or bacterial cells [16-18,26,27] are assessed through spectroscopic methods using “fingerprints” produced by the samples. An advantage of hyperspectral microscope imaging (HMI) is the sensitivity of classifying potential pathogens with only a few cells [16,28,29] and capable of classifying field strains of bacteria with chemometrics [30]. Instead of needing colonies, or concentrated suspensions of bacteria, spectral fingerprints from a few cells imaged by HMI could identify bacteria [26]. Thus, HMI offers a means of detection/identification of bacteria with a cellular-level sensitivity, which enables rapid, early detection with less than 24 hrs needed for enrichment. With an appropriate sampling methods [15,17], quality hyperspectral microscopic images can be collected contiguously from a small amount of bacterial suspension on a microscope slide [31]. Although HMI has the potential to detect/identify non-viable spore-forming Bacillus organisms [28], Enterobacteriaceae [31], STEC [26], and Salmonella serotypes [29] through principal component analysis (PCA), there are still gaps to fully understand the mechanisms of HMI for early (less than 8 hrs including incubation) and rapid (less than 10 min) detection of foodborne pathogenic bacteria for high-throughput practical applications for poultry (food) industry. Thus, more research is needed for the development of robust models with multivariate data analysis (MVDA) methods to classify bacteria when considering various environmental factors that may affect the spectral fingerprints from both pure isolates and bacteria in food matrix. The objective of this research is to develop an acousto-optic tunable filter (AOTF)-based hyperspectral microscopic imaging method to classify Salmonella serotypes with their spectral signatures from the cells. More specifically, to develop methods to classify five Salmonella serotypes including Enteritidis, Typhimurium, Kentucky, Heidelberg and Infantis with different machine learning algorithms of k-nearest neighbor, linear discriminant analysis, quadratic discriminant analysis, support vector machine, and partial least square discriminant analysis.

MATERIALS AND METHODS

Sample preparation

Preparing sample to acquire hyperspectral imagery with a HMI is summarized in a flowchart (Figure 1). Five gram-negative Salmonella serotypes (Enteritidis, Typhimurium, Kentucky, Heidelberg, and Infantis) were obtained from the Poultry Microbiological Safety and Processing Research Unit, U.S. National Poultry Research Center, U.S. Department of Agriculture (USDA), Agricultural Research Service (ARS) in Athens, GA. Bacterial cultures were prepared by inoculating pure isolates into tryptic soy broth (TSB) fluted and incubated at 37 ± 2°C for 18-24 hrs. The overnight grown culture of all five serotypes of Salmonella was centrifuged at 5000 rpm for 10 min. The bacterial pellet was resuspended in deionized (DI) water. From five serotypes of Salmonella, 10-fold serial dilutions were prepared in 0.1% peptone water and 10⁻⁶ final dilutions were plated onto brilliant green sulfa (BGS) agar plates in duplicate. All plates were incubated at 35 ± 2°C for 24 hrs. One colony was picked from BGS plate of each Salmonella serotype as shown in Figure (2) and resuspended in 10 μL of DI water. For hyperspectral microscope, 3 μL of bacterial suspension from all Salmonella serotypes were spread on microscopic glass slides in the center approximately in the area of 20 x 20 mm followed by drying for 10 min in the biosafety cabinet. After drying process has been completed,
additional 0.8 µL DI water was added in the center of the slide to secure a cover slip on the top of the sample.

**Hyperspectral microscope imaging system**

A hyperspectral microscope imaging (HMI) system setup is illustrated in Figure (3). The system consists of a Nikon upright microscope (Eclipse e80i, Lewisville, TX), acousto-optic tunable filters (AOTF) (HSI-400, Gooch & Housego, Orlando, FL), a high performance cooled electron-multiplying charge coupled device (EMCCD) 16-bit camera (iXon, Andor Technology, Belfast, Northern Ireland), and dark-field illumination [27] lighting sources (CytoViva 150 Unit, 24W Metal Halide, CytoViva, Auburn, AL). The AOTF used for this research has a high-speed, high-throughput, random-access solid-state optical filter with an adjustable optical pass-band and exceptionally high rejected light levels. AOTF delivers diffraction limited image quality with variable bandwidth resolution down to within 2 nm in a spectral range from 450 to 800 nm with bandwidths of 1.5 nm at 450 nm and 3 nm at 800 nm, respectively. An AOTF-based hyperspectral microscope is a scanning spectrophotometer employing an instrumental technology with no moving parts, capable of high speed of scan, random access to any number of wavelengths pre-selected prior to scanning to generate linear polarized output for quality image acquisition.

**Hyperspectral microscope image acquisition**

Since hyperspectral image acquisition for the wavelengths between 450 and 800 nm with 4nm increments requires a longer time depending on exposure time than regular microscope imaging, the complete immobilization of live cells is necessary during image acquisition. In this study we used a modified drying method [31] to immobilize live cells completely for quality image acquisition.

The procedure of hyperspectral microscope image acquisition and analysis from live bacterial cells is summarized in a flow diagram (Figure 4). Foodborne bacteria, *Salmonella* isolated from poultry carcass rinsate were used for hyperspectral image data. Images from five *Salmonella* serotypes (Enteritidis, Typhimurium, Heidelberg, Kentucky, and Infantis) were acquired with an AOTF-based HMI system. In this experiment, visible/NIR hyperspectral microscope images were collected with a TIFF format at the wavelength ranges from 450 to 800 nm with

Figure 1 Flowchart for sample preparation to acquire hyperspectral images from HMI.

Figure 2 Sampling from *Salmonella Typhimurium* colony for HMI scan (a) BGS Agar Plate; b) Zoomed Colonies; c) Hyperspectral Microscope Image.

Figure 3 ARS acousto-optic tunable filter (AOTF) hyperspectral microscope imaging system including Microscope, AOTF, Controller, Moving Stage, and Light Source.

Figure 4 Flowchart for HMI image acquisition and data processing.
2 nm bandwidth, 4 nm spectral intervals at individual scanning exposure time with 250 ms and the gain of 3.5% selected from previous study [26,31] for quality image acquisition. All images were acquired with dark-field illumination [26] lighting equipped with a metal-halide by a spectral sweep mode for collecting contiguous spectral images [26]. The acquired images (originally TIFF format) were converted to hyperspectral image format (hypercube) with HSiAnalysis™ software (Gooch & Housego, Orlando, FL).

After image conversion has been completed, a region of interest (ROI) from exclusively bacterial cells from background were created followed by generating spectral data from ROIs for further analysis with ENVI (Exelis Visual Information Solutions, Inc., Boulder, CO) software (version 4.8) as shown in Figure (5).

Data preprocessing for selecting data for training and validation from ROI was completed prior to classification model development. R software (version 3.0.1) was used for developing classification methods to identify different Salmonella serotypes using their spectral signatures collected by a HMI system. CLASSIFICATION METHODS

To develop the optimum model, five classification algorithms including k-nearest neighbor method (kNN) [32], Linear discriminant analysis (LDA) [33], Quadratic discriminant analysis (QDA) [34], Partial least squares discriminant analysis (PLS-DA) [35] and Support vector machine (SVM) [36]. More details regarding classification algorithms used in this research see reference [27].

RESULTS AND DISCUSSION

Color composite hyperspectral microscope images with ROI from Salmonella five serotypes (Enteritidis, Typhimurium, Kentucky, Heidelberg, and Infantis) are illustrated in Figure (5). Most cells stand independently except some cells in S. Kentucky and S. Heidelberg are naturally congregated to each other. For collecting spectral data from ROI in the cells, most cautious trials were employed to pick independent single cell. Since the scattering intensity of spectral image at 546 nm was higher than others, we used this image as a template for ROI selection to generate image data for further classification.

Characteristics of hyperspectral image from salmonella

Among five serotypes of gram-negative bacteria, Salmonella Enteritidis was selected to demonstrate hyperspectral microscopic images and corresponding spectral characteristics based on cell structure or morphology. A typical hyperspectral microscope image with ROIs from Salmonella Enteritidis serotype is shown in Figure (6a). Comparing spectral signatures between inner and outer membrane from the cell, two scattered image ROIs, one from inner membrane (green) and the other from outer membrane (red) were collected from S. Enteritidis bacterial cells. Figure (6b) compares the spectral signatures from inner cell walls (4,956 pixels) and outer cell walls (12,846 pixels) from S. Enteritidis. In this sample, we observed the scattering intensity of outer cells were higher than inner cell walls at the wavelength of 498, 522, 546, 574, and 598 nm. However, the scattering intensity from outer membrane were lower than inner at 462, 670, and 690 nm, which means that possibly less scattering occurred in the outer membrane of Salmonella bacterial cells. As seen in Figure (6b), the intensity between two membranes were similar to each other at 474, 626, and beyond 742 nm, respectively.

Scattering intensity distribution of salmonella bacterial cells

Three-dimensional representation of a typical cell obtained
from each of the five Salmonella serotypes (Figure 7) was observed from a slice of the hypercube at the metal halide excitation peak of 546 nm. It can be difficult to visually determine the boundaries of the outer cell wall from the HMIs, but plotting the data in a three-dimensional surface plot shows a more clearly defined boundary starting at raw scattering intensities around 3,000 - 4,000 a.u. for each Salmonella serotype. We can see that these cells are morphologically similar to the short rod shaped structure. The scattering intensity patterns vary with serotype, S. Kentucky and S. Typhimurium displaying a strong intensity, which was maximum at the polar end of the cell. The images suggest that light scattering is responsible for differentiating serotypes, and the spatial information of serotypes within the same species are similar in structure. Further investigation of hypercube dissection is necessary to determine the biological correlation to the distribution of light scattering patterns.

Principal component score plots from salmonella serotypes

Figure 8 demonstrates visual description of clusters from five Salmonella serotypes using PCA score plots from single pixels in the inner and outer cell wall. The serotype S. Enteritidis (red) is relatively easier to be separated than other serotypes, especially from S. Typhimurium (black). However, score plots from S. Heidelberg (blue) and S. Infantis (cyan) are overlapped, which means that it was difficult to separate. Similarly, separation of S. Kentucky (green) from Typhimurium (black) was not simple to be classified. Thus, additional classification method such as SVM could be useful to separate those serogroups, because a SVM algorithm performs well for nonlinear data.

Classification of salmonella serotypes with euclidean distance

Figure 9 is dendrogram plot to cluster five serotypes based on Euclidean distance considering correlation between serotypes with normalized spectral intensity data. According to this dendrogram, S. Enteritidis and S. Kentucky are highly correlated. Also, S. Heidelberg is highly correlated with S. Infantis, but less correlation with S. Typhimurium. However, both S. Heidelberg and S. Infantis are separable from S. Enteritidis and S. Kentucky with relatively high classification accuracy.

Classification accuracy of salmonella serotypes

Table 1 shows the classification accuracy to identify S. Typhimurium and S. Enteritidis from other serotypes. Overall, a SVM algorithm performed more accurately than other classification algorithms for classification of both S. Typhimurium and S. Enteritidis. Specifically, the classification accuracies to identify S. Typhimurium were 86.4% (k = 0.49) for k-NN, 85.4% (k = 0.03) for PLS-DA, 88.1% (k = 0.35) for LDA, 93.3% (k = 0.49) for QDA, and 93.2% (k = 0.7) for SVM, which is the highest accuracy among five classification algorithms tested. Whereas, for the identification of S. Enteritidis, the classification accuracies increase as 89.5% (k = 0.62) for k-NN, 91.1% (k = 0.64) for PLS-DA, 91.6% (k = 0.68) for LDA, 87.6% (k = 0.58) for QDA, and 93.9% (k = 0.77) for SVM, respectively.

Table 2 shows all classification accuracy and corresponding kappa coefficients of five individual Salmonella serotypes with five classification algorithms. As seen in the table, the overall accuracy for classifying individual serotype was not high enough. Among five serotypes, the classification accuracies ranged from 50% to 75% for S. Enteritidis, from 22% to 86% for S. Typhimurium, from 72% to 86% for S. Kentucky, from 68% to 88% for S. Heidelberg, and from 47% to 78% for S. Infantis, respectively. The mean accuracies of five classification methods were 64.4% (lowest) with PLS-DA, 66.3% with k-NN, 74.6% with QDA, 76.8% with LDA, and 84% (highest) with SVM. In this case,
kappa coefficients varied from 0.52 (PLS-DA) up to 0.79 (SVM). Specifically, S. Heidelberg serotype was identified with 88% accuracy when SVM classification method was applied for five different *Salmonella* serotypes.

Figure 10 demonstrates visual description of classification of two selected *Salmonella* serotypes (S. Typhimurium and S. Enteritidis) to others (S. Kentucky, S. Infantis, and S. Heidelberg) using partial least squares (PLS) score plots. Although clusters between two classes were not perfectly separable, S. Typhimurium (red) can be separated from S. Kentucky (Figure 10a), S. Infantis (Figure 10b), and S. Heidelberg (Figure 10c). Also, S. Enteritidis can be separated from S. Kentucky (Figure 10d), S. Infantis (Figure 10e), and S. Heidelberg (Figure 10f), respectively. In addition to an intuitive processing with relatively simple PLS score plots, further classification methods were applied for obtaining accuracy to identify individual serotype, especially S. Enteritidis and S. Typhimurium as shown in Table (1).

### CONCLUSION

The previous research demonstrated an optical method with acousto-optic tunable filter (AOTF)-based hyperspectral microscope imaging (HMI) has potential for classifying gram-negative from gram-positive foodborne pathogenic bacteria rapidly and nondestructively with minimum sample preparation [27]. In this study, we continued developing HMI methods to identify serotypes of *Salmonella*, most typical gram-negative bacteria at the cell level. We successfully validated the protocol for live-cell immobilization on glass slide to acquire 89 contiguous quality spectral images from five *Salmonella* serotype bacterial cells within 45 sec using 250 ms exposure time and 3.5% of an electron multiplying charge coupled device (EMCCD) camera. Among the spectral imagery from visible/NIR, the scattering intensity was higher at the wavelengths of 454, 542, 550, 582, 630, 690, 710, and 722 nm than other wavelengths for *Salmonella* serotypes. The average accuracy to classify five

### Table 1: Classification Accuracy for SE and ST Identification from Other Serotypes.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy (%)</th>
<th>k</th>
<th>SE vs. ST+SE</th>
<th>SE vs. ST</th>
<th>ST vs. (SE, SH, SK, SI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-NN</td>
<td>86.4</td>
<td>0.49</td>
<td>85.4</td>
<td>0.03</td>
<td>88.1</td>
</tr>
<tr>
<td>PLS-DA</td>
<td>89.5</td>
<td>0.62</td>
<td>91.1</td>
<td>0.64</td>
<td>91.6</td>
</tr>
<tr>
<td>LDA</td>
<td>88.1</td>
<td>0.35</td>
<td>85.3</td>
<td>0.49</td>
<td>93.2</td>
</tr>
<tr>
<td>QDA</td>
<td>91.6</td>
<td>0.68</td>
<td>87.6</td>
<td>0.58</td>
<td>93.9</td>
</tr>
<tr>
<td>SVM</td>
<td>93.2</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: k = kappa coefficient; k-NN: k-nearest neighbor; PLS-DA: Partial Least Square Discriminant Analysis; LDA: Linear Discriminant Analysis; QDA: Quadratic Discriminant Analysis; SVM: Support Vector Machine; k-NN vs. (SE, SH, SK, SI); SE vs. (ST, SH, SK, SI); ST vs. SE; ST vs. (SE, SH, SK, SI)

### Table 2: Classification Accuracy of Five *Salmonella* Serotypes.

<table>
<thead>
<tr>
<th>Method</th>
<th>SE</th>
<th>ST</th>
<th>SK</th>
<th>SH</th>
<th>SI</th>
<th>Mean Accuracy (%)</th>
<th>k-NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-NN</td>
<td>0.62</td>
<td>0.69</td>
<td>0.72</td>
<td>0.68</td>
<td>0.51</td>
<td>66.3</td>
<td>0.56</td>
</tr>
<tr>
<td>LDA</td>
<td>0.65</td>
<td>0.79</td>
<td>0.76</td>
<td>0.86</td>
<td>0.68</td>
<td>76.8</td>
<td>0.70</td>
</tr>
<tr>
<td>QDA</td>
<td>0.71</td>
<td>0.82</td>
<td>0.75</td>
<td>0.79</td>
<td>0.59</td>
<td>74.6</td>
<td>0.67</td>
</tr>
<tr>
<td>SVM</td>
<td>0.75</td>
<td>0.86</td>
<td>0.86</td>
<td>0.88</td>
<td>0.79</td>
<td>84.0</td>
<td>0.79</td>
</tr>
<tr>
<td>PLS-DA</td>
<td>0.50</td>
<td>0.22</td>
<td>0.78</td>
<td>0.87</td>
<td>0.47</td>
<td>64.4</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Note: k-NN vs. (SE, ST, SH, SK, SI); k-NN vs. (SE, ST, SH, SK, SI); SVM: Support Vector Machine; PLS-DA: Partial Least Square Discriminant Analysis; Accuracy and Kappa Coefficient (KC) were obtained from the average of ten replicates. Samples include S. Enteritidis (SE), S. Typhimurium (ST), S. Kentucky (SK), S. Heid...
Salmonella serotypes was 84%. However, S. Typhimurium and S. Enteritidis were classified with 93.2% and 93.9% accuracy using a SVM algorithm. Further research is needed to validate with positively identified colonies using confirmatory testing such as latex agglutination or PCR tests. Also, classification model development with selective spectral images using a random access capability of AOTF-based HMI is needed for faster image acquisition to improve imaging-based optical method. In addition, sampling from micro-colonies which are grown in agar media less than 24hrs incubation time will be helpful for rapid detection protocol in the laboratory. Using spectral library from various bacterial species, serotypes, and strains will be helpful for robust detection method from food matrix. Finally, HMI approach with no-culturing method will be very effective for future foodborne pathogen detection tools in poultry food industry.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Nasreen Bano Quality and Safety Assessment Research Unit in Athens for her assistant for this research.

REFERENCES


33. Dixon SJ, Brereton RG. Comparison of performance of five common classifiers represented as boundary methods: Euclidean Distance to Centroids, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Learning Vector Quantization and Support Vector Machines, as dependent on data structure. Chemometrics and Intelligent Lab Syst. 2009; 95: 1-17.

