IMPLEMENTING BIG DATA ANALYTICS APPROACHES TO IMPROVE FOOD QUALITY AND MINIMIZE FOOD WASTE AND LOSS

by

Aftab Siddique

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy
Auburn, Alabama

5th May, 2023

Keywords: Artificial Intelligence, Supervised Machine Learning,
Food Quality, Food Waste, and loss

Approved by

Amit Morey, Chair, Associate Professor, Poultry Science
Tung-Shi Huang, Professor, Food Science
Jason T Sawyer, Associate Professor, Animal Science
Ashish Gupta, Associate Professor, Business Analytics and Information Systems

ABSTRACT

This study was conducted to explore the application of non-invasive, rapid advanced technological enhancements and to combine big data analytics methods to detect muscle quality issues such as woody breast (WB), white striping (WS), and spaghetti meat (SM) conditions that arise in fast-growing broilers within the poultry industry. Results obtained from the rapid identification of myopathies in chicken breast fillets experiments using BIA (Hand-held and plate BIA) collected data analyzed with supervised and unsupervised machine learning algorithms (Support Vector Machine (SVM), Back Propagation Neural Network (BPNN), K-nearest neighbor (k-NN), Fuzzy C Means (FCM) clustering and K-means clustering) have shown that model developed with SVM separated WB with a higher accuracy of 71.0% for normal , 59.9% for moderate, 81.4% for severe WB. Compared to SVM, the BPNN training model accurately (100%) separated normal WB fillets with and without SM demonstrating the ability of BIA to detect SM. While on the other hand, the modified BIA showed better detection ability for normal chicken breast fillets than the probe BIA setup. In the plate BIA setup, fillets were 80.0% for normal, 66.6% for moderate, and 85.0 % for severe WB. However, hand-held BIA showed 77.78%, 85.71%, and 88.89% for normal, moderate, and severe WB. Plate BIA setup is more effective in detecting WB myopathies and could be installed without slowing the processing line. Breast fillet detection on the processing line can be significantly improved using a modified automated plate BIA.

Radio wave frequencies were also used in detection of these myopathic conditions. Results obtained from this experiment indicates that pre-processed data (False discovery rate, predictor screening, and variable clustering) with identified signature frequencies were used to develop classification-based models using Back Propagation Neural Network (BPNN) and Support vector machines (SVM). The BPNN model effectively predicts bird myopathies with varying accuracy in stages: Live Birds (83% variance, 87.5%-100% accuracy), Pre-Chill WOG (78% variance, 87.5%-100% accuracy), Post-Chill WOG (91% variance, 69.7%-100% accuracy), Deboned Fillets (85% variance, 66.7%-100% accuracy). It remains sensitive despite 26% misclassification rates. Conversely, the SVM model shows lower sensitivity and specificity (54.8%-69.7% accuracy). BPNN surpasses SVM in predicting myopathies across processing stages.

Keywords: Support Vector Machines, Backpropagation Neural Networking, Woody breast, Meat myopathies, Spaghetti meat, Bio-electrical impedance analysis, Machine learning, Artificial intelligence

Dedication

I would like to dedicate my dissertation to my better half, Dr. Sanower Warsi, Dr. Amit Morey and all others who have encouraged me to move forward in this journey. Words cannot convey how grateful I am for all that you have done for me. I could not have made it this far without all the support, advice, and guidance. Thank you “Maa” for giving me the opportunity to follow my dreams and the love to make them a reality. Furthermore, I would also like to dedicate my project to my loving sisters, Nida Shahin, my brothers Ahatisham Siddique and Anurag Singh, and my friends for their love, faith, support and encouragement. Special thank you to Ms. Claudine Jenda and whole library staff for strong support and pushing me to accomplish my academic work at Auburn University.

Acknowledgements

I would like to express my extraordinary gratitude to my academic and dissertation advisor, and committee chair, Dr. Amit Morey, for his strong and unwavering support and guidance during my graduate studies at Auburn University (AU). I am very grateful to him for sharing his experience, expertise, and outstanding intellect in the planning and execution of my Ph.D. dissertation project. I also greatly appreciate Dr. Morey allowing me to pursue my research interests in Implementation of Big Data Analytics approaches in Food Science toward my doctorate research. I am also thankful to Dr. Morey for providing me with financial assistance during my graduate degree work at AU.

I would also like to extend a special appreciation to my committee members, Dr. Ashish Gupta, Dr. Jason T Saywer, and Dr. Tung-Shi Huang for providing their advice and guidance toward my PhD research and writing.

Also, I would like to thank Dr. Laura J Garner, Charles B Herron, Bet Wu, Telah Black, Louis, and Sofia, who helped in many way towards my research project.

Special thanks to my colleagues, my best friend Amrit Pal for their supports, help, and advice that they have given to me throughout my graduate studies at Auburn University.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

WB – Woody Breast

WS – White Stripping

SM – Spaghetti Meat

BDA – Big Data Analytics

SVM – Support Vector Machines

BPNN – Back Propagation Neural Network

K-NN- K-Nearest Neighbor

FCM – Fuzzy C Means

PS – Predictor Screening

BS – Boot Strapping

RF Waves – Radio-Frequency Waves

BIA – Bioelectrical Impedance

Chapter 1 LITERATURE REVIEW

Running Head: Big data analytics in food Industry: A state of art literature review

Aftab siddique[1], Ashish Gupta[2], Jason Sawyer[3], Tung-Shi Huang[1], and Amit Morey[1]*

[1] Department of Poultry Science, Auburn University, Auburn, AL 36849, USA;

azs0242@auburn.edu, Huangtu@auburn.edu (A.S; T.S.H);

[2] Department of Business analytics and Information, Auburn University, AL 36849, USA;

azg0074@auburn.edu (AG);

[3]Department of Animal Sciences, Auburn University, Auburn, AL 36849, USA;

jts0109@auburn.edu (J.T.S.)

***Corresponding Author Address and email**
Dr. Amit Morey
Associate Professor
260 Lem Morrison Drive
202-E Poultry Science Building
Auburn University, AL 36849
Email: azm0011@auburn.edu

1.1 ABSTRACT

Food industry have seen a rapid growth in different advancements in the past two decades, and advancement in technologies have created huge amount of data with a rapid rate including variability, and variety in these collected data. This rapid expansion in volume of data had led the current food industry to explore the option of big data analytics approaches including multivariate analysis, and machine learning based algorithms to solve this problem. Analysis of different format of data which includes multimedia data, rapid identification technological equipment data became possible by machine learning approaches such as support vector machines, Neural network, decision tree, K-means clustering, K-nearest neighbor, and Natural language processing etc. In this presented state of art literature review paper, we have tried to collect information from different food sectors especially from food safety, food processing and food quality and tried to provide one informational document that covers above mentioned field collectively. In spite of the fact that many of these applications are still in their infancy, general and domain related problems and issues linked with machine learning have started to be identified and handled. These are essential to the projected usage and eventual integration of massive datatypes and their related machine learning techniques for food safety, food processing, and food quality evaluation applications.

Keywords: Big data analytics, Machine learning, Natural language processing, Support Vector Machines

1.2 INTRODUCTION

Food is the most basic urgent need for growing population around the world and access to enough, safe and nutritious food is a necessity. The most important task of food industry is to ensure that food is highly nutritious and safe for consumers. Currently, the food sector is facing challenges in developing and implementing systems ensuring food quality, safety, and supply (Manufuture, 2006). To solve these complex challenges, effective, affordable, and environmentally sustainable innovative approaches are needed for the food sector's new processes, products, and tools.

Big data analytics (BDA) is gaining attention and becoming a vital tool for the food industry in addressing the predicted increase in global demand caused by population growth and rising incomes in emerging nations (Gupta et al. 2019). Big data analytics may improve food processing processes in the food industry, resulting in goods with improved qualities and additional functionality. Big data analytics can also help to minimize food waste and keep stores from running out of food by closely looking at inventory levels [Dubey et al. 2020; Gupta et al.2019; Manufuture, 2006). Big data analytics employ several techniques to analyze big data sets, such as machine learning, artificial intelligence (backpropagation neural network, natural language processing, and deep learning), data mining, exploratory data analysis, and graphical and visual approaches (Deshpande and Kumar, 2018).

Total investment in big data and AI increased from 27% in 2018 to 33.9% in 2019 (Cision PR newswire, 2020). The global BDA valued the sector at $169 billion in 2018 and is expected to reach $274 billion in 2022 (Statista, 2022). The increased capital investment has promoted research into big food data and use of BDA in food has significantly increased by $35.8 billion from 2010 to 2019 (Ag-funder Agri-food tech, 2019).

A quick search of "big data analytics in the food" yielded over 700 articles as of 2022 in the Journal of Food Science alone. There are hundreds of researches that have been published in other scholarly journals. Big data analytics have potential applications in all aspects of the food chain, including food quality, safety, and processing. Information on the BDA in food sectors is synthesized from published literature (Margaritis et al. 2022). There have been several studies on the feasibility and necessity of big data in the food sector, but not in depth. This paper mainly discusses the three areas of food sectors i. e., food safety, quality, and processing in depth. The main aim of this paper is to synthesize a wealth of information into a single document.

1.3 **DEFINITION AND CHARACTERISTICS OF BIG DATA**

Big data is a term coined in the 1990s by the computer industry and is now used as a catch-all term for anything negative or positive about the twenty-first-century technological society. Interestingly, before the 2000s, big data was considered a problem (Russom, 2011). Since the advent of computers, a tremendous quantity of data has been produced at an increasingly rapid rate. This condition serves as the primary impetus for both the ongoing and upcoming horizons of research (Bryant et al. 2008). Mobile devices, digital sensors, communications, computation, and storage have all advanced in recent years, which has made it possible to collect data (Russom, 2011). Industrial Development Corporation (IDC) has mentioned that the total amount of world data has expanded nine times in the past five years (Gantz and Reinsel, 2011), and Chen and Liu (2014) have mentioned that this generated will get doubled in the next two years. The necessity for massive corporations like Yahoo, Google, and Facebook to examine large volumes of data gave rise to the relatively new concept known as "big data" (Garlasu et al. 2013).

The term "big data" has been defined in a number of different ways, ranging from the 3V model of Volume, Variety, and Velocity to the 4V model of Volume, Velocity, Variety, and

Veracity (Garlasu et al. 2013; Chen and Liu, 2014; Mazahua et al. 2016). The amount of the information is referred to as its volume, the pace at which it is received and transmitted is referred to as its velocity, and the sources of the data and the sorts of data are referred to as its variety (Mauro et al. 2015). The definition of "big data" was expanded by IBM and Microsoft to include "veracity" or "variability" as the fourth "V." The unpredictability and dependability of data are what is meant by the term "veracity."(Al-Sai et al. 2019). In order to characterize big data, McKinsey & Company included value as the fourth V. The term "value" alludes to the significance of the insights that are buried inside large amounts of data (Chen and Zhang, 2014). A detailed description of the 4 V's is given below for a better understanding of data analytics dependencies used in current scenarios (Katal et al. 2013):

i. **Volume:** The data that is currently being stored is measured in petabytes, which is troublesome in and of itself; it is expected that during the next few years, it will expand to zettabyte (ZB). This is mostly attributable to the increased utilization of smartphones and social media networking platforms.

ii. **Velocity:** The term "velocity" can refer to either the rate at which data is collected or the rate at which it is transferred. The increased reliance on live data presents difficulties for the more conventional methods of data analysis because the data is both too extensive and constantly shifting.

iii. **Variety:** Because the data that is collected does not come from a certain set category or from a primary source, it comes in many different raw data forms. These formats can be obtained through the internet, texts, sensors, emails, and other sources, and they can be either structured or unstructured. The sheer magnitude of the problem renders obsolete conventional analytical approaches useless for handling big data.

iv. **Veracity:** It is the central objective in this category, and the primary source of difficulty within the dataset is often noise or anomalies well within data.

Big data analytics approaches comprises many different approaches such as supervised learning approaches, unsupervised approaches, Artificial intelligence based approaches, and decision based approaches, and machine learning (ML) along with data dimensionality reduction techniques (DRT) including linear discriminant Analysis (LDA) (Figure 1) and Principle component analysis (Rashidi et al. 2019).

Machine learning (ML) can be defined as sub-branch of artificial intelligence that is based on computer algorithms and is significantly used in predictive analysis models which can handle large amounts of data and specific trends and patterns (Rashidi et al. 2019; Khan et al. 2022). Popular ML methods such artificial neural network, support vector machines (Figure 2), backpropagation neural network (Figure 3), decision tree, random forest, k-means clustering (Figure 4), k-nearest neighbor are used in the vast engineering areas for data categorization, data clustering regressive predictive modeling, ensemble methods, clustering, transfer learning, image processing, feature extraction, reinforcement learning, natural language processing, and deep learning. Based on learning approaches there are three types of machine learning methods: supervised, semi-supervised, and unsupervised learning (Rashidi et al. 2019; Zhang et al. 2019; Khan et al. 2022).

1.4 SOURCES OF BIG DATA IN THE FOOD SECTOR

The most common sources of big food data pertaining to food industries (from harvesting to restaurants), government sectors, health care (Misra et al. 2020), and media posts include news, video, pictures, and audio. An analysis of big data with a high level of quality can contribute to the growth of the food industry (Zhang et al. 2019; Misra et al. 2020).

1.4.1 Food regulatory agencies' data

The federal regulatory agencies play a big role in providing big data related to food activities (Metcalf and Crawford, 2016; Misra et al. 2020). Four major food safety regulatory agencies are the Food and Drug Administration (FDA) Department of Health and Human Services (DHHS), the Food Safety and Inspection Service (FSIS) U.S. Department of Agriculture (USDA), the Environmental Protection Agency (EPA) and the National Marine Fisheries Service (NMFS) Department of Commerce. Food Safety and Applied Nutrition (CFSAN) that a part of the FDA ensures that the quality of foods is safe for human consumption (Johnson, 2012; (Metcalf and Crawford, 2016; Misra et al. 2020). Food safety activities consumed around $1.6 billion of the FDA budget in 2021 (President's FY 2022 Budget Request, 2022).

The Food Safety Inspection Services (FSIS) system is used by the US government to share food sample analysis reports (Johnson, 2012). Data about food consumption habits and patterns from across the European Union can be found in the EFSA database (Merten et al. 2010). The PulseNet (Swaminathan et al. 2001), the National Antimicrobial Resistance Monitoring System (NARMS)(Gupta et al. 2004), FoodNet (Scallan and Mahon, 2012), and the National Outbreak Reporting System are all examples of the governmental level in the United States (Hall et al. 2013). Rapid Alert System of Food and Feed (RASFF) is a popular online health and safety repository (EU) for industrial and research work (Postolache, 2020). The other food-related database includes the Import Rejection Report (IRR), Inspection Classification Database (ICD) (U.S), and the State Administration for Market Regulation (SAMR) (China). The inclusion of genetic data on food safety activities has increased the amount of data acquired by several of these networks in the last several years. Whole-genome sequencing (WGS) has largely driven an eruption of freely accessible information in new systems like GenomeTrakr (Jackson et al. 2016;

Timme et al. 2018), EnteroBase (Zhou et al. 2020), and the National Center for Biotechnology Information's Pathogen Detection (Sayers et al. 2021). The widespread application of WGS in public health microbiology has spawned the data-driven field of genomic epidemiology (Deng et al. 2021).

Furthermore, Moy and Vannoort (2013) have mentioned that in 1976, the World Health Organization (WHO) founded the Global Environmental Monitoring System (GEMS/Food), in which active participation organizations submit information about food pollutant concentration levels and develop data centers to assist authorities. In 2015, the WHO combined data from agriculture, food, public health, and economics to create a big data digital infrastructure for food safety vulnerability assessments (Marvin et al. 2017). Although, the building of an intelligent supervisory system for the food supply chain is helped by the collaboration and exchange of data across the authorities and organizations that are responsible for the regulation of food (Martinez et al. 2007; Moy and Vannoort, 2013; Marvin et al. 2017) but still there are several obstacles to overcome. Some of the challenges are limited data share, and lack of the standard (Wieczorek et al. 2012). Multiple analysis of the same product by different departments, and agencies had led to a concerning problem of waste of resources and increased operating costs. Specific global standard, proper sharing of real time data within the department and between departments and in between nations for import and export purposes could be a possible area to explore which might be useful in decreasing some of waste of resources and time. There should be development of data mining and classification models that can easily categorize the same product with different names which are also a point of problem in development of network inspection models (Tao et al. 2021).

1.4.2 Food industries data

The food industry is closely linked to agriculture, fisheries, poultry, dairy, processing, and restaurants. All these sectors use modern machinery to increase operational control and performance (Willemson, 2011 Freudenthal and Willemson, 2017; Snow et al. 2021). Interconnected food supply chains can be created using cloud computing, Wireless sensor Networks, blockchain technologies, and the internet of things (Lezoche et al. 2020). Agricultural production and business management are made more efficient with the use of translational active technologies (Lasi et al. 2014). Sensors and drone robotics are used to collect data on precipitation, topography, animal science, nutrition, agricultural planting, and enhanced growth cycles to assist farmers in optimizing these processes. Smart sensors and developed models can collect data that can be use and make real-time decisions that reduce unplanned equipment downtime (Lasi et al. 2014; Mondino and Andújar; 2019; Lezoche et al. 2020).

In recent years, studies on the Internet of Things (IoT) in food manufacturing have encouraged the expansion of the IoT platform in order to fulfill market needs (Choi et al. 2018; Lee et al. 2018), diverse monitoring models, and unbalanced energy usage (Lasi et al. 2014; Lezoche et al. 2020). Applications that integrate the Internet of Things will assist food industries in the creation of new data sources (Da Xu et al. 2014). Not only does Industry 4.0 encourage the rapid agricultural evolution 4.0, but it also makes it possible for businesses to send real-time data in order to recognize and fulfill the shifting stakeholder requirements (Da Xu et al. 2014; Lasi et al. 2014; Choi et al. 2018; Lee et al. 2018; Mondino and Andújar; 2019; Lezoche et al. 2020).

According to a Eurostat report use of smart agriculture will help in the reduction of agricultural costs by 4-6% and will increase profitability by 3% by 2026 (Brookings, 2019)[52]. Implementation of these approaches will help industries the industries to tackle the food production problems and facilitates the lowing of raw materials. It encourages smart agriculture, which saves resources like water, maintains soil, limits carbon pollution, and improves productivity (Ayaz et al. 2019; Brookings, 2019). Smart agriculture enables producers, network operators, the administration, as well as other stakeholders to exchange their insights enhancing the agro value chain for sustainable development (Lasi et al. 2014; Lee et al. 2018; Mondino and Andújar; 2019). These big data analytics approaches have several challenges including data fairness, process traceability, reusability of shared data, and lack of standard information. The lack of well-developed defined protocols has generated inconsistencies among data managerial platforms (Da Xu et al. 2014; Lasi et al. 2014; Choi et al. 2018; Lee et al. 2018; Mondino and Andújar; 2019; Lezoche et al. 2020). Insecure IoT nodes inside the global food supply also pose a threat and might weaken the system. Many firms employ cloud computing, but its application to massive data on food safety is relatively new (Mondino and Andújar; 2019). Durability, data equality, information security, and legal difficulties remain unresolved (Choi et al. 2018). Blockchain technology could make the food production process safer and much more accessible, but it's underdeveloped and complicated to use (Rana et al. 2021). Currently, blockchain's product safety usage is confined to traceability (Tan et al. 2022). Data validation and information management still need exploration (Da Xu et al. 2014; Lasi et al. 2014; Choi et al. 2018; Lee et al. 2018; Ayaz et al. 2019; Brookings, 2019; Mondino and Andújar; 2019; Lezoche et al. 2020; Rana et al. 2021; Tan et al. 2022).

1.4.3 Interactive media data

Different social media platforms are also one of the important factors in collection big data and data generation. There are 4.65 billion users in 2022 which is equal to 57.8% global population (Data Reportal, 2022). Consumer interacts with foods somewhere at end of the food distribution chain, such through transactions, consumption, evaluation, and exchange of experiences, generating a massive volume of information (Blackwell and Blackwell, 2014). These generated data are progressively being disseminated and made accessible using digital media such as social networking sites, search histories, user ratings, and comments, as well as repositories of sales revenue and usage records (Blazquez and Domenech, 2018). Data mining approaches are well versed in the collection of these data and generate valuable information (Cios et al. 2012; Blazquez and Domenech, 2018).

There is a steady stream of clips, articles, as well as other types of material being disseminated on social networking sites (Blackwell and Blackwell, 2014). The food-related data is also acquired from Facebook and post it on social media (Klassen et al. 2018). Using online data, Fried et al. (2014) have used three million Twitter posts to predict population characteristics, and were also able to design and implemented a real-time online system for the query and visualization of collected datasets. Authors in this study found that their developed model outperformed other baseline existing models for the same work (Fried et al. 2014). In the food industry, Singh et al. (2018) uncovered supply chain management issues by analyzing Twitter data. However, data collected from different media sources and social networking sites have their own set of challenges that are needed to be taken care of in order to improve food safety issues. In this aspect, there are several points that needed to be addressed such as multisource data (Soon, 2020). There is a strong need for an algorithm that can take up all the information and fuse it together and direct it to one source (Widom, 1995). The development of these kind of central point source will ultimately help in decreasing the chance of social media rumors and will be useful in increases the chance of public security and safety. As BDA approaches are still in developmental stage, the technology that currently exists is completely based on simulation. In order to stop rumors in time, it will be required to analyze administration are required to analyze social networking sites, and web pages for the establishment of an effective and authoritative integrated component. These research activities and interactions on social networking sites can utilized to investigate how to promote and influence the public's perspective, attitude, and conduct on rumors pertaining to food, healthcare, or other areas of research (Young et al. 2017).

There is a massive quantity of data pertaining to food both within and without the food system, and the collecting and analysis of this data can encourage businesses to extend their market share (Wang et al. 2016). Financial information gives an accurate history of consumer food consumption records and has been shown to support, enhance, and even reinforce traditional exploration methods in description-based hypotheses about the main causal food vehicle in the process of investigation, and/or identifying the area of contamination in department stores or restaurants and at other in the food distribution chain.

Machine learning algorithms could be used to analyze pooled sales data, and the application of this data in epidemic surveillance and outbreak investigations has been demonstrated in a few situations (Sarker, 2021; Singh and Singh, 2021). Food products with sales data that are more similar to the outbreak spread are thought to be the triggering agent (Todd et al. 2007). A probabilistic model of purchases and maximum-likelihood prediction is used in the methodology to determine a group of potentially contaminated commodities (Cameron, 1988; Todd et al. 2007). The product selection method is a theory-driven probabilistic model; however, classification design learning methods are being used to quantify the approach's efficiency and discover patterns in its effectiveness (Liu et al. 2020). Unsupervised classification algorithms are applied to comparable product spatial spread patterns to identify categories of food products that are difficult to differentiate (ElMasry and Nakauchi, 2016). Kaufman et al. (2014) tested the approach using weekly sales Figures of 580 grocery items in Germany using artificial (simulated) foodborne epidemic characteristics. This method has been adjusted to compensate for time, customer movement, and noise, and has been tested on a real-life epidemic in Norway (Kaufman et al. 2014; Norström et al. 2015).

1.4.4 Food-related text data

In machine learning approaches, text data can be compared as “oil” for the algorithms systems to run these ML based models. Commercial and sales data are attractive but have limited early validated uses. Text data are generally unprocessed and contains natural-language texts capable of providing real-time information about food safety contamination occurrences or hazards (Greis and Nogueira, 2017). Text data providers include customer status updates or assessment sites, webpage data such as media outlets or professional organizations portals, and private corporation web-based data (Greis and Nogueira, 2017; Toa et al. 2021).

Online data mining, text mining, pattern recognition and natural language processing (NLP) have been employed to enhance the standard monitoring systems using notifications alert for foodborne diseases or food safety issues (Toa et al.al. 2021). Text data from consumer posts encompass posts on Twitter (Harris et al. 2017; Devinney et al. 2018; Harrison and Johnson, 2019; Tao et al. 2021), Facebook, Yelp (Effland et al. 2018), and Amazon (Maharana et al. 2019). These text data can contain private content like company comment boards, public forums, and blog posts and query data like Google search history (Harris et al. 2017). Food safety ML based applications can mine and analyze post data. A post's text may comprise natural language writing, a title, and customer hashtags. Articles published through traditional news outlets, as well as websites maintained by academic or professional organizations, are examples of types of content that can be found on the internet (Riff et al. 2014). Web data are often used to develop food safety hazard monitoring systems (e.g., outbreaks, recalls) that analyze, integrate, and evaluate relevant material and terms across various websites and geographical locations (Chen and Zhang, 2014).

Oldroyd et al. (2018) and Tao et al. (2021) provided thorough assessments of user data for foodborne disease monitoring systems and a broad sense of text data that is used in the field of food science. Important features of these online channels over conventional data feeds are that the data is available almost immediately, unlike government statements of illness or outbreaks, which can be pushed back by weeks, and that the data has a wide coverage, which is particularly helpful for getting reports from online generations, who are heavily represented on social networking sites but underreported in nationwide foodborne illness outbreak survey results (Oldroyd et al. 2018). This helps solve the important problem of the underrepresentation of foodborne diseases (Scallan and Mahon, 2012; Oldroyd et al. 2018).

Over a decade text data is being used to monitor and detect food safety issues. Effland et al. (2018) and Harris et al. (2017) started to look at tweets done in Chicago, St. Louis, Las Vegas, and New York City. In a creative and unique approach, integration of ML algorithm and the analysis of text data can be seen in study reported by Sadilek et al. (2018). In this study, a team made up of Google researchers and the health departments of Chicago and Las Vegas integrated the Google keyword used to search feedback along with smartphone location to find restaurants that were having lower reviews to find anomaly (Sadilek et al. 2018). Authors have emphasized that this approach was almost three times better than Twitter-based systems at finding places that might be breaking health codes and regulations. In addition to these approaches, Maharana et al. (2018) used text classification model through coupled with Amazon reviews data for identifying food safety issues in food products using consumer feedback (Maharana et al. 2018). As a result, several of these systems have been created to spot epidemics that might otherwise go unnoticed.

While traditional surveillance methods have did miss a few minor incidents of foodborne disease outbreaks, there is initial evidence that big outbreaks involving national eatery chains are being detected on a broad scale (Kuehn, 2014). Its primary use has been in spotting establishments at risk of foodborne illness, even if an outbreak has not yet occurred. Systems that engage with signage to acquire supplementary details about the foodborne outbreaks being reported, including timestamp of the foodborne illness incidence, restaurant specifics, and use of personal details, have also been employed in the past (Kuehn, 2014; Harris et al.2017). A ML based predictive investigative approach of consumers' food assessments (benefits and drawbacks, reliability, and nutritional content) from social networks like Facebook, Twitter, and Instagram will enables businesses to anticipate future issues and improve the quality of their products and services.

1.5 APPLICATIONS OF BIG DATA ANALYTICS IN THE FOOD SAFETY

Infectious diseases that are spread through food continue to be a significant and persistent threat to public health. Foodborne illnesses are responsible for 128,000 hospitalizations and 3,000 fatalities per year in the United States. According to the World Health Organization (WHO), foodborne pathogens are responsible for the illness of 600 million individuals and the deaths of 420,000 people every single year (WHO, 2015). Campylobacteriosis has been the most common foodborne illness in Europe, according to the European Food Safety Authority (EFSA) and the European Centre for Disease Prevention and Control (ECDC), followed by salmonellosis, yersiniosis, Shiga toxin-producing Escherichia coli-

STEC infections, and listeriosis. In 2020, listeriosis had the largest portion of hospitalization in foodborne illness cases (Authority, 2021).

Early detection of harmful organisms and microbial load could improve food safety and prevent foodborne outbreaks. The current "gold-standard" methodology for characterizing foodborne pathogens in food products is reliable, but time-consuming and labor-intensive, restricting food services companies to releasing their commodity to customers first rather than getting entire microbiological information on a specific lot or batch (especially in fresh products). There are several approaches for diagnosing foodborne pathogens in food products, with each having its own advantages and disadvantages in terms of ease of use, consistency of results, schedule and cost effectiveness, etc. Rapid alternatives like spectroscopic technology have been around for more than a decade, however there remain barriers to replacing conventional pathogen identification methodologies (Gracias and McKillip, 2004; Authority, 2021). The major reasons for these difficulties are the necessity for qualified staff, relatively expensive equipment and pre-analysis procedures for some approaches, such as DNA-based procedures (Gracias and McKillip, 2004; Vanegas et al. 2017).

1.5.1 Data analytics approaches in Food spoilage and Pathogen detection

The difficulty in converting these big data sets stems from the fact that the measurements obtained by different sources contain multiple sources of variability, necessitating the use of varied statistical methods to analyze the data. Multivariate analytics is based on the combined analysis of numerous response variables versus many explanatory variables, allowing for a more comprehensive view of the gathered data and causes of unpredictability in a single run (Granato et al. 2018). When the observations involve large amounts of data, a typical multivariate analysis consists of two steps: (i) data pre-treatment and (ii) modeling (Kemsely et al. 2019). Principal components analysis (PCA), cluster analysis (CA), linear discriminant analysis (LDA), and partial least squares (PLS) are the most used multivariable approaches for evaluating these huge variable-associated datasets (Sârbu et al. 2012). Multivariate statistics (MS) is a subfield of statistics. Even in distinct domains, the ML and MS ideas currently often overlap because of their abilities to analyze high-dimensional datasets, sometimes focusing much more on fundamental relationships between variables (multivariate statistics) as well as the algorithms and their implications (probabilistic statistics) (machine learning) (Sârbu et al. 2012).

For the determination of meat spoilage, Pu et al. (2013) have used fluorescent spectroscopy on the meat samples stored at 4 and 15 ^{o}C with an excitation wavelength of 340 nm. The data was then collected and processed using Multivariate Curve Resolution with Alternating Least-Squares (MCR-ALS) to get statistical information that was connected to fluorescence fluctuations and primarily ascribed to NADH content (Pu et al. 2013).

Based on Laser-Induced Breakdown Spectroscopy (LIBS) and Backpropagation Neural Networks, Marcos-Martinez et al. (2011) developed a technique for identifying Pseudomonas aeroginosa, E. coli, and S. Typhimurium isolates. For this experiment, authors have first cultivated above mentioned cultures in three distinct agar plates, and a base spectrum database was first generated. The spectra were taken in the 200–1000 nm region. Then, using a back-propagation (BP) technique for the training method a three-layer Multilayer Perceptron model was constructed, tested with basic measurements such as sensitivity and specificity, and then externally evaluated. The approach correctly identified both known and unknown samples with an accuracy of 100 % regardless of the culture media; nevertheless, the limited sample size used to create the technique was emphasized (Marcos-Martinez et al. 2011).

A method for the subjective and statistical detection of four bacterial species/strains, including one Staphylococcus aureus strain, one S. Typhimurium strain, and two E. coli strains were developed by Liao et al. (2018). The method is based on the conjunction of Three-Dimensional Surface-Enhanced Raman Scattering (3D SERS) and Laser-Induced Breakdown Spectroscopy (LIBS). For quality diagnosis, the 3D SERS approach was used, while LIBS was used for quantitation. Principal Components Analysis (PCA) and Hierarchical Cluster Analysis (HCA) were used to evaluate the spectral data, which resulted in correct classification in all situations. The LIBS method was then applied to the spectral band 200–800 nm, and the most prominent emission spectra band was at 279.5 nm and related to intracellular magnesium ions, was chosen for detection. The spectrum data were examined for this technique by fitting the emission spectrum to a Voigt profile (a probability distribution) to decrease noise, and then using log-log linear regression to make quantitative estimates between fitted peak area and bacterial concentrations. The quantification limit was estimated to be around $5x10^3$ CFU/mL and the correlation of reliability R^2 was found to be >0.97 (Liao et al. 2018).

Argyri et al. (2010) for the very first time showed that Raman spectroscopy can be a potential rapid approach for the detection of meat spoilage. The method involved analyzing meat samples directly with Raman spectroscopy and conventional microbiological techniques were used to construct a quantitative model following data processing for relationship. During this study, pH variations and organoleptic quality evaluation were also performed and a subjective framework with three classes (fresh, semi-fresh, spoil) was constructed. Half-out cross-validation models were used in association with multivariate analysis techniques and ML algorithms. The statistical results for Raman Spectroscopy analysis were promising and showed that the Radial basis function of support vector machines (SVM), support vector regression (SVR) and support vector machines regression sigmoid function (SVRP) achieved a 70% accuracy rate for all counts. However, genetic algorithm artificial neural network (GA-ANN) analysis showed improved classification results with no classification issues on fresh, semi-fresh, and spoiled meat samples (Argyri et al. 2010).

Lu et al. (2020) have developed a method for the identification of 14 microbial species at different growth stages using convolution neural network (CNN) and Laser Tweezers Raman Spectroscopy (LTRS). They have shown that their model was useful in classifying all 14 microbial strains with an overall average classification accuracy of 95.64%. Although their method was capable enough for classification-based study, the relatively high cost of equipment sure limits its commercial use in the field of food microbiology and its related areas (Lu et al. 2020).

Several studies have employed Fourier-Transform Infrared Spectroscopy (FTIR) to identify microbial deterioration in various food products. Fengou et al. (2019) and Spyrelli et al. (2021) used the same concept as Argyri et al. (2010) evaluated FTIR for its ability to estimate surface spoilage in fish and chicken breast fillets using a combination of conventional and advanced analytics approaches to develop a model for spoilage prediction. Both of the aforementioned studies used partial least square regression (PLS-R) models to evaluate quantitative predictions. Fengou et al. (2019) showed that FTIR might be an effective technique for predicting the total count in fish samples (both whole and fillets), with the root mean square error (RMSE) of the constructed model estimated to be 0.717 log CFU/g [60]. The results of the Spyrelli et al. (2021) chicken breast fillets study indicated that using the PLS-R model and SoftML online platform algorithms, reliable quantitative predictions for the total count and

Pseudomonas spp. at chicken breast fillets could be made using FTIR and Multispectral image analysis (MSI).

Hyperspectral imaging (HSI) is similar to MSI in terms of principles, with the distinction being the number of bands used. Because of the continuous high bandwidth detected by HSI, it has a better spectral resolution but a poorer spatial resolution (Feng et al. 2020). The higher the number of bands, the more in-depth details and precise fingerprints of samples can be obtained. Michael et al. (2019) used HSI to establish a method for rapidly distinguishing isolated Cronobacter sakazakii, Salmonella spp., E. coli, L. monocytogenes, and S. aureus. The procedure entailed isolating various strains of the aforementioned bacteria and immobilizing them in a microscope slide, which was then studied with HSI to build a database. Multivariate approach and data analytics approaches were used to develop classification models including principal component analysis (PCA), and k-nearest neighbor (k-NN) after selecting a wavelength range of 425.57 to 753.84 nm. In the results, authors have shown than classification accuracy of various strains within genera such as C. sakazakii, Salmonella spp., and E. coli was found to be 100 % classified with the exception of strain BAA-894 in C. sakazakii and strains O26, O45, and O121 in E. coli had 66.67% classification accuracy. When evaluated together, only C. sakazakii P1, E. coli O104, O111, and O145, S. Montevideo, and L. monocytogenes showed 100% classification accuracy, whereas E. coli O45 and S. Tennessee showed 0.00% classification accuracy in the developed model.

Bonah et al. (2021) examined variable selection techniques for detecting E. coli O157:H7 and S. aureus in pork samples using Vis-NIR hyperspectral imaging utilizing Variable Combination Population Analysis (VCPA), informative variables (IRIV), and Genetic Algorithm (GA). Before collecting the Vis-NIR HIS spectrum, pork samples were inoculated with pathogen culture. Spectral data were processed and cleaned with "noise-reducing methods," including Savitzky–Golay filtration techniques, Second derivatives, and Standard Normal Variate (SNV). They also have investigated six wavelengths selection method and their combinations to determine representative factors. Root mean square errors of measurement, cross validation, and forecasting on the prediction dataset were used to evaluate the algorithms' prediction accuracy. Based on the results authors have emphasized that Vis-NIR HSI may be a good set of instrumentation along with BDA approaches for detecting foodborne pathogens (Bonah et al. 2020; Bonah et al. 2021). Same authors have also developed detection methods for S. Typhimurium in minced pork using electronic nose for different inoculation levels (102, 104, and 107 CFU/gm). For qualitative classification of infected samples, principal components analysis (PCA) was performed, while SVM techniques were used to build the model for computational predictions. SVM regression models with and without improved hyper parameters were also created (Bonah et al. 2021). In machine learning, hyper parameters are frequently used to design the basic training procedure (Wu et al. 2020). The results showed that SVM with optimal parameters performed well and could be used to estimate S. Typhimurium in pork samples quantitatively, whereas PCA can be employed for subjective discrimination analysis (Bonah et al. 2020).

An electronic tongue was applied by Al Ramahi et al. [106] in order to differentiate between E. coli, S. aureus, and P. aeruginosa that were suspended in nutritional broth. They used principal components analysis (PCA) to analyze the outcomes of their investigation, and they placed a great emphasis on the fact that the created approach was able to effectively differentiate the three isolates after 15 hours of incubation. Ghrissi et al. [107] followed the same methodology as the previous investigation (Ramahi and Khalaf, 2019). In aqueous dilutions, the authors employed e-tongue to distinguish and measure Enterococcus faecalis, S. aureus, E. coli, and P. aeruginosa. Sensors were expected to interact chemically with bacterial cell membrane in this investigation. The authors of this paper developed a model for microbe discrimination using linear discriminant analysis and a simulated annealing technique for variable selection (LDA-SA). They also employed multiple linear regression combined with a simulated annealing technique (MLRSA) to create the quantifying model by selecting the most appropriate sensory data. Leave-one-out cross-validation was used to verify both designs (LOOCV) (Ghrissi et al. 2021).

Due to the incidences of foodborne outbreak, rapid detection of different food products are in the great need to be implemented in food industry. Examples provided in the above paragraphs clearly reflects that there has been some work done by authors to develop rapid detection methods for microbial pathogen detection but at the same time it seems that either these methods are still in developmental stages, very naïve, and new concept in research. Some of the methods needs pre-processing steps to perform these techniques which will require resources and trained personnel to complete the data collection and analysis. New approaches such as use of sensors are still in concept phase, on the other hand, spectroscopic techniques have shown an intermediate status. New ideas are needed to reduce the pre-processing steps, sample preparation and direct implementation of BDA in industry so that time, and resources can be saved and quality of detection can improve.

1.6 BIG DATA ANALYTICS IN FOOD PROCESSING

The term "food processing" refers to a variety of processes, some of which include evaporation, boiling, toasting, freezing, bottling, extruding, encapsulating, fermenting, and modified environment packaging. These techniques are used to increase shelf life and quality of food (Ghoshal, 2018). Because of its enormous economic potential, the food processing industry has grown rapidly (Rabbinge, 1993; Ghoshal, 2018). It is the largest segment of the world's food sector, and is expected to grow even more in the near upcoming near future. According to published reports in 2019, food processing industries are valued at $ 11.7 trillion in 2019, and it is predicted to rise at a compound annual growth rate (CAGR) of 5% from 2020 to 2027 (Size, 2020). This increase in food processing industry is connected to increase in human population, life style changes, pandemic situation, and improved food quality (Pelto and Pelto, 1983). Although it is a growing field, processing industries are not completely problem free, there are many issues related to time, increased raw material cost, increase energy cost, and decline in product quality due to unexpected hurdle in processing plants. Inadequate optimized parameters, erroneous sensors, and not well trained workers, and unidentified patterns cause these challenges in food processing plants (Leistner and Gould, 2002). Many researchers had also developed modelling techniques that can be implemented in of food manufacturing, but due to their dependencies on raw ingredient, final product and involved processes limits the use in practical applicability (Jomaa and Puiggali, 1991; Kiranoudis et al. 1997). To solve these applicability issues, researchers have developed semi-physical and entirely theoretical models such as multi-phase models, and single-phase diffusion models (Jomaa & Puiggali, 1991; Kiranoudis et al. 1997; Leistner and Gould, 2002; Putranto et al. 2011; Mabrouk et al. 2012). Although there are benefits of using these fundamental basic model, there are several key challenges associated to it due to the nature of food system which in heterogeneous, porous, and perishable in nature. Conventional modeling is computationally demanding compared to the advance analytical models. Several investigators have used conventional statistical models, such as Page models (Page, 1949), the Henderson and Pabis model (Hendorson, 1961), the Lewis model (Lewis,

1921) and the Newton model, to anticipate humidity transport during evaporation and roasting (Campos et al. 2018) for frying, and response–surface technique for optimization of canning (Afoakwa et al. 2006). These is no doubt that these models are simple, easy to fit, and relatively cost - effective, but analyzing and maintaining large and complicated datasets is challenging.

Observation-based, classification data-driven models like machine learning (ML) have potential for food manufacturing (Sablani and Rahman; 2008). Using ML-based modeling can help in explaining the nonlinearity in the data, inter-relationship of food manufacturing processes that are difficult to address with conventional modeling.

1.6.1 Big data analytics in different processing steps

Hernandez-Perez et al. (2004) have employed ML-based algorithms that can be used to calculate and determine evaporation rate and moisture spread in samples during the drying process of mango and cassava. Authors have mentioned that good quality simulation of drying process is obtained using artificial neural network and also emphasized that ANN can be implemented in online estimation of the product drying process.

Generally, baking is a simple method yet involve complex inter-relationship of physical (Heat, time, size of oven) and chemical properties (water content, protein content, fat content and others) to develop a good quality baked product. The main challenge that this part of processing industry is to increase production with improved quality of baked food. Several studies have been conducted to solve above mentioned issues using mathematical based models (Standing, 1974; Zanoni et al. 1993; ÖZILGEN and Heil, 1994; Sablani et al. 1998). But due to the complexity of these model and intensive computational process involvement, implementation of these models to the baking industry is not practical. On the other hand, simplicity and ability of ML predictive approaches researchers have tried to use these in the industries such as baking of soft cake (Goyal and Goyal, 2011), milk cake (Emerald, 2020), and Bread (Banooni et al. 2009). Broyart and Trystram (2003) developed two neural networks models to forecast the changes in biscuit texture and color during baking process. Overall average moisture content and temperature, thickness, and surface color throughout the length of the oven as a factor of baking process was considered as output. Inductive modeling techniques based on ANN models have the ability to accurately forecast product thickness and color changes.

Sablani et al. (2002) used an ANN-based prediction model to assess the thermal conductivity of several bread commodities. The ANN model with two hidden layers containing six neurons in each hidden layer design yielded better results with a 10% mean relative error (MRE). Based on the observed results, authors have showed that to forecast thermal conductivity values there developed model might be useful in bread baking industry.

Extrusion is an effective approach to transform raw food resources into finished food products with a specific cross-sectional shape and design. Because this technology is cost-effective, simple, energy-efficient, and environmentally friendly, it has attracted a lot of attention and has grown in popularity over the last two decades. Extrusion is used to make various products such as cereals, pasta, noodles, and nuggets (Aksenova and Alexeev, 2020). Controlling feed material, raw component quality, water content, total proteins, pH value, and characteristics like feed rate, extruder length, and screw geometry are the industry's key issues. Above-mentioned parameters affect extruded product quality (Bhagya Raj and Dash, 2022). There is no developed mathematical model exists that allows predictive modeling to regulate these parameters to improve product quality by optimizing process. Expensive equipment, and changing settings for specific product types is difficult, therefore optimizing parameters with the same settings can improve product quality. Optimal process parameters are crucial for the development of these product. Optimizing feed rate, temperature, and pace can enhance color, appearance, and textural quality (Alam et al. 2016; Alemayehu et al. 2019). There have been several studies reported in which authors have used ANN to predict product quality. Shihani et al. (2006) used an ANN and RSM model with several inputs (temperature, moisture content, and screw speed) and outputs (water solubility, water absorption, specific mechanical energy, sensory scores, and expansion ratio) to characterize extruded goods. Authors compared RSM and ANN models to characterize extruded goods. ANN models forecast extruded products with less inaccuracy than RSM models. Fan et al. (2013) used a feedforward ANN model to solve hardness and gumminess in rice flour-based products. Unknown relationships between input and output factors complicate extrusion operations. Authors employed a multilayer feedforward ANN model to solve complicated food processing prediction problems. The network was trained using BPNN with input and output vectors. The created BPNN model has shown great prediction accuracy and promising outcomes, but these results may be erroneous when interior variables like as moisture content, ripeness, and flavor are considered for the model development and predictive analytics work (Fan et al. 2013).

A low cost, color-based ANN model was developed by León-Roque et al. (2016) to estimate the ratio between fermented and non-fermented products total free amino acids in 120 cocoa beans. Authors in this study have collected the Red Green Blue (RGB) color of the fermented cocoa beans from the surface and central region, in the absorption spectrum range from 400 to 450 nm. The predicted results showed excellent classification results for the classification of fermented beans.

Zhu et al. (2019) have develop a rapid method for the detection of fermentation in black tea using electrical properties and used several ML-based model using multilayer perception

(MLP), random forest (RF), and support vector machines (SVM) along with PCA and hierarchical clustering analysis to predict the quality attributes of the fermentation process of black tea. Based on prediction accuracy results multilayer perception, random forest, and SVM are 88.90%, 100%, and 76.92%, respectively, indicated that the random forest was the most appropriate algorithm for predicting the degree of fermentation of black tea.

Canning is a unique process not only in food processing but also important from food safety prospective. In this process food is sealed in a container and subjected to a heating procedure to increase its shelf life (ranging from 1 to 5 years). The quality of canned food is directly and indirectly affected by a variety of elements, including the kind of solution, concentrations, soaking duration of food items, and processing parameters (temperature and time; container material; and the characteristics of food material, such as moisture content, pH, and thermal diffusivity)(Yildiz, 1994). Increasing the shelf life of canned foods while improving their quality and safety can be accomplished by optimizing these various manufacturing parameters. Various statistical models for forecasting the canning process and optimizing the canning process parameters have been proposed and applied. However, due to the complicated nature of mathematical expressions and direct application in canning operations, no theoretical or purely mechanical model for anticipating canning operations has been produced to date. ML-based technology, such as an ANN prediction-based approach, might be a good way to keep track of the process and improve the many quality parameters. ANN model developed by Kseibat et al. (2004) to predict the operating temperature, duration, and basic minimum deterioration during canning. The authors in this study employed beginning temperature, can size, microorganism sensitivity indicator, and sensitive indicator of quality as input factors, and temperature, duration, and basic minimum deterioration as output components of the model. The model's results demonstrated that the ANN-developed model can accurately forecast the temperature, time, and quality degradation of the canning process with MREs of 0.2 %, 3.9 %, and 1.5%, respectively. Based on the model's findings, authors have concluded that the initial food temperature had little impact on the output parameters used to forecast responses Kseibat et al. 2004).

Zhang et al. (2019) for the first time showed the implementation of a generic hybrid mechanistic modeling and machine learning approach to design new food products. In this study, authors have explained the mechanism for mechanistic models for estimation of food characteristics while the machine learning model predicts the sensory characteristics of the developed product. Mittal and Zhang (2000a;2000b;2000c) have intensively investigated the used of predictive modelling approach for deep-frying process, prediction of moisture and temperature content, and for the prediction of freezing point for different food products using neural network algorithmic model.

In this study, frying duration, moisture of the product and surface, product thickness, oil temperature, food's initial temperature, and other product parameter are used as input for the input layer of ANN. Also, a big dataset using four level ANN networks is used to predict more accurate frying process and validated it with experimental data for forecasting meatball deep-frying process. A backpropagation neural network was developed using initial moisture, relative humidity, average temperature of smoke house were used as input variables and dry basis moisture content, center temperature of product, and average temperature of product were considered as output desired labels. Authors have noticed that shrinkage rate as input variable improved the prediction accuracy. For predicting the effect of several modified pre-treatment processes before frying and its effect on end product on the moisture as well as oil concentration of fried mushrooms, Mohebbi et al. (2011) developed a coupled algorithm employing ANN- and genetic algorithms (GA) using frying temperature, time, osmotic condition, and gum-coating parameters as inputs to the model, with moisture and oil content as outputs. With a goodness of fit (R^2) of 0.93 and 96 % prediction accuracy was obtained. In another study to predict the textural features of potato chips during deep-fat frying, Gouyo et al.(2020) have used ensemble learner to develop a decision tree (DT) based algorithm. In this, study, the deep-fat-fried potato chips were found to be crispier than the air-fried chips. This was most likely owing to the differences in water transport pathways between deep-fat and air frying.

Although these models contributed to a better understanding of the frying, baking, canning, extrusion, and freezing for different food product, they were unable to determine the underlying cause of contraction of food during deep-fat frying. Although crust generation is a common occurrence in frying process, this key factor was overlooked when developing the models. All the aforementioned studies have used ANN-based algorithms in almost every case, there is a vast gap of studies to show the use of other predictive modelling approaches such as support vector machine, k-nearest neighbor, random forest and fuzzy set classification and decision making process. There is also a strong need of studies which can shed some light on the combination of multivariate analysis techniques along with big data analytic specifically in image processing techniques for supervised and unsupervised approaches to solve issues related to different food processing steps.

1.7 BIG DATA ANALYTICS IN FOOD QUALITY AND AUTHENTICITY

Different big data analytics approaches in current world are being used in every aspect of food industry. As seen in above section different ML techniques, big data analytics approaches have been used in different areas of food safety, and in different food processing steps. ML based models have been used in inspection of various food ranging from fresh produce to stored food products. Food quality has also been seen as an important factor that influences the cost of final product starting from initial raw ingredients till reaching to the consumers at retailer stores (Trienekens and Zuurbier, 2008). Before the technological advancements in the assessment of food quality in any field including fresh produce, dairy, fisheries and poultry was labor intensive, more prone to false positive results, and requires experienced employee to complete the task and based on employee experiences lot of quality defects goes unnoticed.

Quality evaluation of food is basically consist of grading the food product based on external feature, morphological character and visual sensory attributes such as color, texture and appearance (Patel et al. 2012). Considering the current demand of food industry there is a great need to explore the use of non-invasive sensor during inline, and online food quality detection systems (Dixit et al. 2021). To solve this problem, different researchers have explored the ideas to use rapid quality detection techniques such as cameras (Sun, 2016), sensors (Ruiz et al. 2010), near-infrared (Pasquini, 2018), hyperspectral imaging, radio-frequency waves, and Fourier transformation infrared techniques (Xu et al. 2015) in the quality evaluation of food product in different food matrixes.

Although, these detection system are proved to be helpful but their practical application gets limited due to complex and large dataset generation. Use of BDA in unwinding these complex data sets for data pattern identification and analysis will provide a great insight to the food industry which will be helpful in maintaining and improving the quality. There have been number of studies which focused on these quality parameters for the grading of food products based physical, and chemical attributes and even sometimes combination of these with ML based models (Ruiz et al. 2010; Xu et al. 2015; Sun, 2016; Pasquini, 2018; Dixit et al. 2021). Rapid detection methods for these quality attributes generates huge datasets but there is always a chance to getting redundant, noisy and inappropriate information associated to the data. This noisy and uncleaned large volume of data has been a major concern for feature extraction that can relate to the problem directly in solving food quality issues.

1.8 DATA ANALYTICS IN FOOD QUALITY

Computer based image analysis system has been used for classification of various fruits and vegetables such as grading of apples using multilayer perceptron model, grading of strawberries using image processing using k-means clustering, Low cost tomatoes grading system coupled with machine learning techniques has been reported by Ireri et al. (2019) have showed that grading of tomatoes were done on the basis of color, size and weight. In this study SVM, ANN, and random forest algorithms were developed for the grading of tomatoes based RBG image analysis. Based on the analysis of collected image data SVM showed 91.26 to 94.67% of classification, ANN have showed 92.99 to 95.83% classification, for the decision tree analysis 91.08 to 94.12% of classification accuracy (Ireri et al. 2019).

Kanade and Shaligram (2018) have reported the use of k-nearest neighbor model in the classification of guava fruits (k-NN) into four different classes of green, ripe, overripe and spoiled. Authors have reported about 90% of classification accuracy. Before the development of analytic approaches, classification of corn seed was a challenging task, the process was labor intensive and need experts to do these quality evaluations.

Prakasa et al. (2017) have developed automated image classification system based on region of interest (ROI) and k-means clustering for the classification of corn seed. Results in this study showed that 90% of the accurate classification. Septiarini et al. (2019) have used SVM along with image processing techniques for the classification oil palm fruit based on level of ripeness and color processing for red, green, blue and gray. Results obtained from this experiments showed that developed model was tested on 160 images with an accuracy of 92.5%. Authors have also reported that error percentage was found to be less than 2.4% and color features is the dominant factor in the analysis. Aiadi et al.(2019) have used Gaussian Mixture Model (GMM), and Expectation-Maximization (EM) algorithm was used for parameters estimation and Davies-Bouldin index was used to automatically and precisely estimate the number of components (i.e., appearances) for the classification of 11 different types of dates. Results obtained from the experimental data showed that developed model had high identification rate of 98.65%.

Yu et al. (2018) have demonstrated that using stacked auto-encoder on visible and near-infrared hyperspectral imaging (HSI) generated data were able to combined classify shrimp based on assigned label of freshness in determining their total volatile basic nitrogen (TVB-N) contents. Addition to calibration set of data 116 samples were used in the experiments. Results obtained in the experiment showed that 93.97% of classification accuracy based of desired output grade of Shrimp. Yu et al. (2019) used successive projections algorithm (SPA) and deep-learning-based stacked auto-encoders algorithms to for the prediction of TVB-N content in Pacific white shrimp. Authors in the study used combination of multivariate analysis and BDA approaches for the prediction and found that results obtained in the study showed that model prediction coefficient had a R^2 value of 0.92.

Big data analytics approaches have also been used in meat and poultry processing industry to maintain the quality of fresh raw poultry, beef, lamb and goat. For classification of poultry meat based on quality such as normal fillets, and myopathic fillets. Barbon et al. (2018) have used a SMV model tell the difference between normal and pale meat. They have shown that

SVM can be used to classify breast fillets with muscle myopathies. The classification accuracy for normal breast fillets was 53.4 %, while it was 72% for pale breast fillets. Geronimo et al. (2019) have used a machine vision system and SVM to categorize fillets images. They have observed that developed model from SVM algorithm for WB classified 91.83 % fillets correctly. These researchers also used multilayer perceptron (MLP) to classify the data set. For WB fillets, the classification performance of model was 90.67% [164]. Yang et al. (2021) analyzed images derived from the expressible fluid to classify WB using SVM and deep learning (DL) algorithms. These researchers found high classification accuracy for both training (100%) and testing set (93.3%). Morey et al (2020) have implanted linear discriminate analysis technique for the classification of normal fillets from the myopathic fillets using electrical sensor. They have observed that these sensor when coupled with LDA techniques were able to classify these fillets up to an accuracy of 68.69% for normal fillets and 57.75% for WB fillets. In another study conducted by Siddique et al (2021a) have showed that use of SVM and backpropagation neural network algorithms performed well in classification these normal and myopathic fillets. Authors have found that SVM model was able to classify 73.28% of normal fillets and 81.48% of WB fillets. Siddique et al. (2021b) have also demonstrated that use of singular value decomposition (SVD) analysis method in the determination of quality of fillets through collection of amplitude and phase. Authors in this study have observed that SVD classified 100% normal fillets and 78% WB fillets based on radio-frequency wave analysis data.

Penning et al (2020) have used eight different ML algorithm for the determination of beef quality attributes using image analysis and mass-spectroscopy data. They have observed that PCA-FS and LDA classified 82% beef for quality grade, FS, and SVM Linear classified 99% of meat for production background, PCA-FS, and SVM—Radial classified 85% for breed type, and

FS and XGBoost classified 91% for muscle tenderness. Alaiz-Rodríguez and Parnell (2020) have used ML algorithms in the detection of lamb meat quality, authors have used decision tree and SVM and compared these model with Partial Least Square (PLS) model and Principal Component Analysis (PCA) regression methods. Results have shown that SVM was able to classify 91.80% of the collected fat data as compared to PCA analysis.

Tampering or adulteration of food is also one of the big problems in the area of food quality area in food industry and accounts for $15-40 billion every year to the food industry such as temperament with food quality, labelling of the food product use of cheap quality ingredients in food processing. We have provided a very small related information about these issues and recent developments to tackle this evolving situation. For example, Al-Sarayreh et al. (2020) have used support vector machines (SVM) and deep convolution neural networks (CNN) algorithm to evaluate the level adulteration in red-meat by using Hyperspectral imaging technique (HIS). Based on the analysis and obtained results from their study, authors have confirmed that the CNN model has the best prediction power with a classification accuracy of 94.4%.

Farah et al (2021) have used differential scanning colorimetry with random forest (RF), gradient boosting machine (GBM) and multilayer perceptron in identification and detection adulterant added for quality evaluation of raw milk. Authors have found that all the developed model for MLP, RF and GBM classified 100% adulterated samples with 100% prediction capability for GBM and MLP and 88.5% prediction capabilities with RF developed models.

Fabris et al. (2010) have used RF and SVM based detection model for quality evaluation of cheese during processing. In this study authors have tried to establish a relationship between the storage condition of milk and final quality of cheese. The results in their study have showed that PCA when coupled with SVM showed better quality identification in cheese samples made in different seasons (summer vs winter). Dankowska and Kowalewski (2019) have studied the classification of olive oil based on shelf life and type (refined/extra virgin) using SVM and k-NN coupled with multivariate analytics approaches (PCA) by analyzing fluorescent data collected from synchronous fluorescence spectroscopic measurements. Authors have found that k-NN and SVM were the best optimized model for the classification of data based on labels. The k-NN and SVM model classified 94.60% and 94.4% labeled data respectively in their assigned groups.

1.9 CONCLUSIONS

The food sector and most of its aspects are expanding rapidly. According to Pitchbook's annual financial report, venture capitalists made an investment of more than $39.3 billion in 2021. The global pandemic of COVID-19 is thought to have a big impact on this trend. Companies and government agencies have spent $6 billion on improving product quality, ingredients, research, food waste, and traceability (Pitch’s Book annual report; 2022). Understanding food system is difficult because it is a sophisticated system with a range of characteristics that are always inter-related and completely reliant on each component of food. For example, moisture influences the texture of food, and it also encourages the growth of microorganisms. Big data analytics ML models may be better suited to solving problems relating to the food business if these connections are better understood.

In the field of food sciences, big data analytics (BDA) is still in a very preliminary phase and has a lot of room to grow. As new, innovative techniques have been made, the amount of data has grown at a faster rate and become more complicated. Using big data analytics in the food industry (processing, safety, and quality) is a new way to make sure that products are of better quality and that people's health is better. But it is complicated to develop a BDA-based model for food processing, quality, and safety applications as it take a lot of scientific experimentation work to use this new technology in the food industry today. In this paper, we have tried to provide a review of the current state of the art of BDA-based modeling approaches up to some extent and a wide range of applications in the food industry. Developing a BDA model to solve problems and improve food quality is dependent on many aspects of food (moisture, protein, fat, physical texture, and appearance), organizations, governmental agencies, multimedia, and relationship between different countries.

Nevertheless, as of right now, the vast majority of the necessary features cannot be obtained because it is difficult to explore these properties through experimental investigation alone. This paper also gives useful information about numerous ML-based models used in food safety, processing, and quality applications. It also shows that extra care may be needed when choosing input and output parameters for ML-based modeling, especially for artificial neural network modeling, image processing, data mining from different sources, and SVM modeling. This is because taking into account too many input parameters that are not needed could make the model building too challenging to solve, which might lead to inaccurate results due to overfitting and under fitting issues. This paper also talked about the different neural network algorithms, SVM, multivariate data analysis classifier, pattern recognition, different sensors, rapid detection techniques, and their limitations, so that a reader could learn more about which algorithms would work best toward their own concerns and experimental objectives. Most classification-based problems have been solved with ANN by different authors in different fields (Processing, quality, and safety) However, ANN has some limitations because it works like a black box with no information on analysis to the user. Even so, researchers have tried to solve this problem by using hybrid approaches that combine neural networks, fuzzy logic, DTs, RF, multivariate analysis, or any combination of these things. More effort should be put into understanding the ANN system and its decision ability for given outputs so that these models can be used on their own, since they are based on human model. When combined with data dimensionality reduction techniques, BDA-based modeling has a lot of potential for use in food industry in reducing the unnecessary data that might affects that overall output of ML based model. It is hoped that the knowledge and information in this paper will significantly improve the understanding of how ML-based approaches can be used in different food sectors.

1.10 REFERENCES

Afoakwa, E. O., Yenyi, S. E., & Sakyi-Dawson, E. (2006). Response surface methodology for optimizing the pre-processing conditions during canning of a newly developed and promising cowpea (Vigna unguiculata) variety. Journal of Food Engineering, 73(4), 346 357.

Ag-funder Agri-food tech (2019). Investing report. https://agfunder.com/research/agfunder agrifood-tech-investing-report-2019/ Accessed on June 14th 2022

Aiadi, O., Kherfi, M. L., & Khaldi, B. (2019). Automatic Date Fruit Recognition Using Outlier Detection Techniques and Gaussian Mixture Models. ELCVIA Electronic Letters on Computer Vision and Image Analysis, 18(1), 52-75.

Aksenova, O. I., & Alexeev, G. V. (2020). The Effect of the Concentration of the Fish Processing Offal Powder, of the Humidity Level and of the Cross-Sectional Area of the Die Molding Channel on the Technological Parameters of the Extrusion Process and on the Quality Characteristics of Potato Snacks. In IOP Conference Series: Materials Science and Engineering (Vol. 753, No. 8, p. 082006). IOP Publishing.

Alaiz-Rodríguez, R., & Parnell, A. C. (2020). A machine learning approach for lamb meat quality assessment using FTIR spectra. IEEE Access, 8, 52385-52394.

Alam, M. S., Pathania, S., & Sharma, A. (2016). Optimization of the extrusion process for development of high fibre soybean-rice ready-to-eat snacks using carrot pomace and cauliflower trimmings. LWT, 74, 135-144.

Alemayehu, H., Emire, S. A., & Henry, C. (2019). Effects of extrusion process parameters on the quality properties of ready-to-eat pulse-based snacks. Cogent Food & Agriculture, 5(1), 1641903.

Al-Sai, Z. A., & Abdullah, R. (2019). Big data impacts and challenges: a review. In 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT) (pp. 150-155). IEEE.

Al-Sarayreh, M., Reis, M. M., Yan, W. Q., & Klette, R. (2020). Potential of deep learning and snapshot hyperspectral imaging for classification of species in meat. Food Control, 117, 107332.

Argyri, A. A., Panagou, E. Z., Tarantilis, P. A., Polysiou, M., & Nychas, G. J. (2010). Rapid qualitative and quantitative detection of beef fillets spoilage based on Fourier transform infrared spectroscopy data and artificial neural networks. Sensors and Actuators B: Chemical, 145(1), 146-154.

Authority, E. F. S. (2021). The European Union one health 2019 zoonoses report. Efsa Journal, 19(2).

Ayaz, M., Ammad-Uddin, M., Sharif, Z., Mansour, A., & Aggoune, E. H. M. (2019). Internet-of Things (IoT)-based smart agriculture: Toward making the fields talk. IEEE access, 7, 129551-129583.

Banooni, S., Hosseinalipour, S. M., Mujumdar, A. S., Taherkhani, P., & Bahiraei, M. (2009). Baking of flat bread in an impingement oven: modeling and optimization. Drying Technology, 27(1), 103-112.

Barbon, S., Costa Barbon, A. P. A. D., Mantovani, R. G., & Barbin, D. F. (2018). Machine learning applied to near-infrared spectra for chicken meat classification. Journal of Spectroscopy, 2018.

Bhagya Raj, G. V. S., & Dash, K. K. (2022). Comprehensive study on applications of artificial neural network in food process modeling. Critical Reviews in Food Science and Nutrition, 62(10), 2756-2783.

Blackwell, R. D., & Blackwell, K. S. (2014). Creating consumer-driven demand chains in food service. Quick Service Restaurants, Franchising, and Multi-Unit Chain Management, 137.

Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. Technological Forecasting and Social Change, 130, 99-113.

Bonah, E., Huang, X., Aheto, J. H., Yi, R., Yu, S., & Tu, H. (2020). Comparison of variable selection algorithms on vis-NIR hyperspectral imaging spectra for quantitative monitoring and visualization of bacterial foodborne pathogens in fresh pork muscles. Infrared Physics & Technology, 107, 103327.

Bonah, E., Huang, X., Hongying, Y., Aheto, J. H., Yi, R., Yu, S., & Tu, H. (2021). Detection of Salmonella Typhimurium contamination levels in fresh pork samples using electronic nose smellprints in tandem with support vector machine regression and metaheuristic optimization algorithms. Journal of Food Science and Technology, 58(10), 3861-3870.

Brookings. Eurostats (2019). https://www.brookings.edu/research/chinas-influence-on-the global-middle-class/ Accessed on 06/09/2022

Broyart, B., & Trystram, G. (2003). Modelling of heat and mass transfer phenomena and quality changes during continuous biscuit baking using both deductive and inductive (neural network) modelling principles. Food and Bioproducts Processing, 81(4), 316-326.

Bryant, R., Katz, R. H., & Lazowska, E. D. (2008). Big-data computing: creating revolutionary breakthroughs in commerce, science and society.

Cameron, T. A. (1988). A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression. Journal of environmental economics and management, 15(3), 355-379.

Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information sciences, 275, 314-347.

Chen, M. M. S., & Liu, Y. (2014). Big Data: A Survey Mobile Networks and Application.

Choi, J., In, Y., Park, C., Seok, S., Seo, H., & Kim, H. (2018). Secure IoT framework and 2D architecture for End-To-End security. The Journal of Supercomputing, 74(8), 3521-3535.

Cios, K. J., Pedrycz, W., & Swiniarski, R. W. (2012). Data mining methods for knowledge discovery (Vol. 458). Springer Science & Business Media.

Cision PR newswire, 2020. The Global Big Data Analytics Market, 2027: A $105+ Billion Opportunity Assessment. https://www.prnewswire.com/news-releases/the-global-big data-analytics-market-2027-a-105-billion-opportunity-assessment-301014418.html Accessed on 06/09/2022

Da Xu, L., He, W., & Li, S. (2014). Internet of things in industries: A survey. IEEE Transactions on industrial informatics, 10(4), 2233-2243.

Dankowska, A., & Kowalewski, W. (2019). Comparison of different classification methods for analyzing fluorescence spectra to characterize type and freshness of olive oils. European Food Research and Technology, 245(3), 745-752.

Data Reportal. Kepios 2022. https://datareportal.com/social-media-users Accessed on 06/07/2022

De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. In AIP conference proceedings (Vol. 1644, No. 1, pp. 97 104). American Institute of Physics.

de Oliveira Campos, B. L., da Costa, A. O. S., de Souza Figueiredo, K. C., & da Costa Junior, E. F. (2018). Performance comparison of different mathematical models in the simulation of a solar desalination by humidification-dehumidification. Desalination, 437, 184-194.

Deng, X., Cao, S., & Horn, A. L. (2021). Emerging applications of machine learning in food safety. Annual Review of Food Science and Technology, 12(1), 513-538.

Deshpande, A., & Kumar, M. (2018). Artificial intelligence for big data: Complete guide to automating big data solutions using artificial intelligence techniques. Packt Publishing Ltd.

Devinney, K., Bekbay, A., Effland, T., Gravano, L., Howell, D., Hsu, D. & Gutelius, B. (2018). Evaluating twitter for foodborne illness outbreak detection in New York city. Online Journal of Public Health Informatics, 10(1).

Dixit, Y., Hitchman, S., Hicks, T. M., Lim, P., Wong, C. K., Holibar, L.,& Reis, M. M. (2021). Non-invasive spectroscopic and imaging systems for prediction of beef quality in a meat processing pilot plant. Meat Science, 181, 108410.

Dubey, R., Gunasekaran, A., Childe, S. J., Bryde, D. J., Giannakis, M., Foropon, C.,& Hazen, B. T. (2020). Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: A study of manufacturing organisations. International Journal of Production Economics, 226, 107599.

Effland, T., Lawson, A., Balter, S., Devinney, K., Reddy, V., Waechter, H.,& Hsu, D. (2018). Discovering foodborne illness in online restaurant reviews. Journal of the American Medical Informatics Association, 25(12), 1586-1592.

ElMasry, G. M., & Nakauchi, S. (2016). Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality–A comprehensive review. Biosystems engineering, 142, 53-82.

Emerald, F., Pushpadass, H. A., Manjunatha, M., Manimala, K., Dejey, D., Salish, K., & Nath, B. S. (2020). Modelling approaches for predicting moisture transfer during baking of chhana podo (milk cake) incorporated with tikhur (Curcuma angustifolia) starch. Journal of Food Measurement and Characterization, 14(6), 2981-2997.

Fabris, A., Biasioli, F., Granitto, P. M., Aprea, E., Cappellin, L., Schuhfried, E., & Endrizzi, I. (2010). PTR-TOF-MS and data-mining methods for rapid characterisation of agro industrial samples: influence of milk storage conditions on the volatile compounds profile of Trentingrana cheese. Journal of mass spectrometry, 45(9), 1065-1074.

Fan, F. H., Ma, Q., Ge, J., Peng, Q. Y., Riley, W. W., & Tang, S. Z. (2013). Prediction of texture characteristics from extrusion food surface images using a computer vision system and artificial neural networks. Journal of Food Engineering, 118(4), 426-433.

Farah, J. S., Cavalcanti, R. N., Guimarães, J. T., Balthazar, C. F., Coimbra, P. T., Pimentel, T. C., & Cruz, A. G. (2021). Differential scanning calorimetry coupled with machine learning technique: An effective approach to determine the milk authenticity. Food Control, 121, 107585.

Feng, X., He, L., Cheng, Q., Long, X., & Yuan, Y. (2020). Hyperspectral and multispectral remote sensing image fusion based on endmember spatial information. Remote Sensing, 12(6), 1009.

Fengou, L. C., Lianou, A., Tsakanikas, P., Gkana, E. N., Panagou, E. Z., & Nychas, G. J. E. (2019). Evaluation of Fourier transform infrared spectroscopy and multispectral imaging as means of estimating the microbiological spoilage of farmed sea bream. Food microbiology, 79, 27-34.

Freudenthal, M., & Willemson, J. (2017). Challenges of federating national data access infrastructures. In International Conference for Information Technology and Communications (pp. 104-114). Springer, Cham.

Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., & Bell, D. (2014). Analyzing the language of food on social media. In 2014 IEEE International Conference on Big Data (Big Data) (pp. 778-783). IEEE.

Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. IDC iview, 1142(2011), 1-12.

Garlasu, D., Sandulescu, V., Halcu, I., Neculoiu, G., Grigoriu, O., Marinescu, M., & Marinescu, V. (2013, January). A big data implementation based on Grid computing. In 2013 11th RoEduNet International Conference (pp. 1-4). IEEE.

Geronimo, B. C., Mastelini, S. M., Carvalho, R. H., Júnior, S. B., Barbin, D. F., Shimokomaki, M., & Ida, E. I. (2019). Computer vision system and near-infrared spectroscopy for identification and classification of chicken with wooden breast, and physicochemical and technological characterization. Infrared Physics & Technology, 96, 303-310.

Ghoshal, G. (2018). Emerging food processing technologies. In Food processing for increased quality and consumption (pp. 29-65). Academic Press.

Ghrissi, H., Veloso, A. C., Marx, Í. M., Dias, T., & Peres, A. M. (2021). A potentiometric electronic tongue as a discrimination tool of water-food indicator/contamination bacteria. Chemosensors, 9(6), 143.

Gouyo, T., Mestres, C., Maraval, I., Fontez, B., Hofleitner, C., & Bohuon, P. (2020). Assessment of acoustic-mechanical measurements for texture of French fries: Comparison of deep-fat frying and air frying. Food Research International, 131, 108947.

Goyal, S., & Goyal, G. K. (2011). Simulated neural network intelligent computing models for predicting shelf life of soft cakes. Global Journal of Computer Science and Technology, 11(14), 29-33.

Gracias, K. S., & McKillip, J. L. (2004). A review of conventional detection and enumeration methods for pathogenic bacteria in food. Canadian journal of microbiology, 50(11), 883 890.

Granato, D., Putnik, P., Kovačević, D. B., Santos, J. S., Calado, V., Rocha, R. S.,& Pomerantsev, A. (2018). Trends in chemometrics: Food authentication, microbiology, and effects of processing. Comprehensive Reviews in Food Science and Food Safety, 17(3), 663-677.

Greis, N. P., & Nogueira, M. L. (2017). A data-driven approach to food safety surveillance and response. In Food protection and security (pp. 75-99). Woodhead Publishing.

Gupta, A., Nelson, J. M., Barrett, T. J., Tauxe, R. V., Rossiter, S. P., Friedman, C. R.,& NARMS Working Group. (2004). Antimicrobial resistance among campylobacter strains, United States, 1997–2001. Emerging infectious diseases, 10(6), 1102.

Gupta, S., Chen, H., Hazen, B. T., Kaur, S., & Gonzalez, E. D. S. (2019). Circular economy and big data analytics: A stakeholder perspective. Technological Forecasting and Social Change, 144, 466-474.

Hall, A. J., Wikswo, M. E., Manikonda, K., Roberts, V. A., Yoder, J. S., & Gould, L. H. (2013). Acute gastroenteritis surveillance through the national outbreak reporting system, United States. Emerging infectious diseases, 19(8), 1305.

Harris, J. K., Hawkins, J. B., Nguyen, L., Nsoesie, E. O., Tuli, G., Mansour, R., & Brownstein, J. S. (2017). Research brief report: Using twitter to identify and respond to food poisoning The food safety stl project. Journal of Public Health Management and Practice, 23(6), 577.

Harrison, S., & Johnson, P. (2019). Challenges in the adoption of crisis crowdsourcing and social media in Canadian emergency management. Government Information Quarterly, 36(3), 501-509.

Hendorson, S. M. (1961). Grain drying theory (I) temperature effect on drying coefficient. Journal of agricultural engineering research, 6(3), 169-174.

Hernandez-Perez, J. A., Garcıa-Alvarado, M. A., Trystram, G., & Heyd, B. (2004). Neural networks for the heat and mass transfer prediction during drying of cassava and mango. Innovative Food Science & Emerging Technologies, 5(1), 57-64.

Ireri, D., Belal, E., Okinda, C., Makange, N., & Ji, C. (2019). A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing. Artificial Intelligence in Agriculture, 2, 28-37.

Jackson, B. R., Tarr, C., Strain, E., Jackson, K. A., Conrad, A., Carleton, H.,& Gerner-Smidt, P. (2016). Implementation of nationwide real-time whole-genome sequencing to enhance listeriosis outbreak detection and investigation. Reviews of Infectious Diseases, 63(3), 380-386.

Johnson, R. (2012). The federal food safety system: a primer. Congressional Research Service.

Jomaa, W., & Puiggali, J. R. (1991). Drying of shrinking materials: modellings with shrinkage velocity. DRYING TECHNOLOGY,, 9(5), 1271-1293.

Kanade, A., & Shaligram, A. (2018). Prepackaging Sorting of Guava Fruits using Machine Vision based Fruit Sorter System based on K-Nearest Neighbor Algorithm. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 3(3), 2456-3307.

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: issues, challenges, tools and good practices. In 2013 Sixth international conference on contemporary computing (IC3) (pp. 404-409). IEEE.

Kaufman, J., Lessler, J., Harry, A., Edlund, S., Hu, K., Douglas, J.,& Filter, M. (2014). A likelihood-based approach to identifying contaminated food products using sales data: performance and challenges. PLoS computational biology, 10(7), e1003692.

Kemsley, E. K., Defernez, M., & Marini, F. (2019). Multivariate statistics: Considerations and confidences in food authenticity problems. Food Control, 105, 102-112.

Khan, M. I. H., Sablani, S. S., Joardder, M. U. H., & Karim, M. A. (2022). Application of machine learning-based approach in food drying: Opportunities and challenges. Drying Technology, 40(6), 1051-1067.

Kiranoudis, C. T., Tsami, E., & Maroulis, Z. B. (1997). Microwave vacuum drying kinetics of some fruits. Drying technology, 15(10), 2421-2440.

Klassen, K. M., Borleis, E. S., Brennan, L., Reid, M., McCaffrey, T. A., & Lim, M. S. (2018). What people “like”: Analysis of social media strategies used by food industry brands, lifestyle brands, and health promotion organizations on Facebook and Instagram. Journal of medical Internet research, 20(6), e10227.

Kseibat, D. S., Mittal, G. S., & Basir, O. A. (2004). Predicting safety and quality of thermally processed canned foods using a neural network. Transactions of the Institute of Measurement and Control, 26(1), 55-68.

Kuehn, B. M. (2014). Agencies use social media to track foodborne illness. JAMA, 312(2), 117 118.

Lasi, H., Fettke, P., Kemper, H. G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. Business & information systems engineering, 6(4), 239-242.

Lee, J., Park, G. I., Shin, J. H., Lee, J. H., Sreenan, C. J., & Yoo, S. E. (2018). SoEasy: A software framework for easy hardware control programming for diverse IoT platforms. Sensors, 18(7), 2162.

Lee, K. M., Runyon, M., Herrman, T. J., Phillips, R., & Hsieh, J. (2015). Review of Salmonella detection and identification methods: Aspects of rapid emergency response and food safety. Food control, 47, 264-276.

Leistner, L., & Gould, G. W. (2002). Hurdle Technologies: Combination Treatments for Food Stability, Safety and Quality: Combination Treatments for Food Stability, Safety, and Quality. Springer Science & Business Media.

León-Roque, N., Abderrahim, M., Nuñez-Alejos, L., Arribas, S. M., & Condezo-Hoyos, L. (2016). Prediction of fermentation index of cocoa beans (Theobroma cacao L.) based on color measurement and artificial neural networks. Talanta, 161, 31-39.

Lewis, W. K. (1921). The rate of drying of solid materials. Industrial & Engineering Chemistry, 13(5), 427-432.

Lezoche, M., Hernandez, J. E., Díaz, M. D. M. E. A., Panetto, H., & Kacprzyk, J. (2020). Agri food 4.0: A survey of the supply chains and technologies for the future agriculture. Computers in industry, 117, 103187.

Liao, W., Lin, Q., Xie, S., He, Y., Tian, Y., & Duan, Y. (2018). A novel strategy for rapid detection of bacteria in water by the combination of three-dimensional surface-enhanced Raman scattering (3D SERS) and laser induced breakdown spectroscopy (LIBS). Analytica Chimica Acta, 1043, 64-71.

Liu, X., Wang, G. A., Fan, W., & Zhang, Z. (2020). Finding useful solutions in online knowledge communities: A theory-driven design and multilevel analysis. Information Systems Research, 31(3), 731-752.

Lu, W., Chen, X., Wang, L., Li, H., & Fu, Y. V. (2020). Combination of an artificial intelligence approach and laser tweezers Raman spectroscopy for microbial identification. Analytical Chemistry, 92(9), 6288-6296.

Mabrouk, S. B., Benali, E., & Oueslati, H. (2012). Experimental study and numerical modelling of drying characteristics of apple slices. Food and Bioproducts Processing, 90(4), 719 728.

Maharana, A., Cai, K., Hellerstein, J., Hswen, Y., Munsell, M., Staneva, V.,& Nsoesie, E. O. (2019). Detecting reports of unsafe foods in consumer product reviews. JAMIA open, 2(3), 330-338.

Manufuture. (2006). Vision 2020 and Strategic Research Agenda of the European Agricultural Machinery Industry and Research Community for the 7th Framework Programme for

Research of the European Community, Brussels, Belgium. http://www.manufuture.org/documents/AET%20Vision%20and%20SRA1.pdf Accessed 06/09/2022.

Marcos-Martinez, D., Ayala, J. A., Izquierdo-Hornillos, R. C., de Villena, F. M., & Caceres, J. O. (2011). Identification and discrimination of bacterial strains by laser induced breakdown spectroscopy and neural networks. Talanta, 84(3), 730-737.

Margaritis, I., Madas, M., & Vlachopoulou, M. (2022). Big Data Applications in Food Supply Chain Management: A Conceptual Framework. Sustainability, 14(7), 4035.

Martinez, M. G., Fearne, A., Caswell, J. A., & Henson, S. (2007). Co-regulation as a possible model for food safety governance: Opportunities for public–private partnerships. Food Policy, 32(3), 299-314.

Marvin, H. J., Janssen, E. M., Bouzembrak, Y., Hendriksen, P. J., & Staats, M. (2017). Big data in food safety: An overview. Critical reviews in food science and nutrition, 57(11), 2286 2295.

Merten, C., Ferrari, P., Bakker, M., Boss, A., Hearty, A., Leclercq, C.,& Arcella, D. (2011). Methodological characteristics of the national dietary surveys carried out in the European Union as included in the European Food Safety Authority (EFSA) Comprehensive European Food Consumption Database. Food Additives & Contaminants: Part A, 28(8), 975-995.

Metcalf, J., & Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. Big Data & Society, 3(1), 2053951716650211.

Michael, M., Phebus, R. K., & Amamcharla, J. (2019). Hyperspectral imaging of common foodborne pathogens for rapid identification and differentiation. Food Science & Nutrition, 7(8), 2716-2725.

Misra, N. N., Dixit, Y., Al-Mallahi, A., Bhullar, M. S., Upadhyay, R., & Martynenko, A. (2020). IoT, big data and artificial intelligence in agriculture and food industry. IEEE Internet of Things Journal.

Mittal, G. S., & Zhang, J. (2000a). Prediction of temperature and moisture content of frankfurters during thermal processing using neural network. Meat Science, 55(1), 13-24.

Mittal, G. S., & Zhang, J. (2000b). Prediction of freezing time for food products using a neural network. Food Research International, 33(7), 557-562.

Mittal, G. S., & Zhang, J. (2000c). Use of artificial neural network to predict temperature, moisture, and fat in slab-shaped foods with edible coatings during deep-fat frying. Journal of Food Science, 65(6), 978-983.

Mohebbi, M., Shahidi, F., Fathi, M., Ehtiati, A., & Noshad, M. (2011). Prediction of moisture content in pre-osmosed and ultrasounded dried banana using genetic algorithm and neural network. Food and Bioproducts Processing, 89(4), 362-366.

Mondino, P., & González-Andújar, J. L. (2019). Evaluation of a decision support system for crop protection in apple orchards. Computers in Industry, 107, 99-103.

Morey, A., Smith, A. E., Garner, L. J., & Cox, M. K. (2020). Application of bioelectrical impedance analysis to detect broiler breast filets affected with woody breast myopathy. Frontiers in Physiology, 11, 808.

Moy, G. G., & Vannoort, R. W. (Eds.). (2013). Total diet studies (pp. 169-177). New York, NY, USA:: Springer.

Norström, M., Kristoffersen, A. B., Görlach, F. S., Nygård, K., & Hopp, P. (2015). An adjusted likelihood ratio approach analysing distribution of food products to assist the investigation of foodborne outbreaks. PloS one, 10(8), e0134344.

Oldroyd, R. A., Morris, M. A., & Birkin, M. (2018). Identifying methods for monitoring foodborne illness: review of existing public health surveillance techniques. JMIR public health and surveillance, 4(2), e8218.

ÖZILGEN, M., & Heil, J. R. (1994). Mathematical modeling of transient heat and mass transport in a baking biscuit. Journal of food processing and preservation, 18(2), 133-148.

Page, G. E. (1949). Factors Influencing the Maximum Rates of Air Drying Shelled Corn in Thin layers. Purdue University.

Panetto, H., & Kacprzyk, J. (2020). Agri-food 4.0: A survey of the supply chains and technologies for the future agriculture. Computers in industry, 117, 103187.

Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new perspectives–A review. Analytica chimica acta, 1026, 8-36.

Patel, K. K., Kar, A., Jha, S. N., & Khan, M. A. (2012). Machine vision system: a tool for quality inspection of food and agricultural products. Journal of food science and technology, 49(2), 123-141.

Pelto, G. H., & Pelto, P. J. (1983). Diet and delocalization: dietary changes since 1750. The Journal of interdisciplinary history, 14(2), 507-528.

Penning, B. W., Snelling, W. M., & Woodward-Greene, M. J. (2020). Machine learning in the assessment of meat quality. IT Professional, 22(3), 39-41.

PitchBook's annual report. Food drive (2022). https://www.fooddive.com/news/food-tech-saw 393b-in-vc-investments-last-year-says-pitchbook/618916/ Accessed on 06/10/2022

Postolache, A. N., Chelmu, S. S., Ariton, A. M., Ciorpac, M. I. T. I. C. Ă., Pop, C., Ciobanu, M. M., & Creangă, Ș. T. E. O. F. I. L. (2020). Analysis of RASFF notifications on contaminated dairy products from the last two decades: 2000-2020. Romanian Biotechnological Letters, 25(2), 1396-1406.

Prakasa, E., Rosiyadi, D., Ni'mah, D. F. I., Khoiruddin, A. A., Lestriandoko, N. H., Suryana, N., & Fajrina, N. (2017). Automatic region-of-interest selection for corn seed grading. In 2017 International Conference on Computer, Control, Informatics and its Applications (IC3INA) (pp. 23-28). IEEE.

President's FY 2022 Budget Request: Key Investments for Food Safety https://www.fda.gov/about-fda/what-we-do-cfsan/presidents-fy-2022-budget-request key-investments-food-safety

Pu, Y., Wang, W., & Alfano, R. R. (2013). Optical detection of meat spoilage using fluorescence spectroscopy with selective excitation wavelength. Applied Spectroscopy, 67(2), 210-213.

Putranto, A., Chen, X. D., Devahastin, S., Xiao, Z., & Webley, P. A. (2011). Application of the reaction engineering approach (REA) for modeling intermittent drying under time varying humidity and temperature. Chemical Engineering Science, 66(10), 2149-2156.

Rabbinge, R. (1993). The ecological background of food production. In Crop protection and sustainable agriculture. Ciba Found. Symp. 177, John Wiley & Sons, Chicester (pp. 2 29).

Rana, R. L., Tricase, C., & De Cesare, L. (2021). Blockchain technology for a sustainable agri food supply chain. British Food Journal.

Rashidi, H. H., Tran, N. K., Betts, E. V., Howell, L. P., & Green, R. (2019). Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. Academic pathology, 6, 2374289519873088.

Riff, D., Lacy, S., & Fico, F. (2014). Analyzing media messages: Using quantitative content analysis in research. Routledge.

Rodríguez-Mazahua, L., Rodríguez-Enríquez, C. A., Sánchez-Cervantes, J. L., Cervantes, J., García-Alcaraz, J. L., & Alor-Hernández, G. (2016). A general perspective of Big Data: applications, tools, challenges and trends. The Journal of Supercomputing, 72(8), 3073 3113.

Rowa'Al Ramahi, A. N. Z., & Abu-Khalaf, N. (2019). Evaluating the potential use of electronic tongue in early identification and diagnosis of bacterial infections. Infection and Drug Resistance, 12, 2445.

Ruiz-Altisent, M., Ruiz-Garcia, L., Moreda, G. P., Lu, R., Hernandez-Sanchez, N., Correa, E. C., & García-Ramos, J. (2010). Sensors for product characterization and quality of specialty crops—A review. Computers and Electronics in agriculture, 74(2), 176-194.

Russom, P. (2011). Big data analytics. TDWI best practices report, fourth quarter, 19(4), 1-34.

Sablani, S. S., & Rahman, S. (2008). Fundamentals of food dehydration. In Food drying: Science and technology (pp. 1-42). DEStech Publications, Inc., Pennsylvania, USA.

Sablani, S. S., Baik, O. D., & Marcotte, M. (2002). Neural networks for predicting thermal conductivity of bakery products. Journal of Food Engineering, 52(3), 299-304.

Sablani, S. S., Marcotte, M., Baik, O. D., & Castaigne, F. (1998). Modeling of simultaneous heat and water transport in the baking process. LWT-Food Science and Technology, 31(3), 201-209.

Sadilek, A., Caty, S., DiPrete, L., Mansour, R., Schenk, T., Bergtholdt, M.,& Gabrilovich, E. (2018). Machine-learned epidemiology: real-time detection of foodborne illness at scale. NPJ digital medicine, 1(1), 1-7.

Sârbu, C., Naşcu-Briciu, R. D., Kot-Wasik, A., Gorinstein, S., Wasik, A., & Namieśnik, J. (2012). Classification and fingerprinting of kiwi and pomelo fruits by multivariate analysis of chromatographic and spectroscopic data. Food Chemistry, 130(4), 994-1002.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN Computer Science, 2(3), 1-21.

Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., ... & Sherry, S. T. (2021). Database resources of the national center for biotechnology information. Nucleic acids research, 49(D1), D10.

Scallan, E., & Mahon, B. E. (2012). Foodborne Diseases Active Surveillance Network (FoodNet) in 2012: a foundation for food safety in the United States. Clinical Infectious Diseases, 54(suppl_5), S381-S384.

Septiarini, A., Hamdani, H., Hatta, H. R., & Kasim, A. A. (2019). Image-based processing for ripeness classification of oil palm fruit. In 2019 5th International Conference on Science in Information Technology (ICSITech) (pp. 23-26). IEEE.

Shihani, N., Kumbhar, B. K., & Kulshreshtha, M. (2006). Modeling of extrusion process using response surface methodology and artificial neural networks. Journal of engineering science and Technology, 1(1), 31-40.

Siddique, A.**,** Freeman, R., and Morey, A., (2021b). Microwave analysis of broiler breast meat in conjunction with singularity value decomposition to categorize myopathic fillets. Poultry sciences association abstract book pp.223

Siddique, A., Shirzaei, S., Smith, A. E., Valenta, J., Garner, L. J., & Morey, A. (2021a). Acceptability of Artificial Intelligence in Poultry Processing and Classification Efficiencies of Different Classification Models in the Categorisation of Breast Fillet Myopathies. Frontiers in Physiology, 1472.

Singh, A., Shukla, N., & Mishra, N. (2018). Social media data analytics to improve supply chain management in food industries. Transportation Research Part E: Logistics and Transportation Review, 114, 398-415.

Singh, R., & Singh, R. (2021). Applications of sentiment analysis and machine learning techniques in disease outbreak prediction–A review. Materials Today: Proceedings.

Size, G. M. (2020). Share & Trends Analysis Report By Source (Biodiesel, Fatty Acids, Fatty Alcohols, Soap), by Type (Crude, Refined) By End Use (Food & Beverage, Pharmaceutical). By Region, And Segment Forecasts, 2027.

Snow, V., Rodriguez, D., Dynes, R., Kaye-Blake, W., Mallawaarachchi, T., Zydenbos, S., & Stevens, D. (2021). Resilience achieved via multiple compensating subsystems: The immediate impacts of COVID-19 control measures on the agri-food systems of Australia and New Zealand. Agricultural Systems, 187, 103025.

Soon, J. M. (2020). Consumers' awareness and trust toward food safety news on social media in Malaysia. Journal of food protection, 83(3), 452-459.

Spyrelli, E. D., Ozcan, O., Mohareb, F., Panagou, E. Z., & Nychas, G. J. E. (2021). Spoilage assessment of chicken breast fillets by means of fourier transform infrared spectroscopy and multispectral image analysis. Current research in food science, 4, 121-131.

Standing, C. N. (1974). Individual heat transfer modes in band oven biscuit baking. Journal of Food Science, 39(2), 267-271.

Statista Research Department (SRD), (2022). Big data market size revenue forecast worldwide from 2011 to 2027 https://www.statista.com/statistics/254266/global-big-data-market forecast/

Sun, D. W. (Ed.). (2016). Computer vision technology for food quality evaluation. Academic Press.

Swaminathan, B., Barrett, T. J., Hunter, S. B., Tauxe, R. V., & Force, C. P. T. (2001). PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerging infectious diseases, 7(3), 382.

Tan, A., Gligor, D., & Ngah, A. (2022). Applying blockchain for halal food traceability. International Journal of Logistics Research and Applications, 25(6), 947-964.

Tao, Q., Ding, H., Wang, H., & Cui, X. (2021). Application research: big data in food industry. Foods, 10(9), 2203.

Timme, R. E., Rand, H., Leon, M. S., Hoffmann, M., Strain, E., Allard, M., & Baugher, J. D. (2018). GenomeTrakr proficiency testing for foodborne pathogen surveillance: an exercise from 2015. Microbial genomics, 4(7).

Todd, E. C., Greig, J. D., Bartleson, C. A., & Michaels, B. S. (2007). Outbreaks where food workers have been implicated in the spread of foodborne disease. Part 3. Factors contributing to outbreaks and description of outbreak categories. Journal of food protection, 70(9), 2199-2217.

Trienekens, J., & Zuurbier, P. (2008). Quality and safety standards in the food industry, developments and challenges. International journal of production economics, 113(1), 107-122.

Vanegas, D. C., Gomes, C. L., Cavallaro, N. D., Giraldo-Escobar, D., & McLamore, E. S. (2017). Emerging biorecognition and transduction schemes for rapid detection of pathogenic bacteria in food. Comprehensive reviews in food science and food safety, 16(6), 1188-1205.

Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. International journal of production economics, 176, 98-110.

WHO. WHO Estimates of the Global Burden of Foodborne Diseases; WHO: Geneva, Switzerland, 2015; ISBN 9789241565165.

Widom, J. (1995, December). Research problems in data warehousing. In Proceedings of the fourth international conference on Information and knowledge management (pp. 25-30).

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R. & Vieglais, D. (2012). Darwin Core: an evolving community-developed biodiversity data standard. PloS one, 7(1), e29715.

Willemson, J. (2011, August). Pseudonymization service for X-road eGovernment data exchange layer. In International Conference on Electronic Government and the Information Systems Perspective (pp. 135-145). Springer, Berlin, Heidelberg.

Wu, J., Chen, S., & Liu, X. (2020). Efficient hyperparameter optimization through model-based reinforcement learning. Neurocomputing, 409, 381-393.

Xu, J. L., Riccioli, C., & Sun, D. W. (2015). An overview on nondestructive spectroscopic techniques for lipid and lipid oxidation analysis in fish and fish products. Comprehensive Reviews in Food Science and Food Safety, 14(4), 466-477.

Yang, Y., Wang, W., Zhuang, H., Yoon, S. C., Bowker, B., Jiang, H., & Pang, B. (2021). Evaluation of broiler breast fillets with the woody breast condition using expressible fluid measurement combined with deep learning algorithm. Journal of Food Engineering, 288, 110133.

Yildiz, F. (1994). Initial preparation, handling, and distribution of minimally processed refrigerated fruits and vegetables. In Minimally Processed Refrigerated Fruits & Vegetables (pp. 15-65). Springer, Boston, MA.

Young, W., Russell, S. V., Robinson, C. A., & Barkemeyer, R. (2017). Can social media be a tool for reducing consumers' food waste? A behaviour change experiment by a UK retailer. Resources, Conservation and Recycling, 117, 195-203.

Yu, X., Tang, L., Wu, X., & Lu, H. (2018). Nondestructive freshness discriminating of shrimp using visible/near-infrared hyperspectral imaging technique and deep learning algorithm. Food analytical methods, 11(3), 768-780.

Yu, X., Wang, J., Wen, S., Yang, J., & Zhang, F. (2019). A deep learning based feature extraction method on hyperspectral images for nondestructive prediction of TVB-N content in Pacific white shrimp (Litopenaeus vannamei). Biosystems Engineering, 178, 244-255.

Zanoni, B., Peri, C., & Pierucci, S. (1993). A study of the bread-baking process. I: A phenomenological model. Journal of food engineering, 19, 389-389.

Zhang, X., Zhou, T., Zhang, L., Fung, K. Y., & Ng, K. M. (2019). Food product design: a hybrid machine learning and mechanistic modeling approach. Industrial & Engineering Chemistry Research, 58(36), 16743-16752.

Zhang, Z., Zhao, Y., Liao, X., Shi, W., Li, K., Zou, Q., & Peng, S. (2019). Deep learning in omics: a survey and guideline. Briefings in functional genomics, 18(1), 41-57.

Zhou, Z., Alikhan, N. F., Mohamed, K., Fan, Y., Achtman, M., Brown, D.,& Agama Study Group. (2020). The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. Genome research, 30(1), 138-152.

Zhu, H., Liu, F., Ye, Y., Chen, L., Liu, J., Gui, A., & Dong, C. (2019). Application of machine learning algorithms in quality assurance of fermentation process of black tea--based on electrical properties. Journal of food engineering, 263, 165-172.

Table 1.1: Summary table of research conducted in food sciences combined with data analytics

Authors	Used Technique	Work
Afoakwa et al. 2006	Stepwise multiple regression analysis	Optimization of pre-processing conditions during canning
Aiadi et al. 2019	Outlier Detection Techniques and Gaussian Mixture Models	Automatic Date fruit recognition technique
Aksenova and Alexeev, 2020.	Kinetics based study with response surface methodology	Fish powder and Potato snack
Alaiz-Rodríguez and Parnell, 2020	Neural network classifiers and different dimensionality reduction techniques	Lamb meat quality assessment
Alam et al. 2016	Multiple regression with response surface methodology	Ready to eat high fibre soybean rice snack
Alemayehu et al. 2019	Central Composite Rotatable design (CCRD) with Regression and RSME	Ready-to-eat pulse-based snacks
Al-Sarayreh et al. 2020	Deep Learning	Classification of species in meat
Argyri et al. 2010	Artificial Neural network	Spoilage in beef fillets
Banooni et al. 2009	Artificial Neural network	Baking of flat bread
Barbon et al. 2018	Decision tree and Support Vector machines	Chicken meat classification
Bonah et al. 2020	variable combination population analysis with genetic algorithm	Quantitative monitoring and visualization of bacterial foodborne pathogens in fresh pork muscles
Bonah et al. 2020	Detection of Salmonella Typhimurium contamination levels in fresh pork samples using electronic nose	support vector machine regression and metaheuristic optimization algorithms
Broyart and Trystram, 2003	Neural Network	Modelling of heat and mass transfer phenomena and quality changes during continuous biscuit baking
Cameron, 1988	maximum likelihood estimation by censored logistic regression	valuing non-market goods using referendum data
Dankowska and Kowalewski, 2019	Principle component Analysis followed by LDA, QDA, RDA, KNN, SVM and Random Forest	Characterize type and freshness of olive oils
de Oliveira Campos, 2018	Different mathematical models	Simulation of a solar desalination by humidification-dehumidification
Devinney et al. 2018	Text mining Algorithms	Foodborne illness outbreak detection in New York city using tweets

Dixit et al. 2021	Multivariate Analysis	Prediction of beef quality
Effland et al. 2018	Text mining	Discovering foodborne illness in online restaurant reviews
Emerald et al. 2020	Neural network with Fuzzy Inference System	Predicting moisture transfer during baking of milk cake incorporated with starch
Fabris et al. 2010	Random forest (RF),Penalized discriminant analysis (PDA), Support vector machine (SVM), Discriminant partial least squares(DPLS)	Influence of milk storage conditions on the volatile compounds profile of Trentingrana cheese
Fan et al. 2013	Artificial Neural network (ANN)	Prediction of texture characteristics from extrusion food surface
Farah et al. 2021	Random forest, gradient boosting machine, and multilayer perceptron	Determination of the milk authenticity
Fengou et al. 2019	Partial least square regression Analysis (PLSR)	Microbiological spoilage of farmed sea bream
Fried et al. 2014	Text mining	Social media data on food
Geronimo et al. 2019	Decision Tree Modeling (DT)	Classification of woody breast chicken fillets
Gouyo et al. 2020	Principle component Analysis with Decision tree	Comparison of deep fat frying and air frying
Goyal and Goyal 2011	Neural Networks (NN)	Soft cake shelf life prediction
Greis and Nogueira, 2017	North Carolina Foodborne Events Data Integration and Analysis (NCFEDIA) system	Food safety surveillance and response
Harris et al. 2017	Text mining	Using twitter to identify and respond to food poisoning
Hernandez-Perez et al. 2004	Neural network (NN)	Heat and mass transfer prediction during drying of cassava and mango
Ireri et al. 2019	Support Vector Machines, Artificial neural network, Random Forest	Grading of tomatoes
Kanade and Shaligram, 2018	k-Nearest Neighbor model	Prepackaging Sorting of Guava Fruits
Kaufman et al. 2014	likelihood-based approach	Contaminated food products idendification using sales data
Kemsley et al. 2019	Multivariate Analysis	Food authenticity
Kseibat et al. 2004	Neural network (NN)	Predicting safety and quality of thermally processed canned foods
Kuehn, 2014	Text mining (TM)	Social media to track foodborne illness

León-Roque et al. 2016	Artificial Neural Network (ANN)	Prediction of fermentation index of cocoa beans
Lu et al. 2020	Artificial Neural Network (ANN)	Microbial Identification
Mabrouk et al. 2012	Mathematical Modeling (MM)	Drying of apple slices
Marcos-Martinez et al. 2011	Neural network (NN)	Identification and discrimination of bacterial strains
Michael et al. 2019	Principle component Analysis and ANN	Rapid identification and differentiation for foodborne pathogens
Mittal and Zhang, 2000a	Neural networks	Prediction of temperature and moisture content of frankfurters during thermal processing
Mittal and Zhang, 2000b	Neural networks	Prediction of freezing time for food products
Mittal and Zhang, 2000c	Neural networks	Prediction of temperature, moisture, and fat in slab-shaped foods with edible coatings during deep-fat frying
Mohebbi et al. 2011	Genetic Algorithm (GA) and Neural network (NN)	Prediction of moisture content in pre-osmosed and ultrasounded dried banana
Mondino and Andújar, 2019	Decision and Support system (DSS)	Crop protection in apple orchards
Morey et al. 2020	Linear Discriminant Analysis (LDA)	Woody breast Identification in chicken
Norström et al. 2015	Adjusted likelihood ratio	Distribution of food products to assist investigation of foodborne outbreaks
Penning et al. 2020	Partial least square analysis, SVM, RF, k-NN, PDA, XGBoost, LogitBoost, LDA	Beef meat quality assessment
Pu et al. 2013	Multivariate Curve Resolution with Alternating Least-Squares (MCR-ALS)	Optical detection of meat spoilage
Ramahi and Khalaf, 2019	Principle component Analysis	Early identification and diagnosis of bacterial infections using electronic tongue
Sablani et al. 2002	Neural Network	Predicting thermal conductivity of bakery products
Sablani et al. 1998	Mathematical modeling	Modeling of simultaneous heat and water transport in the baking process
Sârbu et al. 2012	Multivariate Analysis	Classification and fingerprinting of kiwi and pomelo fruits
Scallan and Mahon, 2012	FoodNet	Foodborne Diseases Active Surveillance Network
Septiarini et al. 2019	Support Vector Machines (SVM)	Ripeness classification of oil palm fruit

Shihani et al. 2006	Response surface methodology and Artificial Neural Networks	Wheat flour and wheat–black soybean blend
Siddique et al. 2021a	Support Vector Machines (SVM) and Back Propagation Neural Network (BPNN)	Classification of woody breast chicken breast fillets
Siddique et al. 2021b	Singular Value decomposition (SVD)	Classification of woody breast chicken fillets
Spyrelli et al. 2021	SVM, RF, ANN, k-NN, PCA, Least-angle regression	Spoilage assessment of chicken breast fillets
Swaminathan et al. 2001	PulseNet	Foodborne bacterial disease surveillance
Tan et al. 2022	Block chain	Halal food traceability
Timme et al. 2018	GenomeTrakr	Proficiency testing for foodborne pathogen surveillance
Yang et al. 2021	Deep learning Algorithm	Evaluation of broiler breast fillets with the woody breast condition
Yu et al. 2018	Deep learning Algorithm	Shrimp freshness
Yu et al. 2019	Deep Learning Algorithm	For nondestructive prediction of TVB-N content in Pacific white shrimp
Zhang, X. et al. 2019	Hybrid machine learning and mechanistic modeling approach	Development of cookie ingredients
Zhang, Z. et al. 2019	Deep learning Algorithm	Omics studies
Zhu et al. 2019	PCA, Clustering, SVM, RF, and MLP	Quality assurance of fermented black tea

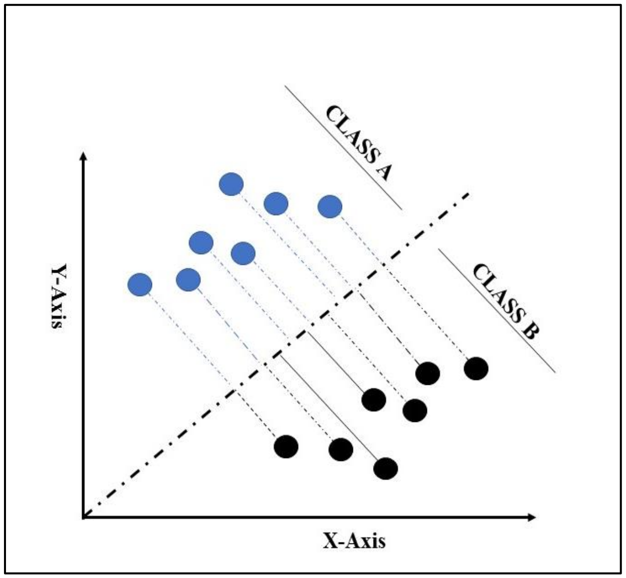


Figure 1-1: Linear discriminant analysis taken from Siddique et al. [167], adopted from Fisher (1986). Two class data in dimensional space for LDA analysis to maximize the classifiable data on the hyper-plane.

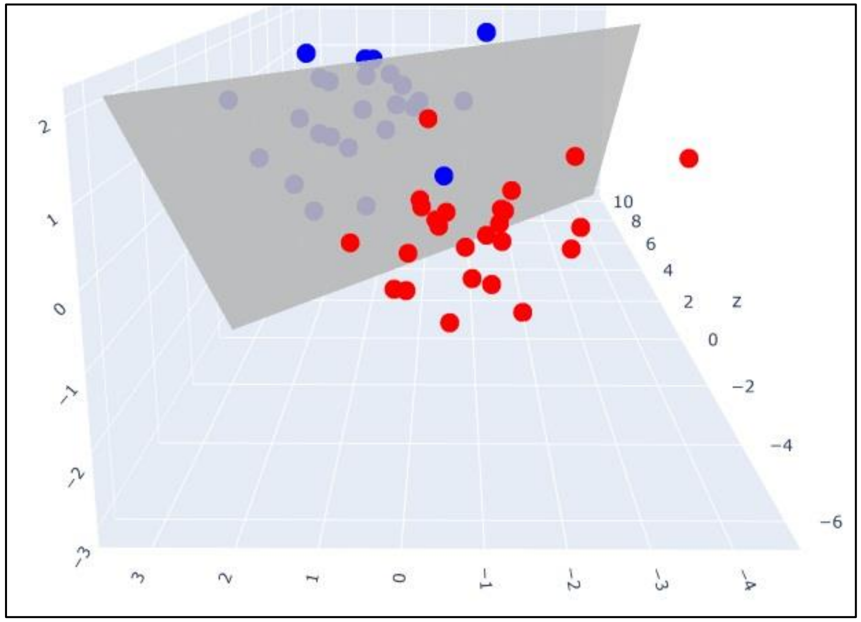


Figure 1-2: Support vector machines adopted from Siddique et al. [167], red dot represents class A data in the front space and blue dot represents class B data behind the shaded gray area (hyperplane)

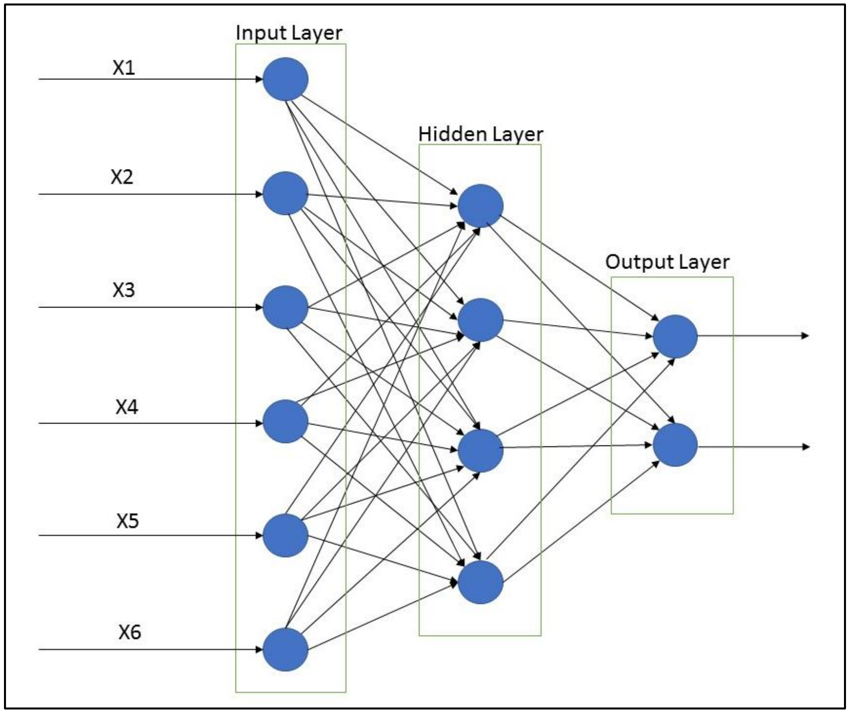


Figure 1-3: Backpropagation neural network from Siddique et al [167]

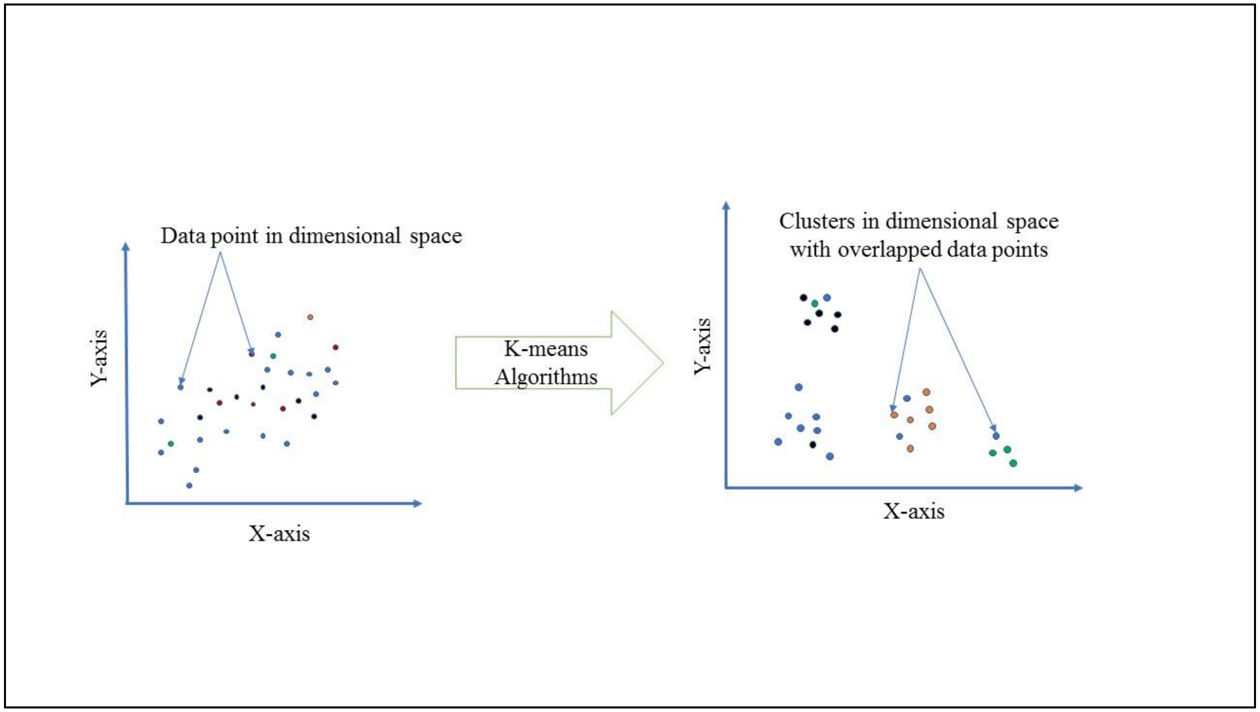


Figure 1-4: General diagrammatic representation of k-means clustering for cluster formation from Siddique et al. 2022

Chapter 2

ACCEPTABILITY OF ARTIFICIAL INTELLIGENCE IN POULTRY PROCESSING AND CLASSIFICATION EFFICIENCIES OF DIFFERENT CLASSIFICATION MODELS IN THE CATEGORIZATION OF BREAST FILLET MYOPATHIES

AUTHORS: Aftab Siddique[1], Samira Shirzaei[2], Alice E. Smith[2], Jaroslav Valenta[3], Laura J. Garner[1], Amit Morey[1]

[1]Auburn University, Department of Poultry Science, Auburn, AL, United States.

[2]Auburn University, Department of Industrial and Systems Engineering and Department of Computer Science and Software Engineering, Auburn, AL, United States.

[3]Czech University of Life Sciences, Department of Animal Science, Prague, Czech Republic.

(Published in Frontiers in Avian Physiology, doi: http://dx.doi.org/10.3389/fphys.2021.712649)

2.1 **ABSTRACT**

Breast meat from modern fast-growing big-birds is affected with myopathies such as woody breast (WB), white striping (WS) and spaghetti meat (SM). The detection and separation of the myopathy affected meat can be carried out at processing plants using technologies such as the bioelectrical impedance analysis (BIA). However, BIA data from myopathy affected raw is extremely complicated, especially due to the overlap of these myopathies in individual breast fillets and the human error associated with the assignment of fillet categories. Previous research has shown that traditional statistical techniques such as ANOVA and regression, among others, are insufficient in categorizing fillets affected with myopathies using BIA. Therefore, more complex data analysis tools can be used such as, support vector machines (SVM) and backpropagation neural network (BPNN) to classify raw poultry breast myopathies using their BIA patterns, such that the technology can be beneficial for the poultry industry in detecting myopathies. Freshly deboned (3-3.5 h post-slaughter) breast fillets (n=100 x 3 flocks) were analyzed by hand-palpation for WB category (0-normal; 1-mild; 2-moderate; 3-Severe) and SM (presence and absence). BIA data (resistance and reactance) was collected on each breast fillet, the equipment's algorithm calculates protein and fat index. Data were analyzed using linear discriminant analysis (LDA), SVM, and BPNN with 70:30 :: training : test data set. Compared to LDA analysis, SVM separated WB with a higher accuracy of 71.04% for normal (data for normal and mild merged), 59.99% for moderate, 81.48% for severe WB. Compared to SVM, the BPNN training model accurately (100%) separated normal WB fillets with and without SM demonstrating the ability of BIA to detect SM. Supervised learning algorithms such as SVM and

BPNN can be combined with BIA and successfully implemented in poultry processing to detect breast fillet myopathies.

Keywords: Support Vector Machines, Backpropagation Neural Networking, Woody breast, Meat myopathies, Spaghetti meat, Bioelectrical impedance analysis, Machine learning, Artificial intelligence

2.2 INTRODUCTION

Globally, consumers are choosing meat and meat products for their higher nutritional value, especially protein (Heinz and Hautzinger, 2009). There has been a drastic increase in consumption of these products worldwide in the last couple of decades. In developing countries, the per capita consumption of poultry has increased from 1.2 kg in the 1960s to 10.5 kg in the 2000s and will reach up to 14.0 kg by 2030 (FAO, 2003). In the US, more than 9 billion broilers were raised in 2018, with a total live weight of 27.1 billion kg and in 2020 per capita consumption of chicken was 44.23 kg (National Chicken Council, 2020). Chicken is a popular consumer choice because of the various physicochemical and sensorial attributes of texture, color, and flavor (Petracci et al. 2013). To supply the increasing demand for breast meat, breeders have increased growth rate of the birds through genetics, in turn increasing total carcass yield (Petracci and Cavani, 2012). Markets are continuously changing due to consumer's preference and demands, which is presently driving the market towards cut-up chicken parts and further processed products. Fast-growing chickens with increased breast meat yield, have developed breast muscle myopathies, leading to the meat quality defects, such as woody breast (WB). In the past 10 years, WB has been more prominently found in heavier birds (Zampiga et al. 2020). Woody breast affected fillets are characterized by an intricate and dull appearance (Sihvo et al. 2014; Kuttappan et al. 2017), and tough texture due to collagen deposition (Soglia et al. 2016). These breast myopathies also affect meat quality parameters such as pH, color, water holding capacity (WHC), proximate composition, cook loss, and texture, which ultimately influence the quality of further processed products (Kuttappan et al. 2012). Due to the lower meat quality, WB meat is sorted out at the processing plants using manual hand-palpation (**Figure 2-4**) and different grading scales based on severity levels (**Table 2.1**) , however, this method is unreliable and subjective leading to potential misclassification of the breast meat (Morey et al. 2020). By setting specific standards to accurately separate WB fillets, poultry processors will be able to reduce fillet misclassification and ultimately losses related to it.

We investigated bioelectrical impedance analysis (BIA) as a potential objective method to detect WB fillets. Bioelectrical impedance analysis technology has been used in many species to measure physical composition and properties including, body water content and fat content. Nyboer et al. (1950) and Hoffer et al. (1969) introduced the four-electrode, whole body, bioelectrical impedance methods in clinical studies for the measuring bodily fluid from hand to foot. Since its inception, the use of BIA has expanded beyond clinical studies. In the food sector, BIA parameters can be calibrated to specific species and has been used in fish for rapid detection of proximate composition and the pre-harvest condition of fish (Cox et al. 2011). Morey et al. (2020) demonstrated that BIA can be successfully used to detect WB fillets as an alternative to hand-palpation which can reduce classification errors. Further, BIA can potentially detect other overlapping breast muscle myopathies such as spaghetti meat which is affecting the poultry industry.

Classification accuracies of BIA data can be improved through the use of modern data analytics techniques such as machine learning (ML), which includes data mining, artificial neural networks (ANN), deep learning (DL), and artificial intelligence (AI; Tufféry, 2011). Machine learning is a complex field with a wide range of frameworks, concepts, approaches, or a combination of these methods. Machine learning is commonly used in the manufacturing sector for process optimization, tracking and management applications in production and predictive maintenance (Alpaydin, 2010; Gardner and Bicker, 2000). These techniques have been widely applied to enhance quality control in production processes (Apte et al. 1993), particularly in complex production processes where predicting the causes of problems is challenging (Kusiak, 2006). Over the last few decades, automated product inspection systems incorporating ML have been used in a wide variety of food industries such as potato and apple (Tao et al. 1995), oil palm fruit (Abdullah et al. 2002), rice and grains (Carter et al. 2005), beef fat (Chen et al. 2010) and color in bakery applications (Nashat et al. 2011).

The use of machine learning models has increased in recent years due to circumstances such as, the availability of complex data with little accountability (Smola and Vishwanathan, 2008) and will become more critical in the future. Although several ML algorithms are available, such as ANN, support vector machines (SVM), and distributed hierarchical decision trees, their ability to deal with large data sets varies significantly (Bar-Or et al. 2005; Do et al. 2010). In the production sector, only specific ML algorithms are capable of handling high-dimensional data sets and having the ability to deal with high dimensionality is considered a benefit of using ML in the processing industry. One of the main benefits of ML algorithms is finding previously unknown (hidden) information and recognizing its associations in large data sets. The available information criteria can depend mainly on the ML algorithm's characteristics (supervised/unsupervised or Reinforcement Learning [RL]). Nevertheless, the ML method's general process of producing outcomes in a production environment has been conclusively proven (Alpaydin, 2010; Filipič and Junkar, 2000; Guo et al. 2008; Kala, 2012). The use of BIA in poultry processing provides complex data with high dimensionality which can be used to train the SVM algorithms for classification of WB (based on severity) and SM fillets. Support vector machines (SVM), with a kernel-based procedure, has emerged in machine learning, as one strategy for sample classification (Pardo and Sberveglieri, 2005). The implication of SVM in machine learning as a supervised learning technique provides good generalization ability and more minor overfitting tendencies. Using kernel functions in SVM's makes the original input values linearly separable in higher-dimensional space. Moreover, SVMs can simultaneously reduce estimation errors and model dimensions (Singh et al. 2011). The main objective of this research was to determine the accuracy of linear discriminant analysis (LDA), SVM and backpropagation neural networks (BPNN) to classify WB and SM using the multi-dimensional BIA data. The LDA, SVM and BPNN methods are discussed in detail, their accuracies were compared and the reasons for the differences in the classification accuracies are discussed. The research will help the poultry industry, technology companies and scholars to investigate into the use of ML to analyze WB and other myopathy/meat quality related complex datasets.

2.3 MATERIAL AND METHODS

2.3.1 Data Collection

Freshly deboned breast filets from 56-day old broilers (Ross 708) were analyzed in a commercial poultry processing facility after deboning. Breast fillets (n=300, 3 replications or flocks) were randomly selected from the processing line 3 to 3.5 hours post slaughter. Deboned breast fillets were analyzed for WB incidence through hand palpation by an experienced team member (**Figure 2-4**). Breast fillets were classified into normal, mild (for data analysis mild was grouped with normal), moderate, and severe WB fillets (Tijare et al. 2016) and SM presence was evaluated by observing the turgor in the cranial-ventral portion of the breast fillets with a decrease in turgor indicating a presence of SM and an increase in turgor representing an absence of SM. Collected chicken breast fillets from the processing line were subjected to BIA by utilizing a hand-held CQ Reader (Seafood Analytics, Clinton Town, MI, United States; Morey et al. 2020), equipped with 4 spring-loaded electrodes (RJL Systems, Detroit, MI, United States). All 4 electrodes were placed to make contact with the ventral surface of the breast fillet. Once the electrodes were in contact with the breast fillets, the circuit was complete and linked. Then the device measured the data for resistance, reactance, fat index, and protein index and the stored data was downloaded for analysis later (Seafood Analytics Certified Quality Reader, Version 3.0.0.3, Seafood Analytics, MI, United States). Individual weights of fillets were also collected by using weighing balance (OHAUS Corporation, Pine Brook, New Jersey, United States) for the analysis and used to train the SVM and BPNN models.

2.3.2 Linear Discriminant Analysis

Linear discriminant analysis (LDA) is one of the conventional data mining algorithms used in supervised and unsupervised learning contemplated by Fischer (1936) for resolving the issue related to flower classification (Xanthopoulos et al. 2013). The LDA model is used to project an imaginary hyper-plane that minimizes the interclass variance and maximizes the distance between class means. Additionally, it produces a transformation in the data that is discriminative in some data cases (Fukunaga, 2013). LDA is more appropriate for data where unequal within-class frequencies are given, and their classification performances have been randomly examined on generated test data. This approach maximizes the ratio of between-class variance to within-class variance with maximum separability. Data sets used in LDA analysis can be transformed, and related test vectors can be classified in the imaginary hyper-plane by class-dependent transformation and class independent transformation (Balakrishnama and

Ganapathiraju, 1998). The class-dependent transformation approach maximizes the ratio between-class variance to within-class variance. This kind of class transformation helps in maximizing class separability (Tharwat et al. 2017). The main objective for implementing LDA analysis is to create a subspace of lower-dimensional data points compared to the sample data set, in which the original data points from the data set can be easily separable (**Figure 2**; Fisher, 1936). Data analytics and statistical analysis classifiable can be defined as the measure of mean value and variance. The use of LDA provides a solution that can be implemented in a generalized eigenvalue system which provides huge and fast data optimization. The original LDA algorithm was used to solve binary classification taxonomic problems, however, Rao (1948) had also proposed multi-class generalizations. In this paper, both class classification and multi-class case classification derivation were provided to better understand the concept from the simple two-class case (Xanthopoulos et al. 2013).

Let "a1,..., $a_p \in R^m$" be a set of "q" data sets related to the two separate classes, A and B. For each class defined sample means are:

$$a_A = IN_A \sum a \in A_{a,}, \bar{a}B = 1N_B \sum a \in B_{a.} \quad (1.1)$$

N_A, N_B is the total number of samples in data set A and set B. Scatter matrices for the data set by the equations:

$$S_A = \sum a \in A(a - \bar{a}_A)(a - \bar{a}_A)^T, S_B = \sum a \epsilon B(a - \bar{a}_B)\ (a - \bar{a}_B)^T. \quad (1.2)$$

Each of these matrices mentioned above is used for the imaginary hyper-plane, which is defined by the vector (φ), the variance for the calculation is minimal and can be explained by the equation:

$$\text{Min } \varphi(\varphi^T s_A \varphi + \varphi^T s_B \varphi) = \min \varphi\varphi^T(s_A + s_B)\varphi = \min\varphi\varphi^T s_\varphi. \quad (1.3)$$

Where S $=S_A+S_B$ by definition and from equation 1.2, the scatter matrix for supposed two matrixes for the two classes are:

$$s_{AB} = (\bar{a}_A - \bar{a}_B)(\bar{a}_A - \bar{a}_B)^T. \tag{1.4}$$

According to Fisher's projection on LDA for a hyper-plane is the expression to maximize the distance between the means and to minimize the variance of each considered class. Mathematically this can be described by Fisher's criterion equation as:

$$\mathrm{Max}\varphi J(\varphi) = \max \varphi\varphi^T s_{AB\varphi}\varphi^T S_{\varphi.} \tag{1.5}$$

There could be several solutions for the optimization-related problem with the same function value. For a solution φ,∗ all the vectors c·φ∗ will give the same value, and considering no loss in generality, we select only one best possible solution by substituting the denominator with an equality constraint. Then the problem becomes:

$$\mathrm{Max}\ \varphi\varphi^T S_{AB}\varphi, \tag{1.6a}$$

$$\mathrm{s.t} \cdot \varphi^T s_\varphi = 1\ . \tag{1.6b}$$

The Lagrangian mechanism associated with this problem is:

$$\mathrm{LLDA}(a_, \lambda) = \varphi^T S_{AB}\varphi - \lambda(\varphi T S_\varphi - 1) \tag{1.7}$$

where λ is the LaGrange multiplier associated with the equation 1.6b. Since S_{AB} is positive and the nature of the problem is convex, and the global minimum will be at the point for which:

$$\partial \mathrm{LLDA}(x_, \lambda)\ /\ \partial x = 0 \Leftrightarrow S_{AB}\varphi - \lambda S_\varphi = 0\ . \tag{1.8}$$

The optimal φ obtained as the eigenvector that corresponds to the smallest value for the generalized eigensystem:

$$S_{AB}\varphi = \lambda S_{\varphi.} \tag{1.9}$$

Multi-class LDA is only the extension of the two-class classification problem. Given x classes, the matrices will be redefined, and the intra-class matrix becomes:

$$S = S_1 + S_2 + \cdots . S_{n,} \tag{1.10}$$

while the inter-class scatter matrix is annotated by,

$$S_{1,\ldots}n = n\Sigma i = 1pi(\bar{a}_i - \bar{a})(\bar{a}_i - \bar{a})^T, \tag{1.11}$$

where the number of samples (p_i) in the i^{th} class, $\bar{a}_i$ is the mean, and $\bar{a}$ is mean vector given in equation,

$$\bar{a} = 1pn\sum i = 1pi\bar{}ai.$$

The linear transformation φ can be achieved by solving the above equation:

$$S1, \ldots, n\,\varphi = \lambda S\varphi.$$

To achieve a better classification by projection hyper-plane. Once the transformation φ achieved, the class of a new point "y" is determined by:

$$\mathrm{class}(y) = \arg\min n\{d(y\varphi, \bar{}\ an\varphi)\}, \tag{1.12}$$

where $\bar{a}_n$ is the centroid of n^{th} class. The calculation reflects that all classes' centroids were defined first and the unknown points on the subspace defined by φ and the closest class concerning D.

2.3.3 Support Vector Machines

Vapnik (2013) first contemplated the support vector machine in 1995, and recently it has enticed an enormous level of endeavor in the machine learning applications community. Several studies have mentioned that the SVM has immense performance in classification accuracy compared to other data classification algorithm methods (Maji et al. 2008; Shao et al. 2012; Vijayarani et al. 2015). SVM generates a line between the two or more classes known as a hyper-plane for data set classification. Input data Q that can fall on either side of the hyper-plane ($Q^T \bullet W - b$) > 0 are labeled as +1, and those that fall on the other side, ($Q^T \bullet W - b$) < 0, are labeled as -1 (**Figure 1A**; Lee and To, 2010); let {Qi, yi} ∈ R^n be training data set, y_i ∈ {1, -1}, i = 1, 2, …, n.

There exits hyper-plane,

$$P = \{Q \in Rn \mid QT \bullet W + b = 0\}. \qquad (2)$$

The equation for the training data set can be written as:

$$QiT\,W + b \geq 1, yi = 1, \qquad (2.1)$$

$$QiT\,W + b \geq -1, yi = -1.$$

Above mentioned equations can be written as:

$$yi\ (QiT\,W + b - 1) \geq 0.$$

Another definition for the hyper-plane considering P^- and P^+, let {Qi, yi} ∈ R^n be training data set, y iÎ {1, -1}, i = 1, 2,…, n,

$$P+ = \{Q \in Rn \mid QTi\,W + b = 1\}, \qquad (2.2)$$

$$P- = \{Q \in Rn \mid QTi\,W + b = -1\}.$$

The optimization mentioned above is a form non-convex optimization problem that relies on absolute value of |W| and is difficult to solve than convex optimization problems. The equation for W's absolute value can be replaced by using 1/ 2 | |W| $|^2$ without having any change in the final solution. So, the representation of the SVM related problem in quadratic programming (QP) form is as follows (Osuna et al. 1997):

$$\text{Min } 1/2\,|\,|W|\,|2,$$

$$s\,t.\,yi\left(Q_i^T\,W + b\right) - 1 \geq 0, 1 \leq i \leq n\,. \qquad (2.3)$$

After solving the SVM optimization problem using Lagrange multipliers (a_i), the Wolfe dual of the optimization problem was achieved (Craven, 1989):

$$L(w, b) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} aiyi\left[\left(Q_i^T W + b\right) - 1\right]. \qquad (2.4)$$

After solving for the value for W and b,

$$\frac{\partial L(w,b)}{\partial w} = 0, \frac{\partial L(w,b)}{\partial b} = 0, \quad (2.5)$$

the solution in (2.5) is the following condition,

$$w = \sum_{i=1}^{n} \alpha_i yi\, Qi, \quad (2.6)$$

$$\sum_{i=1}^{n} \alpha_i yi = 0,$$

putting the value of 2.6 into equation 2.4, we get the dual form of SVM,

$$W(a) = \sum_{i=1} \alpha_i - \left(\sum_{i=1}^{n} \alpha_i \alpha_j \quad y_i y_j \left(Q_i \cdot Q_j\right)\right) / 2,$$

$$s\,t. \sum_{i=1}^{n} \alpha_i y_i \left(Q_i \cdot Q_j\right) = 0, \quad (2.7)$$

$$0 \leq ai \leq c, i = 1,2, \ldots\ldots n.$$

The number of variables in the equation derived is equivalent to the total number of data cases (n). The training set data with $a_i > 0$ represents the position of support vectors for the classification, and Qi p+ or Q_i p-.

The equation for hyper-plane decision can be written as (Pontill and Verri, 1998):

$$f(x) = \pm\left(\sum_{i=1}^{n} \alpha_i^* yi(q \cdot q_i) - b^*\right). \quad (2.9)$$

Where q is the unknown input data that need to be classified. SVM has been employed in a considerable range of real-world problems associated with the different field of automation, forensics, biotechnology, agriculture statistics, and now is being in the food sciences for the classification of bakery products, fresh produce, and meat product classifications (Liu et al. 2013; Asmara et al. 2017; Chen et al. 2017; Arsalane et al. 2018). It has been proven that SVMs are persistently most appropriate for diverse supervised learning methods. Despite this, the performance of SVM is very receptive to the cost parameter, and kernel frameworks are set. As a result, research industries want to conduct ample cross-validation to determine the most influential parameter setting (Durgesh and Lekha, 2010).

2.3.4 Backpropagation Neural Networking

According to Lippmann (1987), here were no practical algorithms available for interconnecting weight values to achieve an overall minimum training error in multilayer networks. Rumelhart et al. (1986) proposed a generalized rule for backpropagation neural networking, an iterative, gradient descent training procedure. The input data, in the form of the vector, is a pattern to be learned, and the desired output is in the form of a vector produced by the network, upon recall of the input training pattern (Paola and Schowengerdt, 1995). The training's main aim is to minimize the overall error between the test set data and training set data outputs of the network (Paola and Schowengerdt, 1995). Multilayer perceptions are also recognized as BPNN, one of the multiple layers forward neural networks. BPNN comprises of one input layer, one or more hidden layers, and one output layer (Bharathi and Subashini, 2011; Lui et al. 2013). Consideration of distinct factors plays a fundamental role when developing a BPNN that consists of the structure of network, initialization, and switch functions in each hidden and output layer, the training way and algorithm, the learning rate, the error-goal (ε), and preprocessed input data. BPNN has some advantages, such as easy architecture, ease of assembling the mannequin, and fast calculation speed. However, BPNN have some issues, such as (i) possible to contain in local extremum, (ii) poor generalization ability, (iii) lack of strict format packages with a theoretical foundation, and (iv) challenging to manage the learning and training method (Yao, 1999).

In spite of these problems, BPNN has been successfully implemented in a range of fields. Users have applied their experiences and prior knowledge during designing a BPNN to overcome these problems (Lui et al. 2013). A supervised BPNN learning algorithm consists of an input layer, one or more hidden layers, an output layer, and the nodes of hidden layers primarily affect the neural network's classification efficiency (**Figure 3**). Parameters that are required to be defined by the users are learning rate ($0 < \eta < 1$) and momentum ($0 < \eta < 1$).

BPNN program training procedure (Lee and To, 2010; Yang et al. 2011):

1. Design and input for network.
2. Normalize the initial input weights W and threshold values (θ).
3. Define the training and testing data set and input the training matrix X and output matrix Y.
4. Estimate the output vector of each neural synaptic unit.

   (a) Evaluate the output vector (Z) for the hidden layer:

$$\mathrm{net_k} = \Sigma \mathrm{w_{ik} x_i} - \theta_{\mathrm{k}}, \tag{3.1}$$

$$\mathrm{Zk} = \mathrm{f}\,(\mathrm{net\ k}), \tag{3.2}$$

$$\mathrm{net_j} = \Sigma \mathrm{w_{kj} Z_i} - \theta_{\mathrm{j}}, \tag{3.3}$$

$$Y_j = f\,(\mathrm{net}_j). \tag{3.4}$$

   (b) The root of the mean square:

$$\mathrm{RMS} = \sqrt{\frac{\Sigma\left(y_j - T_j\right)^2}{n}}\,. \tag{3.5}$$

5. Estimate distance δ for the output layer and hidden layer from the equation (3.6) and (3.7):

$$\delta_j = \left(T_j - y_i\right) - f\left(\mathrm{net}_j\right), \tag{3.6}$$

$$\delta_k = \left(\sum\nolimits_j \delta_j w_{kj}\right) - f'^{\left(\mathrm{net}_j\right)}. \tag{3.7}$$

6. Evaluate modifications for initial weights (W) and distance (δ) (η is the learning rate, α is the momentum) for both output layer (equation 3.8, and 3.9) and hidden layer (equation 3.10, and 3.11):

$$\Delta w_{kj(n)} = \eta\delta_j z_k + \alpha\Delta w_{Kj}(n-1), \quad (3.8)$$

$$\Delta\theta_j(n) = -\eta\delta_j + \alpha\Delta\theta_j(n-1), \quad (3.9)$$

$$\Delta w_{ik(n)} = \eta\delta_j X_i + \alpha\Delta w_{ik}(n-1), \quad (3.10)$$

$$\Delta\theta_k(n) = -\eta\delta_k + \alpha\Delta\theta_k(n-1). \quad (3.11)$$

7. Redefine initial weight (W) and the threshold value (θ), redefine W and θ of the output and hidden layer:

$$w_{kj}(p) = w_{kj}(P-1) + \Delta w_{k_j}, \quad (3.12)$$

$$\theta_j(p) = \theta_j(p-1) + \Delta\theta_j, \quad (3.13)$$

$$w_{ik}(p) = w_{ik}(P-1) + \Delta w_{ik}, \quad (3.14)$$

$$\theta_k(p) = \theta_k(p-1) + \Delta\theta_k. \quad (3.15)$$

After modifying output and hidden layer, the steps will be renewed, and the step from 3-7 will be repeated until converge.

BPNN program-testing process (Lee and To, 2010, Yang et al. 2011):

1. Input parameters related to the network.
2. Input the initial weights (W) and the threshold value (θ).
3. Unknown data entry for data matrix X.
4. Evaluate output vector (Z) for the output and hidden layer:

$$net_k = \Sigma W_{ik} x_i - \theta_k, \quad (3.16)$$

$$z_k = f(net_k), \quad (3.17)$$

$$net_j = \Sigma w_{kj} z_i - \theta_j, \quad (3.18)$$

$$Y_j = f(net_j). \quad (3.19)$$

2.4 STATISTICAL ANALYSIS

The collected data were analyzed using SAS software (Version 9.4) for linear discriminant analysis using proc discrim to classify data. For the analysis SVM and BPNN of collected data, R software (Version 4.0.0, Arbor Day) was used by using the caret package in the analysis to classify various chicken breast fillet myopathies. The data sets collected for the different conditions were divided into 70::30 training set and testing set. The caret package algorithm calculated the best-suited tuning parameter or value of cost (C) for both training and test data sets. A seed value was set for 3,000 for the SVM analysis. For BPNN classification of fillets, Neural net and BBmisc packages were used to classify the collected data sets (WB and SM), and the data sets were divided into 70::30 training and testing data sets. Low learning rate (0.01), the threshold value (0.01), number of maximum steps (10,000), and 4 hidden layers were used in the BPNN classification algorithm for the analysis.

2.5 RESULTS

The classification experiment was conducted on two different data sets containing WB fillets classified as normal, moderate, severe, and the data containing a classification of normal fillets with and without SM. All data sets were analyzed by LDA, SVM and BPNN algorithms for classification of WB and SM conditions. In SVM classification experiments, 10-fold cross-validation with 3 replications was used to determine the best-suited value of cost (C), linear kernel function, and various combinations of these parameters defined the data classification problem-related prediction accuracy. The results obtained from the SVM classification algorithm

(**Table 2.2**) showed that the percent accuracy for the classification of normal WB fillets in the training data was 63.86% and in the test data set was 71.04%, for moderate fillet classification the training set showed an accuracy of 49.88% and 59.99% accuracy for test data set, and classification accuracy for the severe WB fillets for the training set was 71.78% and the test data set was 81.48%. Compared to SVM, the results obtained from BPNN classification analysis showed many misclassified data points (lower classification accuracy) for different WB severity levels (**Table 2.2**). However, the BPNN classification analysis of the SM data show a classification accuracy of 100% in the training set (**Table 2.2**). Due to fewer observations, the test data set classifications only show a 52.94% accuracy for normal without SM fillets and a 75% classification accuracy for normal with SM fillets (**Table 2.2**). Results obtained from LDA (**Table 2.2**) classification, the training set data showed a classification efficiency of 72.31% for normal, 43.75% for moderate, and 75% for severe WB breast fillets, and the validation (or test) set of data underperformed and showed lower classification efficiency for normal (52.63%), moderate (29.41%), and severe WB breast fillets (59.09%).

2.6 DISCUSSION

Using only visual and hand palpation characteristics to identify WB and SM muscle myopathies poses various challenges when classification is performed on a processing line, such as misclassifications, processing inefficiencies and an increase in labor costs. Woody breast is found primarily in the superficial area of the breast fillet and many times includes the visual presence of surface hemorrhages, a light-yellow surface appearance, a rigid bulged fillet, and by mechanically palability of the muscle (**Figure 2-5**; Mazzoni et al. 2015; Mudalal et al. 2015). Additionally, normal breast fillets have smaller cross-sectional areas as compared to WB fillets (Huang and Ahn, 2018), with higher collagen content and elevated post processing pH (Petracci et al. 2015; Chatterjee et al. 2016; Clark and Velleman, 2016; Soglia et al. 2016). SM, on the other hand, is related to immature intramuscular connective tissues in the breast meat, and it has a lower muscular cohesion that breast meat from unaffected fillets (**Figure 2-6**; Bowker and Zhuang, 2016; Radaelli et al. 2017; Sihvo et al. 2017). The thickness of connective tissues in the breast fillets showing SM decreases gradually in the endomysium and perimysium, causing the different muscle fibers to deteriorate or have mushy texture (Baldi et al. 2018). Therefore, using an assortment of already available complex data, we were able to make improvements to the classification of fillets among the WB and SM myopathies.

Results obtained by training accuracy for LDA (70::30) classification was 72.31%, 43.75%, and 75.00% for normal, moderate and severe WB (**Table 2.2**) fillet classification, respectively using the BIA and fillet weight data set (n=300). The testing set was lower in accuracy than our training set with only 52.63% normal classified, 29.41% moderate classified, and 59.09% severe WB classified (n=300; **Table 2.2**). The testing set data was lower in accuracy compared to the training set data, possibly due to low sample size and non-linear data set. The non-linear data is likely due to human error during the manual hand-palpation of the breast fillets, however, in future studies larger data sets could be implemented to increase the accuracy of the BIA method combined with conventional algorithms. Morey et al. (2020) also used LDA (60::40) with a BIA data set (n=120) and reported 68.69 – 70.55% accuracy for normal fillets and 54.42 – 57.75% accuracy for severe WB fillets classification in the testing set. Wold et al. (2019) analyzed a near infrared spectroscopy (NIR) data set (n=102) using a LDA (50::50) classification algorithm with 100% accuracy for fillet classification in the training set and 96% accuracy in the testing set, for a rapid on-line detection method for the WB myopathy in processing plants. LDA is a well-recognized technique to reduce the dimensionality of data in a dataset. However, LDA can only be used for single-label multi-class categorizations and cannot explicitly be extended to multi-label multi-class classification systems. The LDA technique is used to convert high-dimensional data into a low-dimensional data space, maximizing the ratio of between-class variation to within-class variance, thereby ensuring optimal class separation (Pan et al. 2014). The LDA technique works by projecting the initial data matrix onto a lower-dimensional region. For the dimensionality reduction, three steps were required: (i) the inter-class difference or between-class matrix is used to measure the separability across multiple categories (i.e., the distance between the means of different classes), (ii) the within-class variance, also known as the within-class matrix, is calculated as the difference between the mean and the class samples, and (iii) the creation of a lower-dimensional space that maximizes between-class variance while minimizing within-class variance (Mandal et al. 2009). In our current research and Morey et al. (2020), the low performance of data collected and analyzed using LDA compared to the data collected may have two key factors: small sample size and data linearity issues. Su et al. (2017) also found low performance in data sets with small sample size and non-linear data.

The LDA technique is used to find a linear transformation that discriminates between various groups. However, LDA cannot find a lower-dimensional space if the groups are non-linearly separable. In other words, where discriminatory knowledge is not in the means of classes, LDA fails to locate the LDA space. One of the significant issues with the LDA methodology is the singularity, also known as small sample size or under-sampling. This issue arises due to high-dimensional trend classification problems or a low number of training samples available for each class compared to the sample space's dimensionality (Huang et al. 2002; Lu et al. 2005; Zhuang and Dai, 2005; Su et al. 2017; Tharwat et al. 2017).

The ML theory lays the groundwork for SVM, this algorithm has gained widespread attention because of its unique performance efficiency, ability to accomplish pinpoint accuracy, and managing high-dimensional, multi-variate data sources. Cortes and Vapnik (1995) implemented SVMs as a new ML technique for two-group classification problems. Researchers have reported that SVMs are economical, sensitive, and easy to use classifier that can be implemented in organized evaluation assignments. Inspection of large collected data sets during production is a significant application of SVM (Burbidge et al. 2001; Chinnam, 2002). SVM is frequently used in various food production environments, including a product monitoring systems, mechanical fault detection, and dimensional accuracy (Ribeiro, 2005; Azadeh et al. 2013; Salahshoor et al. 2011; Çaydaş and Ekici, 2012). SVMs are used in different processing areas, including drug designing and discovery, surgery, and cancer treatment, in addition to the food product processing industry (Vapnik, 2013). Product quality control (Borin et al. 2006), polymer recognition, and other applications are also possible (Li et al. 2009) areas that SVM can be incorporated. These examples from different industries demonstrate that the SVM algorithms have a broad range of applicability and versatility (Kotsiantis et al. 2007). Present research demonstrates the ability for the implementation of SVM and BPNN in combination with BIA and fillet weight data to classify WB and SM fillets.

Statistical Learning Theory (SLT) is a robust and an appropriate supervised learning algorithm for production research problems. Under SLT, the algorithmic learning allows it to use an achieving function, representing the relationship between different components without being directly connected (Evgeniou et al. 2000). The algorithm enquires about the problem concerning how well the selected method resolves the problem, and accuracy prediction performance for previously unknown inputs, is the subject of SLT (Evgeniou et al. 2000). A few more realistic techniques, such as autoencoders, SVM, and Bayesian optimization, are based on the theories of SLT (Battiti et al. 2002). SVM is considered a mathematical expression in its most basic form, a method (or algorithm) for optimizing alphanumeric equations, with a given set of data (Noble, 2006).

The fundamental idea of SVM algorithmic expression can be easily understood by four fundamental concepts: (i) the imaginary hyper-plane, (ii) the margin of hyper-plane, (iii) the soft margin, and (iv) the kernel function (Tharwat, 2019). A solid line splits the region in half in two dimensions (**Figure 2-1A**), but we require a hypothetical plane to split the area into three dimensions. A hyper-plane is a collective term for a straight line in a high-dimensional region, and the dividing hyper-plane is the line that separates the pieces of data (Kecman, 2001; Tharwat, 2019). The SVM, on the other hand, differs from other hyper-plane-based classifiers based on how the hyper-plane is chosen. Consider the grouping shown in **Figure 2-1**. By implementing SLT, it is easier to find the best possible plane to create the hyper-plane that will be used in the data classification (Vapnik, 1963). The capability of the SVM to classify the correct data points between given classes can be improved by using an imaginary hyper-plane in the space. The SLT theorem implies that the data used to train the SVM originates from the same data set as the data used to test it. For example, if a SVM algorithm is trained on sensory property of product cannot be used to train the data collected for subjective response of consumers. Furthermore, we cannot expect the SVM to work well if training is conducted with a SM breast fillet data set, but a WB data set is used for testing. At the same time, the SLT principle does not assume two data sets came from the same class of distributions. For example, an SVM does not assume that the training data values follow a normal distribution.

For better understanding of SVM and its function, we have concluded an imaginary data that for classes A and B which can be divided using a straight line. When the values in a data set are closer together or intersected (**Figure 2-1B**), the SVM will manage this overlapping of data by inserting a soft margin. In essence, this causes specific data points to pass across the dividing hyper-plane's margin, without influencing the outcome. Use of the soft margin provides the solution to the problem of misclassification (shown in **Figure 2-1B**) by considering the data point as outlier (shown in **Figure 2-1C**). Another essential function for the SVM classification is the kernel function (shown in **Figure 2-1D** and 2-**1E**), a mathematical trick that allows the SVM to perform a two-dimensional classification of a one-dimensional data set. In general, a kernel function projects data from a low-dimensional space to a space of higher dimension.

SVM classification efficiency (**Table 2.2**) for the separation of high dimensionality data showed better classification efficiency for normal (training efficiency 63.86%, testing efficiency 71.04%), moderate WB (training efficiency 49.88 %, testing efficiency 59.99%), and severe WB (training efficiency 71.78%, testing efficiency 81.48 %) compared to the LDA algorithm used by Morey et al. (2020). The BIA and fillet weight data set used in training the SVM performed well due to the higher dimensionality of the data set. When data is highly dimensional and the sample sets are relatively small, SVM analysis is more accurate to classify data and has been used by other authors to help classify multi-dimensional data. Barbon et al. (2018) used a relatively small data set (n=158) of NIR results combined with SVM (75::25) to classify normal and pale meat as it relates to pale, soft and exudative poultry breast meat. They demonstrated the use of SVM as a classification tool for breast fillets with muscle myopathies where classification accuracy for normal fillets was 53.4% and 72.0% for pale fillets. Geronimo et al. (2019) using an NIR system equipped with an image acquisition system found 91.83% classification efficiencies (fillet images) using SVM to analyze a WB fillet sample set (sample size is unclear) using a 70::30 model. These researchers also used multilayer perceptron (a feed forward network differing from the back propagation network in BPNN) to classify the data set and classification accuracy was 90.67% for WB. Yang et al. (2021) analyzed images derived from expressible fluid of breast meat to classify WB using SVM (training and testing ratio is unreported) and DL (training to testing is 2 to 1). These researchers found fewer classification efficiencies for SVM algorithms in the testing set (38.25 - 63.89%), compared to the training set (40.41 - 81.94%) for 3 out of the 4 SVM classification methods used. In their DL classification (a type of ANN) to classify WB and reported 100% accuracy in the training set 100% and 93.30% accuracy in the testing set

Connection of random different nodes or units in a computing system to solve the problems that are impossible to solve by conventional statistical methods are known as artificial neural network and are based on the human brain circuitry. When applied to a processor framework, the subconscious network can execute unique functions (perception, speech synthesis, image recognition), which have proven to be useful in industrial applications (Alpaydin, 2010). Neural networks allow an automated artificial skill to operate unsupervised reinforcement and classification algorithms functions (neural networks) by simulating the central nervous platform's decentralized "data analysis" capabilities through neural networks (Pham and Afify, 2005; Corne et al. 2012). Decentralization employs many necessary, interconnected neurons or nodes and the capacity to process data through the complex response of these endpoints and their links to exogenous variables (Akay, 2011). These algorithms are crucial in today's modern machine learning development (Nilsson, 2005) and can be classified into two categories: interpretation and algorithm. Neural networks are used in a variety of industrial sectors for a range of problems (Wang et al. 2005) e.g., process control emphasizing their key benefit and overall predictive validity (Pham and Afify, 2005). However, ANN (similar to SVM) requires a large sample size to attain maximum precision (Kotsiantis et al. 2007). Overfitting, which is linked to high-variance implementations, is universally acknowledged as a disadvantage of the ANN Algorithm (Kotsiantis et al. 2007). Other difficulties with using neural networks include the sophistication of the generated models, the aversion for missing values, and often the lengthy data set training method (Kotsiantis et al. 2007; Pham and Afify, 2005).

For BPNN, the data was pre-processed and consisted of just two dimensions with a lower level of classification complexity (Panchal et al. 2011). Classification efficiencies for the WB fillets using BPNN (**Table 2.2**) shows that testing data set for normal (47.77%), and moderate fillets (23.33%) did not perform well, compared to the classification efficiency for severe WB fillets (28.88%). BPNN classification algorithm for the WB fillets did not perform well due to the complexity of the data after the pre-processing, and overfitting of the learning model due to uneven distribution of weight on the input neuron layer. The BPNN classification algorithm for the SM data set (**Table 2.2)** performed well for the training data set for normal (training 100%, testing 52.95%) and SM fillets (training 100%, testing 75.00%), however, due to the complexity of pre-processed data, overfitting of BPNN, and small data set the classification efficiency of the testing set was lower than the training set. These studies all use SVM and ANN algorithms to classify small sample data sets, where the results always show that the accuracy in the training set data was higher than the testing set data, indicating that the training of the model is not performing well. Collection of larger data set for the supervised learning methods of classification provides the chances for getting lower error rates and better learning ability for the machine learning algorithms.

In BPNN, the input data vector represents the pattern to be trained, and the output data vector represents the optimal set of output values that the network can generate when the training pattern is recalled. The aim of BPNN training is to reduce the total error between the network's expected and real outputs (Panchal et al. 2011). To generate a reduction in error, the residual differences in the weights at each iteration must be unmeasurable. A learning rate metric, which reflects the rate of the move taken toward minimal error, must be defined in order to accomplish a reasonable training period. Learning will take too much time if this amount is too small, and if it is too high, the loss function will degenerate and errors will rise (Ganatra et al. 2011). When using neural networks to analyze woody breast data, overlearning or overfitting happened when the algorithm took too long to run, and the network was too complicated for the problem or the amount of data available. Whereas, to classify SM in a group of fillets, BPNN was used, and data is processed differently.

2.7 CONCLUSIONS

This project demonstrates the application of ML in poultry production processes to categorize chicken breast fillets into groups based on severity of myopathy. The use of SVM and BPNN can be combined with BIA and fillet weight data to more accurately classify breast fillet myopathies, such as, WB and SM from normal breast fillets in real-time on-line, compared to the subjective hand palpation method. With the implementation of other meat quality parameters, such as water content, classification accuracy of the SVM and BPNN could be improved. To obtain a well-trained model for classification efficiency and to reduce overfitting and underfitting problems related to classification, future research will include larger data sets for breast fillet myopathies to avoid the overlapping of conditions caused by human error in the sorting of the fillets. The innovative combination of these tools has the potential to improve poultry processing efficiencies and downgrades of breast fillets affected by undesirable myopathies, while reducing customer complaints.

2.8 CONTRIBUTION TO THE FIELD STATEMENT:

There has been a drastic increase in consumption of animal protein worldwide in the last couple of decades. Consumption per capita had increased from 10 kg in the 1960s to 26 kg in the 2000s and is predicted to reach 37 kg by 2030. The high demand for chicken meat is due to physicochemical and sensory characteristics including texture, color, and taste. Hand-palpation is the only low cost tool for categorizing the severity of WB fillets, however, it is arbitrary, problematic, and has a large error in classifications. Industries must now develop strict criteria for improved fillet sorting in order to minimize losses due to misidentification. Machine learning has been used effectively in process optimization, output monitoring and control, and predictive maintenance. These methods of artificial intelligence have been commonly used to improve quality management in manufacturing processes in other sectors of the food industry. Through this investigation, supervised machine learning techniques, such as, SVM and BPNN applications have a strong ability to accurately classify breasts fillets into myopathy categories.

2.9 REFERENCES

Abdullah, M. Z., Guan, L. C., Mohamed, A. M. D., and Noor, M. A. M. (2002). Color vision system for ripeness inspection of oil palm Elaeis guineensis. J. Food Process Preserv. 26, 213-23.

Akay, D. (2011). Grey relational analysis based on instance-based learning approach for classification of risks of occupational low back disorders. Saf. Sci. 49, 1277-1282.

Alpaydin, E. (2010). Introduction to machine learning. Cambridge, MA: MIT Press.

Apte, C., Weiss, S., and Grout, G. (1993). Predicting defects in disk drive manufacturing: A case study in high-dimensional classification. Artificial intelligence for applications. Abstract retrieved from Proceedings of 9th IEEE conference.

Arsalane, A., El Barbri, N., Tabyaoui, A., Klilou, A., Rhofir, K., and Halimi, A. (2018). An embedded system based on DSP platform and PCA-SVM algorithms for rapid beef meat freshness prediction and identification. Comput. Electron Agric. 152, 385-392.

Asmara, R. A., Rahutomo, F., Hasanah, Q., and Rahmad, C. (2017). Chicken meat freshness identification using the histogram color feature. Sustainable Information Engineering and Technology (SIET). Abstract retrieved from IEEE.

Azadeh, A., Saberi, M., Kazem, A., Ebrahimipour, V., Nourmohammadzadeh, A., and Saberi, Z. (2013). A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyper-parameters optimization. Appl. Soft Comput. 13, 1478-1485.

Baldi, G., Soglia, F., Mazzoni, M., Sirri, F., Canonico, L., Babini, E., and Petracci, M. (2018). Implications of white striping and spaghetti meat abnormalities on meat quality and histological features in broilers. Animal, 12, 164-173.

Balakrishnama, S., and Ganapathiraju, A. (1998). "Linear discriminant analysis-a brief tutorial," in Institute for Signal and Information Processing (Piscataway, NJ: IEEE), 1-8.

Battiti, R., Brunato, M., and Villani, A. (2002). Statistical learning theory for location fingerprinting in wireless LANs. http://eprints.biblio.unitn.it/238/ Accessed on [15 April, 2021].

Barbon, S., Costa Barbon, A. P. A. D., Mantovani, R. G., and Barbin, D. F. (2018). Machine learning applied to near-infrared spectra for chicken meat classification. J. Spectrosc. 2018. https://doi.org/10.1155/2018/8949741

Bar-Or, A., Schuster, A., Wolff, R., and Keren, D. (2005). Decision tree induction in high dimensional, hierarchically distributed databases. Society for Industrial and Applied Mathematics. Abstract retrieved from Proceedings SIAM International.

Bharathi, P. T., and Subashini, P. (2011). Optimization of image processing techniques using neural networks: a review. WSEAS Trans. Inf. Sci. Appli. 8, 300-328.

Borin, A., Ferrao, M. F., Mello, C., Maretto, D. A., and Poppi, R. J. (2006). Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. Anal. Chim. Acta. 579, 25-32.

Bowker, B. and Zhuang, H. (2016). Impact of white striping on functionality attributes of broiler breast meat. Poult Sci., 95, 1957–1965. doi: 10.3382/ps/pew115

Burbidge, R., Trotter, M., Buxton, B., and Holden, S. (2001). Drug design by machine learning: support vector machines for pharmaceutical data analysis. Comput. Chem. 26, 5-14.

Carter, R. M., Yan, Y., and Tomlins, K. (2005). Digital imaging-based classification and authentication of granular food products. Meas. Sci. and Technol. 17, 235.

Çaydaş, U., and Ekici, S. (2012). Support vector machines models for surface roughness prediction in CNC turning of AISI 304 austenitic stainless steel. J. Intell. Manuf. 23, 639 650.

Chatterjee, D., Zhuang, H., Bowker, B.C., Rincon, A.M., and Sanchez-Brambila G (2016). Instrumental texture characteristics of broiler *Pectoralis major* with the wooden breast condition. Poult Sci. 95, 2449–2454. doi: 10.3382/ps/pew204

Chen, K., Sun, X., Qin, C., and Tang, X. (2010). Color grading of beef fat by using computer vision and support vector machine. Comput. Electron. Agr. 70, 27-32.

Chen, C. P., and Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Inf. Sci. 275, 314-347.

Chen, Z., Cao, S., and Mao, Z. (2017). Remaining useful life estimation of aircraft engines using a modified similarity and supporting vector machine (SVM) approach. Energies. 11, 1 14.

Chinnam, R. B. (2002). Support vector machines for recognizing shifts in correlated and other manufacturing processes. Int. J. Prod. Res. 40, 4449-4466.

Clark, D.L., and Velleman, S.G. (2016). Spatial influence on breast muscle morphological structure, myofiber size, and gene expression associated with the wooden breast myopathy in broilers. Poult Sci. 95, 2930–2945. doi: 10.3382/ps/pew243

Corne, D., Dhaenens, C., and Jourdan, L. (2012). Synergies between operations research and data mining: The emerging use of multi-objective approaches. Eur. J. Oper. Res. 221, 469-479.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. Mach. Learn. 20, 273-297.

Cox, M. K., Heintz, R., and Hartman, K. (2011). Measurements of resistance and reactance in fish with the use of bioelectrical impedance analysis: sources of error. Fishery Bulletin, 109, 34-47.

Craven, B. D. (1989). A modified Wolfe dual for weak vector minimization. Numer. Funct. Anal. Optim. 10, 899-907.

Do, T. N., Lenca, P., Lallich, S., and Pham, N. K. (2010). Classifying very-high-dimensional data with random forests of oblique decision trees. In Advances in knowledge discovery and management. Berlin, Heidelberg: Springer.

Durgesh, K. S., and Lekha, B. (2010). Data classification using support vector machine. J. Theor. Appl. 12, 1-7.

Evgeniou, T., Pontil, M., and Poggio, T. (2000). Statistical learning theory: A primer. Int. J. Comput. Vis. 38, 9-13.

FAO (2003). Livestock commodities. http://www.fao.org/3/y4252e/y4252e00.htm#TopOfPage [Accessed April 17, 2021].

Filipič, B., and Junkar, M. (2000). Using inductive machine learning to support decision making in machining processes. Comput. Ind. 43, 31-41.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Ann. Eugen. 7, 179-188.

Fukunaga, K. (2013). Introduction to statistical pattern recognition. San Diego: Academic Press, INC.

Gardner, R., and Bicker, J. (2000). Using machine learning to solve tough manufacturing problems. Int. J. Ind. Eng. Theory Appl. Pract. 7, 359-364.

Ganatra, A., Kosta, Y. P., Panchal, G., and Gajjar, C. (2011). Initial classification through back propagation in a neural network following optimization through GA to evaluate the fitness of an algorithm. *Int. J*. Comput. Sci. Appl. 3, 98-116.

Geronimo, B. C., Mastelini, S. M., Carvalho, R. H., Júnior, S. B., Barbin, D. F., Shimokomaki, M., and Ida, E. I. (2019). Computer vision system and near-infrared spectroscopy for identification and classification of chicken with wooden breast, and physicochemical and technological characterization. Infrared Phys. Technol. 96, 303-310.

Guo, X., Sun, L., Li, G., and Wang, S. (2008). A hybrid wavelet analysis and support vector machines in forecasting development of manufacturing. Expert Syst. Appl. 35, 415-422.

Hoffer, E. C., Meador, C. K., and Simpson, D. C. (1969). Correlation of whole-body impedance with total body water volume. J. Appl. Physiol. 27, 531-534.

Huang, X., and Ahn, D. U. (2018). The incidence of muscle abnormalities in broiler breast meat A review. Korean J. Food Sci. Anim. Resour. 38, 835.

Huang, R., Liu, Q., Lu, H., and Ma, S. (2002). "Solving the small sample size problem of LDA" in Object recognition supported by user interaction for service robots (Piscataway, New Jersey: IEEE), 29-32.

Jiang, H., Yoon, S. C., Zhuang, H., Wang, W., and Yang, Y. (2017). Evaluation of factors in development of Vis/NIR spectroscopy models for discriminating PSE, DFD and normal broiler breast meat. Br. Poult. Sci. 58, 673-680.

Kala, R. (2012). Multi-robot path planning using co-evolutionary genetic programming. Expert Syst. Appl. 39, 3817-3831.

Kecman, V. (2001). Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models. Massachusetts: MIT Press.

Kotsiantis, S. B., Zaharakis, I., and Pintelas, P. (2007). "Supervised Machine Learning: A Review of Classification Techniques" in Emerging artificial intelligence applications in computer engineering. (Amsterdam: IOS Press), 3-24.

Kuttappan, V., Lee, Y., Erf, G., Meullenet, J., McKee, S., and Owens, C. (2012). Consumer acceptance of visual appearance of broiler breast meat with varying degrees of white striping. Poul. Sci. J. 91, 1240–1247. doi: 10.3382/ps.2011-01947

Kuttappan, V. A., Owens, C. M., Coon, C., Hargis, B. M., and Vazquez-Anon, M. (2017). Incidence of broiler breast myopathies at 2 different ages and its impact on selected raw meat quality parameters. Poult. Sci. J. 96, 3005-3009.

Kusiak, A. (2006). Data Mining in Manufacturing: A Review. J. Manuf. Sci. Eng. 128, 969

Lee, M. C., and To, C. (2010). Comparison of support vector machine and back propagation neural network in evaluating the enterprise financial distress. IJAIA, 31-43.

Li, H., Liang, Y., and Xu, Q. (2009). Support vector machines and its applications in chemistry. Chemom. Intell. Lab. Syst. 95, 188-198.

Lippmann, R. (1987). "An introduction to computing with neural nets" in IEEE Assp magazine. (New York: IEEE), 4-22.

Liu, M., Wang, M., Wang, J., and Li, D. (2013). Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. Sensors Actuat. B Chem. 177, 970-980.

Lu, J., Plataniotis, K. N., & Venetsanopoulos, A. N. (2005). Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition. Pattern Recognit. Lett. 26, 181-191.

Mandal, T., Wu, Q. J., and Yuan, Y. (2009). Curvelet based face recognition via dimension reduction. Signal Process. 89, 2345-2353.

Mazzoni, M. Petracci, M. Meluzzi, A. Cavani, C. Clavenzani, P. and Sirri, F. (2015). Relationship between *Pectoralis major* muscle histology and quality traits of chicken meat. Poult Sci. 94, 123-30. doi: 10.3382/ps/peu043

Maji, S., Berg, A. C., and Malik, J. (2008). Classification using intersection kernel support vector machines is efficient. Computer vision and pattern recognition. Abstract retrieved from IEEE.

Monostori, L. (2003). AI and machine learning techniques for managing complexity, changes and uncertainties in manufacturing. Eng. Appl. Artif. Intell. 16, 277-291.

Morey, A., Smith, A. E., Garner, L. J., and Cox, M. K. (2020). Application of Bioelectrical Impedance Analysis to Detect Broiler Breast Filets Affected with Woody Breast Myopathy. Front. Physiol. 11, 808. doi: 10.3389/fphys.2020.00808

Mudalal, S., Lorenzi, M., Soglia, F., Cavani, C., and Petracci, M. (2015). Implications of white striping and wooden breast abnormalities on quality traits of raw and marinated chicken meat. Animals. 9, 728–734. doi: 10.1017/S175173111400295X

National Chicken Council (2020). Per Capita consumption of Poultry and Livestock, 1965 to estimated 2021. https://www.nationalchickencouncil.org/about-the-industry/statistics/per capita-consumption-of-poultry-and-livestock-1965-to-estimated-2021-in-pounds/ [Accessed March 30, 2021].

Nashat, S., Abdullah, A., Aramvith, S., and Abdullah, M. Z. (2011). Support vector machine approach to real-time inspection of biscuits on moving conveyor belt. Comput. Electron. Agr. 75, 147-158.

Nilsson, N. J. (2005). Introduction to machine learning. Standford: Standford University Press.

Noble, W. S. (2006). What is a support vector machine? Nat. Biotechnol., 24, 1565-1567.

Nyboer, J., Kreider, M. M., and Hannapel, L. (1950). Electrical impedance plethysmography: A physical and physiologic approach to peripheral vascular study. Circulation. 2, 811-821.

Osuna, E., Freund, R., and Girosi, F. (1997). “An improved training algorithm for support vector machines”. In Neural networks for signal processing VII. (New York : IEEE), 276-285.

Pan, F., Song, G., Gan, X., and Gu, Q. (2014). Consistent feature selection and its application to face recognition. J. Intell. Inf. Syst. 43, 307-321.

Panchal, G., Ganatra, A., Shah, P., and Panchal, D. (2011). Determination of over-learning and over-fitting problem in back propagation neural network. Int. J. Soft Comput. 2, 40-51.

Paola, J. D., and Schowengerdt, R. A. (1995). A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. Int. J. Remote Sens. 16, 3033-3058.

Pardo, M., and Sberveglieri, G. (2005). Classification of electronic nose data with support vector machines. Sensors Actuat. B-Chem. 107, 730-737.

Petracci, M., Bianchi, M., Mudalal, S. and Cavani, C. (2013). Functional ingredients for poultry meat products. Trends Food Sci. Technol. 33, 27-39.

Petracci, M., and Cavani, C. (2012). Muscle growth and poultry meat quality issues. Nutrients. 4, 1-12.

Petracci, M., Mudalal, S., Soglia, F., and Cavani, C. (2015). Meat quality in fast-growing broiler chickens. Worlds Poult. Sci. J. 71, 363–374. doi: 10.1017/S0043933915000367

Pham, D. T., and Afify, A. A. (2005). Machine-learning techniques and their applications in manufacturing. Proc. IMechE., Part B: J. Eng. Manuf. 219, 395-412.

Pontil, M., and Verri, A. (1998). Properties of support vector machines. Neural Comput. 10, 955 974. doi: https://doi.org/10.1162/089976698300017575

Radaelli, G., Piccirillo, A., Birolo, M., Bertotto, D., Gratta, F., Ballarin, C., Vascellari, M., Xiccato, G., and Trocino, A. (2017). Effect of age on the occurrence of muscle fiber degeneration associated with myopathies in broiler chickens submitted to feed restriction. Poult. Sci. 96, 309–319. doi: 10.3382/ps/pew270

Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. J. R. Stat. Soc. Series B Stat. Methodol. 10, 159-193.

Ribeiro, B. (2005). Support vector machines for quality monitoring in a plastic injection molding process. IEEE T. Syst. Man. Cy. C. 35, 401-410.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). "Learning internal representations by error propagation" in Parallel Distributed Processing: Explorations in the Microstructures of Cognition, (Cambridge, MA: MIT Press) 318-362.

Salahshoor, K., Khoshro, M. S., and Kordestani, M. (2011). Fault detection and diagnosis of an industrial steam turbine using a distributed configuration of adaptive neuro-fuzzy inference systems. Simul. Model Pract. Th. 19, 1280-1293.

Sihvo, H., Immonen, K., and Puolanne, E. (2014). Myodegeneration with fibrosis and regeneration in the pectoralis major muscle of broilers. Vet. Pathol. 51, 619–623. doi:10.1177/0300985813497488

Sihvo, H. K., Lindén, J., Airas, N., Immonen, K., Valaja, J., and Puolanne, E. (2017). Wooden breast myodegeneration of *Pectoralis major* muscle over the growth period in broilers. Vet. Pathol. 54, 119–128. doi: 10.1177/0300985816658099

Shao, Y., and Lunetta, R. S. (2012). Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. ISPRS J. Photogramm. Remote Sens. 70, 78-87.

Singh, K. P., Basant, N., and Gupta, S. (2011). Support vector machines in water quality management. Anal. Chimica Acta. 703, 152-162.

Smola, A., and Vishwanathan, S. V. N. (2008). Introduction to machine learning. Cambridge: Cambridge University Press.

Soglia, F., Mudalal, S., Babini, E., Di Nunzio, M., Mazzoni, M., Sirri, F., and Petracci, M. (2016). Histology, composition, and quality traits of chicken Pectoralis major muscle affected by wooden breast abnormality. Poult. Sci. J. 95, 651-659.

Su, B., Ding, X., Wang, H., and Wu, Y. (2017). Discriminative dimensionality reduction for multi-dimensional sequences. IEEE. PAMI. 40, 77-91. doi: 10.1109/TPAMI.2017.2665545

Tao, Y., Heinemann, P. H., Varghese, Z., Morrow, C. T., and Sommer III, H. J. (1995). Machine vision for color inspection of potatoes and apples. Trans ASAE, 38, 1555-1561.

Tharwat, A. (2019). Parameter investigation of support vector machine classifier with kernel functions. Knowl. Inf. Syst. 61, 1269-1302.

Tharwat, A., Gaber, T., Ibrahim, A., and Hassanien, A. E. (2017). Linear discriminant analysis: a detailed tutorial. AI Commun. 30, 169-190.

Tijare, V. V., Yang, F. L., Kuttappan, V. A., Alvarado, C. Z., Coon, C. N., and Owens, C. M. (2016). Meat quality of broiler breast fillets with white striping and woody breast muscle myopathies. Poult. Sci. J. 95, 2167–2173. doi: 10.3382/ps/pew129

Tufféry, S. (2011). Data Mining and Statistics for Decision Making. New Jersey: John Wiley & Sons.

Vapnik, V. (1963). Pattern recognition using generalized portrait method. Autom. Remote. Control. 24, 774-780.

Vapnik, V. (1995). The Nature of Statistical Learning Theory. New York: Springer-Verlag.

Vapnik, V. (2013). The Nature of Statistical Learning Theory. Berlin: Springer Science & Business Media.

Vijayarani, S., Dhayanand, S., and Phil, M. (2015). Kidney disease prediction using SVM and ANN algorithms. Int. J. Comput. Mark. Res. 6, 1-12.

Wang, K. J., Chen, J. C., and Lin, Y. S. (2005). A hybrid knowledge discovery model using decision tree and neural network for selecting dispatching rules of a semiconductor final testing factory. Prod. Plan. Control. 16, 665-680.

Wiendahl, H. P., and Scholtissek, P. (1994). Management and control of complexity in manufacturing. Cirp. Ann.-Manuf. Techn. 43, 533-540.

Wold, J. P., Måge, I., Løvland, A., Sanden, K. W., and Ofstad, R. (2019). Near-infrared spectroscopy detects woody breast syndrome in chicken fillets by the markers protein content and degree of water binding. Poult. Sci. J. 98, 480-490.

Xanthopoulos, P., Pardalos, P. M., and Trafalis, T. B. (2013). "Linear discriminant analysis", in Robust Data Mining (Springer, New York, NY), 27-33.

Yang, Y., Zhu, J., Zhao, C., Liu, S., and Tong, X. (2011). The spatial continuity study of NDVI based on Kriging and BPNN algorithm. Math. Comput. Model. 54, 1138-1144.

Yang, Y., Wang, W., Zhuang, H., Yoon, S. C., Bowker, B., Jiang, H., & Pang, B. (2021). Evaluation of broiler breast fillets with the woody breast condition using expressible fluid measurement combined with deep learning algorithm. J. Food Eng. 288, 110133.

Yao, X. (1999). Evolving artificial neural networks. Proceedings IEEE. 87, 1423-1447.

Zampiga, M., Soglia, F., Baldi, G., Petracci, M., Strasburg, G. M., and Sirri, F. (2020). Muscle abnormalities and meat quality consequences in modern turkey hybrids. Front. Physiol, 11:554. doi: 10.3389/fphys.2020.00554

Zhuang, X. S., and Dai, D. Q. (2005). Inverse Fisher discriminate criteria for small sample size problem and its application to face recognition. Pattern Recogniti. 38, 2192-2194

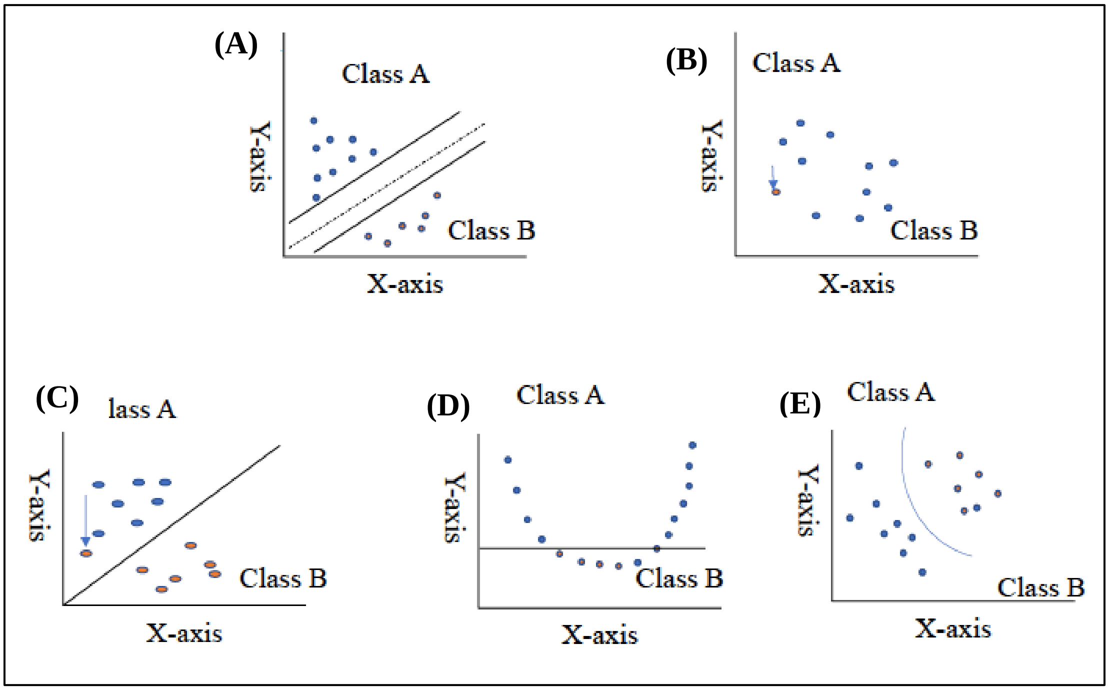


Figure 2-1: Representation of two class data using hyper-plane for support vector machine (Siddique et al. 2021)

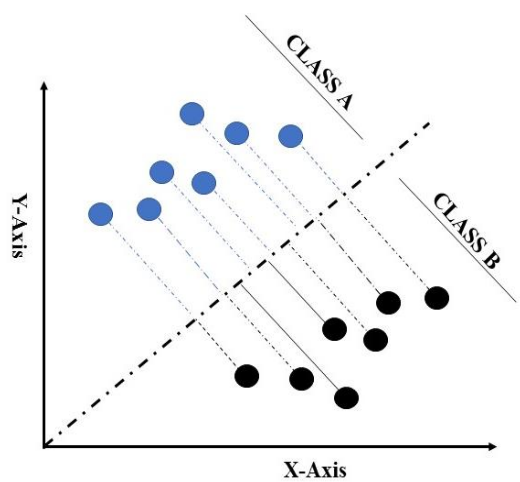


Figure 2-2: Representation of two class data in dimensional space for LDA analysis to maximize the classifiable data on the hyper-plane. This Figure adapted from Fisher (1963).

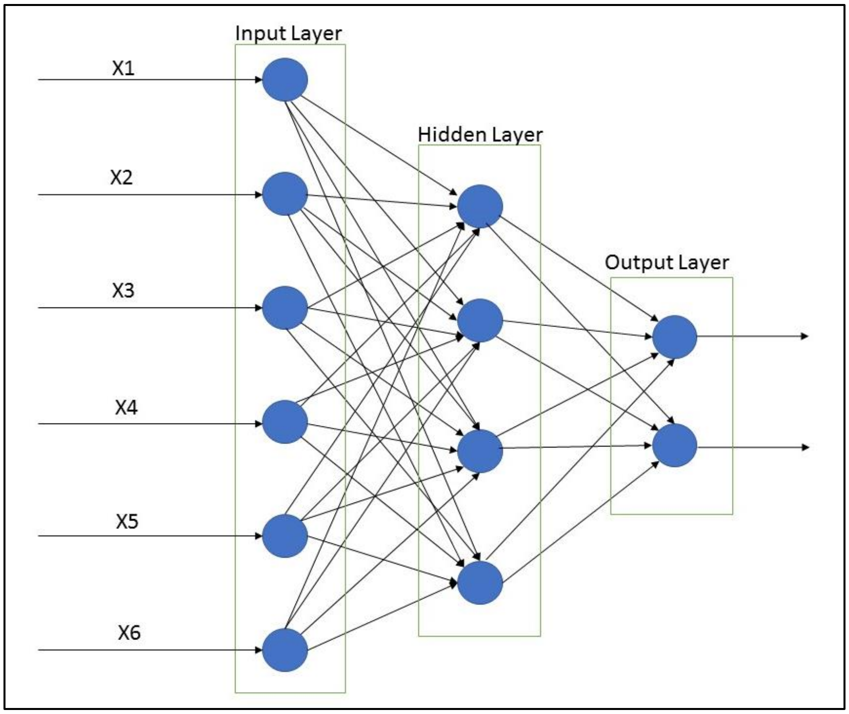


Figure 2-3: Back propagation neural network classification for input, hidden and output layer. This Figure was adapted from Rumelhart et al. (1986).

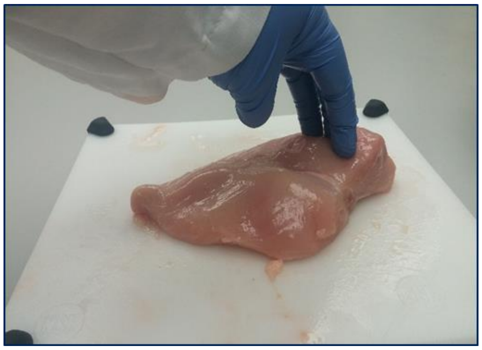


Figure 2-4: Hand-palpation method for identifying severity of woody breast myopathy in breast fillets.

Figure 2-5: Spaghetti meat condition in chicken breast fillets.

Figure 2-6: Severe woody breast fillet in the collected samples.

Figure 2-7: Hand-held Bioelectrical impedance device to measure the severity level of fillets

Table 2.1: Different subjective scales used for the classification of woody breast meat

Woody Breast Subjective Classification Scale[1]	Condition	Description
2 Point Scale	Normal	No toughness or Hardness
	Severe	Tough fillets
3 Point Scale	Normal	No toughness or Hardness
	Moderate	Medium toughness up to 50%
	Severe	More than 50% toughness
4 Point Scale	Normal	No toughness or hardness
	Mild	Hardness at cranial region
	Moderate	Filets extremely hard and rigid through from cranial region of caudal tip filets that were hard throughout but flexible in mid-to caudal region
	Severe	More than 50% of fillet area is woody

[1]2 point scale (Sihvo et al. 2014), and 3 point scale (Sihvo et al. 2014), 4 point scale (Tijare et al. 2016)

Table 2.2: Percentage classification efficiency for various supervised machine-learning algorithms (Linear Discriminant Analysis, Support Vector Machines, and Back Propagation Neural Networking) for breast fillets with woody breast.

Classification Method	Subjective Classification	Accuracy	
		Training (%)	**Testing (%)**
Woody Breast Meat			
Linear Discriminant Analysis	Normal	72.31	**52.63**
	Moderate	43.75	**29.41**
	Severe	75.00	**59.09**
Support Vector Machines	Normal	63.86	**71.04**
	Moderate	49.88	**59.99**
	Severe	71.78	**81.48**
Back Propagation Neural Networking	Normal	50.00	**47.77**
	Moderate	29.04	**23.33**
	Severe	20.95	**28.88**
Spaghetti Meat			
Support Vector Machines	Normal fillet without spaghetti	69.38	**50.00**
	Normal fillet with spaghetti	53.33	**50.00**
Back Propagation Neural Networking	Normal fillet without spaghetti	100.00	**52.95**
	Normal fillet with spaghetti	100.00	**75.00**

[1]n=300 (Normal=148, Moderate=82, Severe=70)

[2]n=84

Chapter 3

CLASSIFICATION AND FEATURE EXTRACTION USING SUPERVISED AND UNSUPERVISED MACHINE LEARNING APPROACH FOR BROILER WOODY BREAST MYOPATHY DETECTION

AUTHORS: Aftab Siddique[1], Charles B. Herron[1], Jaroslav Valenta[2], Laura Garner[1], Ashish Gupta[3], Jason T. Sawyer[4] and Amit Morey[1]

[1]Department of Poultry Science, Auburn University, Auburn, AL 36849, USA

[2]Department of Animal Science, Czech University of Life Sciences Prague, Prague, Czechia

[3]Department of Business analytics and Information, Auburn University, AL 36849, USA

[4]Department of Animal Sciences, Auburn University, Auburn, AL 36849, USA

(Published in MDPI Foods https://doi.org/10.3390/foods11203270**)**

3.1 ABSTRACT:

Bioelectrical impedance analysis (BIA) was established to quantify diverse cellular characteristics. This technique has been widely used in various species, such as fish, poultry, and humans for compositional analysis. This technology was limited to offline quality assurance/detection of woody breast (WB); however, inline technology that could be retrofitted on the conveyor belt would be more helpful to processors. Freshly deboned (n = 80) chicken breast fillets were collected from a local processor and analyzed by hand-palpation for different WB severity levels. Data collected from both BIA setups were subjected to supervised and unsupervised learning algorithms. The modified BIA showed better detection ability for regular fillets than the probe BIA setup. In the plate BIA setup, fillets were 80.00% for normal, 66.67% for moderate (data for mild and moderate merged), and 85.00 % for severe WB. However, hand-held BIA showed 77.78%, 85.71%, and 88.89% for normal, moderate, and severe WB. Plate BIA setup is more effective in detecting WB myopathies and could be installed without slowing the processing line. Breast fillet detection on the processing line can be significantly improved using a modified automated plate BIA.

Keywords: bioelectrical impedance; hand palpation; in-line processing; supervised learning; unsupervised learning; woody breast

3.2 INTRODUCTION

Poultry is a widely consumed form of protein in the United States, followed by beef and pork (NCC, 2020). According to the National Chicken Council, approximately 113.58 lbs. of total poultry meat will be consumed per person by 2022, which is more than almost any other country (NCC, 2020). Consumers primarily choose poultry meat because of the nutritional and functional characteristics, including consistency, appearance, and taste (Petracci et al. 2013). Additionally, markets are constantly shifting due to customer preferences and expectations, guiding the industry to produce greater volumes of trimmed chicken meat. Poultry breeders have boosted growth rates of birds through combination of feeding programs, genetic selections, and improvement in animal husbandry practices to meet the growing market pressure for white meat. Consumer demand for improved quality poultry white meat has increased overall carcass output (Petracci and Cavani, 2012). Nonetheless, fast-growing birds with greater breast meat yields have developed myopathies in their breast muscles, resulting in meat quality problems such as WB (**Figure 3-1**). According to Lui et al. (2020), the global occurrence of WB is estimated at 20% and may be increasing. With many factors influencing meat quality, the most important factors focus on appearance, water retention capacity, color, and texture of the meat (ElMasry et al.2012). Fresh poultry meat characteristics can influence the sensory qualities or consuming behavior of a product and its acceptability by consumers (Fletcher, 2002).

Woody breast (WB) is a myopathic condition that has plagued the broiler meat industry for years (Morey et al.2020). Woody breast in chicken breast fillets can be found at various levels of severity (Kuttappan et al. 2016). In a recent survey conducted by de Almeida Mallmann (2020), moderate and severe WB increased by 10% (from 25% to 35%) in chicken breeds with increased yield, and a 5% increase was observed (25% to 30%) in mild WB fillets (Morey et al.2020). Barbut (2020) have estimated that severe WB instances have detrimental to the poultry industry resulting in annual losses of $200 million and are estimated to reach $1 billion each year in North America (Barbut, 2020). Therefore, various systems have been implemented to identify, characterize, and classify chicken breast fillets based on WB myopathic severity (Chatterjee et al. 2016; Tijare et al. 2016; Petracci et al. 2019; Geronimo et al. 2019). To limit time and effort, inline non-destructive identification of these myopathic conditions is needed for both small and large-scale poultry processors (Wold et al. 2019). There have been several studies conducted that evaluate different approaches for rapid and accurate identification of myopathic fillets including computer image recognition systems, Near-infrared (NIR) spectrographic analysis, and a combination of low-coherence interferometry and hyperspectral imaging (HSI) (Petracci et al. 2019; Wold et al. 2017; Yoon et al. 2016; Kyle et al. 2004). The major advantage of these systems is that they can detect myopathy fillets without contacting the fillets unlike BIA wherein the electrodes have to touch the meat. However, computer vision systems separate myopathy meat based on the image of the fillet but analyzing the biochemical component of the meat would be a better predictor for detecting meat myopathies. Technologies such as HIS and NIR have shown promise as they analyze the biochemical characteristics of the breast fillets but they require equipment with large footprint, need complex data pre-processing prior to data analysis. The BIA technology used in the current study is a simple 4-electrode hand-held device which can be retrofitted in conveyor belts (as studied) so the device has a versatile form factor, it is easy to use and generates resistance and reactance data for myopathy fillet classification.

The primary working principle of BIA is based on the resistance property of a conducting material or wire, which is inversely proportional to its cross-sectional area and directly proportional to its length. However, the differences in material attributes, such as their form, shape, density, and composition, alter the rate of impulse conductance (Wold et al. 2019). Conductivity of an electric current is determined not only by the physiological portion of the cytosol, but also by the frequency response used in the prediction. Thus, signals can shift due to minor changes in muscle anatomy (Wold et al. 2017; Kyle et al. 2004). Several recent exome sequencing, meta-proteomics, and proteomics investigations have discovered several variables such as oxidative stress and changes in intracellular calcium that may contribute to the development of WB (Abasht et al. 2019; Abasht et al. 2016; Greene et al. 2019; Lee et al. 2015; Wang et al. 2020).

Bioelectrical impedance analysis technology has been utilized in many species to determine physical composition and qualities, such as total body fluid content and fat content, by measuring electrical resistance or impedance. In the food industry, BIA parameters can be adjusted to specific species, and it has been utilized in fish for quick detection of proximate composition and pre-harvest conditions to improve fish quality (Nyboer et al. 1950; Hoffer et al. 1969; Cox et al. 2011; Hartman et al. 2015). Morey et al.(2020) and Siddique et al. (2020) have reported that BIA may be used to successfully detect WB fillets as an alternative to hand-palpation, reducing hand classification mistakes and increasing accuracy of detecting WB.

Furthermore, BIA can detect other overlapping breast muscle myopathies, such as spaghetti meat, which is a myopathic condition impacting the chicken industry. Classification using BIA data can be improved using advanced data analytic technologies such as machine learning (ML), which includes data mining, artificial neural networks (ANN), deep learning (DL), and artificial intelligence (Siddique et al.2021). Siddique et al. (2021) reported that machine learning techniques such as Support Vector Machines (SVM) and backpropagation neural network (BPNN) classification algorithms could be another approach to classify the myopathic fillets based on severity levels (Siddique et al. 2021).

This research aims to assess the performance of two different bioelectrical impedance analysis (BIA) (**Figure 3-2 & 3-3**) device setups as a prospective quantitative approach to detect WB fillets with differing severity degrees to develop an inline WB detection system. The presented article is divided into subsections on detailed descriptions materials and methods used in section 3.3; section 3.4 describes Data analysis; section 3.5 describes results and discussion, and section 3.6 concludes the research.

3.3 MATERIALS AND METHODS

For the proposed experiment, 56-day-old broilers (Ross 708) deboned chicken breast fillets (n = 80) were obtained from a commercial poultry processor and transported to department of poultry sciences at Auburn University. Chicken breast fillets were categorized (**Figure 3-1**) by an experienced team member based on WB severity (normal, moderate, and severe) using hand palpation. During the hand-palpation technique, an experienced team member inspected each chicken breast fillet based on the perceived hardness in the three WB categories. The CQ Reader (Seafood Analytics, Version 3.0.0.3, Clinton Town, MI, USA), with four spring-loaded electrodes (2 receiver probes and two electrical signals sending probes), was used to collect data from an inbuilt algorithm of the equipment (RJL Systems, Detroit, MI, USA). Four electrodes were inserted into the geometric center of each breast fillet along the ventral surface (**Figure 3-2**). The device then measured the response for fat indices and protein indices data, resistance, and reactance and retrieved the encrypted data for assessment (Seafood Analytics Certified Quality Reader, Version 3.0.0.3, Seafood Analytics, MI, United States). Weighing balance (OHAUS Corporation, Pine Brook, New Jersey, United States) was used to measure the weight of separate fillets, and the acquired data was used in the analysis.

3.3.1 k-Means Clustering (k-Means)

Clustering is a powerful approach to information retrieval and machine learning algorithms for predicting and summarizing data. It has been used effectively in diverse industries, including product differentiation, social platform analysis, document classification, and image classification in the food industry. The goal of Clustering aims to put similar observations together and separate dissimilar ones (Pérez et al. 2013; Wang and Fleury, 2011; Liu et al. 2011; Baker, 2010; Dehariya et al. 2010; McCallum et al.2010; Aliguliyev, 2009; Van and Hoijtink, 2009; DeSarbo et al. 1988).

Clustering algorithms are valuable for many applications, but their capabilities are severely constrained. If each observation is assigned to a single cluster, then the data is explained by all aspects of k and distinct clusters. As a rule of thumb, discontinuous data splitting is not always the most accurate way to represent it. It is possible to have numerous clusters in the same observation area (Jain, 2010). The discontinuous segmentation is not always the best representation of the analyzed information because the data can have a significantly larger and more sophisticated hidden interpretation (Mustafi and Sahoo, 2019; Celebi et al. 2013; Singh et al. 2013; Wu and Kumar, 2009; Jain et al. 1999). In a given a data set, k-means algorithms (**Figure 3-5**) were applied to the points in a d-dimensional space in which X = {x1; x2 . . . xn}, in RD, i.e., N points (vectors) each with different D components in the data frame, partitional algorithms used in k-means clustering divides X into K full attributes, and mutually referenced clusters P= {P1, P2; . . . ; PK}, $\mathrm{U}_{\mathrm{i}-1}^{\mathrm{k}}\mathrm{P}_{\mathrm{i}} = \mathrm{X}, \mathrm{P}_{\mathrm{i}} \cap \mathrm{Pj} = 0$ for$1 \leq \mathrm{i} \neq \mathrm{j} \leq \mathrm{K}$. This algorithm generates the clusters by optimizing function and the sum of squared error can be explained as (Wu and Kumar, 2009; Singh et al. 2013):

$$\mathrm{SSE} = \sum_{\mathrm{i}=1}^{\mathrm{k}} \sum_{\mathrm{x_j} \in \mathrm{P}} \left\| \mathrm{x_j} - \mathrm{C_i} \right\|_2^2 \qquad \textbf{(1)}$$

Where $||x_j\text{-}c_i||^2$ *represents* the Euclidean norm and $c_i = 1 / |\mathrm{P_i}| \sum_{\mathrm{x_j} \in \mathrm{P_i}} \mathrm{x_j}$ denotes the cluster centroid (Pi), whose property of the data group is |Pi|. The minimal optimization of the above equation is also well known as the problem of minimum sum of squared error (SSE) clustering. The resulting k-means algorithm clusters for each data point get clustered into “one and only one of the k partitions”. Data points with the same cluster-ID are in the same clusters and vise-versa (Singh et al. 2013). For the initialization of k-means clustering algorithms, the initial k value is needed based on prior knowledge about the data, how many clusters are needed for the whole data set, and the number of clusters found by exploration data analysis (EDA) (Singh et al. 2013). The k-means algorithms represented in equation 1 start to cluster in a repetitive process in two alternative manners: (i) cluster-ID is assigned to each data point in the vector space, and (ii) updating the clusters as the data point changes in dimensional space. This process continues until there is no change in the data point position. The number of continuous repetitive steps may depend on vector points (N). Due to the linear comparison of data, the nature of k-means algorithms in the dimensional space also shows the linear dimensionality of data. The working algorithm of the k-means cluster follows these steps:

Input: Data set (D) and initial cluster values depending on the prior knowledge, i.e., based on experience or exploratory data analysis. Output: Set a cluster representative P in the vector space.

Repeat: Assign the data points to the closest cluster mean and update the cluster ID of j^{th} point in the data set.

Relocation of means: Update P so that Pj is the mean of the j^{th} cluster until the whole algorithm converges in equation 1 with minimum local optima.

3.3.2 Fuzzy c means (FCM)

Using fuzzy logic principles makes it possible to group highly dimensional data into clusters, assigning each point a percentage of membership in each cluster center between 0 and 100% (Jiang et al. 2004). Compared to typical hard-threshold clustering (*k-means*), in which every point is allocated a clear, accurate identity, FCM could provide improved clustering outcomes in certain types of datasets that don't lend themselves to traditional clustering approaches. This approach operates by assigning participation to each data point belonging to the nearest centroid on the basis of the distance between both the nearest cluster and the target value, calculated using a distance matrix (Swanson et al. 1998). The closer the data is to the nearest cluster, the greater the likelihood that it will belong to those specific centroids. The sum of each data point is entirely participation-based and should always be equal to one (Yong et al. 2004).

FCM is an unsupervised clustering approach used extensively for selecting features, clustering, and classifier designing challenges in astronomy, chemistry, geophysics, and medical diagnosis (Rao and Vidhyavathi, 2010; Fix and Hodges, 1989). A clustering technique built on the Ruspini Fuzzy clustering concept was introduced in the 1980s due to the evolution of fuzzy theory (Dai et al. 2012). This technique is used to analyze the data coordinates based on their distance from each other. For each grouping, the centroids are constructed depending on the spacing amongst sample points in the original data set. Developed algorithm for the fuzzy c-means clustering is as follows (Fix and Hodges, 1989; Dai et al. 2012):

Step 1. Calculate the center of the given data set

$$v_{ij} = \sum_{k=1}^{n} (u_{ik})^m x_{kj} \Big/ \sum_{k=1}^{n} \left(u_{ij}\right)^m \tag{1}$$

Step 2. Calculate matrix distance ($D_{[c,n]}$)

$$D_{ij} = \left(\sum_{j=1}^{m} \left(x_{kj} - v_{ij}\right)^2\right)^{1/2} \tag{2}$$

Step 3. Updating of partition matrix for nth step as

$$u_{ij}^{n-1} = \left(\frac{1}{\sum_{j-1} \left(d_{ik}^{n}/d_{jk}^{n}\right)^{\frac{2}{m}-1}} \right) \tag{3}$$

Termination of fuzzy c means the algorithm only takes place when the algorithm reaches $||U^{(k+1)}-U^{(k)}||<\delta$, if not achieved, the algorithm returns to step 2 and re-executes until the conditions are being met by updating the centroid continuously (Fix and Hodges, 1989).

3.3.3 K-Nearest Neighbor (k-NN)

The K-nearest neighbor (k-NN) technique was developed to perform statistical techniques when valid parametric estimates of probability densities are unknown or difficult to calculate. Fix and Hodges (1989) presented a non-parametric design data classification approach in 1951 through an unpublished paper by the US Air Force School of Aviation Medicine, which had established as the k-nearest neighbor rule (Sun and Huang, 2010) and was further developed by Thomas Cover in 1967. A new observation is classified depending on how comparable it is to other observations already analyzed (Bernal et al. 2021). According to its neighboring labels, this classification is performed. Several industries, including cyber and information securities, aviation industry, valuable life forecast, defect categorization, nephropathy diagnosis in children, and infiltration prevention systems, have been implemented using the k-NN method (Zhang et al. 2021; Hu et al. 2016; Vapnik, 2019; Maji et al. 2008).

For understanding the concept and working principle of k-NN, let us suppose that we have a data set for different conditioned myopathic chicken breast fillets represent resistance and reactance (Figure 3-6, blue points represent resistance and red points represents reactance. Now suppose that we added a new data point for resistance and reactance and were told that the new fillets value in the dataset in class is a severe woody breast (value represented by square box,

Figure 3-6, let's see if the k-NN algorithm will be able to identify the class by input data. For the classification of new data point k-NN general rule must be followed as (Bernal et al. 2021; Romeo et al. 2020; Dai et al.2012; Sun and Huang, 2010).

Step 1. Input: V, V_l, c, i.e., V = training data set, V_l = labels of data set, c = sample data point that needs to be classify

Step 2. For v to training data set size do: Compute the distance between training data set and sample data point d (V_i, c).

Step 3. End For loop: Select the desired number of clusters of nearest neighbors, arrange the computed distance in increasing order, and count the number of occurrences for each label in top k-neighbors.

Step 4. Output: Assign c to the most occurring label (l)

The k-nearest-neighbor classifier is based on the Euclidean distance between a test sample and the specified training samples. The Euclidean distance between sample Vi and Vl (l=1,2,…,n) is defined as $d(Vi,Vl)= \sqrt{}(Vi1-Vl1)^2+(Vi2-Vl2)^2+\cdots+(Vip-Vl_N)^2$ [51]. After the computation of the new data set and evaluation of distance between the original training data set and test dataset, we can decide based on the neighbors that where the new data point belongs to (Figure 3-6. In Figure 3-6 the new observation is classified as severe fillets with minimum numbers of k-neighbors (k=2). For k=2 closest neighbors are in small circle. It can be clearly observed that if k=2, two of the neighbors are severe, so the new data point would also be classified as severe. So, for the classification of new data point added into the data set will get classified as severe fillet. However, this does not imply that K=2 is the highest performance quantity for the dataset; more additional observations should be classified with other K values to identify the quantity with the optimum overall performance from the data set.

3.3.4 Support Vector Machines (SVM) Algorithms

Vapnik (2019) proposed the SVM method for the first time in 1995, and it has received a tremendous amount of attention from the machine learning applications community. Many studies have found that the SVM approach outperforms other data classification algorithms depending on the data type regarding classification accuracy compared to other methods (Murphy and Monteiro, 2013; Lukaski, 1987). For data set categorization, SVM generates a line between two or more classes, referred to as hyper-planes. SVM aims to discover a hyperplane that can separate two classes of provided data with a maximum margin while still offering the best generalization capability for a two-class linearly separable learning assignment. It provides highly accurate results on the training dataset and high predicting accuracy for the new dataset from the same population as the training dataset (Siddique et al.2021). A detailed working principle for SVM algorithm can be found in Siddique et al.(2021).

3.3.5 Bioelectrical impedance Analysis

Samples collected from local processors were placed on non-conducting surface for BIA analysis. The resistance and reactance qualities of myopathic conditioned fillets were measured with a hand-held BIA device and Plate BIA device (Seafood Analytics, Clinton Town, Michigan, United States) on the upper surface of the chicken breast filet. Both BIA unit is made up of four electrodes: two signal electrodes and two detecting electrodes. These electrodes are connected to an AC current of 800 μA and 50 kHz, and they are able to produce voltage fluctuations ranging from 3.75 V to 10.60 V (Morey et al. 2020). Electrodes that are used in the collection of data were made of stainless steel and are used to complete the circuit between the chicken breast fillets and the four electrodes (RJL Systems, Detroit, MI, United States). In hand-held BIA, device was held in hand to perform compression-based analysis (**Figure 3-2**) while on the other hand samples were place on plate BIA to collect the data (**Figure 3-3**). As soon as the electrodes come into contact with the item being tested, the circuit is completed, and the instrument begins to take two sets of measurements.

3.4 DATA ANALYSIS

Data were analyzed using two different BIA configurations (Hand-held and Plate setup) and analyzed for descriptive summary of collected values on plate BIA for Resistance, Reactance, and Breast fillet weight using SAS software (Version 9.4, Cary, NC, USA) using Tukey HSD means value. Data collected from both BIA setups was subjected to R Software and jmp16 Pro (Version 16.0, Cary, NC, USA) by implementing the unsupervised (k-means clustering, Fuzzy c-means clustering) and supervised learning algorithms (K-nearest neighbor, SVM) to evaluate the accuracy of equipment for the detection of WB in chicken breast fillets. Clustering method was implemented in our data set to observe the identification efficiency of both BIA in classification of myopathic fillets without using labels or using any supervision. The collected data used in the analysis had 3 dimensions i.e., resistance, reactance, and weight of individual fillets. For k-means and FCM analysis for the BIA collected data R software (Version 4.2.0, Vigorous Calisthenics) was used with set replicability of 3000 using set seed command. For the development of supervised learning algorithms using k-NN models the data were randomly divided into 55::45 training and testing sets using validation options in jmp 16 Pro. For the SVM analysis of collected data, caret package was used in R software (Version 4.2.0, Vigorous Calisthenics). The caret package algorithm calculated the best-suited tuning parameter or cost (C) for both the training and testing data sets. A seed value (replicability) was set for 3,000 for the SVM analysis.

3.5 RESULTS AND DISCUSSION

Table 3.1 illustrates the differences in multiple factors determined by different BIA setups among the different severity categories of WB meat. In **Table 3.1**, there are statistically significant differences ($p < 0.05$) in the resistance and reactance for normal chicken breast fillets and moderate breast fillets, but no there were no significant differences observed between normal fillets and severe WB fillets in hand-held BIA collected data (**Table 3.1**). Our results for plate BIA and hand-held BIA setup agree with the study conducted and reported by Morey et al. (2020), in which authors have found lower resistance and reactance value for normal chicken breast fillets as compared to severe chicken breast fillets and vice-versa. Reactance and resistance are affected by the compositional content of a product. They are measurements of a substance's capacity to accommodate a charge and carry or conduct electrical charges, respectively (Maji et al. 2008). Histological changes in growing birds can alter the water distribution within the muscle structure, which influences or changes the electrical properties of chicken breast fillets. In our presented data, there were no statistically significant differences observed in resistance and reactance values for normal chicken breast filets (72.89 Ω and 25.76 Ω, respectively) compared to severe WB filets (70.60 Ω and 21.76 Ω) for hand-held BIA collected data.

On the other hand, data collected from plate BIA showed that normal chicken breast fillets have a resistance of 103.34 Ω and severe WB fillets have an average resistance of 112.02 Ω. Kyle et al. (2004) have mentioned that at lower frequencies, muscles and their other components, such as fatty tissues and peptide residues, acts as an insulator and do not allow the flow of electrical charges, which in results forces these electrical charges to flow from alternative components of the cell, i.e., extracellular fluid. Authors have also mentioned that increasing non-conducting suspended elements in water will also increase the resistance capacity of the fluid system. Severe WB fillets have a high amount of extracellular fluid. They have these non-conducting suspended components, which ultimately increase the resistance of WB fillets as compared to normal chicken breast fillets (Siddique et al.2021; Morey et al. 2020; Kyle et al. 2008; Lukaski, 1987).

The classification experiment was conducted on two different BIA configurations generating data sets for WB fillets classified as normal, moderate, and severe. In the k-means clustering algorithm (unsupervised method), three clusters were formed using resistance, reactance, and weight of fillets. The k-means clustering results for the plate BIA data showed that the average distance value of each observation from the center of clusters one, two, and three ranged from 0.14 to 2.50 for normal fillets, 0.45 to 3.83 for moderate fillets, and 0.46 to 5.09 for severe fillets (**Table 3.1**). The first cluster means for the plate BIA method for resistance, reactance, and weight is 101.25 ± 13.13 Ω, 27.06 ± 5.67 Ω, and 426.68 ± 44.17 gm, respectively; the second cluster means for resistance, reactance, and weight is 109.62 ± 11.66 Ω, 34.66 ± 6.71 Ω, and 556 ± 44.84 gm, respectively; the third cluster means for resistance, reactance, and weight were 143.21 ± 8.73 Ω, 56.25 ± 5.21 Ω, and 542.28 ± 67.95 gm, respectively (**Table 3.5**).

Hand-held BIA cluster data classification showed that the average distance value of each observation from the center of clusters one, two, and three ranged from 0.04 to 4.73 for normal fillets, 0.07 to 2.22 for moderate, and 0.20 to 7.49 for severe fillets. The first cluster means for the hand-held BIA method for resistance, reactance, and weight is 64.52 ± 4.88, 19.78 ± 4.71, and 571.36 ± 38.29, respectively; the second cluster means for resistance, reactance, and weight is 81.71 ± 5.91 Ω, 31.80 ± 3.33 Ω, and 444.12 ± 52.44 gm, respectively; the third cluster means for resistance, reactance, and weight were 70.75 ± 3.04 Ω, 21.63 ± 4.47 Ω, and 475.26 ±

33.10gm, respectively (**Table 3.6**). Generally average distance values in clustering are not very informative but it can provide us with an idea that how much the clusters are separated from each other. More the average distance between the clusters better is the clustering results. In our presented data for k-means clustering results for hand-held BIA showed better average distance value for clusters of normal fillets and severe fillets as compared to the data collected for plate BIA. Data obtained from plate BIA indicates that 18.75% normal, 13.75% moderate, and 22.50% severe WB fillets were placed in cluster 1; 6.25% normal, 7.50% moderate, 23.75% severe fillets were placed in cluster 2; and 1.25% normal, 0.00% moderate, and 6.25% of severe fillets were placed in cluster 3 (**Table 3.5**).

However, data obtained for hand-held BIA concluded that 2.50% normal, 6.25% moderate, and 26.25% severe WB fillets were clustered in cluster 1; 8.75% normal, 2.50% moderate, 6.25% severe WB fillets were clustered in cluster 2; 15.00% normal, 12.50% moderate, and 18.75% severe fillets were clustered in cluster 3 (**Figure 3-4**). Based on the optimal number of clusters (k=3) needed for better clustering average Silhouette Index value for plate BIA and hand-held BIA was 0.7688 and 0.8036, respectively. Indicating that k-means clustering analysis performed on both BIA data set creates dense and compact clusters with less overlapping between cluster data points. Separate and compact clusters clearly indicated that both BIA devices are capable of identifying myopathic conditions in non-labeled data with some degree of overlapping.

Fuzzy c means membership analysis showed that hand-held BIA data for chicken breast fillets were clustered as 15% normal, 5% moderate, 12.5% severe in cluster one, 2.5% normal, 5% moderate, 15% severe in cluster two, and 10% normal, 13.75% moderate, and 25% severe in cluster three compared to plate BIA data in which chicken breast fillets were clustered as 7.5%

normal, 12.5% moderate, 25% severe in cluster one, 2.5% normal, 5% moderate, 15% severe in cluster two and 15% normal, 5% moderate, and 12.5% severe in cluster three (**Figure 3-4**). Data analyzed for hand-held BIA showed that the probability of normal fillets in cluster 1 to cluster 3 ranges from 9.09% to 54.5%; probability percentage for moderate fillets in cluster 1, 2, and 3 ranges from 21.0% to 57.8% and severe fillets in three distinct clusters were ranged from 23.8% to 47.6%. On the other hand, plate BIA analyzed data for FCM clustering showed that probability percentage for normal, moderate and severe conditioned myopathic fillets ranges from 10% to 60% (**Table 3.2**). Fuzzy c means is a soft clustering method in which it keeps all the observation values in each cluster although they belong to different clusters at the same time. Hand-held BIA collected data showed much better separated clusters Dunn's Coefficient value compared to plate BIA collected value. Dunn's coefficient value for hand-held BIA data was 0.775, and for plate BIA data was 0.741. Clusters formed by the hand-held BIA dataset (Supplementary file, FIGURE 3-7) showed better cluster formation than plate BIA data (Supplementary file, FIGURE 3-8). Cluster achieved reduction for hand-held BIA observation was 82.08%, and for the plate BIA data was 77.87%. Hand-held BIA data showed a better fuzzy Silhouette Index value at k=3 (as compared to plate BIA data (Hand-held BIA fuzzy Silhouette Index value = 0.803 and Plate BIA fuzzy Silhouette Index value =0.768) (supplementary file, FIGURE 3-9 & FIGURE 3-10).

The k-NN is an occurrence-based classification algorithm that distinguishes a new data point by analyzing it with an already stored value in the system that has been employed as the training set to analyze it with a screening set value. As indicated in **table 3.3**, k-NN successfully classified both BIA data sets with superior efficiency in the testing set despite overlapping resistance, reactance, and fillets weight data displayed in density (Supplementary file, FIGURE

3-11, FIGURE 3-12 & FIGURE 3-13). The k-NN analysis for hand-held BIA setup showed better classification performance based on instance in the testing data set (normal = 50%, moderate = 42.90%, and severe = 87.50%) especially in comparison with the training data set (normal= 38.50%, moderate = 10.0%, and severe = 67.70%), and for plate BIA the categorization efficiency for testing set was 40.0% for normal, 25% for moderate, and 78.60% for severe fillets. Both classification efficiencies were reported in the k-NN algorithm. The trained model was observed to be efficient in classification during the testing stage. Results obtained by training accuracy of both BIA methods for k-NN classification were 38.50%-31.30% (Hand-held BIA-Plate BIA) for normal fillets, 10.00%-7.71% (Hand-held BIA-Plate BIA) for moderate, and 57.70% -57.10% (Hand-held BIA-Plate BIA) for severe fillets classification (**Table 3.3**), respectively using the BIA parameters and fillet weight data set (n=80). The testing set was found to be higher in accuracy as compared to our training set with 40.00%-50.00% (Plate BIA - hand-held BIA) normal classified, 25.00%-42.00% (Plate BIA - hand-held BIA) moderate classified, and 78.60% - 87.50% (Plate BIA - hand-held BIA) severe WB classified (n=80; **Table 3.3**). The testing set data was higher in accuracy than the training set data, possibly due to the nature of myopathic fillets i.e., presence of extracellular fluid in severe WB fillets and also the nature of algorithm used. The k-NN algorithm, which is completely based on instance-based counting method for the number of observations used in the training set and then assigning the next upcoming value-based to a most voted group called as neighbors. A higher classification was observed in the both BIA setup for the k-NN algorithm for the testing. Normal fillets showed 50.00%, moderate fillets showed 42.90%, and Severe fillets showed 87.50% of classification than in the training set classification labels (**Table 3.3**).

The results obtained from the SVM classification algorithm showed that the percent accuracy for the classification of normal WB fillets in the test data set was 80.00%, for moderate fillet classification, the testing set showed an accuracy of 66.67%, and classification accuracy for the severe WB fillets for test data set was 85.00% (**Table 3.4**). On the other hand, the data collected for the hand-held BIA system, analyzed using SVM analysis, showed the following classification efficiency in testing data sets for normal fillets were 77.78%, for moderate fillets were 85.71%, and severe fillets were 88.89% respectively (**Table 3.4**). Due to the nature of uneven random distribution of data points in training and testing split method, we have also performed the k-fold cross validation technique to determine the overall accuracy of the developed SVM model for two different BIA datasets. Our results indicated that at the cost function of C=1.5, with repeated cross-validation accuracy for the hand-held BIA data was 92.05% (Supplementary file, FIGURE 3-14) while on the other hand for plate BIA was 93.11% (Supplementary file, FIGURE 3-15) with the cost function of C= 0.5. Confusion matrix table showed the number of test fillets correctly classified based on the labels (**Table 3.7**). A total number of 34 fillets were randomly selected for testing data set which includes 9 normal fillets, 7 moderate fillets, and 18 severe WB fillets in hand-held BIA data while on the other hand, 6 normal fillets, 9 moderate fillets, and 20 severe WB fillets were selected for plate BIA data. Confusion matrix table (**Table 3.7**) shows that from hand-held BIA testing data set two fillets were mis-classified from normal category, one fillet from moderate condition was mis-classified in normal category, and two fillet from severe WB condition to normal and moderate category. In plate BIA collected data, one fillet was mis-classified in moderate category, for moderate fillets, two fillets were classified in severe fillets, and one fillet was classified in normal category. For severe WB fillets, 3 fillets were mis-classified in normal category.

Based on analyzed data for hand-held BIA and plate BIA data, the sensitivity of normal, moderate, and severe fillets was 100%, 80%, and 100% respectively, i.e., if a breast fillet had a diagnosis of normal fillets, the probability that this myopathic fillet was correctly assigned to the normal category was 1.0. The specificity of the normal fillets classification was 94.7%, moderate was 100%, and for severe fillets was 100%, implying that 94.7% of the normal fillets were correctly classified as normal on the basis of their collected parameters such as resistance, and reactance. Results obtained in this study is in agreement with Siddique et al. (2021), where the researchers have found that the classification efficiency of the SVM algorithm in the categorization effectiveness of SVM for the partitioning of high dimensionality data showed improved classification efficiency for normal (training performance 63.86%, testing performance 71.04%), moderate (training performance 49.88%, testing performance 59.99%), and severe WB (training performance 49.88 %, testing performance 59.99%). Authors have also found that the results for overall repeated cross-validation accuracy percentage for hand-held BIA and plate BIA were above 90% which agrees with the study published by Geronimo et al. (2019), where authors have reported the accuracy percentage of SVM model over 90%. The lower cost function of the plate BIA collected data in conjunction with SVM model implies that the developed model can be easily implemented in the real-world scenario without much pre-processing of data. The architecture of the plate BIA plays a significant role in the collection and accuracy of data. Long separate probes in plate BIA setup allow the flow of electrical impulse throughout the fillets, while on the other hand-held probe BIA covers less area on fillets surface. Additionally, Siddique et al. (2021) found that conventional classification analysis methods such as LDA did not perform as well with the hand-held BIA setup, with an accuracy of 61.70% for normal chicken breast fillets, 31.30% for moderate WB fillets, and 68.50% for severe WB fillets in the training set, compared to 75.60% normal, 33.33% moderate, and 56.30% for severe WB in testing data set (Siddique et al. 2021). When kernel functions are used in SVMs, the original input variables become axially separable in the higher-dimensional domain, allowing them to be classified. Furthermore, SVMs can reduce both the estimating dimension and error of the system simultaneously (Siddique et al. 2021).

3.6 CONCLUSIONS

Results from the current work demonstrated that using a plate BIA in real-time inline chicken production systems to classify chicken breast fillets, depending on the severity of myopathy, is conceivable. Plate BIA detection for muscle histopathologic classification is superior to a hand-held BIA device when used in conjunction with SVM and k-nearest neighbor analysis. Input variables can provide a more efficient separation capability than the hand palpation method or other unsupervised algorithms used in BIA collected data for evaluating chicken breast fillets. In terms of rapid detection of woody breast myopathies in chicken breast fillets without having pre-processing steps for data analysis is an important part of our research. For the collection and analysis of our data no data pre-processing step is involved for clustering analysis using FCM and k-means. The results obtained in supervised learning techniques are following the same pattern of accuracy with an overall accuracy of 90% that are reported in previous research done using NIR, image analysis and HSI imaging techniques where intensive pre-data processing is needed before the classification of woody breast fillets can be performed. The overall cross validation accuracy with lower cost function indicates that the developed SVM model for plate BIA setup can be directly implemented in actual processing plant for classification of fillets without implementing intensive data pre-processing steps.

Future studies are needed to collect and classify larger volumes of data on myopathic breast fillet for conditions such as white stripping, spaghetti meat condition, fillet time, processed bird age, different variation in used frequencies, and fillet temperature at throughout the processing procedures. In addition, the use of different variables avoids overlapping circumstances induced by human counterparts during the inline processing of the fillets. It is apparent that through use of this novel technology chicken processing efficiency and classification of chicken breast fillets with undesirable WB myopathy can be improved.

3.7 REFERENCES

Abasht B, Zhou N, Lee WR, Zhuo Z, Peripolli E. The metabolic characteristics of susceptibility to wooden breast disease in chickens with high feed efficiency. Poultry Sci., **2019,** 98:3246–56. Doi: http://dx.doi.org/10.3382/ps/pez183

Abasht, B., Mutryn, M. F., Michalek, R. D., Lee, W. R. Oxidative stress and metabolic perturbations in wooden breast disorder in chickens. PloS one, **2016,** 11(4), e0153750. doi: http://dx.doi.org/10.1371/journal.pone.0153750

Aliguliyev, R. M. A new sentence similarity measure and sentence based extractive technique for automatic text summarization. Expert Systems with Applications, **2009,** 36(4), 7764-7772.doi: http://dx.doi.org/10.1016/j.eswa.2008.11.022

Baker, R. S. J. D. (2010). Data mining for education. International encyclopedia of education, **2010,** 7(3), pp. 112-118.

Barbut, S. Understanding the woody breast syndrome and other myopathies in modern broiler chickens. In Australian Poultry Science Symposium. Sydney, Australia**, 2020** (pp. 99-102).

Bera, T. K. Bioelectrical impedance methods for noninvasive health monitoring: a review. J. Med. Eng., **2014**, 1-28. doi: http://dx.doi.org/10.1155/2014/381251

Celebi, M. E., Kingravi, H. A., Vela, P. A. A comparative study of efficient initialization methods for the k-means clustering algorithm. Expert systems with applications, **2013** 40(1), 200-210.doi: http://dx.doi.org/10.1016/j.eswa.2012.07.021

Chatterjee, D., Zhuang, H., Bowker, B.C., Rincon, A.M., Sanchez-Brambila G. Instrumental texture characteristics of broiler *Pectoralis major* with the wooden breast condition. Poult Sci. 2016, 95, 2449–2454. Doi: http://dx.doi.org/10.3382/ps/pew204

Cox, M. K., Heintz, R., Hartman, K. Measurements of resistance and reactance in fish with the use of bioelectrical impedance analysis: sources of error. Fishery Bulletin, **2011,** 109, 34-47.

Dai, M., Dong, F., Hirota, K. Fuzzy Three-Dimensional Voronoi Diagram and its Application to Geographical Data Analysis. Journal of Advanced Computational Intelligence and Intelligent Informatics, **2012,** 16(2), 191-198.doi: http://dx.doi.org/10.20965/jaciii.2012.p0191

Dehariya, V. K., Shrivastava, S. K., Jain, R. C. (2010). Clustering of image data set using k-means and fuzzy k-means algorithms. In 2010 International conference on computational intelligence and communication networks, **2010,** (pp. 386-391). IEEE.doi: http://dx.doi.org/10.1109/CICN.2010.80

DeSarbo,W. S., Cron, W. L. A maximum likelihood methodology for cluster wise linear regression. Journal of classification, **1988,** 5(2), 249-282.

Elmasry, G., Barbin, D. F., Sun, D. W., Allen, P. Meat quality evaluation by hyperspectral imaging technique: an overview. Critical reviews in food science and nutrition, **2012,** 52(8), 689-711. Doi: http://dx.doi.org/10.1080/10408398.2010.507908

Fix, E., Hodges, J.L. Discriminatory analysis, nonparametric discrimination: Consistency properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas, **1989**.doi: http://dx.doi.org/10.2307/1403797

Fletcher, D. L. Poultry meat quality. World's Poultry Science Journal, **2002,** 58(2), 131-145. Doi: http://dx.doi.org/10.1079/WPS20020013

Geronimo, B. C., Mastelini, S. M., Carvalho, R. H., Júnior, S. B., Barbin, D. F., Shimokomaki, M., Ida, E. I. Computer vision system and near-infrared spectroscopy for identification and classification of chicken with wooden breast, and physicochemical and technological characterization. Infrared Physics & Technology, **2019,** 96, 303-310.doi: http://dx.doi.org/10.1016/j.infrared.2018.11.036

Greene, E., Flees, J., Dadgar, S., Mallmann, B., Orlowski, S., Dhamad, A., Dridi, S. Quantum blue reduces the severity of woody breast myopathy via modulation of oxygen homeostasis-related genes in broiler chickens. Front Physiol., **2019,** 10:1251.doi: http://dx.doi.org/10.3389/fphys.2019.01251

Hartman, K. J., Margraf, F. J., Hafs, A. W., Cox, M. K. (2015). Bioelectrical impedance analysis: a new tool for assessing fish condition. Fisheries, **2015,** 40(12), 590-600.doi: http://dx.doi.org/10.1080/03632415.2015.1106943

Hoffer, E. C., Meador, C. K., Simpson, D. C. Correlation of whole-body impedance with total body water volume. J. Appl. Physiol. **1969,** 27, 531-534.doi: http://dx.doi.org/10.1152/jappl.1969.27.4.531 http://dx.doi.org/10.3390/nu4010001

Hu, L. Y., Huang, M. W., Ke, S. W., Tsai, C. F. The distance function effect on k nearest neighbor classification for medical datasets. Springer Plus, **2016,** 5(1), 1-9.doi: http://dx.doi.org/10.1186/s40064-016-2941-7

Hu, M. H., Dong, Q. L., Liu, B. L., Malakar, P. K. The potential of double K-means clustering for banana image segmentation. Journal of Food Process Engineering, **2014,** 37(1), 10-18. Doi: http://dx.doi.org/10.1111/jfpe.12054

Jain, A. K. Data clustering: 50 years beyond K-means. Pattern recognition letters, **2010,** 31(8), 651-666.doi: http://dx.doi.org/10.1016/j.patrec.2009.09.011

Jain, A. K., Murty, M. N., Flynn, P. J. Data clustering: a review. ACM computing surveys (CSUR), **1999,** 31(3), 264-323.doi: http://dx.doi.org/10.1145/331499.331504

Jiang, D., Tang, C., Zhang, A. Cluster analysis for gene expression data: a survey. IEEE Transactions on knowledge and data engineering, **2004,** 16(11), pp. 1370-1386.

Khanmohammadi, S., Adibeig, N., Shanehbandy, S. (2017). An improved overlapping k means clustering method for medical applications. Expert Systems with Applications,**2017,** 67, pp.12-18.doi: http://dx.doi.org/10.1016/j.eswa.2016.09.025

Kuttappan, V. A., Hargis, B. M., Owens, C. M. (2016). White striping and woody breast myopathies in the modern poultry industry: a review. Poultry Science, **2016,** 95(11), 2724-2733.doi: http://dx.doi.org/10.3382/ps/pew216

Kyle, U. G., Bosaeus, I., De Lorenzo, A. D., Deurenberg, P., Elia, M., Gómez, J. M., Composition of the ESPEN Working Group. Bioelectrical impedance analysis part I: review of principles and methods. Clin Nutr, **2004,** 23, 1226 1243.doi: http://dx.doi.org/10.1016/j.clnu.2004.06.004

Kyle, U. G., Bosaeus, I., De Lorenzo, A. D., Deurenberg, P., Elia, M., Gómez, J. M., Composition of the ESPEN Working Group. Bioelectrical impedance analysis—part I: review of principles and methods. Clinical nutrition, **2004**, 23(5), 1226-1243. doi:http://dx.doi.org/10.1016/j.clnu.2004.06.004

Lee, W. R., Brannick, E. M., Mutryn, M. F., Fu, W., Abasht, B. Characterization of a novel chicken muscle disorder through differential gene expression and pathway analysis using RNA-sequencing. BMC Genomics. **2015,** 16:399. doi: http://dx.doi.org/10.1186/s12864-0151623

Liu, J., Puolanne, E., Schwartzkopf, M., Arner, A. Altered sarcomeric structure and function in woody breast myopathy of avian pectoralis major muscle. Frontiers in physiology, **2020,** 11, 287. Doi: http://dx.doi.org/10.3389/fphys.2020.00287

Liu, Y. C., Wu, C., Liu, M. (2011). Research of fast SOM clustering for text information. Expert Systems with Applications, **2011,** 38(8), 9325-9333.doi: http://dx.doi.org/10.1016/j.eswa.2011.01.126

Lopez-Bernal, D., Balderas, D., Ponce, P., Molina, A. Education 4.0: Teaching the Basics of KNN, LDA and Simple Perceptron Algorithms for Binary Classification Problems. Future Internet, **2021,** 13(8), 193.doi: http://dx.doi.org/10.3390/fi13080193

Lukaski, H. C. Methods for the assessment of human body composition: traditional and new. The American journal of clinical nutrition, **1987**, 46(4), 537-556.doi: http://dx.doi.org/10.1093/ajcn/46.4.537

Maji, S., Berg, A. C., Malik, J. Classification using intersection kernel support vector machines is efficient. In 2008 IEEE conference on computer vision and pattern recognition, **2008,** pp. 1-8. IEEE.doi: http://dx.doi.org/10.1109/CVPR.2008.4587630

Mallmann, B. (2019). Prediction and evaluation of breast myopathy. Poultry Science, University of Arkansas. Available online at: https://core. ac. uk/download/pdf/250604484. pdf [Accessed on July 7, 2022].

McCallum, A., Nigam, K., & Ungar, L. H. Efficient clustering of high dimensional data sets with application to reference matching. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, **2010,** (pp. 169-178). Doi: http://dx.doi.org/10.1145/347090.347123

Morey, A., Smith, A. E., Garner, L. J., Cox, M. K. Application of Bioelectrical Impedance Analysis to Detect Broiler Breast Filets Affected with Woody Breast Myopathy. Front. Physiol. 11, **2020,** 808. doi: http://dx.doi.org/10.3389/fphys.2020.00808

Murphy, R. J., Monteiro, S. T. (2013). Mapping the distribution of ferric iron minerals on a vertical mine face using derivative analysis of hyperspectral imagery (430–970 nm). ISPRS Journal of Photogrammetry and Remote Sensing, **2013,** 75, 29-39.doi: http://dx.doi.org/10.1016/j.isprsjprs.2012.09.0

Mustafi, D., Sahoo, G. A hybrid approach using genetic algorithm and the differential evolution heuristic for enhanced initialization of the k-means algorithm with applications in text clustering. Soft Computing, **2019,** 23(15), 6361-6378. doi: http://dx.doi.org/10.1007/s00500-018-3289-4

National Chicken Council. (2020). Per Capita consumption of Poultry and Livestock, 1965 to estimated 2021. https://www.nationalchickencouncil.org/about-the industry/statistics/per capita-consumption-of-poultry-and-livestock-1965-to-estimated 2021-in-pounds/ [Accessed July 30, 2021].

Nyboer, J., Kreider, M. M., Hannapel, L. (1950). Electrical impedance plethysmography: A physical and physiologic approach to peripheral vascular study. Circulation, **1950,** 2, pp. 811-821.

Pérez-Suárez, A., Martínez-Trinidad, J. F., Carrasco-Ochoa, J. A., Medina-Pagola, J. E. OClustR: A new graph-based algorithm for overlapping clustering. Neurocomputing, **2013,** 121, 234-247.doi: http://dx.doi.org/10.1016/j.neucom.2013.04.025

Petracci, M., Cavani, C.. Muscle growth and poultry meat quality issues. Nutrients.**2012,** 4, pp. 1-12. Doi:

Petracci, M., Soglia, F., Madruga, M., Carvalho, L., Ida, E., Estévez, M. Wooden-breast, white striping, and spaghetti meat: causes, consequences and consumer perception of emerging broiler meat abnormalities. Comprehensive Reviews in Food Science and Food Safety, **2019,** 18(2), 565-583. Doi: http://dx.doi.org/10.1111/1541-4337.12431

Petracci, M.; Bianchi, M.; Mudalal, S.; Cavani, C. Functional ingredients for poultry meat products. Trends Food Sci. Technol. **2013**, 33, 27-39. Doi: http://dx.doi.org/10.1016/j.tifs.2013.06.004

Rao, V. S., Vidyavathi, D. S. Comparative investigations and performance analysis of FCM and MFPCM algorithms on iris data. Indian Journal of Computer Science and Engineering, **2010,** 1(2), pp. 145-151.

Romeo, L., Loncarski, J., Paolanti, M., Bocchini, G., Mancini, A., Frontoni, E. Machine learning-based design support system for the prediction of heterogeneous machine parameters in industry 4.0. Expert Systems with Applications, **2020,** 140, 112869.doi: http://dx.doi.org/10.1016/j.eswa.2019.112869

Siddique, A., Shirzaei, S., Smith, A. E., Valenta, J., Garner, L. J., Morey, A. (2021). Acceptability of Artificial Intelligence in Poultry Processing and Classification Efficiencies of Different Classification Models in the Categorisation of Breast Fillet Myopathies. Frontiers in Physiology, **2021,** 1472.doi: http://dx.doi.org/10.3389/fphys.2021.712649

Sihvo, H. K., Immonen, K., & Puolanne, E. Myodegeneration with fibrosis and regeneration in the pectoralis major muscle of broilers. Veterinary pathology, **2014,** 51(3), 619-623. doi: http://dx.doi.org/10.1177/0300985813497488

Singh, A. K., Purohit, N., Varma, S. Fuzzy logic-based clustering in wireless sensor networks: a survey. International Journal of Electronics, **2013,** 100(1), 126-141.doi: http://dx.doi.org/10.1080/00207217.2012.687191

Soglia, F., Mudalal, S., Babini, E., Di Nunzio, M., Mazzoni, M., Sirri, F., Petracci, M. Histology, composition, and quality traits of chicken Pectoralis major muscle affected by wooden breast abnormality. Poultry Science, **2016**, 95(3), 651-659.doi: http://dx.doi.org/10.3382/ps/pev353

Springer, Berlin, Heidelberg. Doi: http://dx.doi.org/10.1007/978-3-642-25501-4_18

Sun, S., Huang, R. An adaptive k-nearest neighbor algorithm. In 2010 seventh international conference on fuzzy systems and knowledge discovery, **2010,** Vol. 1, pp. 91-94, IEEE. Doi: http://dx.doi.org/10.1109/FSKD.2010.5569740

Swanson, M. D., Kobayashi, M., Tewfik, A. H. Multimedia data-embedding and watermarking technologies. Proceedings of the IEEE, **1998,** 86(6), pp. 1064-1087.doi: http://dx.doi.org/10.1109/5.687830

Tijare, V. V., Yang, F. L., Kuttappan, V. A., Alvarado, C. Z., Coon, C. N., Owens, C. M.(2016). Meat quality of broiler chicken breast fillets with white striping and woody breast muscle myopathies. Poult. Sci. J. **2016,** 95, 2167–2173. doi: http://dx.doi.org/10.3382/ps/pew129

Van Hattum, P., Hoijtink, H. Market segmentation using brand strategy research: Bayesian inference with respect to mixtures of log-linear models. Journal of classification, **2009,** 26(3), 297-328.doi: http://dx.doi.org/10.1007/s00357-009-9040-1

Vapnik, V., Izmailov, R. Rethinking statistical learning theory: learning using statistical invariants. Machine Learning, **2019,** 108(3), 381-423.doi: http://dx.doi.org/10.1007/s10994-018-5742-0

Wang, Q., Fleury, E. Uncovering overlapping community structure. In Complex networks, **2011,** (pp. 176-186).

Wang, Y., Yang, Y., Pan, D., He, J., Cao, J., Wang, H., Ertbjerg, P. Metabolite profile based on 1H NMR of broiler chicken breasts affected by wooden breast myodegeneration. Food chem.,**2020,** 310, 125852. doi: http://dx.doi.org/10.1016/j.foodchem.2019.12585223

Wold, J. P., Måge, I., Løvland, A., Sanden, K. W., Ofstad, R. Near-infrared spectroscopy detects woody breast syndrome in chicken fillets by the markers protein content and degree of water binding. Poultry science, **2019,** 98(1), 480-490.doi: http://dx.doi.org/10.3382/ps/pey351

Wold, J. P., Veiseth-Kent, E., Høst, V., Løvland, A. Rapid on-line detection and grading of wooden breast myopathy in chicken fillets by near-infrared spectroscopy PLoS One, **2017,** 12(3), e0173384.doi: http://dx.doi.org/10.1371/journal.pone.0173384

Wu, X., Kumar, V. (Eds.). The top ten algorithms in data mining, **2009,** CRC press.doi: http://dx.doi.org/10.1201/9781420089653

Yong,Y. Chongxun, Z., Pan, L. A Novel Fuzzy C-Means Clustering Algorithm for Image Thresholding. Measurement Science Review, **2004,** vol. 4, no.1, 2004.

Yoon, S. C., Bowker, B. C., Zhuang, H. Toward a fusion of optical coherence tomography and hyperspectral imaging for poultry meat quality assessment. Electronic Imaging, **2016**, (14), 1-5.doi: http://dx.doi.org/10.2352/ISSN.2470-1173.2016.14.IPMVA-380

Zhang, P., Wang, R., Shi, N. IgA Nephropathy Prediction in Children with Machine Learning Algorithms. Future Internet, **2020,** 12(12), 230.doi: http://dx.doi.org/10.3390/fi12120230

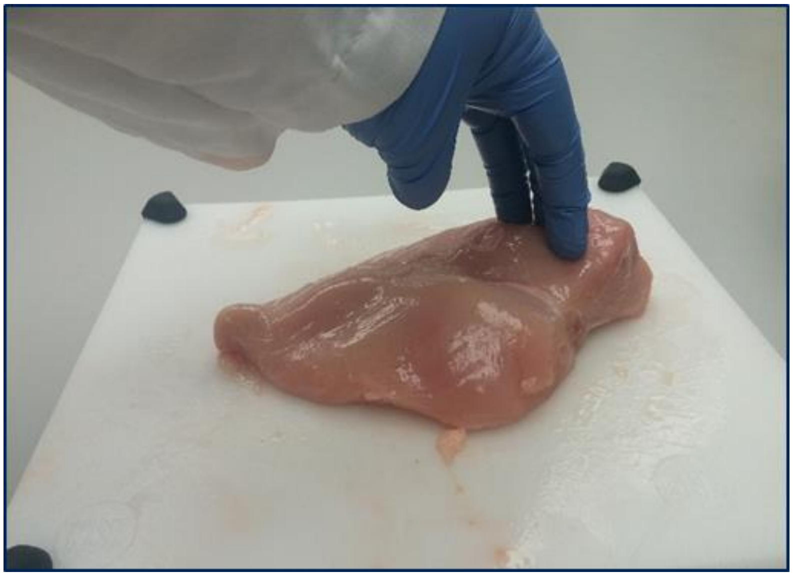


Figure 3-1: Hand palpation of fillets for manual classification based on severity level

Figure 3-2: Hand-held BIA setup for the classification of fillets based on different myopathic conditions

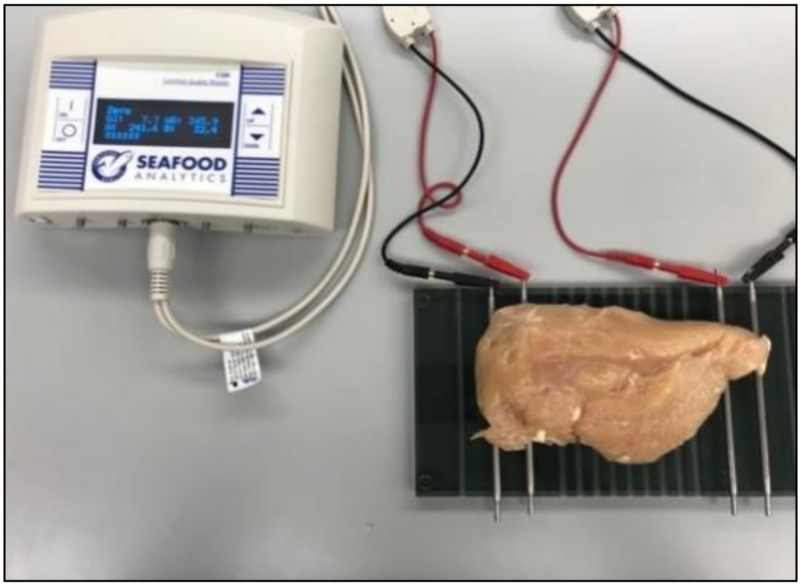


Figure 3-3: Plate BIA device for the classification of fillets based on severity level

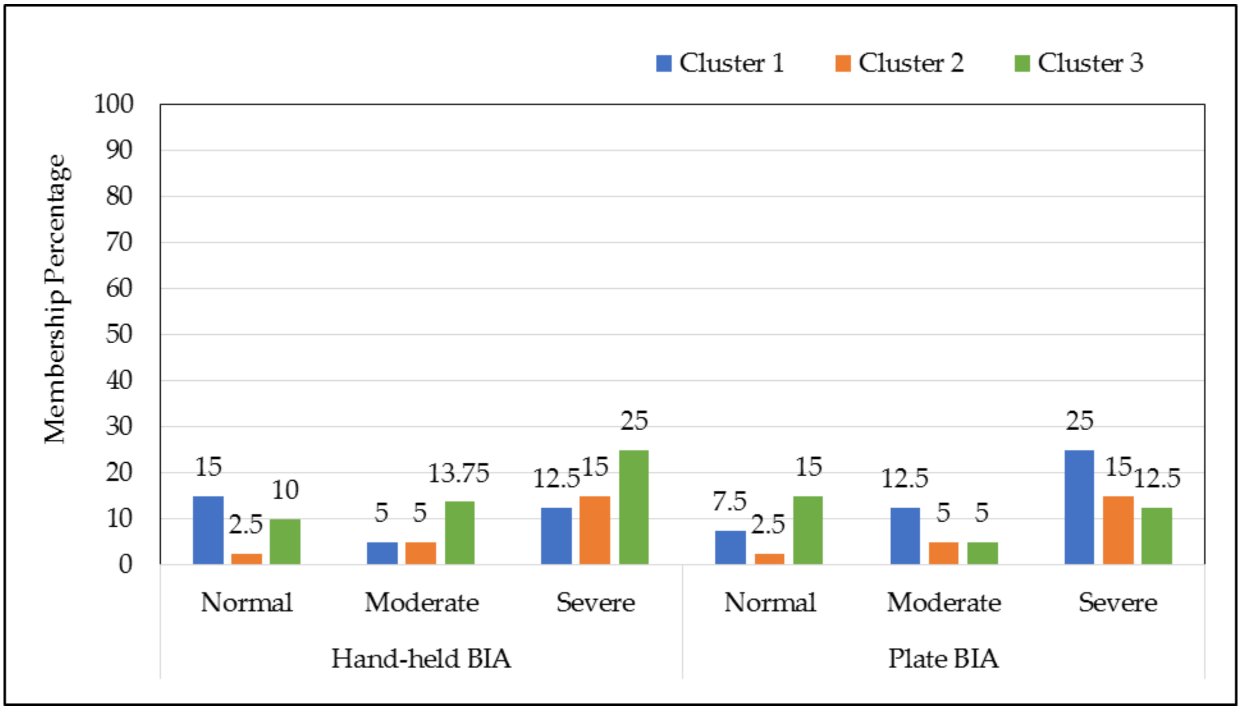


Figure 3-4: Membership cluster percentage formed in fuzzy c means analysis for different BIA setup

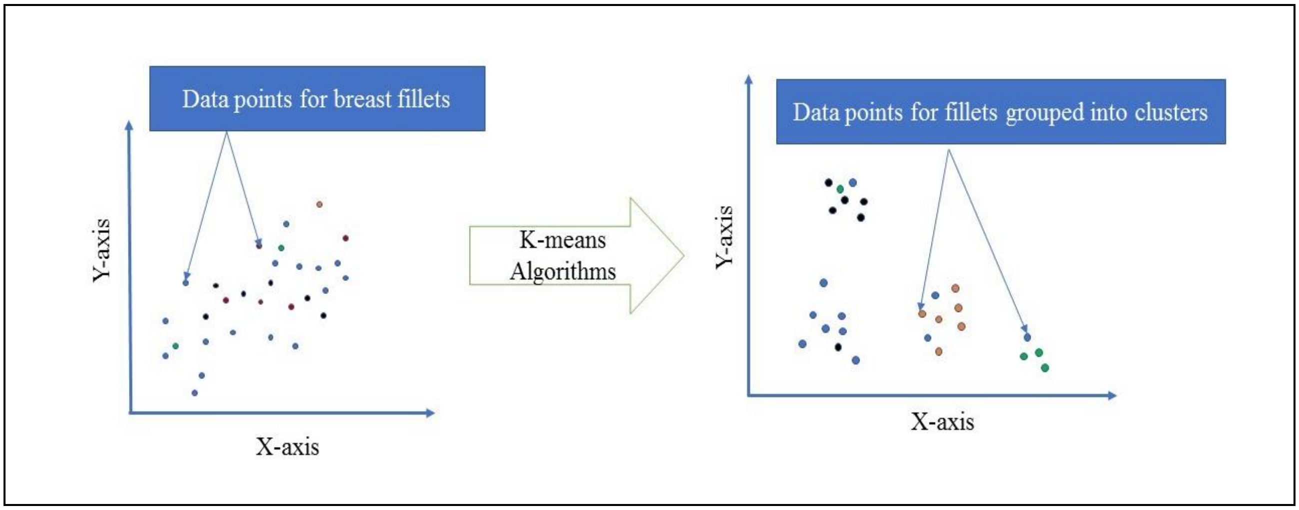


Figure 3-5: Diagrammatic representation of k-means clustering

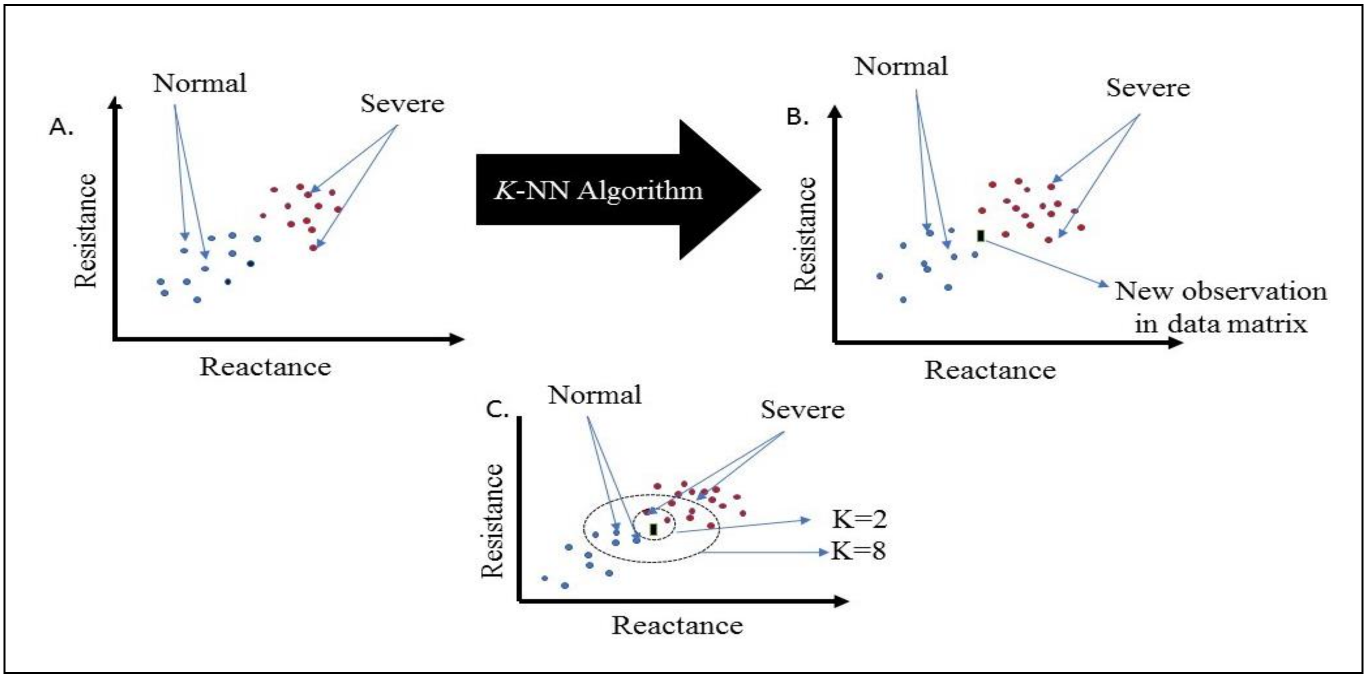


Figure 3-6: Diagrammatic representation of steps in k- nearest neighbor clustering analysis

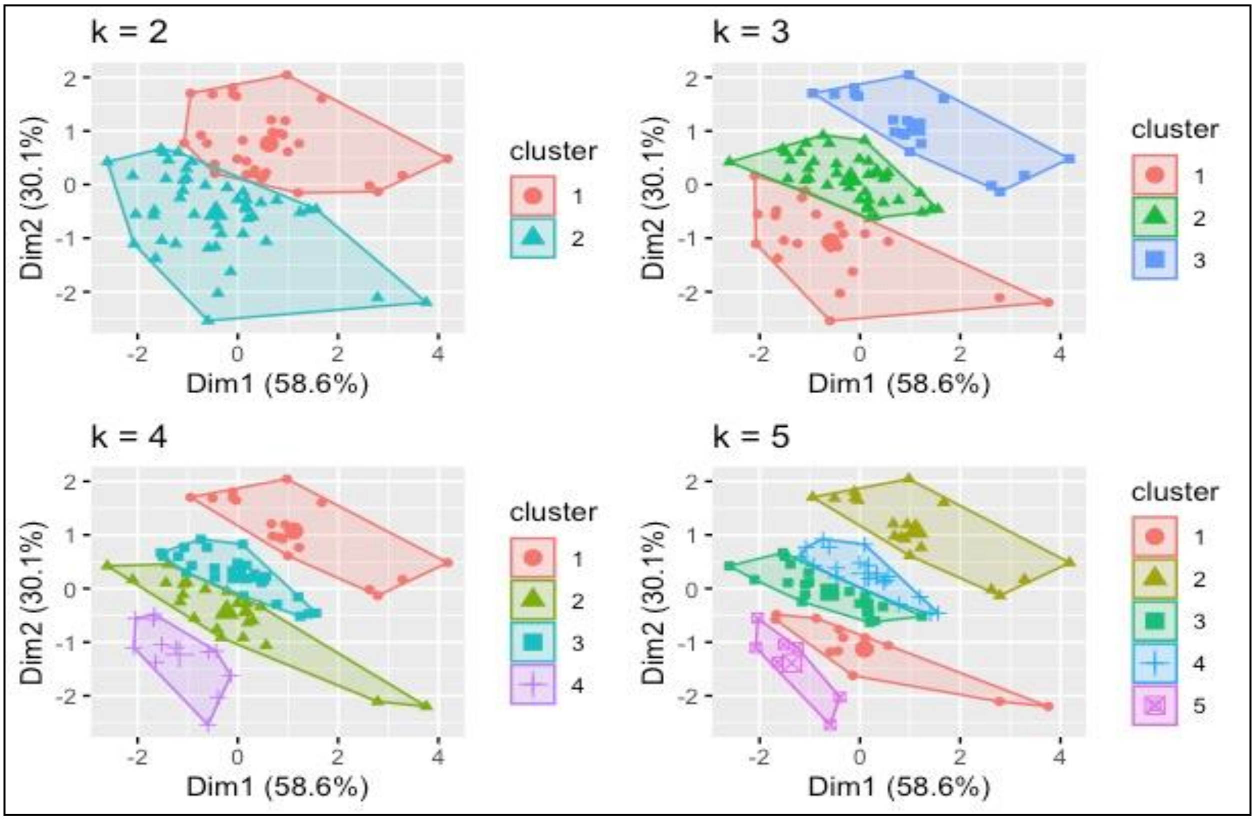


Figure 3-7: Different clusters formed for hand-held BIA data collected for different severity level of fillets

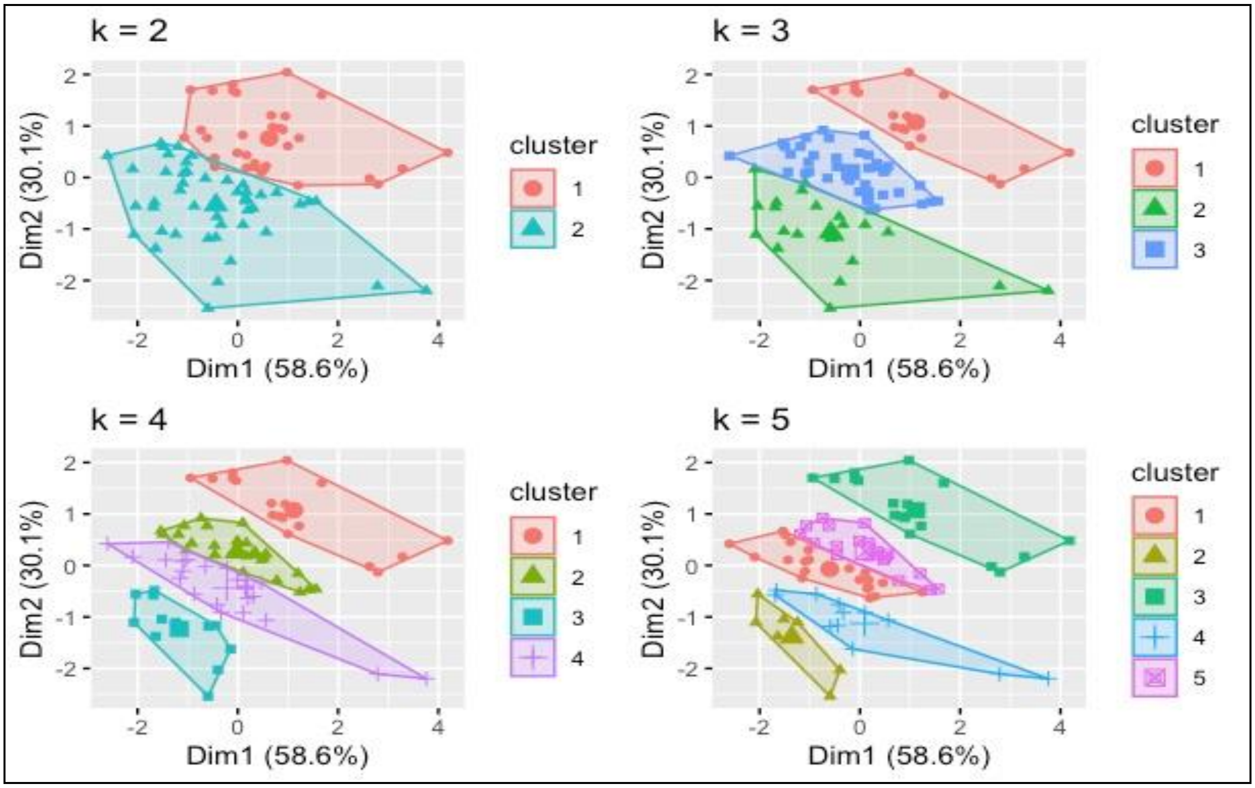


Figure 3-8: Different clusters formed for plate BIA data collected for different severity level of fillets

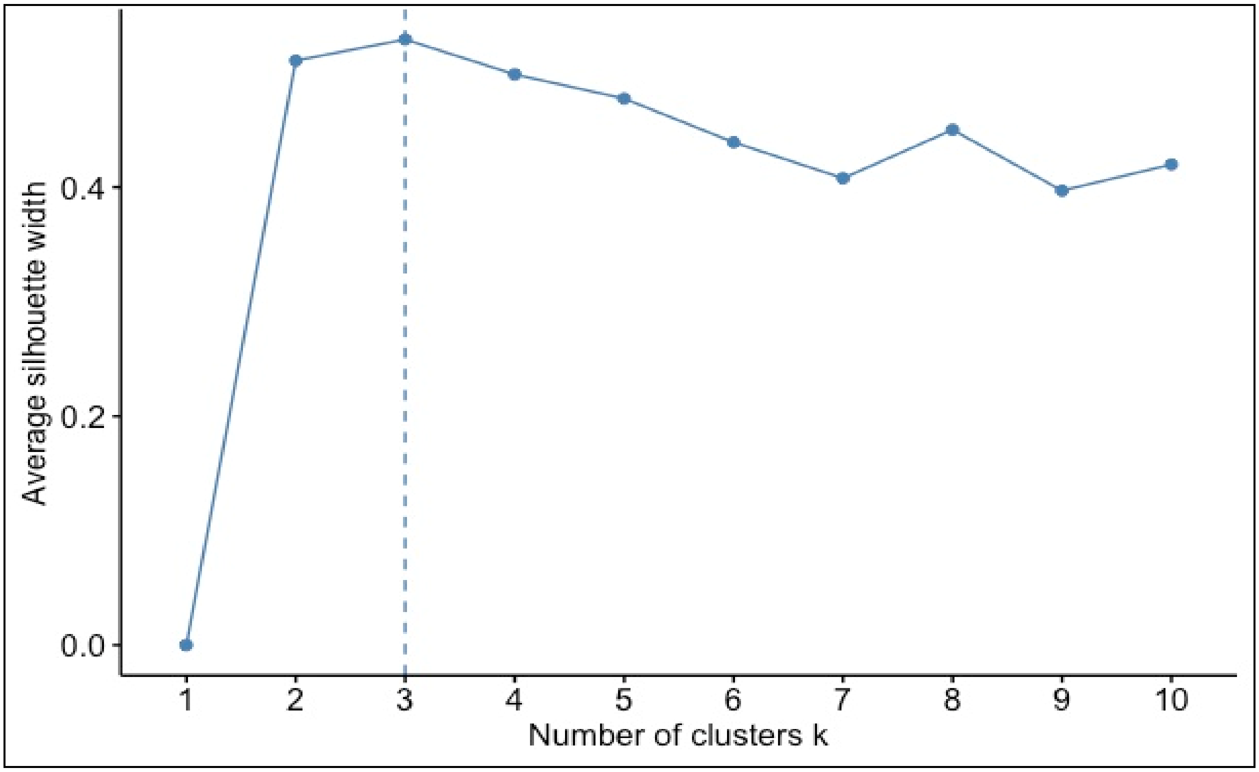


Figure 3-9: Graph for optimal number of clusters based on Silhouette Index value for handheld BIA dataset.

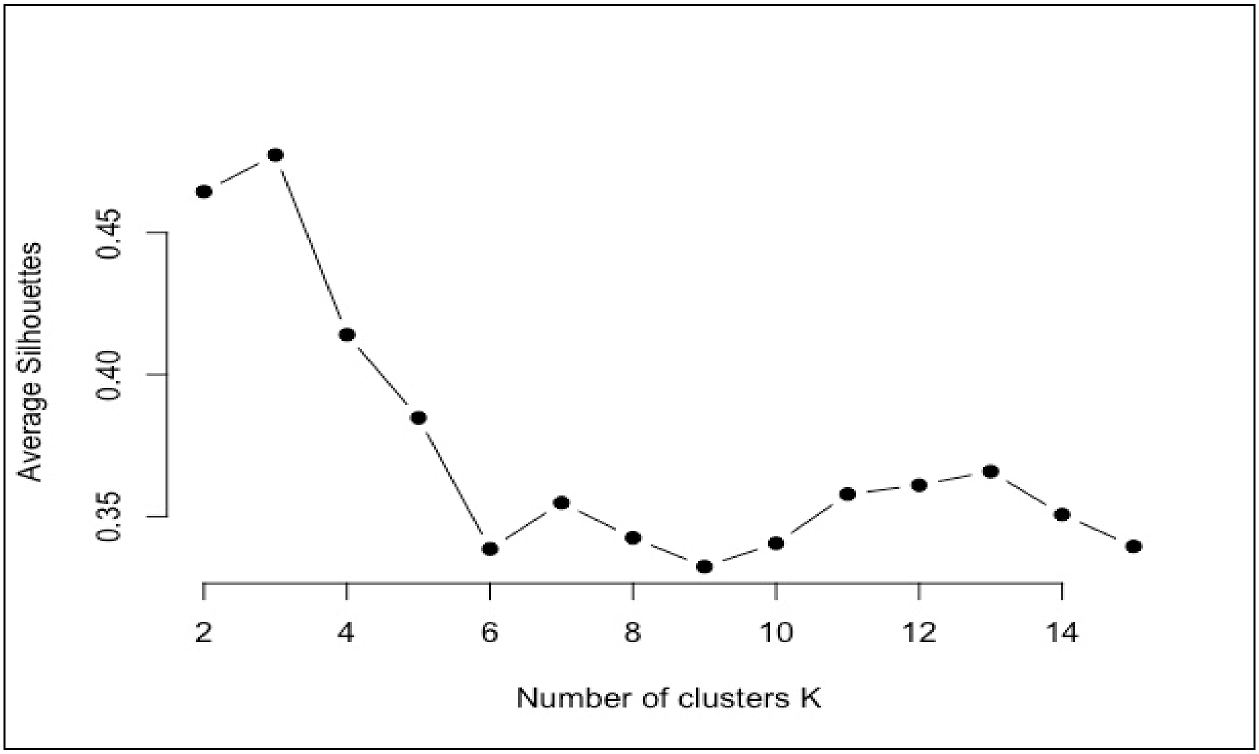


Figure 3-10: Graph for optimal number of clusters based on Silhouette Index value for plate BIA dataset.

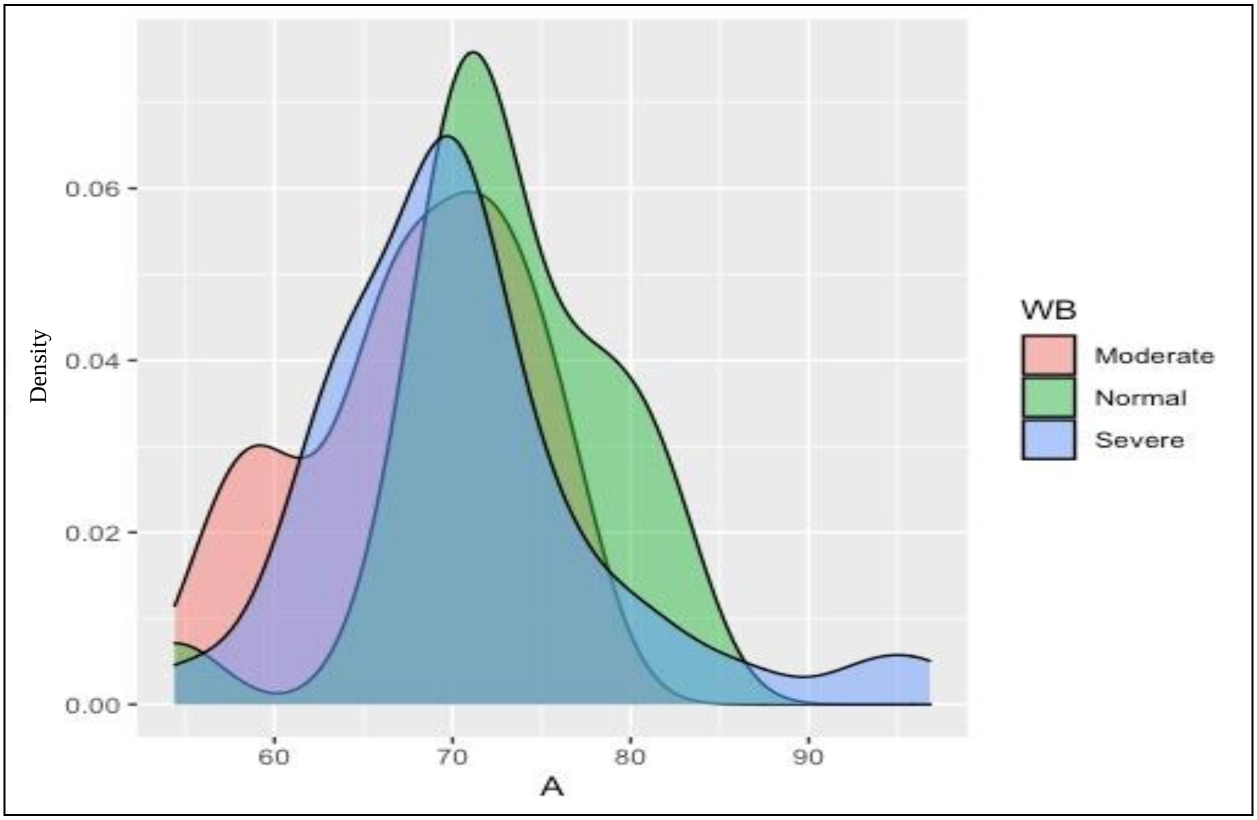


Figure 3-11: Density graph representing average distribution and overlapping of resistance (A) based on different severity levels

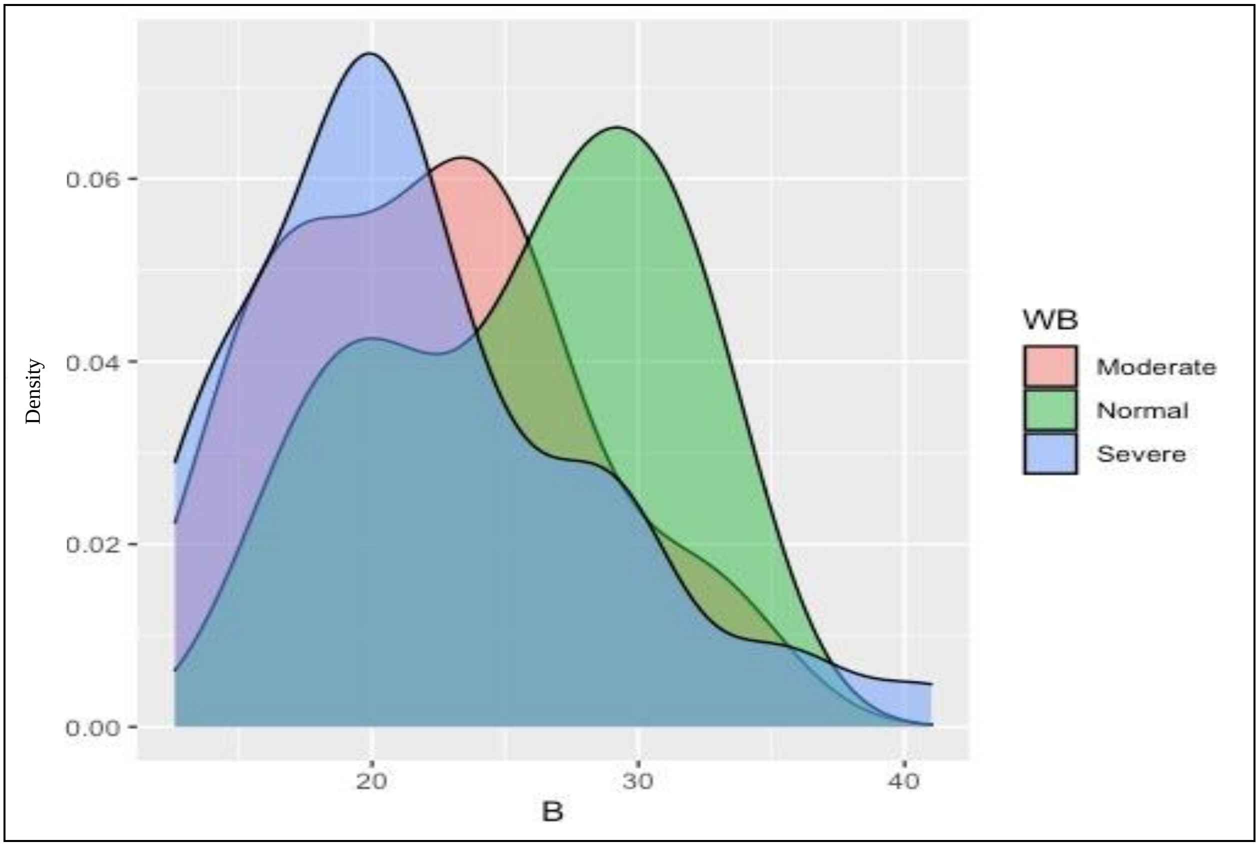


Figure 3-12: Density graph representing average distribution and overlapping of reactance (B) based on different severity levels

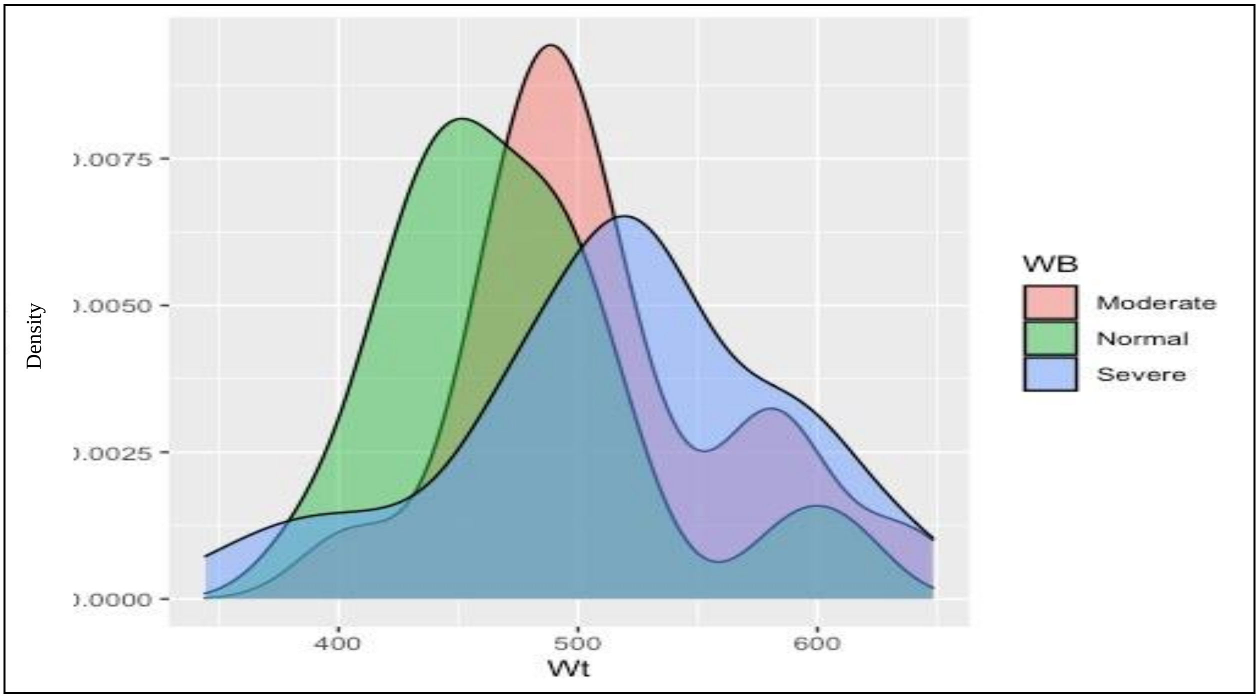


Figure 3-13: Density graph representing average distribution and overlapping of fillets weights (Wt) based on different severity levels

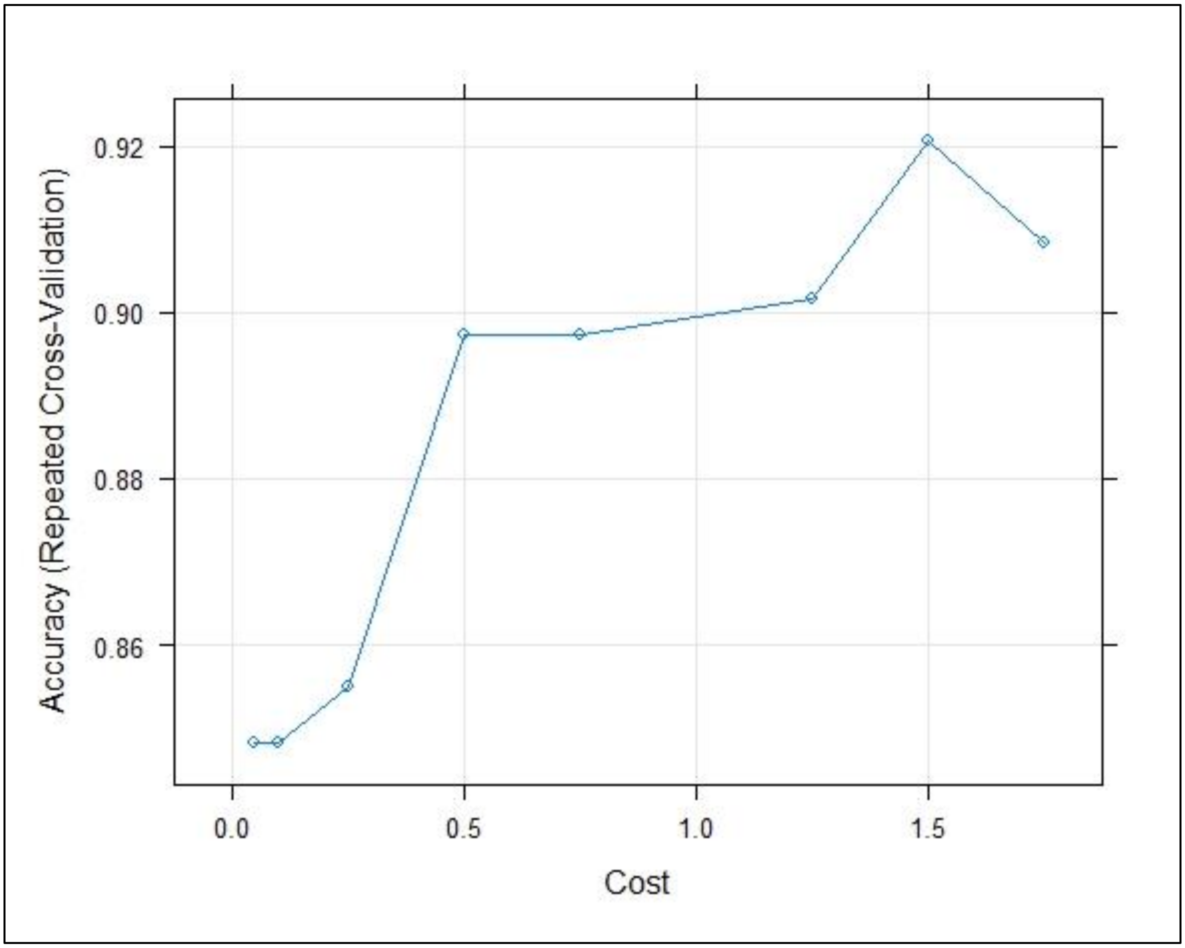


Figure 3-14: Hand-held BIA collected data graph for repeated cross validation accuracy for SVM model with cost function

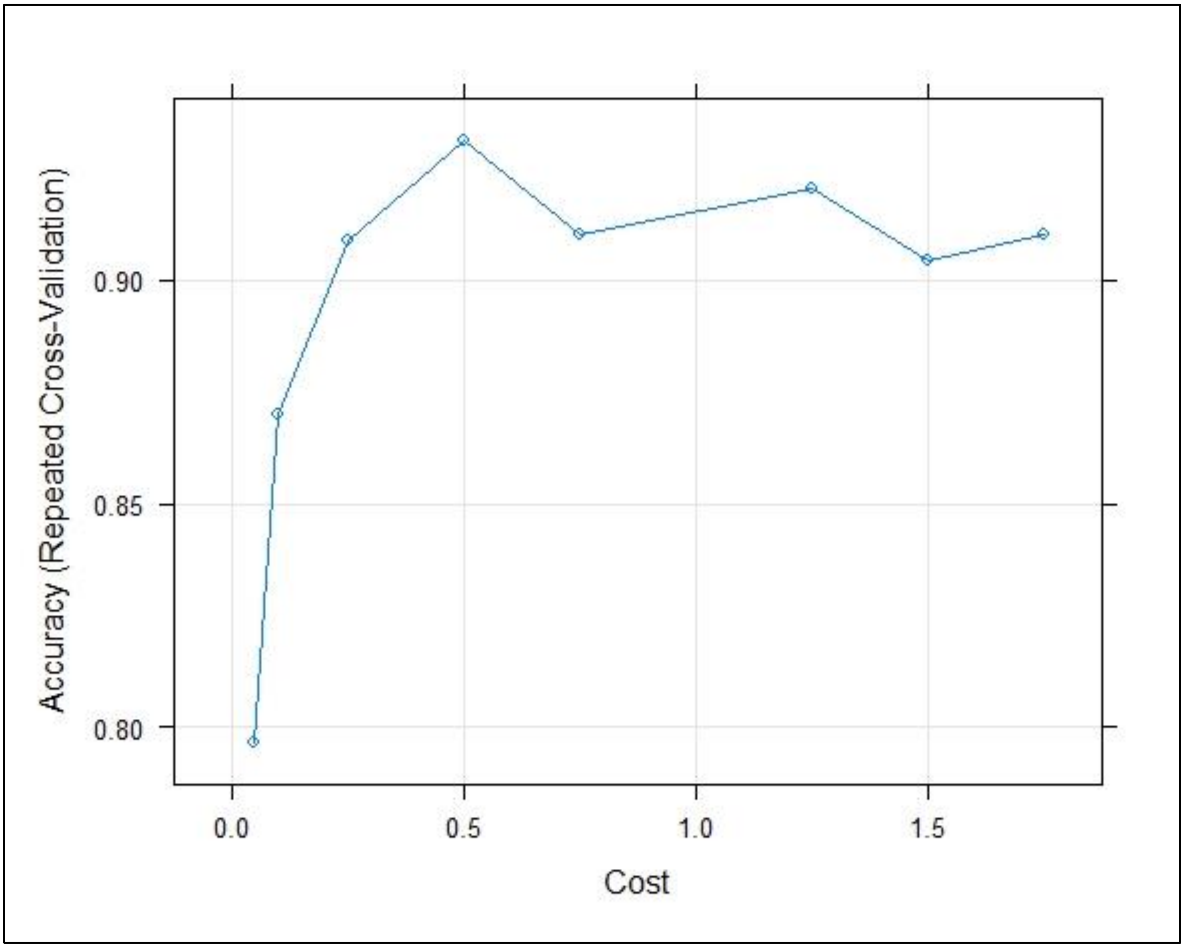


Figure 3-15: Repeated cross validation accuracy graph for plate BIA collected data in SVM model with cost function

Table 3.1: Summary table for two different bioelectrical impedance device setup for resistance, reactance and fillet weights among woody chicken breast fillets with varying severity levels.

		Hand-held BIA		Plate BIA	
WB Type	**Fillet Weight (g)**	**Resistance (R; Ω)**	**Reactance (X_c, Ω)**	**Resistance (R; Ω)**	**Reactance (X_c, Ω)**
Normal	473.54 ± 54.56^{b}	72.89 ± 6.25^{a}	25.76 ± 5.50^{a}	103.34 ± 15.89^{ab}	31.02 ± 8.91^{a}
Moderate	510.10 ± 57.26^{ab}	67.88 ± 6.19^{b}	22.14 ± 5.51^{b}	101.61 ± 14.50^{b}	28.78 ± 9.02^{a}
Severe	514.96 ± 69.33^{a}	70.60 ± 8.24^{ab}	21.76 ± 6.48^{b}	112.02 ± 16.68^{a}	33.98 ± 10.47^{a}

[a,b] Means with different superscript in columns are significantly ($p < 0.05$) different from each other.

Table 3.2: Percentage probability of fillets grouped into three different cluster

	Cluster probability percentage for Hand-held BIA			Cluster probability percentage for plate BIA		
Fillets Type	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 1**	**Cluster 2**	**Cluster 3**
Normal	54.5	9.09	36.3	30	10	60
Moderate	21.0	21.0	57.8	55.5	22.2	22.2
Severe	23.8	28.5	47.6	47.6	28.5	23.8

Table 3.3: K-nearest neighbor classification table for hand-held BIA, and Plate BIA collected parameters occurred with different severity levels of woody breast myopathies.

WB Classification	No. of Fillets	Hand-held BIA		Plate BIA	
		Training (%)	**Testing (%)**	**Training (%)**	**Testing (%)**
Normal	21	38.50	**50.00**	31.30	**40.00**
Moderate	17	10.0	**42.90**	7.70	**25.00**
Severe	42	57.70	**87.50**	57.10	**78.60**

Table 3.4: Summary table of classification accuracies for hand-held BIA and plate BIA setup for normal, moderate, and severe woody chicken breast fillets using SVM algorithms.

WB Classification	No. of Fillets	Hand-held BIA		Plate BIA	
		Testing (%)	**No. of classified fillets/Total fillets**	**Testing (%)**	**No. of classified fillets/Total fillets**
Normal	21	**77.78**	**7/9**	**80.00**	**4/5**
Moderate	17	**85.71**	**6/7**	**66.67**	**6/9**
Severe	42	**88.89**	**16/18**	**85.00**	**17/20**

*No. of classified fillets and total number of fillets are based on testing data set

Table 3.5: Summary table for the average cluster distance values from the center of the clusters for Hand-held BIA and Plate BIA

WB Type	Cluster Number	Average distance range for Plate BIA	Average distance range for Hand-held BIA
Normal	1	0.14-2.50	0.04-4.73
Moderate	2	0.45-3.83	0.07-2.22
Severe	3	0.46-5.09	0.20-7.49

Table 3.6: K-means clustering table for average cluster means for 3 different clusters for conventional BIA and plate BIA collected parameters occurred with woody breast myopathies.

	Hand-held BIA			Plate BIA		
Clusters	**Resistance Means ±SD**	**Reactance Means ±SD**	**Weight Means ±SD**	**Resistance Means ± SD**	**Reactance Means ± SD**	**Weight Means ± SD**
1	64.52±4.88	19.78±4.71	571.36±38.29	101.25±13.13	27.06±5.67	462.68±44.17
2	81.71±5.91	31.80±3.33	444.12±52.44	109.62±11.66	34.66±6.71	556.27±44.84
3	70.75±3.04	21.63±4.47	475.26±33.10	143.21±8.73	56.25±5.21	542.28±67.95

Table 3.7: Confusion matrix table for number of fillets classified in each labeled categories in testing data set using split data set method

Fillets Type	Hand-held BIA			Plate BIA		
	Normal	Moderate	Severe	Normal	Moderate	Severe
Normal	7	1	1	4	1	0
Moderate	1	6	0	1	6	2
Severe	1	1	16	3	0	17

Chapter 4

EFFECT OF AGE, DEBONING TIME OF CARCASS, AND DIFFERENT COOKING CONDITIONS ON THE WOODY BREAST MYOPATHIES IN CHICKEN: A META-ANALYSIS

Aftab Siddique [1], Micah T. Black[1], Bet W. Alvarado[1], Laura Garner[1], Tung-Shi Huang[1], Ashish Gupta[2], Alan E. Wilson[3], Jayson T. Sawyer[4], Amit Morey[1,*]

[1]*Department of Poultry Science, Auburn University, Auburn, AL 36849, USA;*
azs0242@auburn.edu, bsw0033@auburn.edu, mtb0042@auburn.edu, bauerlj@auburn.edu, huangtu@auburn.edu (A.S; B.W.A, M.T.B.; LG, T.S.H);

[2]*Department of Business analytics and Information, Auburn University, AL 36849, USA;*
azg0074@auburn.edu (AG);

[3]*School of Fisheries, Aquaculture, & Aquatic Science, Auburn, Alabama, 36849, USA;*
aew0009@auburn.edu, (AEW)

[4]*Department of Animal Science, Auburn University, Auburn, Alabama, 36849, USA;*
jts0109@auburn.edu (JTS)

*****Correspondence: azm0011@auburn.edu; (AM) Tel: +1-229-395-9837

Amit Morey, Ph.D.

Associate Professor

Address: 260 Lem Morrison Drive

202-E Poultry Science Building

Auburn University, AL 36849

4.1 ABSTRACT

A Meta-analysis review was performed to examine the different approaches for identification of myopathic fillets and evaluate the effects of age of bird, deboning time, different cooking and storage conditions on woody breast (WB) myopathic conditions in broilers deboned fillets. Data we collected from 20 different articles based on inclusion criteria searched and extracted from different databases and online resources. Deboning time has shown significant effect ($p<0.001$) on MORS, BMORS and descriptive analysis values. This quantitative analysis identifies that there is a strong impact of instrumentation techniques such as Compression force, shear force different cooking conditions on BMORS shear force values (R^2=86.80%) were observed with significance level ranging from 0.01 to 0.001. Deboning time showed strong evidence to have impact on the MORS shear force values (R= 64.03 %). There was minimal effect of deboning time, age of bird and cooking conditions on descriptive sensory evaluation when compared with woody breast fillets (Age of birds: R^2=26.53 %; Cooking conditions: R^2=32.57 %; Deboning time: R^2=10.06%). The overall effect of age of birds have shown significant difference for the evaluated parameters on the chicken breast meat quality (Hedges'g [95% CI] =-0.72 [0.17, 1.26], I^2 = 93%, p<0.01). Sous vide method of cooking for woody breast fillets have significant effect for different analyzed shear force energies and sensory descriptive sensory evaluation (Hedges'g [95% CI] =5.30 [-50.30, 83.40], I^2 = 98%, p<0.01).

Keywords: BMORSE, MORS, Woody breast fillets, Sous Vide

4.2 INTRODUCTION

Different types of meat and meat products that are nutrient-filled are always the first and foremost choice for protein across the world (Heinz and Hautzinger, 2007). There is a drastic increase in meat and meat products consumption worldwide since the last couple of decades. Into which the consumption per capita had increased from 22.04 lbs in the 1960s to 57.3 lbs in 2000 yearly and will reach up to 81.5 lbs by the year 2030 (NCC, 2022). Consumption of meat-based protein is one of the main sources for almost every consumer in the United States as well as at the global level. According to National Chicken Council (NCC), nearly nine billion broilers were raised in U.S and as per estimates, per capita consumption of chicken is nearly 94.5 pounds of chicken every year in the United States (NCC, 2022). The popularity of chicken meat is in high demand because of different organoleptic attributes such as texture, color, and flavors (Petracci et al. 2013). High consumer demand of better-quality chicken breast meat is increasing which is having an impact on the industry to produce fast growing birds, feed efficiency, and the measurement of the breast muscle (Petracci and Cavani. 2012).

To meet the excessive demands for boneless white meat, the broiler growers and processors have successfully incorporated and utilized better genetic breed selections which resulted in improvement in nutritional diet to obtain weight gain in an average chicken, increased growth rate and also increase in total carcass yield. In the response of continuously changing markets demands that is completely guided as per end users' preferences and demands which inclined more towards cut-up processed chicken parts rather than whole chicken carcasses. Despite of having fast growing chickens and increase in white breast meat yield there has been an increase in the cases of breast myopathies, one of the muscle abnormalities that has been discovered in broiler breast meat is referred to as "woody breast" and is more dominant in bigger and heavier birds. This woody breast (WB) (figure 1) condition can be easily identified by their faded-pale color appearance with swollen caudal part of the breast fillet which is consist of varying level of hard appearance . WB myopathy condition can be identified by stiffness in the breast muscle, which may have on its surface faded pale color and exudate (Sihvo et al. 2014; Kuttappan et al. 2017). Velleman et al.(2017) stated a theory on the hardness in chicken breast meat may be due to fibrosis which is the outcome of cross-linked collagen fibrils accumulation. Soglia et al. (2016) reported that collagen may be the one of the reasons for increased firmness related to the development of this condition (Soglia et al. 2016). These changes in the woody breast muscles can also influence different physical and chemical meat quality attributes such as pH, color, water holding capacity (WHC), cook loss and texture profile attributes mainly associated with the pectoralis major muscle (Kuttappan et al. 2012).

4.3 CLASSIFICATION ACCURACY

In simple terms, classification accuracy can be defined as the number of right predictions divided by the total number of predictions highlights a categorization effectiveness of the models. It is the most used statistic for assessing binary classifiers (Chicco and Jurman, 2020). Classification models have been used in food sciences and related fields for more than a decade (Huang et al. 2007). There have been several classification algorithms established and are being used for classification-based studies such as Support Vector Machines (SVM)(Cortez et al. 2006; Asmara et al. 2017; Ning et al. 2020), Back Propagation Neural Networks (BPNN) (Rumelhart et al. 1988), Linear Discriminant Analysis (LDA) (Siddique et al. 2021). For more details on these classification accuracy models and their implementation, readers can refer to the article Siddique et al. (2021).

4.4 COMPRESSION FORCE (CF) AND SHEAR FORCE (SF)

Compression force (CF) can be defined as the force that is being generated by compressing an object or substance. In other words, when SF aligned to each other are defined as compression force to the object surface resulting in some degree of deformation. In the poultry processing industry, the textural characteristics of raw chicken breast meat are used as a set of criteria for WB characterization, several instrumental texture measurements have been used to evaluate the level of WB state in raw chicken fillets, including CF. According to published research, there was a substantial difference in CF between WB and regular normal fillets (Dalgaard et al. 2018; Sun et al. 2018). Mudalal et al. (2015) and Soglia et al. (2017) found that CF measurements of raw broiler breast meat when compared with the WB state were significantly greater than normal fillets. Since 1930's, Shear force (SF) test has been the most widely used instrumental approach for measuring meat tenderness (Destefanis et al. 2008). The force (N) as a function of knife movement (mm) and compressed to cut off a sample of tissue is measured in this test (MPa) and is determined by hardness or toughness of sample (Dar and Light, 2014). Shear force denotes to the movement of muscle parallel towards the axis of immediate contact while applying tangential force on the section. Nonetheless, within the food industry, this term is widely used to describe any cutting technique that separates a product into smaller fragments (Berk, 2018).

4.5 MEULLENET-OWENS RAZOR SHEAR (MORS) AND BLUNT MEULLENET-OWENS RAZOR SHEAR (BMORS)

Meullenet-Owens Razor Shear (MORS) was developed and first introduced by Cavitt et al.(2005) by the name of Razor Blade Shear which was renamed later as Meullenet-Owens Razor Shear (MORS) test. They have reported that the use of a razor blade in determination of texture of cooked chicken was much easier and more efficient. In addition to shear force use of a razor blade on meat samples, it also provides with one more additional parameter called as "shear energy". Use of MORS test on a sample reduces the chances of experimental error, no time is needed to prepare a sample and MORS is independent of sample size (Cavitt et al. 2005). Meullenet et al. (2004) have developed a modified version of MORS that provides better comparison between tough portions of meat (ASTM, 1992; Lee at al. 2008) . Reliability and the effectiveness of BMORS and MORS was demonstrated by Lee et al.(2008) on the tenderness of chicken breast meat. Instrumental analysis of breast meat by using BMORS test had shown much better correlation to the tenderness as reported by consumer panel (ASTM, 1992).

4.6 DESCRIPTIVE SENSORY ANALYSIS

Descriptive analysis one of the methods that elaborate the quality and intensity of specific end user product (Meullenet et al. 2004). A wide range of descriptive analysis techniques have been developed by using basic principles of sensory science. In conventional descriptive techniques, such as food attributes profiling methods and quantitative descriptive analysis, involve a trained person to objectively quantify the sensory attributes of samples (Lee et al. 2008). Due to the versatility of tasks that were completed with descriptive sensory analysis and the amount of generated data, this method had become the valuable source of product information, not just limited to research settings, but also for further processed food product development industries and government agencies (ASTM, 1992). Descriptive analysis for quality evaluation of product was first implemented for food products and beverages (Meullenet et al. 2004). The implementation of descriptive sensory evaluation is not only limited to evaluate different attributes of product, but now is also being used to monitor product lifecycle, mapping market graph, variety of product development, value optimization, and quality control of existing line product (Meullenet et al. 2004). Descriptive sensory analysis is more importantly used in various product design and development when sensory data is linked to consumer response through hedonic data and instrumental analysis data for physico-chemical attributes. Relative study of both generated data allows professionals and developers to easily understand the consumer preference trend which helps companies to design their product and to enhance quality attributes (Kemp et al.2018). As the poultry industry is a fast growing business and can also be considered as one of the important contributing factor in providing food to growing population, creating jobs, and consumer driven in nature, the presented work in this article will provide an insight to the readers about the important parameters to consider for designing of their research with set parameters as different authors have used different parameters of deboning time, storage conditions along with some insight on the use of big data analytic approaches for the classification of myopathic fillets during inline processing. This paper tries to fill the gap for the literature evidences that can be helpful in designing the experiments, and will also be helpful for the poultry processing industries to implement novel techniques for their processes and optimizing steps to reduce losses and increasing profitability.

4.7 MATERIALS AND METHODS

4.7.1 Inclusion and Exclusion criteria

We included studies if they (1) were comparing raw chicken normal fillets with myopathic conditioned fillets; (2) Use of approaches such as MORS, BMORS, compression force, shear force; (3) Used big data analytics approaches such as Support Vector machine (SVM), Multilayer Perceptron (MLP), Back Propagation Neural Network (BPNN), and Linear discriminant analysis (LDA);(4) Used different cooking and storage conditions; (5) had used different processed products made from normal and myopathic fillets;(6) published in English.

Additionally, we excluded studies with no comparison between normal and myopathic fillets, papers with subjective analysis (WB scoring by hand palpation) because the subjective analysis is outside the aims of the present study.

4.7.2 Databases and search criteria

The Web of Science database, Google Scholar, and publications from Poultry Sciences related journals were searched for articles that had examined the detection of woody breast condition different cooking condition, storage conditions, and deboning time for the normal and woody breast chicken fillets (from January 2011 to December 2021). These databases were selected on the merits of having full text articles that were published in English language. The following search string was used to locate plausible studies on woody breast myopathic conditions are “woody breast” OR “woody breast myopathies” OR “muscle abnormalities” OR “abnormalities in fast growing broilers” OR “BMORS (Blunt Meullenet-Owens Razor Shear)” OR “MORS (Meullenet-Owens Razor Shear)” OR “TPA” (texture profile analysis) OR “compression force” OR “shear force” OR “classification accuracy” combined with different cooking and storage methods such as “Raw Frozen” OR “Frozen thawed”, “Cooked” OR “Grilled” OR “Baked” OR “Boiled”. These searches led to a total publication 630, from which 200 duplicate articles were removed. Other information that was appeared during search process such as studies related to spaghetti meat, white stripping meat, and woody breast scoring by hand palpation have been excluded to ensure convenient search of papers related to the review questions of the paper. Full texts that were downloaded have been inspected in detail. A total number of 20 complete research articles were selected on the basis of classification accuracy, toughness, tenderness, and descriptive sensory evaluation parameters such as hardness, cohesiveness, gumminess, and chewiness as analyzed using big data analytics approaches as

Support Vector Machines (SVM), Back Propagation Neural Network (BPNN), Random Forest (RF), Multilayer Perceptron (MLP), BMORS, MORS, Textural profile analysis (TPA) by descriptive sensory evaluation for age of bird, deboning time, cooking methods and storage condition on woody breast myopathic conditions. Every selected research article for this paper was thoroughly reviewed by other authors (BW, TB, CH) included in this papers using standard procedure as described below. Complete references were collected and information that was extracted from the publication are crossed verified whether the collected information was extracted from primary experimental research, or from a review or meta-analysis (PRISMA flow diagram 1).

4.7.3 Effect size calculations

Effect size in meta-analysis can be defined as the difference between two experimentally created groups (control and treatment group) (Morris and DeShon, 2002) . For this paper, the standardized mean difference (differences in means; Hedges' g) was used to measure the difference between mean values in control (normal breast fillets) and treatment group (woody breast fillets) relative to the pooled standard deviation. This standard statistic measures that how much the treatment affects the outcome on average relative to the control (Schmid et al. 1998).

4.7.4 Publication bias

Rosenthal's fail, safe approach, and funnel plot analysis were used to address the potential publication bias that influenced the outcomes of studies (Rothstein, 2008). Rosenthal's approach suggests that how many studies would need to be published before retrieving and performing meta-analysis data to nullify the effect size (Braver et al. 2014). This non-significant number of studies were calculated through logarithm using software. If this approach suggests requirement of up to ten studies to void the effect, then it would be considered that true effect was insignificant but if it showed the higher number of studies, for example 20,000, then there would be little to no reason of concern. Funnel plot is generated with effect size on x-axis and variance on y-axis. Clusters of the larger studies generally appear on the top and smaller studies appear at the base of the plot. In addition, the trim and fill method was also used to estimate the number of missing studies (Duval and Tweedie, 2000).

4.8 DATA ANALYSIS

For the analysis of collected data, R language software (Version 4.2.0; Vigorous Calisthenics) was used. Random or mixed effect models were used because fixed effect model analyzes true effect size based on differences between studies other than one true effect size as assumed in fixed effect model. Heterogeneity was also calculated to understand the variances in studies. Meta-regression model was used to determine the variation in effect sizes in studies that attributed to differences in classification accuracy, compression force, shear force, MORS and BMORS due to different deboning time, and age of birds. Heterogeneity is explained by the moderator (Q_M) and ominous (Q_E) heterogeneity (Table1).

4.9 RESULTS

4.9.1 Effect of Deboning Time

The overall effect of deboning time had significant impact on the different parameters evaluated for woody breast compared to control group using standardized mean difference (Hedges'g [95% CI] =1.30 [0.26, 2.34], I^2 = 95%, p< 0.01) showed strong relationship between deboning time of chicken carcasses on different parameters analyzed . The overall effect of deboning time on BMORS values are significantly different (Hedges'g [95% CI] =0.49 [0.09, 0.89], 2.88], I^2 = 73%, p< 0.01) . The BMORS value for deboning time at 3 hours showed small positive effect on myopathic fillets (Hedge's g = 0.36 [-0.23, 0.95], I^2 = 71%, p<0.01), for 2-hour deboning (Hedge's g = 1.11 [0.30, 1.93], I^2 = *NA*), and for 8 hours medium positive effect observed (Hedge's g = 0.60 [-0.39, 1.58], I^2 = 83%, p<0.01) with 83% of heterogeneity. Overall standardized mean difference for MORS analysis value (Hedge's g = 0.70 [-0.70, 2.09], I^2 = 95%, p<0.01) showed medium effect (g≥0.5) on the effect of deboning time on MORS value, high effect (g≥0.8) relationship was observed for 3 hour deboning time (Hedge's g = 3.23 [-2.20, 8.66], I^2 = 92%, p<0.01) showing that MORS analysis provides better results for , negative medium effect (g≥0.5) relationship was observed on 6 hour deboning time (Hedge's g = -0.71 [-1.97, 0.55], I^2 = 83%, p<0.01) and MORS value, and positive effect on 6 hour deboning time (Hedge's g = 0.36 [-0.23, 0.95], I^2 = 71%, p<0.01). For classification accuracy-based studies, analysis showed small effect (g≤0.2) (Hedges'g [95% CI] =0.20[-1.35, 1.74], I^2 = 98%, p< 0.01) , with positive small effect (g≤0.5) for 3 hour (Hedges'g [95% CI] =0.49 [-0.67, 1.65], I^2 = 82%, p< 0.01), indicating that techniques employed for the classification work performed well up to some extent for 3 hours of deboning time. Overall shear force value showed negative small effect (g≤0.2) for deboning time (Hedges'g [95% CI] =-0.23 [-1.43, 0.96], I^2 = 97%, p< 0.01) , 3-hour deboning time favors 68.30% of studies for normal fillets analysis using shear force method (Hedges'g [95% CI] = -0.39 [-2.24, 1.45], I^2 = 97%, p< 0.01). Overall, descriptive TPA analysis showed better meat qualities for normal fillets (g≤0.2) (Hedge's [95% CI] = -0.11 [-2.17, 1.94], I^2 = 79%, p< 0.01). Majority of studies (72.30%) comparing descriptive sensory analysis for myopathic fillets with normal fillets favored 3-hours of deboning time for normal fillets (Hedge's [95% CI] = -0.41 [-3.54, 2.72], I^2 = 84%, p< 0.01). Overall, textural profile analysis performed on normal and woody breast fillets for the effect of deboning time showed significant difference in TPA values for different textural attributes (Hedges'g [95% CI] =-0.82 [-0.14, 1.79], I^2 = 83%, p<0.01). TPA values when analyzed separately for 2 hr (Hedges'g [95% CI] =-0.04 [-0.13, 0.21], I^2 = 29%, p= 0.19) was not significantly different when compared to 3hr (Hedges'g [95% CI] =1.11 [-0.20, 2.42], I^2 = 85%, p<0.01) deboning time.

4.9.2 Effect on the Age of Birds

Breast fillets that are analyzed for MORS (pooled Hedges' g [95% CI] =0.70 [-0.70, 2.09], I^2 = 95%, p< 0.01), BMORS (pooled Hedges'g [95% CI] =0.49 [0.09, 0.89], I^2 = 73%, p< 0.01), shear force (pooled Hedges'g [95% CI] =-0.23 [-1.43, 0.96], I^2 = 97%, p< 0.01), classification accuracy (pooled Hedges'g [95% CI] =0.20 [-1.35, 1.74], I^2 = 98%, p< 0.01) and for the descriptive analysis (TPA) (pooled Hedges'g [95% CI] =-0.09 [-2.13, 1.94], I^2 = 79%, p< 0.01) were significantly different for processing age of birds. From the analysis, large effects (g>0.8) were observed for compression force for all ages of birds ranging from 34 days old to 56 days old. Small effects (g<0.5) were observed for 56 days old birds when performed classification accuracy for rapid detection approaches, conventional BMORS analysis, and MORS analysis on the effect of age of birds. Overall, negative Hedge's g values for shear force and descriptive analysis showed small effect (g<0.2), based on individual age of birds 45 days old birds showed large effect (g>0.8) for shear force and descriptive sensory analysis. The overall effect of age shows significant effect on the breast meat quality (Hedges'g [95% CI] =1.30 [0.26, 2.34], I^2 = 95%, p<0.01). When analyzed together for classification accuracy, compression force, shear force, BMORS, MORS and TPA (descriptive analysis), the Hedges g values for age of birds at 34 days old (Hedges'g [95% CI] =1.43 [-2.06, 4.92], I^2 = 91%, p<0.01), at 38 days old (Hedges'g [95% CI] =11.05 [-108.57, 132.28], I^2 = 95%, p=0.58), at 42 days old birds (Hedges'g

[95% CI] =0.39 [0.09, 0.68], I^2 = 0.00%, *p*=0.86), at 45 days old (Hedges'g [95% CI] =2.29 [-0.49, 5.06], I^2 = 86%, *p*<0.01), 46 days old (Hedges'g [95% CI] =1.05 [0.73, 1.37], I^2 = *NA*), 48 days old (Hedges'g [95% CI] = 0.27 [-1.02, 1.57], I^2 = 96%, *p*<0.01), 52 day old (Hedges'g [95% CI] =0.63 [-15.62, 16.87], I^2 = 98%, *p*<0.01),56 days old (Hedges'g [95% CI] =1.10[-0.63, 2.83],, I^2 = 96%, *p*<0.01), and at 60 days old (Hedges'g [95% CI] =0.03 [-0.10, 0.17], I^2 = 0.00%, *p*=0.58) respectively, indicating that birds at the ages of 34d, 38d, 45d, 46d, and 56d showed a large effect (g>0.8), birds at the age of 52d showed medium (g>0.5), and birds at the age of 42d, 48d and 60 days showed small effect on different parameters evaluated on bird's age. Birds that are processed at the age of 45 days old (Hedges'g [95% CI] =2.66 [-0.86, 6.18], I^2 = 92%, *p*<0.01) and 52 (Hedges'g [95% CI] =0.53 [-0.27, 1.33], I^2 = 74%, *p*<0.01) are significantly different from other processed birds at the age of 42, 56, 60 days respectively.

4.9.3 Effect of Different Storage and Cooking Condition

Different storage conditions of raw and further processed fillets and their products are key factors in affecting the quality parameters such as texture (toughness, tenderness, juiciness, and chewiness), appearance (color), odor, and overall acceptability of product. It also affects the chemical parameters related to the meat quality [34-37]. The overall effect of different cooking conditions (Hedges'g [95% CI] =0.72 [0.17, 1.26], I^2 = 93%, *p*<0.01) have significant effects on breast fillet quality. Cooked breast fillets (Hedges'g [95% CI] =0.44 [0.21, 0.67], I^2 = 54%, *p*<0.01) showed a significant effect on shear force energies values obtained from MORS, BMORS and sensory descriptive evaluation values. There was not any significant differences observed between cooked hot served (Hedges'g [95% CI] =-0.09 [-0.44, 0.26], I^2 = 41%, *p*<0.17) and cooked cold served (Hedges'g [95% CI] =0.17 [0.13, 0.21], I^2 = 0%, *p*<0.99) breast fillets to the sensory panel for descriptive sensory evaluation. BMORS shear force values for cooked breast fillets were not significant when compared to raw breast fillets (Hedges'g [95% CI] =0.69 [-0.22, 1.60], I^2 = 98%, p<0.01). Overall, BMORS shear force values for the cooking conditions were significantly different (Hedges'g [95% CI] =1.07 [-0.73, 2.88], I^2 = 97%, p<0.01). MORS shear force value cooked samples (Hedges'g [95% CI] =0.93 [-0.10, 7.87], I^2 = 85%, p=0.01) were significantly different for various cooking methods in work done by Combs [12] such as Baked, Cooked frozen, Sous vide, Grill, and Raw frozen. Overall, the sous vide method of cooking for woody breast fillets showed a significant effect for different analyzed shear force energies and sensory descriptive sensory evaluation (Hedges'g [95% CI] =5.30 [-50.30, 83.40], I^2 = 98%, p<0.01).

4.9.4 Publication Bias

To completely avoid publication bias, a funnel plot was studied between effect size and sample size. Number of studies that are not in symmetrical order, which is interpreted to mean, with publication bias is present. The study which was conducted between the effect of deboning time on different shear force energies and descriptive sensory evaluation, shows that 58.4 % of the studies do not have publication bias, whereas in 41.6 % of the studies, publication bias is present. In addition, the effect of different conditions on BMORS, around 37.5 % of studies shows no publication bias. There is no publication bias observed in 42.8 % of the studies evaluated for MORS analysis. Descriptive sensory evaluation funnel plot shows no publication bias for 47.22 % of the studies.

4.10 DISCUSSION

The main aim of this conducted study was to evaluate the effect of deboning time, age of bird, and different cooking conditions on classification accuracy, compression force, shear force, and descriptive sensory evaluation for control group (normal breast fillets) as compared to woody breast fillets. From the analyzed data, it can be concluded that different cooking conditions have significant effects on the analyzed shear force value and sensory descriptive analysis. This quantitative analysis evidence shows that there was no impact on the age of birds, and deboning time on BMORS shear force values for wood breast fillets when compared to normal breast fillets. A strong impact of different cooking conditions on BMORS shear force values were observed with ranging *p* value from 0.01 to 0.001. There was no impact observed for the age of birds, and cooking condition on the analyzed MORS shear force value. Deboning time showed strong evidence to have impact on the MORS shear force values. There was minimal effects of deboning time, age of birds, and cooking conditions on descriptive sensory evaluation when compared with woody breast fillets. In current conducted study, the effect of different deboning times and age of birds have no such significance in shear force values analyzed by MORS and BMORS. As reported in other published articles, meat texture is noticeably influenced by cooking due to denaturation of different components, such as protein and fats (Forrest et al. 1975). In most cases, these modifications result in toughening of the muscle and an increment in shear values in cooked muscle products. In this analysis, cooking had a significant influence on the effectiveness of the shear approaches to distinguish between distinct WB groups (Moller, 1981). Each analyzed value at different conditions have showed high levels of heterogeneity, which may be due to other factors such as marination time, type of marination technique, different cooking conditions, age of processed birds, and deboning time. Classification accuracy for the identification of myopathic fillets is affected by the deboning time and different methods of analysis to analyze the collected data. Deboning time and methods for classification accuracy may have shown some level of heterogeneity due to the physiochemical characteristic of fillets such as the amount of collagen, fat, water, and some output variables such as reflectance, resistance, and reactance values generated from classification techniques. Classification results also depends on the selection of model for the analysis as different model uses different algorithms such as LDA and Random forest uses data dimensionality reduction techniques (Aydadenta and Adiwijaya, 2018) , while on the other hand, big data analytic approaches such as SVM, and MLP do not require the reduction of the data dimensionality (Chen et al. 2020; Siddique et al. 2021). Different techniques for the classification of fillets uses different algorithms to generate results and it is completely dependent on the researchers to set the acceptable range of the results. Several factors that are considered in these classification kind problems are linearity of data, amount of preprocessing of data, noise in the generated dataset, and different unknown confounding variables. Analysis in this paper showed that classification accuracy depends on the age of birds and methods used in classification study. Compression force analysis, and shear force analysis based on deboning time, the age of birds, different storage, and various conditions showed some level of heterogeneity due to the fact that heavy fast-growing broilers have more chances to show woody breast condition due to the deposition of collagen protein. Sihvo et al. (2014) have reported that increase in age and weight of broilers are directly associated with WB incidences [51]. A study conducted by Souza et al. (2005) suggested that early deboning of broiler carcass at 3 h and 4 h postmortem have caused a higher shear force value as compared to 24 h deboning time. During this study, several questions were raised, which need to be addressed with complete explanation, such as how effective is the treatment of different cooking conditions on analyzed shear force values or what would be the best average deboning time for chicken carcasses without affecting the speed to processing line within a profitable range.

4.11 CONCLUSIONS

In conclusion, this meta-analysis provides evidence that there are very smaller numbers of published research available for a comparative study between normal and woody breast fillets due to the fact that there is not a fixed quantitative method for the classification of myopathic fillets and methods that are available for the classification is completely based on employee experiences which are more susceptible to give deviated false results during the processes. Other factors that contribute to these results are unexperienced employees, speed of processing lines, stress on employees, and levels of fatigue. In our observation, those studies that have used big data analytic approaches such as regression model, LDA, computer vision systems have mainly focused on the identification techniques and reported whether the implemented techniques are able to detect the myopathic conditions in fillets or not without performing comparisons by how much the new technique is able to detect these conditions. More studies are encouraged to be performed to explore different methods to classify the fillets based on quantitative methods, rather than qualitative approaches, to set a standardized parameter with new innovative technologies in poultry processing plants that can be placed to reduce the losses associated with the misclassification of these fillets and will also be helpful in maintaining the quality and keep up with the speed that can reduce the incidences of misclassification. Interestingly, studies that have used myopathic fillets in further processing products that utilize the woody breast fillets have agreed to the fact that further processing steps for different products made from these myopathic fillets do not differ from the normal fillets product and consumer panel found these product as acceptable in nature.

Author Contributions: Conceptualization, AS; ; Methodology, A.S, AM, AW ; Validation, LG, A.S; MTB,AW; formal analysis, A.S.; MTB, B.W investigation, A.S; A.M., B.W.A and MTB ; data curation, A.S..; writing—original draft preparation, A.S.; writing—review and editing, A.M, A.G ,L.G, J.T.S, and T.S.H.; supervision, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement:

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

4.12 REFERENCES

Asmara, R. A., Rahutomo, F., Hasanah, Q., & Rahmad, C. (2017). Chicken meat freshness identification using the histogram color feature. In *2017 International Conference on Sustainable Information Engineering and Technology (SIET)* (pp. 57-61). IEEE. Doi: http://dx.doi.org/10.1109/SIET.2017.8304109

ASTM. (1992). Manual on Descriptive Analysis Testing for Sensory Evaluation. *ASTM Manual Series: MNL 13. American Society for Testing and Materials*.

Berk, Z. (2018). *Food process engineering and technology*. Academic press. ISBN: 9780128120187

Cavitt, L. C., Meullenet, J. F., Gandhapuneni, R. K., Youm, G. W., & Owens, C. M. (2005). Rigor development and meat quality of large and small broilers and the use of Allo Kramer shear, needle puncture, and razor blade shear to measure texture. *Poultry science*, *84*(1), 113-118.doi: http://dx.doi.org/10.1093/ps/84.1.113

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, *21*(1), 1 13.doi: http://dx.doi.org/10.1186/s12864-019-6413-7

Cortez, P., Portelinha, M., Rodrigues, S., Cadavez, V., & Teixeira, A. (2006). Lamb meat quality assessment by support vector machines. *Neural Processing Letters*, *24*(1), 41-51.doi: http://dx.doi.org/10.1007/s11063-006-9009-6

Dalgaard, L. B., Rasmussen, M. K., Bertram, H. C., Jensen, J. A., Møller, H. S., Aaslyng, M. D., & Young, J. F. (2018). Classification of wooden breast myopathy in chicken pectoralis major by a standardised method and association with conventional quality assessments.

International Journal of Food Science & Technology, *53*(7), 1744-1752.doi: http://dx.doi.org/10.1111/ijfs.13759

Dar, Y. L., & Light, J. M. (Eds.). (2014). *Food texture design and optimization*. John Wiley & Sons. Doi: http://dx.doi.org/10.1002/9781118765616

Destefanis, G., Brugiapaglia, A., Barge, M. T., & Dal Molin, E. (2008). Relationship between beef consumer tenderness perception and Warner–Bratzler shear force. *Meat science*, *78*(3), 153-156. doi: http://dx.doi.org/10.1016/j.meatsci.2007.05.031

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463. Doi: http://dx.doi.org/10.1111/j.0006-341X.2000.00455.x

Forrest, J. C., Aberle, E. D., Hedrick, H. B., Judge, M. D., & Merkel, R. A. (1975). Principles of meat processing. *Principles of meat science*, 190-226.

Heinz, G., and P. Hautzinger, 2007. Meat Processing Technology. For Small-To Medium scale Producers. Food and Agriculture Organization of the United Nations Regional Office for Asia and the Pacific. Retrieved on 10th April 2020, from https://agris.fao.org/agris search/search.do?recordID=XF2016077667

Huang, Y., Kangas, L. J., & Rasco, B. A. (2007). Applications of artificial neural networks (ANNs) in food science. *Critical reviews in food science and nutrition*, *47*(2), 113 126.doi: http://dx.doi.org/10.1080/10408390600626453

Kemp, S. E., Ng, M., Hollowood, T., & Hort, J. (2018). Introduction to descriptive analysis. *Descriptive Analysis in Sensory Evaluation*, *1*. Doi: http://dx.doi.org/10.1002/9781118991657.ch1

Kuttappan V, Lee Y, Erf G, Meullenet J, McKee S, Owens C. 2012. Consumer acceptance of visual appearance of broiler breast meat with varying degrees of white striping. En: Poultry science 91 (5), pág. 1240–1247. DOI: http://dx.doi.org/10.3382/ps.2011-01947

Kuttappan, V. A., Owens, C. M., Coon, C., Hargis, B. M., & Vazquez-Anon, M. (2017). Incidence of broiler breast myopathies at 2 different ages and its impact on selected raw meat quality parameters. *Poultry Science*, *96*(8), 3005-3009. Doi: http://dx.doi.org/10.3382/ps/pex072

Lee, Y. S., Owens, C. M., & Meullenet, J. F. (2008). The Meullenet-Owens razor shear (MORS) for predicting poultry meat tenderness: its applications and optimization. *Journal of texture studies*, *39*(6), 655-672. Doi: http://dx.doi.org/10.1111/j.1745-4603.2008.00165.x

MEULLENET, J. F., Jonville, E., Grezes, D., & Owens, C. M. (2004). Prediction of the texture of cooked poultry pectoralis major muscles by near-infrared reflectance analysis of raw meat. *Journal of Texture Studies*, *35*(6), 573-585. Doi: http://dx.doi.org/10.1111/j.1538 7836.2004.01165.x-i1

Møller, A. J. (1981). Analysis of Warner-Bratzler shear pattern with regard to myofibrillar and connective tissue components of tenderness. *Meat Science*, *5*(4), 247-260.doi: http://dx.doi.org/10.1016/0309-1740(81)90015-2

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological methods*, *7*(1), 105. Doi: http://dx.doi.org/10.1037/1082-989X.7.1.105

Mudalal, S., Lorenzi, M., Soglia, F., Cavani, C., & Petracci, M. (2015). Implications of white striping and wooden breast abnormalities on quality traits of raw and marinated chicken meat. *Animal*, *9*(4), 728-734. Doi: http://dx.doi.org/10.1017/S175173111400295X

National Chicken Council. 2022. Statistics: Broiler chicken industry key facts 2022. http://www.nationalchickencouncil.org/about-the-industry/statistics/broiler-chicken industry key- facts/ Accessed February 01, 2022.

Ning, X., Mu-Hua, L., Hai-Chao, Y., Shuang-Gen, H., Xiao, W., Jin-Hui, Z., & Yi-Xin, S. (2020). Classification of sulfadimidine and sulfapyridine in duck meat by surface enhanced Raman spectroscopy combined with principal component analysis and support vector machine. *Analytical Letters*, *53*(10), 1513-1524. Doi: http://dx.doi.org/10.1080/00032719.2019.1710524

Petracci, M. and C. Cavani. 2012. Muscle growth and poultry meat quality issues. Nutrients. 4:1 12. World's Poultry Science Journal. 71:363-374. Doi: http://dx.doi.org/10.3390/nu4010001

Petracci, M., Bianchi, M., Mudalal, S. and Cavani, C. 2013. Functional ingredients for poultry meat products. Trends in Food Science and Technology. 33:27-39. Doi: http://dx.doi.org/10.1016/j.tifs.2013.06.004

Rothstein, H. R. (2008). Publication bias as a threat to the validity of meta-analytic results. *Journal of Experimental Criminology*, *4*(1), 61-81. Doi: http://dx.doi.org/10.1007/s11292-007-9046-9

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1988). *Parallel distributed processing* (Vol. 1, pp. 354-362). New York: IEEE. Doi: http://dx.doi.org/10.7551/mitpress/5236.001.0001

Schmid, C. H., Lau, J., McIntosh, M. W., & Cappelleri, J. C. (1998). An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Statistics in medicine*, *17*(17), 1923-1942. Doi: http://dx.doi.org/10.1002/(SICI)1097-0258(19980915)17:17%3C1923::AID SIM874%3E3.0.CO;2-6

Siddique, A., Shirzaei, S., Smith, A. E., Valenta, J., Garner, L. J., & Morey, A. (2021). Acceptability of Artificial Intelligence in Poultry Processing and Classification Efficiencies of Different Classification Models in the Categorisation of Breast Fillet Myopathies. *Frontiers in Physiology*, 1472.doi: http://dx.doi.org/10.3389/fphys.2021.712649

Sihvo H, Immonen K, Puolanne E. 2014. Myodegeneration with fibrosis and regeneration in the pectoralis major muscle of broilers. En: Veterinary pathology 51 (3), pág. 619–623. doi: http://dx.doi.org/10.1177/0300985813497488

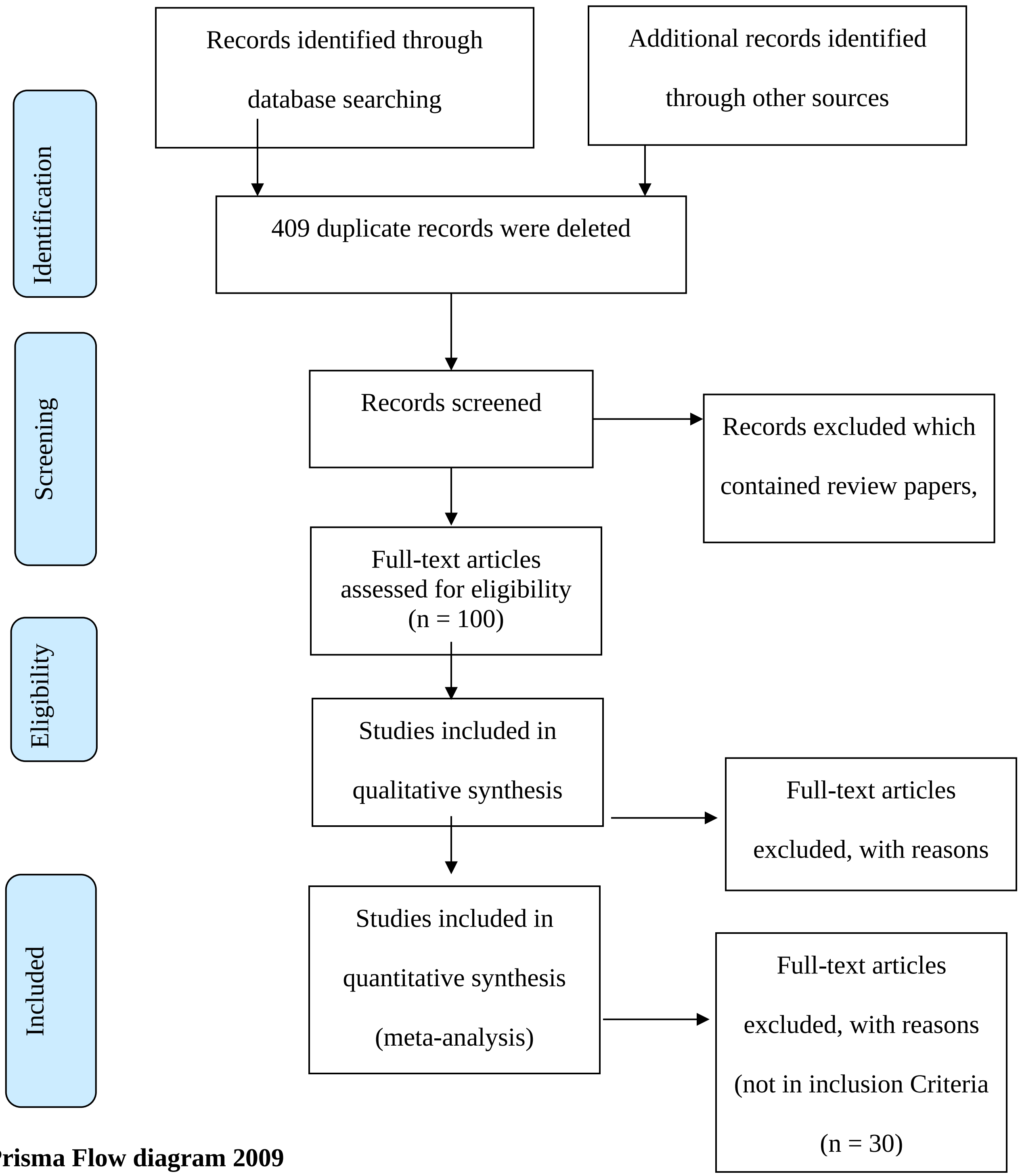
Soglia, F., Gao, J., Mazzoni, M., Puolanne, E., Cavani, C., Petracci, M., & Ertbjerg, P. (2017). Superficial and deep changes of histology, texture and particle size distribution in broiler wooden breast muscle during refrigerated storage. *Poultry Science*, *96*(9), 3465-3472.doi: http://dx.doi.org/10.3382/ps/pex115

Soglia, F., Mudalal, S., Babini, E., Di Nunzio, M., Mazzoni, M., Sirri, F., & Petracci, M. (2016). Histology, composition, and quality traits of chicken Pectoralis major muscle affected by wooden breast abnormality. *Poultry science*, *95*(3), 651-659.doi: http://dx.doi.org/10.3382/ps/pev353

Sun, X., Koltes, D. A., Coon, C. N., Chen, K., & Owens, C. M. (2018). Instrumental compression force and meat attribute changes in woody broiler breast fillets during short term storage. *Poultry science*, *97*(7), 2600-2606.doi: http://dx.doi.org/10.3382/ps/pey107

Velleman, S. G., Clark, D. L., & Tonniges, J. R. (2017). Fibrillar collagen organization associated with broiler wooden breast fibrotic myopathy. *Avian diseases*, *61*(4), 481 490.doi: http://dx.doi.org/10.1637/11738-080217-Reg.1

PRISMA 2009 Flow Diagram 1



Prisma Flow diagram 2009

Figure 4-1: Hand palpation of fillets for manual classification based on severity level.

Table 4.1:Meta-regression of moderators including age of bird, deboning time, and condition (Cooking and Storage) on the shear force of MORS, BMORS and other textural parameters analyzed by descriptive sensory evaluation.

		Q_E	df	*P*	Q_M	df	*P*	τ^2	I^2
BMORS	**Age of Bird**	335.02	8	<0.01	0.54	5	0.99	14.55	99.65
	Deboning time	274.59	9	<0.01	1.02	4	0.90	12.17	99.69
	Cooking and storage	330.30	7	<0.01	60.60	7	<0.001	1.16	98.01
MORS	**Age of Bird**	181.02	15	<0.01	3.37	5	0.64	0.52	91.34
	Deboning time	64.85	15	<0.01	32.13	5	<0.0001	0.17	77.97
	Cooking and storage	273.78	14	<0.01	4.84	6	0.56	0.51	93.44
Descriptive Sensory	**Age of Bird**	176.44	31	<0.01	14.33	4	0.0063	0.33	83.75
	Deboning time	202.39	34	<0.01	3.10	1	0.07	0.44	86.93
	Cooking and storage	129.79	27	<0.01	51.83	8	<0.0001	0.33	83.81

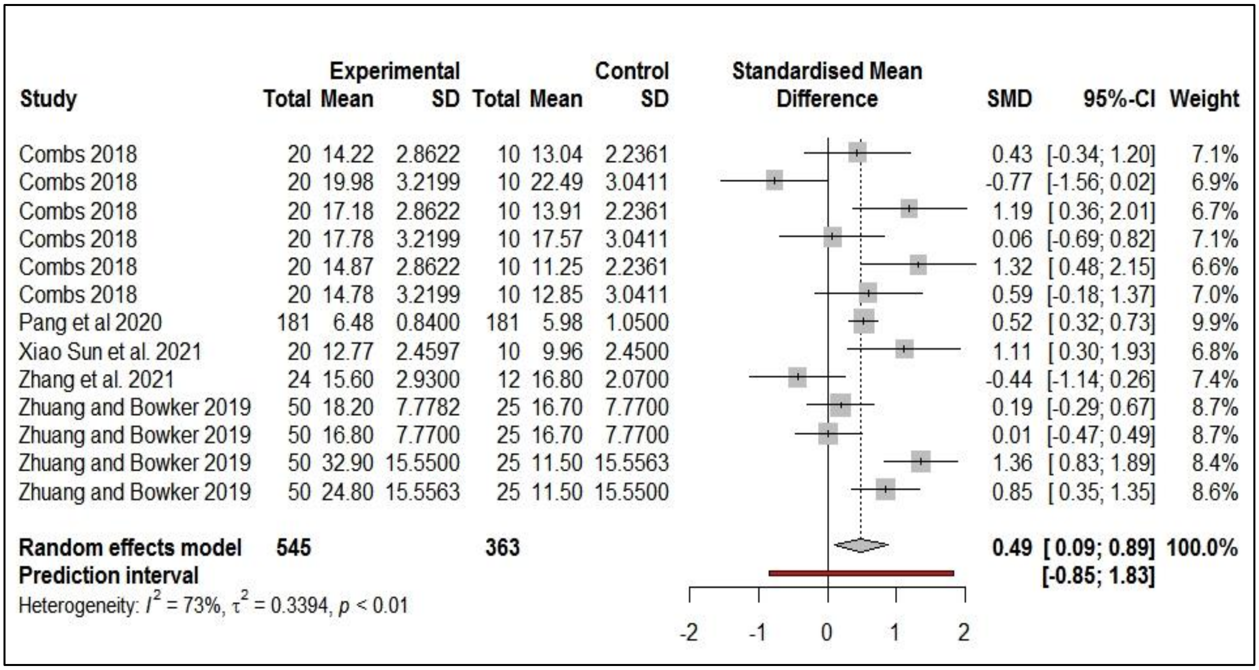


Figure 4-2: Forest plot analyzed for overall BMORS value for deboning time

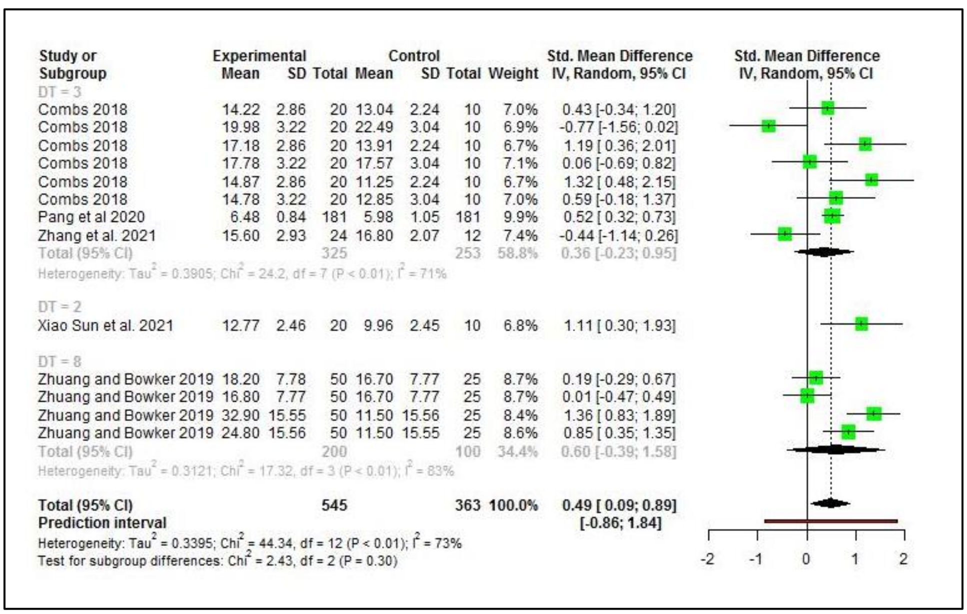


Figure 4-3: Forest plot for BMORS value for different deboning time periods

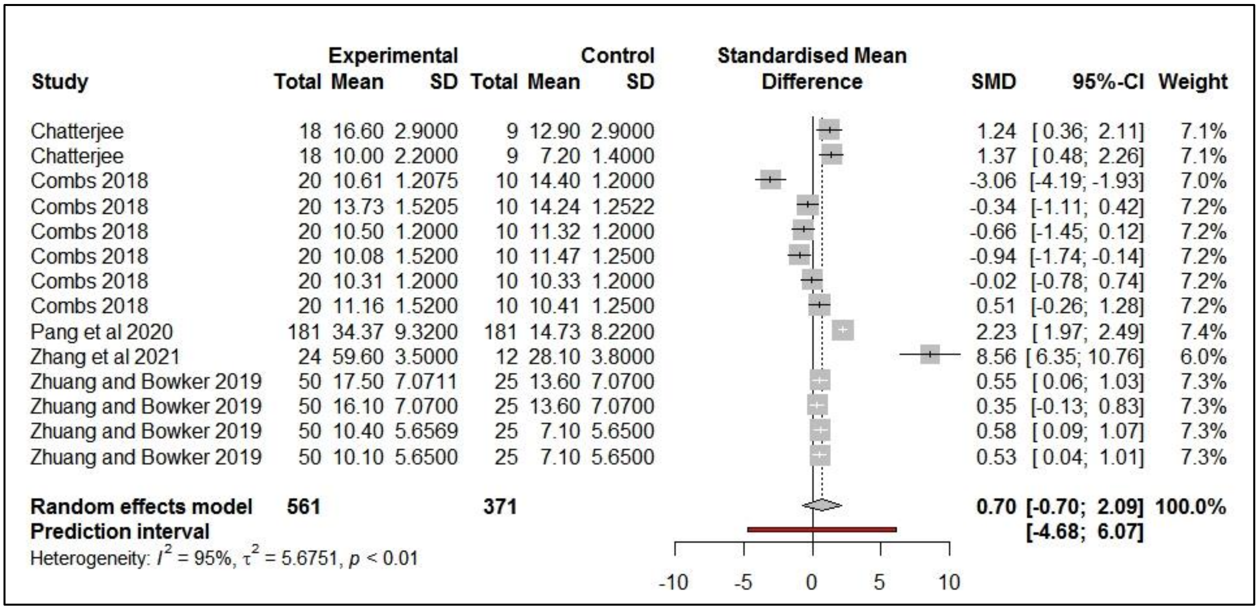


Figure 4-4: Forest plot for overall MORS value for deboning time

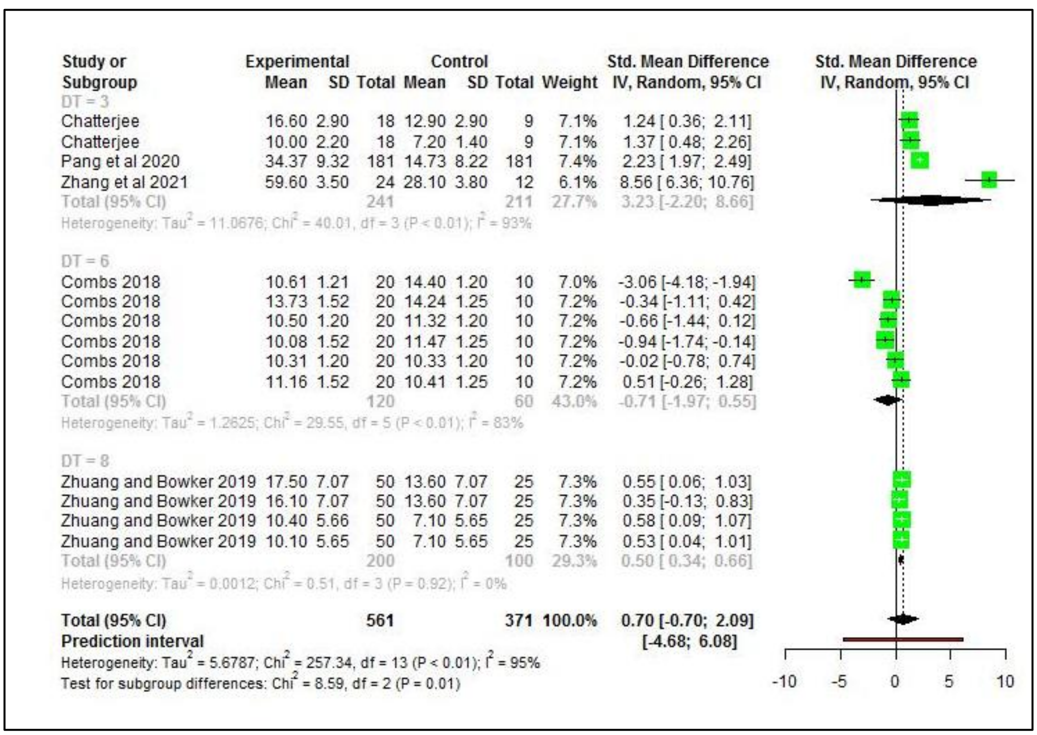


Figure 4-5: Forest plot for BMORS value for different deboning time periods

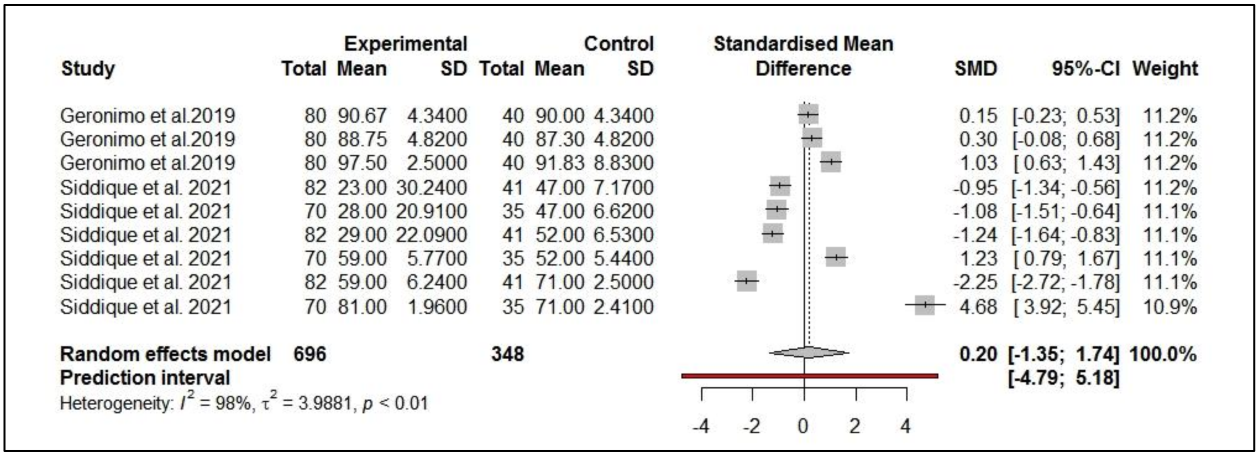


Figure 4-6: Forest plot for overall classification accuracy results for deboning time

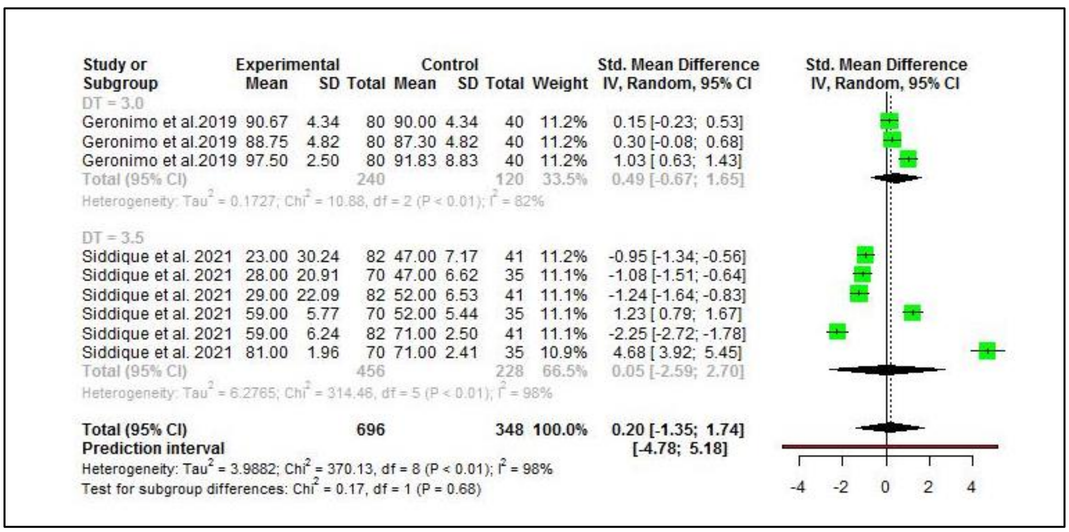


Figure 4-7: Forest plot for classification accuracy values for different deboning time periods

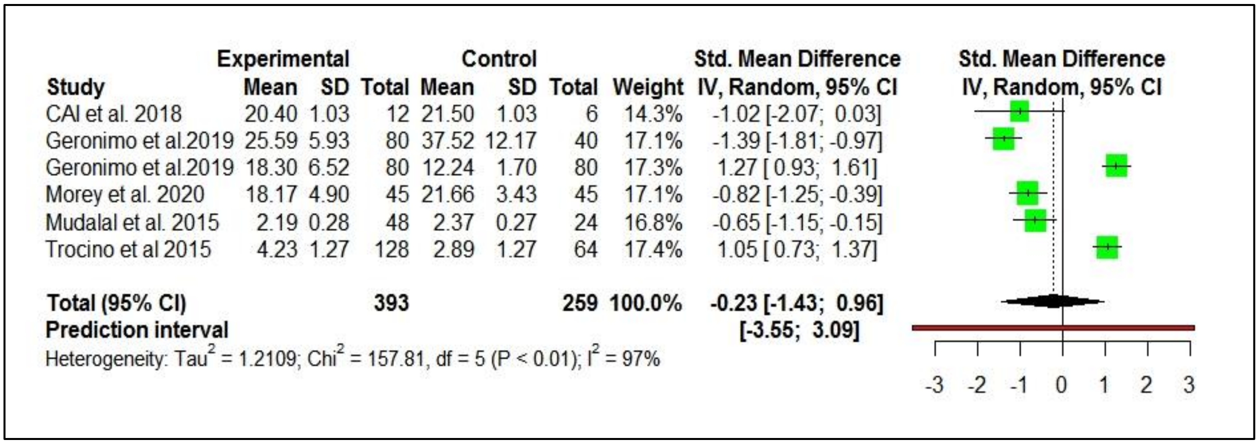


Figure 4-8: Forest plot for overall shear force value results for deboning time

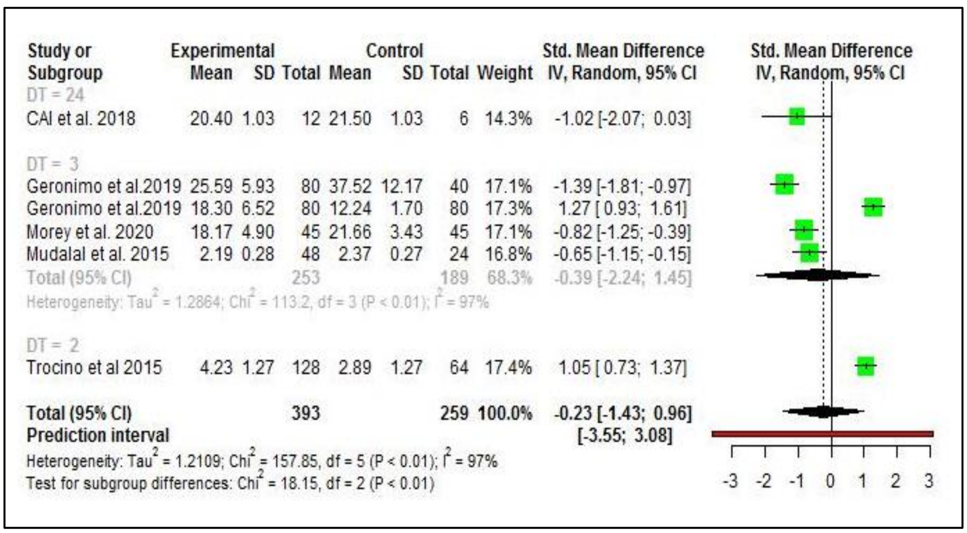


Figure 4-9: Forest plot for shear force values for different deboning time periods

Study	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Aguirre 2018	5.85	0.16	16	2.50	0.16	8	2.3%	20.21 [13.93; 26.50]
Brambila 2017	5.70	11.08	24	5.20	11.08	12	2.8%	0.04 [-0.65; 0.74]
Brambila 2017	4.00	1.30	24	4.50	1.10	12	2.8%	-0.39 [-1.09; 0.30]
Brambila 2018	5.70	2.00	24	5.40	2.10	12	2.8%	0.14 [-0.55; 0.84]
Chatterjee	5.43	3.36	18	3.44	1.20	9	2.8%	0.68 [-0.15; 1.50]
Chatterjee	3.85	3.06	18	2.39	1.11	9	2.8%	0.54 [-0.27; 1.36]
Chen et al	17.86	0.72	5	36.70	0.27	5	1.1%	-31.28 [-48.62; -13.94]
Solo 2016	5.62	2.26	70	5.27	2.26	35	2.8%	0.15 [-0.25; 0.56]
Solo 2016	5.13	2.26	70	5.70	2.26	35	2.8%	-0.25 [-0.66; 0.16]
Aguirre 2018	5.31	0.04	16	4.96	0.04	8	2.7%	8.45 [5.71; 11.18]
Brambila 2017	7.10	1.10	24	7.10	1.20	12	2.8%	0.00 [-0.69; 0.69]
Brambila 2017	4.50	1.40	24	4.80	1.40	12	2.8%	-0.21 [-0.90; 0.49]
Brambila 2018	7.10	1.90	16	6.20	2.20	8	2.8%	0.43 [-0.42; 1.29]
Chatterjee	0.58	0.05	18	0.55	0.10	9	2.8%	0.42 [-0.39; 1.22]
Chatterjee	0.47	0.12	18	0.55	0.05	9	2.8%	-0.75 [-1.58; 0.08]
Chen et al	0.49	0.02	5	0.67	0.01	5	2.3%	-10.28 [-16.09; -4.46]
Solo 2016	6.30	2.76	70	5.90	2.76	35	2.8%	0.14 [-0.26; 0.55]
Solo 2016	6.13	2.76	70	5.76	2.76	35	2.8%	0.13 [-0.27; 0.54]
Aguirre 2018	13.95	2.08	16	8.13	3.40	8	2.8%	2.18 [1.10; 3.27]
Brambila 2017	6.90	6.37	24	6.20	8.31	12	2.8%	0.10 [-0.60; 0.79]
Brambila 2017	5.00	0.80	24	5.10	0.80	12	2.8%	-0.12 [-0.82; 0.57]
Brambila 2018	6.50	1.10	24	6.00	1.30	12	2.8%	0.42 [-0.28; 1.12]
Chatterjee	12.76	3.78	18	9.97	2.27	9	2.8%	0.80 [-0.03; 1.64]
Chatterjee	9.32	4.56	18	4.99	1.96	9	2.8%	1.07 [0.21; 1.93]
Chen et al	43.63	0.46	5	60.50	0.20	5	0.7%	-42.94 [-66.71; -19.16]
Solo 2016	6.64	1.92	70	6.30	1.92	35	2.8%	0.18 [-0.23; 0.58]
Solo 2016	6.64	1.92	70	6.30	1.92	35	2.8%	0.18 [-0.23; 0.58]
Solo 2016	5.76	1.92	70	6.20	1.92	35	2.8%	-0.23 [-0.63; 0.18]
Solo 2016	5.76	1.92	70	6.20	1.92	35	2.8%	-0.23 [-0.63; 0.18]
Aguirre 2018	4.32	0.48	16	3.87	0.52	8	2.8%	0.88 [-0.01; 1.77]
Brambila 2017	7.30	1.90	24	8.00	2.20	12	2.8%	-0.34 [-1.04; 0.36]
Brambila 2017	10.00	1.90	24	8.60	1.60	12	2.8%	0.76 [0.04; 1.47]
Brambila 2018	9.30	2.00	24	8.20	1.60	12	2.8%	0.57 [-0.13; 1.28]
Chatterjee	0.69	0.09	18	0.63	0.07	9	2.8%	0.69 [-0.13; 1.52]
Chatterjee	0.86	0.11	18	0.85	0.10	9	2.8%	0.09 [-0.71; 0.89]
Chen et al	0.84	0.01	5	0.91	0.01	5	2.6%	-6.32 [-10.03; -2.61]
Solo 2016	6.30	1.92	70	5.90	1.92	35	2.8%	0.21 [-0.20; 0.61]
Solo 2016	4.01	2.76	70	3.86	2.76	35	2.8%	0.05 [-0.35; 0.46]
Total (95% CI)			**1208**			**614**	**100.0%**	**-0.11 [-2.17; 1.94]**
Prediction interval								**[-14.42; 14.20]**

Heterogeneity: $Tau^2 = 48.4375$; $Chi^2 = 177.66$, df = 37 ($P < 0.01$); $I^2 = 79\%$

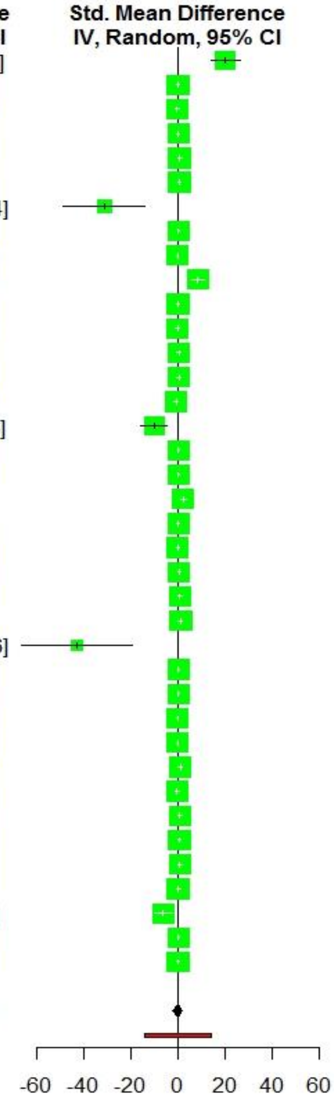


Figure 4-10: Forest plot for overall Textural profile analysis value for deboning time

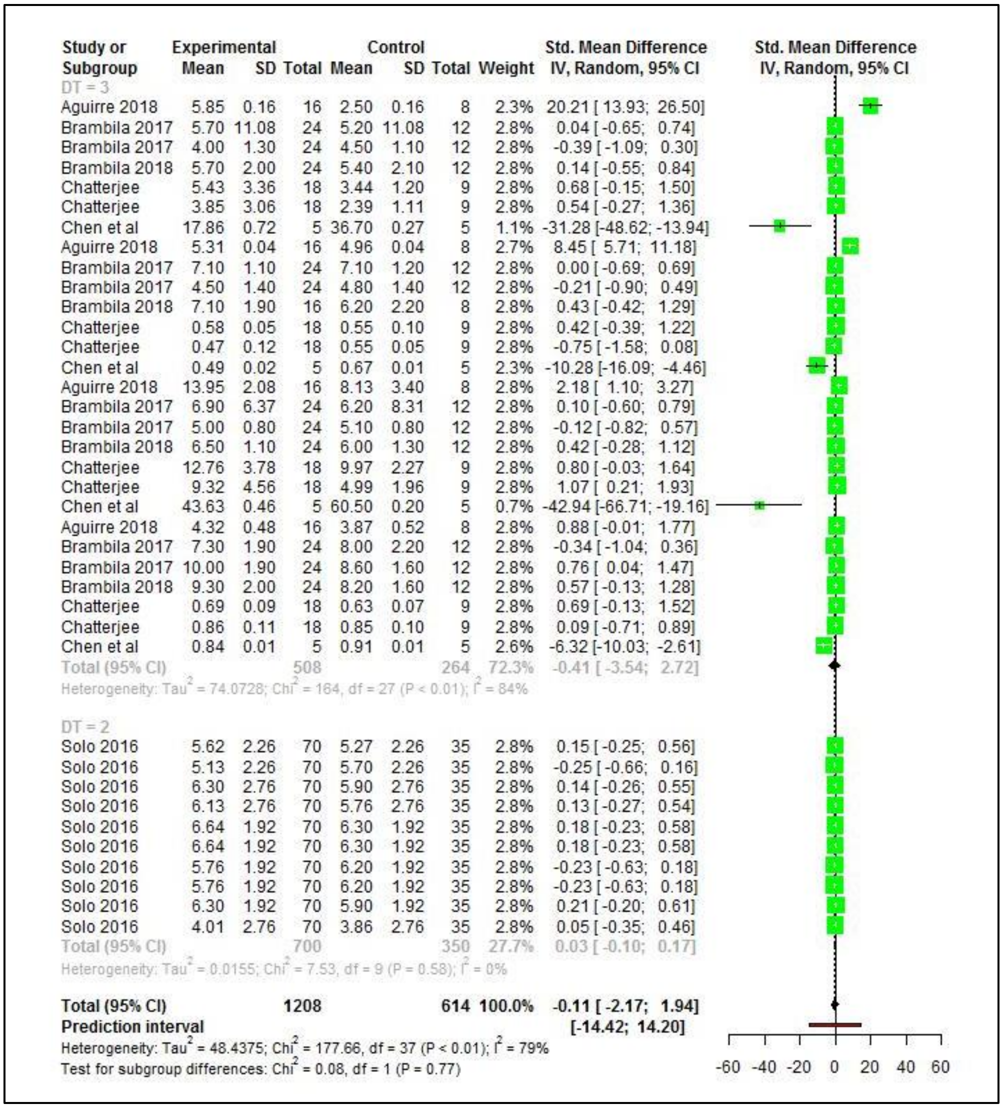


Figure 4-11: Forest plot for textural profile analysis values for different deboning time periods

Chapter 5

RAPID DETECTION AND CLASSIFICATION OF WOODY BREAST MYOPATHIES AT DIFFERENT PROCESSING STEPS USING S-BAND TO KU-BAND RADIO-FREQUENCIES WAVES

Aftab Siddique[1], Ashish Gupta[2], Jason Sawyer[3], Amit Morey[1]*

Affiliation:

[1]*Auburn University, Department of Poultry Science, Auburn, AL, United States.*

[2]*Auburn University, Department of Business Analytics, Auburn, AL, United States.*

[3]*Auburn University, Department of Animal Sciences, Auburn, AL, United States.*

***Corresponding Author:**

Amit Morey, Ph.D.

azm0011@auburn.edu

Article Type: Original Article

Number of Words: 5,301

Number of Figures: 3

Number of Tables: 4

5.1 ABSTRACT

Rapid change in consumer preferences and demand for high quality animal-based protein has shifted the poultry industry for the identification of rapid non-invasive technologies that can be implemented during in-line processing for rapid detection of muscle meat quality defects. At production plant, technologies such as Radiofrequency waves (RF-waves) can be used to identify and separate myopathy-conditioned meat. This can be done in order to prevent the mis-classification errors caused by factors associated with human counterpart such as fatigue, and less experience. Previous studies have demonstrated that advance developed diagnostic tools in conjunction with complex data analytics tools such as support vector machines (SVM) and backpropagation neural network (BPNN) can be used to classify chicken breast myopathies with usability only after the deboning process. The present study demonstrates the use of RF-waves in detection of myopathies at four different processing steps. A total of 107 (48 days old) broilers were used in the experiment. RF-wave data in form of amplitude and phase was collected on live birds, pre-chill WOG's, post-chill WOG's, and on freshly deboned fillets (3-3.5 h after slaughter) were examined by hand-palpation for WB category (1-normal; 2-moderate; 3-Severe). The collected data was preprocessed using false discovery rate, predictor analysis. For identification of specific signature frequencies and development of classification model using supervised machine learning algorithms variable clustering analysis was used. Variable clustering approach resulted in the identification of 7 to 8 different frequencies at different processing steps. Pre-processed data with identified signature frequencies were used to develop classification based models using Back Propagation Neural Network (BPNN) and Support vector machines (SVM). BPNN showed better classification accuracy as compared to SVM when it came to separating WB conditions at different processing steps, with an accuracy ranging from 90.0% to 96.1% for

live birds, 78.9% to 97.1% for pre-chill WOG's, 82.1% to 95.9% for post-chilled WOG's, and 94.2% to 98.2% for deboned fillets. In poultry processing industry, use of specific Radio frequency range devices or sensors in combination with supervised machine learning algorithms like SVM and BPNN can be successfully integrated in order to detect muscle myopathies at different processing steps during in-line processing.

Keywords: Support vector machines, Back propagation neural network, radio waves, Variable clustering

5.2 INTRODUCTION

The chicken industry in the United States is extremely popular with customers and is the most extensively consumed product with consumption of 44.08 Kg in 2020; it is projected that by the year 2022, on an average people will consume around 44.80 Kg of chicken (NCC, 2022). According to a survey released by National chicken council on November 2020, during the time of COVID-19 pandemic retail sales of chicken was increased $1.3 billion, and was up by 19.5% from last year same time period scale (NCC,2022). Over the past 70 years, increased consumption of chicken was observed as compared to beef and pork (NCC, 2022; **Figure 5-1**). Due to rich source of protein, less fat, easy in preparation, low cost and associated health benefits of chicken meat had played a significant role in purchasing behavior of consumers for boneless chicken fillets (Resurreccion, 2004). In a survey, consumers have preferred white meat over dark meat and consumer in US eats chicken 10 times a month when prepared at home. To complete this increased demand of chicken meat, researchers have moved towards the genetic selection fast growing of chicken breeds (Petracci et al. 2015). As results to this method, genetic selection had also lead us to the chicken muscle myopathic conditions such as Woody breast (WB), White striping (WS), Spaghetti meat (SM). Based on literature and published report, Huang and Ahn

(2018) reported that the prevalence rate of muscle myopathies in broilers can ranged from 5 to 50 % and could reach up to 100 % in controlled dietary studies (Cruz et al. 2017).

Woody breast condition is a quality defect of chicken meat which affects basic nutrient content, cook loss, pH, texture, water holding capacity (WHC), dull appearance and histologically characterized by deposition of extracellular collagen in breast part of birds (Sihvo et al. 2014; Soglia et al. 2016; Kuttappan et al. 2017). These conditions not only inflicts the quality of chicken meat but also responsible for the economic loss worth of $1 billion annually (Kuttappan et al. 2016). Because of the inferior quality of WB meat, it is separated out in processing factories by manual hand-palpation (**Figure 5-2**) and multiple grading scales dependent on the degree of the condition (**Table 5.1**). However, because this method is inconsistent and biased, there is a risk of misinterpretation of the chicken breast in some conditions (Morey et al. 2020). During the past several decades, researchers have investigated several approaches and attempted to put these strategies into practice for the identification of these aberrations through the production of chicken fillets. Several recent studies have investigated near-infrared spectroscopy (NIR) (Wold et al. 2019), computer vision system (CVS) (Geronimo et al. 2019), and hyperspectral imaging (HS-Imaging) (Jiang et al. 2017). According to the findings of molecular investigations, there is a positive association between the conditions of WB and the bird's age, with a larger occurrence of WB detected in aged birds (Petracci et al. 2015; Papah et al. 2017; Radaelli et al. 2017; Sihvo et al. 2017; Kuttappan et al. 2017).

5.3 RADIO-FREQUENCY WAVES SPECTRUM

History of radio-wave dated back to 1867, when Scottish physicist J. C. Maxwell first proposed the theory of electromagnetism later this theory was recognized as Maxwell’s equation (Nappo, 2021). Later in 1887, German physicist, Heinrich Hertz, created the EM waves in laboratory proving the nature of radio waves similar to the properties of light including refraction, reflection, polarization and diffraction (Mulligan, 1987). Radio waves represent a specific component of the electromagnetic spectrum (EM) that typically have frequencies ranging from low frequencies (30 Kilo Hertz) to extremely high frequencies (300 gigahertz)(Romanenko et al. 2017) (**Figure 5-3**), and contains the widest wavelengths of any form of EM waves in the spectrum series (Johnson and Guy, 1972). Based on mathematical expression frequency is inversely proportional to the wavelength. As the frequency increases the wavelength decreases (equation 1).

$$\vartheta = c/\lambda \tag{1}$$

Where $\vartheta =$ frequency, c = speed of light and $\lambda =$ Wavelength

For example, 300 GHz can only travels a distance of 1 mm, while a distance 33 km is achieved from frequency at 30 Hz. These waves are capable to travel in vacuum space with the speed of light. These radio waves are generated by electronic devices charged particles that are subjected to acceleration, such as those found in time-varying electrical impulses through transmitters and generated waves are detected by receivers. These transmitters and receivers antennas are designated to operate on a set standard limited frequency ranges (Ellingson, 2016). Different frequency waves have different properties such as long distance waves are affected by the environmental distraction while on the other hand shot distance waves are unaffected by these distraction and travels in a straight line of path.

Ultra-high frequency is a designation given by International Telecommunication Union (ITU) to the frequency ranging from 300 MHz to 3GHz. In the context of electromagnetic radiation, UHF band has a wavelength that ranges from one meter to one hundred millimeters, and is also generally known as the decimeter band. The UHF frequency band are unaffected by environmental factors and is used frequently for multichannel transmission in TV and radio broadcasting.

The present study investigated the use of radio-frequency wave as an alternative non-destructive technique (RF-NDT) to detect woody breast myopathies identification during different processing steps from live birds to deboned fillets considering different variables using frequency ranged from 2GHz to 18 GHz. These frequencies ranged between Ultra-high frequency and Super-high frequency with the wavelength ranges from 100 mm to 10 mm. Radio-frequency waves are electromagnetic waves ranging from 300 kHz to 300 MHz. Frequencies of radio waves can range from a few kilohertz to several billion cycles per second, depending on their wavelength. Radio waves have an inverse relationship between their wavelength and frequency. Different sources use different methods to specify the frequency range (Fleming, 1919; Ghirardi, 1932). RF waves have the ability to heat materials, and when compared to microwave heating, they have higher heating uniformity and penetration depth (Jiao et al. 2017). The warming time is decreased and the thermal spread is more even as a result of electromagnetic heating (Duan et al. 2005; Song et al. 2009). There has been number of studies conducted on different materials for the use of microwave sensors to determine the moisture content, and density of cereal grains (Trabelsi et al. 1997), for monitoring of moisture content of wheat (Nelson and Stetson, 1976; Kraszewski et al. 1977; Trabelsi et al. 1998), for determination of moisture and bulk density content in sand and rice (Li and Zhang, 2015), moisture determination in coal (Klein, 1980).

5.4 MATERIAL AND METHODS

5.4.1 Data Collection

Ross 708 broilers (n =108), 40 days old birds were analyzed in poultry sciences slaughtering and deboning facility in Auburn, Alabama. A portable radio-wave transmitter device (Compass tech. LCC., Georgia) was employed to capture data using radio waves with a frequency range of 2 GHz to 18 GHz on live birds, pre-chilled WOG's, post-chilled WOG's (4^{o}C), and deboned fillets, frequency data in the form of amplitude and phase were gathered in two different positions (waves perpendicular to muscle fibers and waves parallel to muscle fibers). The deboned breast fillets were hand palpated and were classified into WB severity level (Tijare et al. 2016) and used to train the ML algorithm for classification of live birds, pre-chilled WOG's, post-chilled WOG's, and fillets. Weight of individual live birds, pre-chill WOG’s, post-chill WOG’s and fillets were recorded using a weighing balance (Ohaus Corporation, Pine Brook, NJ, United States) for the analysis. Additionally, a false discovery rate (FDR) analysis was done on the obtained data to eliminate false-positive results and avoid data complexity of the model.

5.5 DATA ANALYSIS

Raw data collected from RF-device in form of amplitude and phase were subjected to R software, and jmp16 pro (Version 16.0) for pre-processing and screening procedure. The data sets collected for RF ranges with different myopathic conditions for various processing steps were subjected into false discovery rate (FDR) analysis for the separation of false positive results to avoid error and predictor screening (PS) (Bootstrap Forest) method was used to separate out the actual frequency range responsible for detection of myopathies at different processing steps.

Both FDR analysis, and feature extraction was performed using response screening, and predictor screening option provided in JMP16 pro software. Top 100 isolated frequencies obtained after the FDR and predictor screening analysis were subjected to variable clustering approach to identify different signature frequencies for each myopathic condition at various processing steps. In our datasets, the various class labels corresponding to the four processing stages (Normal, Moderate, and Severe) were not uniformly distributed. It is widely recognized that prediction models may suffer from suboptimal performance when confronted with significantly imbalanced datasets. To address this issue of dataset imbalance, we employed the Synthetic Minority Over-sampling Technique (SMOTE) package in combination with a 10-fold nested cross-validation approach for multi-class classification, which facilitated the replication of instances belonging to the underrepresented class. For the data preprocessing in our multi-class classification model, we utilized the R software and leveraged the capabilities of the caret package. This ensured a more balanced representation of the class labels across the four processing steps, ultimately leading to improved prediction performance.

5.5.1 False Discovery Rate Analysis

In the field of statistics, the false discovery rate, also known as the FDR, is a means for conceptualizing the rate of type I errors in the process of evaluating the null assumption when multiple comparisons are performed. The Erroneous Discovery Rate (EDR), or FDR, is the expected fraction of "discoveries" (rejected null hypotheses) that are false. FDR-controlling strategies are designed to control the FDR (incorrect rejections of the null) (Bogomolov et al. 2017). The FDR analysis is another way of referring to the expected ratio of the number of erroneous positive classifications, also known as false discoveries, to the total number of positive classifications (rejections of the null) (Benjamini and Hochberg, 1995; Alzu'bi et al. 2021;Tay et al. 2022).

Both total false positives (TFP) and the total true positives (TTP) are considered toward the number of null hypothesis rejections (TP) (TP = TFP + TTP) (Tay et al. 2022). To put it in mathematical way, FDR is TFP divided by TTP plus TFP. When compared to family wise error rate (FWER) controlling procedures (such as the Bonferroni correction), which control the probability of at least one Type I error, FDR-controlling procedures offer a less stringent control of Type I errors (Alzu'bi et al. 2021). This is because FDR-controlling procedures use a false discovery rate (FDR). Therefore, FDR-controlling techniques have more power, but at the expense of producing more Type I error. The FDR refers to the proportion of times a feature (here, frequency bands) is found to be statistically and practically insignificant.

The FDR includes features that can be helpful. So, if the FDR is 5%, then 5% of the features are insignificant. If we assume that there are a thousand distinct frequency bands in our sample, then the p-value for the middle band, at 15.89 GHz, is 0.0005 and thus the q-value is 0.03. It's likely the test statistic is really out of the ordinary here, and that there are, in fact, differentially expressed frequencies with test values less extreme than 15.89 GHz. When using q-value of 0.03 tells us that 3% of the radiofrequencies as extreme as frequency of 15.89 GHz are false positives (i.e., the bands with lower p-values). By calculating q-values, we can set a threshold for the percentage of insignificant features we're ready to accommodate. This comes in handy when we're willing to accept a broad range of frequencies for later verification. In the event that all null hypotheses are correct (there are no meaningfully different outcomes), then the FDR will equal the FWER.

When there are enough plausible alternate explanations, controlling for the FWER also controls the FDR by definition. Implementation of FDR analysis in the larger data set reduces the chance of making inaccurate pronouncements of significance (Storey and Tibshirani, 2003).

5.5.2 Predictor Screening

In the analysis of our large data set and identification of specific frequency ranges for the categorization of myopathic conditions at different processing steps feature extraction from the larger data set is one of the important steps for the training of developed ML models (Rahman et al. 2018). For the feature extraction purposes, we implemented predictor screening method offered by JMP 16 pro software. This method screens many potential predictors for the significant effect in a given response. In another words, The Predictor Screening tool gives users a way to evaluate many potential predictors based on their capacity to accurately forecast a result.

The predictor screening is not the same as the response screening method as describe above. Response screening for an outcome uses several parameters, one at a time, to make a prediction about a certain response. On the other hand, predictor screening in based on the method of bootstrap forest partitioning, in this process of predictor screening to assess the impact of predictors on the response. The models for the partitioning are constructed using various predictors. Screening for predictors can help to uncover variables that stand on their own as being unreliable but are powerful in expression of response when used in conjunction with other predictors (variables) (Klimberg and McCullough, 2016).

5.5.3 Variable Clustering

Variables reduction is an essential stage in the process of model generation without compromising the potential predictive value of the data. One such method that contributes to the reduction of variables is known as variable clustering (Bogomolov et al. 2018). The more dimensions a dataset contains, the more complicated it is to analyze. It lengthens the time required to perform the data processing, hinders the capability to investigate the model relationship, reduces the accuracy of model scoring, and adds more redundant information to the dataset (Adebiyi et al. 2022). Variable clustering is one of the corrective measures that may be taken. It locates a group of variables that, inside a cluster, have the highest possible correlation between themselves and the lowest possible correlation with variables from other clusters. When working on a modeling assignment that contains a large number of variables, an analyst may find themselves in a position to reduce unnecessary variables before they can develop a model. During this process there is a higher chance that the variables will act as the relationship between the goals of the modeling effort is not known (Lee, 1973). In addition, larger variables numbers makes it difficult to determine the relationship between each other (Lee, 1973; Bagozzi, 2007).

The use of an excessive number of variables makes the created model less efficient and also increases the amount of time needed for analysis (Graham, 2007). The parameters then would be used as inputs in various predictive analysis and classification approaches, such as support vector machines, and neural networks for developing models. It is not common to find a detailed explanation of the variable clustering algorithm method in the majority of textbooks that cover multivariate approaches. However, the variable cluster track the progress primarily as an application in statistical analysis software (Anderberg, 2014).

A set of nominal variables can be clustered using either a hierarchical structure or a disjunct clustering using the variables clustering technique (Michalski et al. 1981). There is a linear integration of the variables that make up each cluster that is connected to each cluster. This linear combination might be the initial principal component or it could be the centroid component. The rule stipulates that the variable with the lowest 1-R^2 score should be chosen to serve as the cluster representative (Ramchandran et al. 1994).

The formula for calculating the 1-R^2 is as follows:

$$1-R^2 = (1-R^2)_{(own)} / (1 - R^2)_{(nearest)} \quad (2)$$

Intrinsically, we want to get the group representative to have the strongest possible correlation to its own cluster and the weakest possible correlation to the cluster that is geographically closest to it. As a result, a variable is considered to be the best indicator of a cluster if it is one in which 1-R^2 trends toward zero. In the literature on clustering, there is generally a rule for choosing the sample of the cluster, which is referred to as the 1-R^2. In addition to this criterion, organizational "knowledge from subject matter experts" should also be used to guide the choice of factors. As a result of this, we might settle on the idea of employing multiple variables (in our case the various frequencies) in each cluster.

5.5.4 Back Propagation Neural Networking (BPNN)

The most important part of training a neural network is called backpropagation.
It is the process of fine-tuning the weights of an artificial neural networks based on the error rate achieved in the preceding epoch of the network's training (i.e., iteration). By fine-tuning the weights in the right way, it is possible to lower the overall error rate and improve the model's reliability by expanding its applicability. The term "backward propagation of errors" is what's meant to be referred to when discussing about backpropagation in neural networks. It is a tried-

and-true approach to the development of artificial neural networks. The gradient of a loss function in relation to all of the weights in the network can be more easily computed with the assistance of this method. Detailed explanation on the working principle of BPNN can be found in Siddique et al. (2021).

5.5.5 *Support Vector Machines (SVM) Algorithms*

Although the SVM approach was initially developed by Vapnik (1995), it has recently gained a lot of traction in the field of machine learning because of its many potential practical applications. Many research have indicated that the SVM methodology surpasses alternative data classification models depending on the information type concerning classification accuracy when particularly in comparison to other approaches (Maji et al. 2008). When classifying data sets, SVM creates a line (hyper-plane) between the classes. SVM seeks to locate a higher dimensional space that can distinguish different classes of available dataset with a largest margin while still delivering the best generalization capabilities for a two or more class linearly differentiated (Siddique et al. 2021). The SVM algorithm's underlying logic is described in detail by Siddique et al (2021, 2022).

5.6 RESULTS AND DISCUSSIONS

A total number of 535 different significant frequencies ($p \leq 0.05$) were identified for live chicken data using FDR analysis. These frequencies ranged from 2.00 GHz to 2.38 GHz; 6.01 GHz to 6.29 GHz; 8.06 GHz to 10.78 GHz; 12.44 GHz to 13.16 GHz; and 16.80 GHz to 18.00 GHz. There were not significant frequencies were identified for the FDR analysis performed on pre-chilled WOG's collected data. For post-chilled WOG's data, a total of 719 significant ($p \leq 0.05$) frequencies were found ranging from 9.27 GHz to 16.45 GHz. A total number of 308 different significantly different frequencies were obtained from FDR analysis for deboned fillets data. These frequencies ranged from 9.64 GHz to 10.60 GHz; 14.70 GHz to 15.74 GHz; 16.95 GHz to 18.00 GHz.

After the identification of specific frequency ranges for different processing steps, cleaned frequency data was subjected into predictor screening analysis for the identification of top 100 different frequencies for each processing steps. These top 100 separated frequencies were used in variable clustering method to identify the most representative frequencies for each myopathic conditioned fillets. Most representative frequencies for each condition for different processing steps are presented in **table 5-2**. For normal condition in live chicken, we have identified 7 specific radio-frequencies (2.14 GHz, 2.33 GHz, 6.06 GHz, 8.73 GHz, 10.21 GHz, 12.60 GHz, and 16.95 GHz); 7 frequencies for moderate conditioned live birds (2.16 GHz, 6.06 GHz, 8.41GHz, 9.27GHz, 10.36GHz, 12.61GHz, and 16.95 GHz) and 6 different frequencies for severe condition (2.16 GHz, 6.06 GHz, 8.77 GHz, 10.21 GHz, 12.61 GHz, and 16.95 GHz). A total of 2-3 overlapping radio waves were also identified in analyzed frequency ranges that have contributed into the identification of myopathic condition in live birds. Proportion of variation explained by radio-frequencies normal, moderate and severe live birds conditioned radiofrequencies were 97.4%, 97.1% and 97.8% respectively. For the pre-chilled WOGs, variable cluster analysis for normal WOG's resulted in identification of 6 frequencies (3.21 GHz, 4.60 GHz, 7.50 GHz, 10.06 GHz, 11.16 GHz, 16.09 GHz), for moderate condition WOG's, 7 frequencies were identified (3.15 GHz, 3.92 GHz, 4.85 GHz, 6.93 GHz, 7.69 GHz, 9.97 GHz, 16.07 GHz) and severe pre-chilled WOG's resulted in identification of 6 radio-frequency waves (3.33 GHz, 4.75 GHz, 5.90 GHz, 8.52 GHz, 10.00 GHz, and 15.99 GHz) respectively with over lapping of 2-3 frequency waves. Proportion of variation explained by the radiofrequencies for pre-chilled WOG's normal, moderate and severe conditioned WOG's were 94.7%, 95.9% and 94.9% respectively. Variable clustering results for post chilled WOG's showed 6 different frequencies waves for both normal (9.40 GHz, 9.90 GHz, 11.67 GHz, 12.48 GHz, 13.01 GHz, 14.79 GHz) and moderate condition (9.40 GHz, 9.90 GHz, 11.98 GHz, 12.48 GHz, 13.01 GHz, 14.79 GHz) followed by 5 different frequencies for severe conditioned (9.40 GHz, 9.90 GHz, 12.24 GHz, 12.86 GHz, 14.86 GHz) post-chilled WOG's. The results for post-chilled WOGs also showed overlapping of 2-4 frequencies in all post chilled WOG's with different myopathic conditions. Variable clustering for deboned fillets shows 5 different frequencies for normal (9.54 GHz, 9.93 GHz, 10.37 GHz, 14.86 GHz, 17.92 GHz) explaining 97.30% of proportion of variations, for moderate condition 6 frequencies (9.41 GHz, 9.71 GHz, 10.12 GHz, 14.96 GHz, 17.14 GHz, 17.91 GHz) were identified with 97.60% variance explained, and severe deboned chicken breast fillets data resulted in 5 different radio-frequencies (9.52 GHz, 10.03 GHz, 14.99 GHz, 17.63 GHz) 98.30% of variation explained with no overlapping frequencies.

The performance of the Back Propagation Neural Network (BPNN) model (a supervised machine learning model) in predicting the myopathic condition (Normal, Moderate, and Severe) of poultry during different stages of processing was evaluated. The model was developed using a 10-fold cross-validation technique with the nested k-fold cross-validation method to deal with an unbalanced post-processed dataset with a set seed value of 2000. Different processing steps include Live Birds, Pre-Chill Whole Bird Without Giblets (WOG), Post-Chill WOG, and Deboned Fillets. The model's performance is assessed using various statistical measures, such as Generalized R2, Entropy R2, RASE (Relative Approximation Squared Error), Misclassification Rate, and Receiver Operating Characteristics (ROC).

In the Live Birds stage, the BPNN model demonstrates strong performance, explaining 83% of the variance in the data with a Generalized R2 value of 0.83. The Entropy R2 of 0.64 suggests a moderate ability to predict the myopathic conditions. A RASE value of 0.33 denotes a relatively low error rate, and a Misclassification Rate of 0.12 implies that 12% of the model's predictions are incorrect. The testing accuracy for the different myopathic conditions ranges from 87.5% to 100%. Classification accuracy for testing data results showed 87.5 % accuracy for normal, 83.3% accuracy for moderate, and 100 % accuracy for severe myopathic conditioned birds (**Table 5.3**). The high ROC values in this stage reflect the model's capacity to identify the myopathic condition with high sensitivity.

During the Pre-Chill WOG stage, the model accounts for 78% of the data's variability, as shown by a Generalized R2 of 0.78. The model has a moderate predictive ability, indicated by an Entropy R2 of 0.56. A RASE value of 0.37 suggests a somewhat low error rate, while a Misclassification Rate of 0.06 reveals that 6% of the predictions are incorrect. For pre-chill WOGs, classification accuracy analysis showed 100 % accuracy for normal, 87.5% for moderate, and for severe 100% respectively (**Table 5.3**). The model's sensitivity for identifying the correct condition is 100% in this stage, as demonstrated by the ROC values.

The model explains 91% of the variance for the Post-Chill WOG stage, denoted by a Generalized R2 of 0.91. The Entropy R2 of 0.78 indicates strong predictive ability. A RASE value of 0.26 implies a low error rate, while a Misclassification Rate of 0.10 shows that 10% of predictions are incorrect. The testing accuracy for the three conditions ranges from 69.7% to 100%. The model's sensitivity for identifying the correct condition remains at 100% for this stage (**Table 5.3**).

In the Deboned Fillets stage, the model accounts for 85% of the data's variability, as evidenced by a Generalized R2 of 0.85. The Entropy R2 of 0.66 signifies moderate to strong predictive ability. A RASE value of 0.33 indicates a relatively low error rate. However, a Misclassification Rate of 0.26 reveals that 26% of predictions are incorrect. The testing accuracy for the conditions is between 66.7% and 100% (**Table 5.3**). Despite the higher misclassification rate, the model's sensitivity for identifying the correct condition remains 100%, as shown by the ROC values. In overall obtained accuracy results, separated signature frequencies for all different processing steps performed well, with accuracy ranging from 66.7% to 100%.

In SVM analysis, Live birds under normal conditions, the SVM classification model showed a 68.4% sensitivity and a 55.5% specificity rate, leading to a 61.9% testing accuracy **(Table 5.4)**. The sensitivity of the SVM model dropped to 45.5% in moderate conditions, while the specificity increased to 73.0%, resulting in a 59.2% testing accuracy **(Table 5.4)**. In severe conditions, the sensitivity dipped slightly to 42.8%, but the specificity rose considerably to 96.6%, yielding a 69.7% testing accuracy **(Table 5.4)**. When analyzing pre-chill whole bird carcasses (WOGs), the detection model demonstrated a 34.6% sensitivity rate and a 75.0% specificity rate in normal conditions, culminating in a 54.8% testing accuracy **(Table 5.4)**. In moderate conditions, the sensitivity rose to 50.0%, and the specificity marginally increased to 77.7%, achieving a 63.8% testing accuracy **(Table 5.4)**. In severe conditions, the sensitivity rate climbed to 70.0%, while the specificity fell to 67.5%, leading to a 68.7% testing accuracy **(Table 5.4)**.

For post-chill WOGs, the detection model revealed a 38.4% sensitivity rate and an 84.0% specificity rate in normal conditions, amounting to a 61.2% testing accuracy **(Table 5.4)**. In moderate conditions, the sensitivity grew to 53.3%, and the specificity remained consistent at

77.7%, attaining a 65.5% testing accuracy **(Table 5.4)**. The sensitivity rate rose to 60.0% in severe conditions, but the specificity declined to 63.4%, generating a 61.7% testing accuracy (Table 5.4).

Finally, in deboned fillets, the detection model exhibited a 46.1% sensitivity rate and a 92.3% specificity rate in normal conditions, resulting in a 69.2% testing accuracy **(Table 5.4)**. In moderate conditions, the sensitivity substantially increased to 75.0%, while the specificity marginally dropped to 72.2%, yielding a 73.6% testing accuracy **(Table 5.4)**. In severe conditions, the sensitivity rate further enhanced to 80.0%, and the specificity also grew to 80.9%, accomplishing an 80.4% testing accuracy **(Table 5.4)**.

In summary, the detection model's performance varied based on the chicken product type and condition, with the highest accuracy observed in severe conditions of deboned fillets. The model's effectiveness in recognizing normal and moderate conditions demonstrated room for improvement, indicating that further research and optimization may be necessary to enhance its accuracy in these areas.

SVM analysis has been shown higher accuracy level with high dimensional data used for data classification with relatively smaller sample set. This method has also been utilized by other authors to assist in the classification of multi-dimensional data. Barbon et al. (2018) have used SVM as a classification technique for breast fillets with muscular myopathies in which they have classified normal and pale meat in relation to pale, soft, and exudative poultry breast meat using NIR results. The classification accuracy for normal breast fillets was 53.4%, while the classification accuracy for pale breast fillets was 72.0%. Geronimo et al. (2019), have also implemented SVM technique in conjunction with image acquisition system with NIR output and observed 91.83% classification efficiencies (fillet images). The classification accuracy for WB was 90.67% when these researchers classified the data set using multilayer perceptron, which is a feed forward network rather than the back propagation network employed in BPNN. Expressible fluid images of breast flesh were studied by Yang et al. (2021), who classified WB using SVM (the training and testing ratio was not provided) and DL (training to testing is 2 to 1). These researchers have reported lower testing accuracies (38.25 - 63.89%) of developed model compared to training set (40.41 - 81.94%). In their DL classification, which is a form of ANN, they classified WB and claimed an accuracy of 93.30% in the testing set despite having achieved an accuracy of 100% in the training set.

The SVM method has garnered considerable attention due to its exceptional performance efficiency, capacity to achieve pinpoint accuracy, and management of high-dimensional, multi-variable data sets. The ML theory is responsible for laying the framework for the SVM algorithm. SVMs were implemented by Cortes and Vapnik (1995) as a new machine learning technique for the problem of group classification. Researchers have revealed that support vector machines (SVMs) are an inexpensive, sensitive, and simple classifier that may be utilized in organized evaluation tasks. SVM has a number of important applications, one of which is the inspection of enormous data sets generated during production (Burbidge et al. 2001; Chinnam, 2002). SVM is frequently used in a variety of food industrial operations, including as a product surveillance system, robotic fault, and dimensional tolerances (Ribeiro, 2005; Azadeh et al. 2013; Salahshoor et al. 2011; Cydaş and Ekici, 2012). In addition to their application in the food commodity industry worldwide, SVMs are also utilized in the pharmaceutical and medical research industries, as well as in surgical procedures and the treatment of cancer (Vapnik, 2013). In addition to product quality monitoring (Borin et al. 2006) and molecular level polymer identification (Li et al. 2009), other industries are also feasible domains in which SVM might be implemented. These examples, taken from a variety of industries, show that the SVM algorithms are applicable to a wide variety of settings and offer a high degree of adaptability (Kotsiantis et al. 2007). The current study reveals that SVM and BPNN, in conjunction with RF waves and different processing step weight data, can be used to classify broilers at different processing steps.

After carcasses have been processed in deboning facilities, WB can be identified based on visual and hand palpation characteristics. Misclassifications owing to human error, inefficient processing, and higher labor expenses all arise while classifying muscle myopathies on a production chain. Woody breast manifests itself largely in the breast fillet's superficial area and is identified by the proximity of surface intracranial hemorrhage, a yellowish color breast surface appearance, a rigidly bulging fillet, and physical palpability of the tissue (**Figure 5-2**) (Mazzoni et al. 2015; Mudalal et al. 2015). In addition, the connective tissue content (Collagen content) and pH of WB fillets are higher than those of normal chicken breasts, and the cross-sectional areas of the two are significantly different (Huang and Ahn, 2018). (Petracci et al. 2015; Chatterjee et al. 2016; Clark and Velleman, 2016; Soglia et al. 2016).To best of our knowledge, there is no supporting documents available for investigation on the categorization of WB myopathies at various stages of processing steps.

Thus, we were able to categorize live birds, pre-chill WOG's, post-chill WOG's, and deboned fillets among the WB myopathies by employing a variety of complicated information accessible from an RF wave transmitter in the form of amplitude and phase with distinguishing separable signature frequency signals.

Artificial neural networks, which are modeled based on the actual human neural complex, connect seemingly unrelated nodes or units in a computer system to address problems that cannot be addressed by more traditional statistical methods. The implementation of the subconscious system to a processing architecture allows for the execution of specialized operations (perception, voice synthesis, picture recognition) that have proven effective in the manufacturing sector (Kecman, 2001; Alpaydin, 2010). To interpret data through the intricate response of these terminals and their connections to exogenous stimuli, decentralization requires a large number of interlinked hubs or neurons. (Akay, 2011). These approaches, which fall within the domains of perception and computation, are essential in the advancement of contemporary machine learning techniques (Nilsson, 2005). Process control is just one example of an area where neural networks have been put to use in industry (Pham and Afify, 2005; Wang et al. 2005). However, a larger sample size is necessary for acceptable accuracy when using ANN (Kotsiantis et al. 2007).

The problem of over-fitting and under-fitting, which is associated with high-variance algorithms, has been widely recognized as a drawback of ANN Algorithms (Kotsiantis et al. 2007). The complexity of the resulting models, the avoidance of missing values, and the time required to develop a neural network on a collected data are just a few of the issues that arise when utilizing this technique (Kotsiantis et al. 2007; Pham and Afify, 2005).

5.7 CONCLUSIONS

This study is intended to illustrate the application of radio waves in conjunction with supervised machine learning (ML) in the operations of poultry manufacturing by grouping myopathic chicken into categories based on the severity level at different processing steps. In comparison to the qualitative approach, SVM and BPNN combined with RF-waves and processing steps weight data can be used to classify myopathies, such as identification of WB in real-time in-line processing. This results in a more accurate classification of myopathies at different processing steps. It may be possible to increase the classification accuracy of both the SVM and the BPNN by include other meat quality criteria, such as the number of dissolved solids, fat, and protein in the product. Future studies will also include larger amounts of information for myopathies to mitigate the interleaved of conditions resulting from human error during processing. This will allow researchers to obtain a well-trained framework for categorization at different processing steps. The novel combination of these techniques has the potential to increase the efficiency of poultry processing and reduce the number of downgrades of breast fillets caused by undesired myopathies, all while simultaneously lowering the number of customer complaints.

5.8 DATA AVAILABILITY STATEMENT

The data sets that were collected for this study will be made available upon request.

5.9 CONTRIBUTIONS MADE BY AUTHORS

AM is the principal investigator and was responsible for conceiving of the idea, securing funding, and carrying out the research. AS was responsible for the gathering of Radio-wave analysis data, and performing SVM analysis and BPNN analysis, and production of the text. AM and LJG both contributed to the writing process by reviewing it and modifying it.

5.10 THE CONTRIBUTION THAT YOU MADE TO THE FIELD STATEMENT:

Over the course of the past few of decades, there has been a significant rise in the intake of animal-based food on a global scale. The average annual consumption was only 10 kg in the 1960s, but by the 2000s it had risen to 26 kg, and projections indicate that it will reach 37 kg by the year 2030. The physicochemical and sensory features of chicken meat, such as its texture, color, and flavor, are largely responsible for the strong market for chicken meat. Hand-palpation is the only low-cost method for categorizing the severity of WB fillets after processing and deboning; nevertheless, it is arbitrary, difficult, and has a substantial error in categorization.

Despite these drawbacks, it is still the only method currently available. In order for businesses to reduce the amount of money lost due to incorrect identification, stringent criteria for enhanced fillet sorting need to be developed right away. The processes of optimization, output tracking and control, and forecasting have all benefited significantly from the application of machine learning. In other areas of the food industry, quality assurance in production industry has seen significant improvements thanks to the widespread implementation of artificial intelligence techniques such as these. According to the findings of this study, radio-wave sensors in conjunction with machine learning algorithms during in-line processing had a strong ability to effectively categorize myopathies at different processing steps.

5.11 REFERENCES

Adebiyi, M. O., Arowolo, M. O., Mshelia, M. D., and Olugbara, O. O. (2022). A Linear Discriminant Analysis and Classification Model for Breast Cancer Diagnosis. Appl. Sci., 12, 11455.Doi: http://dx.doi.org/10.3390/app122211455

Akay, D. (2011). Grey relational analysis based on instance-based learning approach for classification of risks of occupational low back disorders. Saf. Sci. 49, 1277-1282.Doi: http://dx.doi.org/10.1016/j.ssci.2011.04.018

Alpaydin, E. (2010). Introduction to machine learning. Cambridge, MA: MIT Press.

Alzu'bi, A., Albalas, F., Al-Hadhrami, T., Younis, L. B., and Bashayreh, A. (2021). Masked face recognition using deep learning: A review. Electronics, 10, 2666.Doi: http://dx.doi.org/10.3390/electronics10212666

Anderberg, M. R. (2014). Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks (Vol. 19). Academic press.

Azadeh, A., Saberi, M., Kazem, A., Ebrahimipour, V., Nourmohammadzadeh, A., and Saberi, Z. (2013). A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyper-parameters optimization. Appl. Soft Comput. 13, 1478-1485.Doi: http://dx.doi.org/10.1016/j.asoc.2012.06.020

Bagozzi, R. P. (2007). The legacy of the technology acceptance model and a proposal for a paradigm shift. J. Assoc. Info. Sys. 8, 3.Doi: http://dx.doi.org/10.17705/1jais.00122

Barbon, S., Costa Barbon, A. P. A. D., Mantovani, R. G., and Barbin, D. F. (2018). Machine learning applied to near-infrared spectra for chicken meat classification. J. Spectrosc. https://doi.org/10.1155/2018/8949741

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B (1), 289-300. Doi: http://dx.doi.org/10.1111/j.2517 6161.1995.tb02031.x

Bogomolov, M., Peterson, C. B., Benjamini, Y., and Sabatti, C. (2017). Testing hypotheses on a tree: new error rates and controlling strategies. arXiv preprint arXiv:1705.07529.Doi: http://dx.doi.org/10.1093/biomet/asaa086

Borin, A., Ferrao, M. F., Mello, C., Maretto, D. A., and Poppi, R. J. (2006). Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. Anal. Chim. Acta. 579, 25-32.Doi: http://dx.doi.org/10.1016/j.aca.2006.07.008

Burbidge, R., Trotter, M., Buxton, B., and Holden, S. (2001). Drug design by machine learning: support vector machines for pharmaceutical data analysis. Comput. Chem. 26, 5-14.Doi: http://dx.doi.org/10.1016/S0097-8485(01)00094-8

Çaydaş, U., and Ekici, S. (2012). Support vector machines models for surface roughness prediction in CNC turning of AISI 304 austenitic stainless steel. J. Intell. Manuf. 23, 639 650.

Chatterjee, D., Zhuang, H., Bowker, B.C., Rincon, A.M., and Sanchez-Brambila G (2016). Instrumental texture characteristics of broiler *Pectoralis major* with the wooden breast condition. Poult Sci. 95, 2449–2454. doi: 10.3382/ps/pew204

Chinnam, R. B. (2002). Support vector machines for recognizing shifts in correlated and other manufacturing processes. Int. J. Prod. Res. 40, 4449-4466.Doi: http://dx.doi.org/10.1080/00207540210152920

Clark, D.L., and Velleman, S.G. (2016). Spatial influence on breast muscle morphological structure, myofiber size, and gene expression associated with the wooden breast myopathy in broilers. Poult Sci. 95, 2930–2945. doi: 10.3382/ps/pew243

Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p values. R. Soc. Open Sci., 1, 140216.Doi: http://dx.doi.org/10.1098/rsos.140216

Cortes, C., and Vapnik, V. (1995). Support-vector networks. Mach. Learn. 20, 273-297.Doi: http://dx.doi.org/10.1007/BF00994018

Cruz, R. F. A., Vieira, S. L., Kindlein, L., Kipper, M., Cemin, H. S., and Rauber, S. M. (2017). Occurrence of white striping and wooden breast in broilers fed grower and finisher diets with increasing lysine levels. Poul. Sci., 96, 501-510. Doi: http://dx.doi.org/10.3382/ps/pew310

Duan, Z., Zhang, M., Hu, Q., & Sun, J. (2005). Characteristics of microwave drying of bighead carp. *Dry. Technol.*, *23*(3), 637-643. Doi: http://dx.doi.org/10.1081/DRT 200054156

Ellingson, S. W. (2016). Radio Systems Engineering. Cambridge University Press.

Fleming, J. A. (1919). The principles of electric wave telegraphy and telephony. Longmans, Green.

Geronimo, B. C., Mastelini, S. M., Carvalho, R. H., Júnior, S. B., Barbin, D. F., Shimokomaki, M., and Ida, E. I. (2019). Computer vision system and near-infrared spectroscopy for identification and classification of chicken with wooden breast, and physicochemical and technological characterization. Infrared Phys. Technol. 96, 303-310.Doi: http://dx.doi.org/10.1016/j.infrared.2018.11.036

Ghirardi, A. (1932). Radio Physics Course, 2nd ed. New York: Rinehart Books, p. 249

Graham, J. W., Olchowski, A. E., and Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev. Sci.,8, 206 213.Doi: http://dx.doi.org/10.1007/s11121-007-0070-9

Huang, X., and Ahn, D. U. (2018). The incidence of muscle abnormalities in broiler breast meat A review. Korean J. Food Sci. Anim. Resour. 38, 835. Doi: http://dx.doi.org/10.5851/kosfa.2018.e2

Johnson, C. C., and Guy, A. W. (1972). Nonionizing electromagnetic wave effects in biological materials and systems. Proceedings of the IEEE, 60, 692-718.Doi: http://dx.doi.org/10.1109/PROC.1972.8728

Jiang, H., Yoon, S. C., Zhuang, H., Wang, W., and Yang, Y. (2017). Evaluation of factors in development of Vis/NIR spectroscopy models for discriminating PSE, DFD and normal broiler breast meat. Br. Poult. Sci. 58, 673-680.Doi: http://dx.doi.org/10.1080/00071668.2017.1364350

Jiao, S., Sun, W., Yang, T., Zou, Y., Zhu, X., and Zhao, Y. (2017). Investigation of the feasibility of radio frequency energy for controlling insects in milled rice. Food Bioproc. Tech., 10(4), 781-788.Doi: http://dx.doi.org/10.1007/s11947-017-1865-8

Kecman, V. (2001). Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models. Massachusetts: MIT Press.

Klein, A. (1980). Microwave moisture determination of coal.-A comparison of attenuation and phase measurement. In 1980 10th European Microwave Conf. (pp. 526-530). IEEE. Doi: http://dx.doi.org/10.1109/EUMA.1980.332853

Klimberg, R., and McCullough, B. D. (2016). Fundamentals of predictive analytics with JMP. SAS Institute.

Kotsiantis, S. B., Zaharakis, I., and Pintelas, P. (2007). Supervised Machine Learning: A Review of Classification Techniques in Emerging artificial intelligence applications in computer engineering. (Amsterdam: IOS Press), 3-24.

Kraszewski, A., Kulinski, S., and Stosio, Z. (1977). A preliminary study on microwave monitoring of moisture content in wheat. J. Microw. Power, 12(3), 241-252.Doi: http://dx.doi.org/10.1080/16070658.1977.11689052

Kuttappan, V. A., Hargis, B. M., and Owens, C. M. (2016). White striping and woody breast myopathies in the modern poultry industry: a review. Poult. Sci., 95, 2724-2733.Doi: http://dx.doi.org/10.3382/ps/pew216

Kuttappan, V. A., Owens, C. M., Coon, C., Hargis, B. M., and Vazquez-Anon, M. (2017). Incidence of broiler breast myopathies at 2 different ages and its impact on selected raw meat quality parameters. Poult. Sci. J. 96, 3005-3009.Doi: http://dx.doi.org/10.3382/ps/pex072

Lee Jr, D. B. (1973). Requiem for large-scale models. *J Am Plann Assoc*, 39(3), 163-178. Doi: http://dx.doi.org/10.1080/01944367308977851

Li, H., Liang, Y., and Xu, Q. (2009). Support vector machines and its applications in chemistry. Chemom. Intell. Lab. Syst. 95, 188-198. Doi: http://dx.doi.org/10.1016/j.chemolab.2008.10.007

Li, C., Han, B., and Zhang, T. (2015). Free-space reflection method for measuring moisture contentand bulk density of particulate materials at microwave frequency. Rev. Sci. Instrum, 86, 034712.Doi: http://dx.doi.org/10.1063/1.4916262

Maji, S., Berg, A. C., and Malik, J. (2008). Classification using intersection kernel support vector machines is efficient. In 2008 IEEE conference on computer vision and pattern recognition (pp. 1-8). IEEE. Doi: http://dx.doi.org/10.1109/CVPR.2008.4587630

Mazzoni, M. Petracci, M. Meluzzi, A. Cavani, C. Clavenzani, P. and Sirri, F. (2015).Relationship between *Pectoralis major* muscle histology and quality traits of chicken meat. Poult Sci. 94, 123-30. doi: 10.3382/ps/peu043

Michalski, R. S., Stepp, R. E., and Diday, E. (1981). A recent advance in data analysis: Clustering objects into classes characterized by conjunctive concepts. Pattern Recognit., 33-56.Doi: http://dx.doi.org/10.1016/B978-0-444-86325-6.50005-9

Morey, A., Smith, A. E., Garner, L. J., and Cox, M. K. (2020). Application of Bioelectrical Impedance Analysis to Detect Broiler Breast Filets Affected with Woody Breast Myopathy. Front. Physiol. 11, 808. doi: 10.3389/fphys.2020.00808

Mudalal, S., Lorenzi, M., Soglia, F., Cavani, C., and Petracci, M. (2015). Implications of white striping and wooden breast abnormalities on quality traits of raw and marinated chicken meat. Animals. 9, 728–734. doi: 10.1017/S175173111400295X

Mulligan, J. F. (1987). The influence of Hermann von Helmholtz on Heinrich Hertz's contributions to physics. Am. J. Phys., *55*, 711-719.Doi: http://dx.doi.org/10.1119/1.15062

Nappo, F. (2021). The double nature of Maxwell's physical analogies. Stud Hist Philos Sci Part A, 89, 212-225.Doi: http://dx.doi.org/10.1016/j.shpsa.2021.08.003

National Chicken Council (NCC). 2020. Statistics: Capita Consumption of Poultry and Livestock, 1960 to Forecast 2021, in Pounds. Available from https://www.nationalchickencouncil.org/about-the-industry/statistics/per-capita consumption-of-poultry-and-livestock-1965-to-estimated-2012-in-pounds/ [accessed May. 2022].

Nelson, S. O., and Stetson, L. E. (1976). Frequency and moisture dependence of the dielectric properties of hard red winter wheat. J. Agric. Eng. Res., 21, 181-192.Doi: http://dx.doi.org/10.1016/0021-8634(76)90073-1

Nilsson, N. J. (2005). Introduction to machine learning. Standford: Standford University Press.

Papah, M. B., Brannick, E. M., Schmidt, C. J., and Abasht, B. (2017). Evidence and role of phlebitis and lipid infiltration in the onset and pathogenesis of Wooden Breast Disease in modern broiler chickens. Avian Pathol. 46, 623-643.Doi: http://dx.doi.org/10.1080/03079457.2017.1339346

Petracci, M., Mudalal, S., Soglia, F., and Cavani, C. (2015). Meat quality in fast-growing broiler chickens. Worlds Poult. Sci. J. 71, 363–374. doi: 10.1017/S0043933915000367

Pham, D. T., and Afify, A. A. (2005). Machine-learning techniques and their applications in manufacturing. Proc. IMechE., Part B: J. Eng. Manuf. 219, 395-412.Doi: http://dx.doi.org/10.1243/095440505X32274

Radaelli, G., Piccirillo, A., Birolo, M., Bertotto, D., Gratta, F., Ballarin, C., Vascellari, M., Xiccato, G., and Trocino, A. (2017). Effect of age on the occurrence of muscle fiber degeneration associated with myopathies in broiler chickens submitted to feed restriction. Poult. Sci. 96, 309–319. doi: 10.3382/ps/pew270

Rahman, M. S., Al-Farsi, K., Al-Maskari, S. S., and Al-Habsi, N. A. (2018). Stability of electronic nose (e-nose) as determined by considering date-pits heated at different temperatures. Int. J. food prop., 21, 850-857.Doi: http://dx.doi.org/10.1080/10942912.2018.1463540

Ramachandran, L., Gajski, D. D., & Chaiyakul, V. (1994). An algorithm for array variable clustering. In Proceedings of European Design and Test Conference EDAC ETC EUROASIC, pp. 262-266.Doi: http://dx.doi.org/10.1109/EDTC.1994.326867

Resurreccion, A. V. A. (2004). Sensory aspects of consumer choices for meat and meat products. Meat Sci., 66, 11-20.Doi: http://dx.doi.org/10.1016/S0309-1740(03)00021-4

Ribeiro, B. (2005). Support vector machines for quality monitoring in a plastic injection molding process. IEEE T. Syst. Man. Cy. C. 35, 401-410.Doi: http://dx.doi.org/10.1109/TSMCC.2004.843228

Romanenko, S., Begley, R., Harvey, A. R., Hool, L., and Wallace, V. P. (2017). The interaction between electromagnetic fields at megahertz, gigahertz and terahertz frequencies with cells, tissues and organisms: risks and potential. J. R. Soc. Int., 14, 20170585.Doi: http://dx.doi.org/10.1098/rsif.2017.0585

Salahshoor, K., Khoshro, M. S., and Kordestani, M. (2011). Fault detection and diagnosis of an industrial steam turbine using a distributed configuration of adaptive neuro-fuzzy inference systems. Simul. Model Pract. Th. 19, 1280-1293.Doi: http://dx.doi.org/10.1016/j.simpat.2011.01.005

Siddique, A., Shirzaei, S., Smith, A. E., Valenta, J., Garner, L. J., & Morey, A. (2021). Acceptability of Artificial Intelligence in Poultry Processing and Classification Efficiencies of Different Classification Models in the Categorisation of Breast Fillet Myopathies. Front. Avian Physiol., 1472. Doi: https://doi.org/10.3389/fphys.2021.712649

Siddique, A.; Herron, C.B.; Valenta, J.; Garner, L.J.; Gupta, A.; Sawyer, J.T.; Morey, A.(2022). Classification and Feature Extraction Using Supervised and Unsupervised Machine

Learning Approach for Broiler Woody Breast Myopathy Detection. *Foods*, *11*, 3270. Doi: https://doi.org/10.3390/foods11203270

Sihvo, H. K., Immonen, K., & Puolanne, E. (2014). Myodegeneration with fibrosis and regeneration in the pectoralis major muscle of broilers. Vet. Pathol., 51, 619-623.Doi: http://dx.doi.org/10.1177/0300985813497488

Sihvo, H. K., Lindén, J., Airas, N., Immonen, K., Valaja, J., and Puolanne, E. (2017). Wooden breast myodegeneration of *Pectoralis major* muscle over the growth period in broilers. Vet. Pathol. 54, 119–128. doi: 10.1177/0300985816658099

Soglia, F., Mudalal, S., Babini, E., Di Nunzio, M., Mazzoni, M., Sirri, F., and Petracci, M. (2016). Histology, composition, and quality traits of chicken Pectoralis major muscle affected by wooden breast abnormality. Poult. Sci. J. 95, 651-659.Doi: http://dx.doi.org/10.3382/ps/pev353

Song, X. J., Zhang, M., Mujumdar, A. S., and Fan, L. (2009). Drying characteristics and kinetics of vacuum microwave–dried potato slices. Dry. Technol., 27, 969-974.Doi: http://dx.doi.org/10.1080/07373930902902099

Storey, J. D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences, 100, 9440-9445.Doi: http://dx.doi.org/10.1073/pnas.1530509100

Tay, S. H., Toh, M. M. X., Thian, Y. L., Vellayappan, B. A., Fairhurst, A. M., Chan, Y. H., and Wong, A. S. C. (2022). Cytokine Release Syndrome in Cancer Patients Receiving Immune Checkpoint Inhibitors: A Case Series of 25 Patients and Review of the Literature. Front. immunol. , 13.Doi: http://dx.doi.org/10.3389/fimmu.2022.807050

Trabelsi, S. A. M. I. R., Kraszewski, A. W., and Nelson, S. O. (1997). Simultaneous determination of density and water content by microwave sensors. Electr. Letters, 33, 874-876.

Trabelsi, S., Krazsewski, A. W., and Nelson, S. O. (1998). New density-independent calibration function for microwave sensing of moisture content in particulate materials. IEEE Trans Instrum Meas, 47, 613-622.Doi: http://dx.doi.org/10.1109/19.744310

Vapnik, V. (1995). The Nature of Statistical Learning Theory. New York: Springer-Verlag.Doi: http://dx.doi.org/10.1007/978-1-4757-2440-0

Wang, K. J., Chen, J. C., and Lin, Y. S. (2005). A hybrid knowledge discovery model using decision tree and neural network for selecting dispatching rules of a semiconductor final testing factory. Prod. Plan. Control. 16, 665-680. Doi: http://dx.doi.org/10.1080/09537280500213757

Wold, J. P., Måge, I., Løvland, A., Sanden, K. W., and Ofstad, R. (2019). Near-infrared spectroscopy detects woody breast syndrome in chicken fillets by the markers protein content and degree of water binding. Poult. Sci. J. 98, 480-490. Doi: http://dx.doi.org/10.3382/ps/pey351

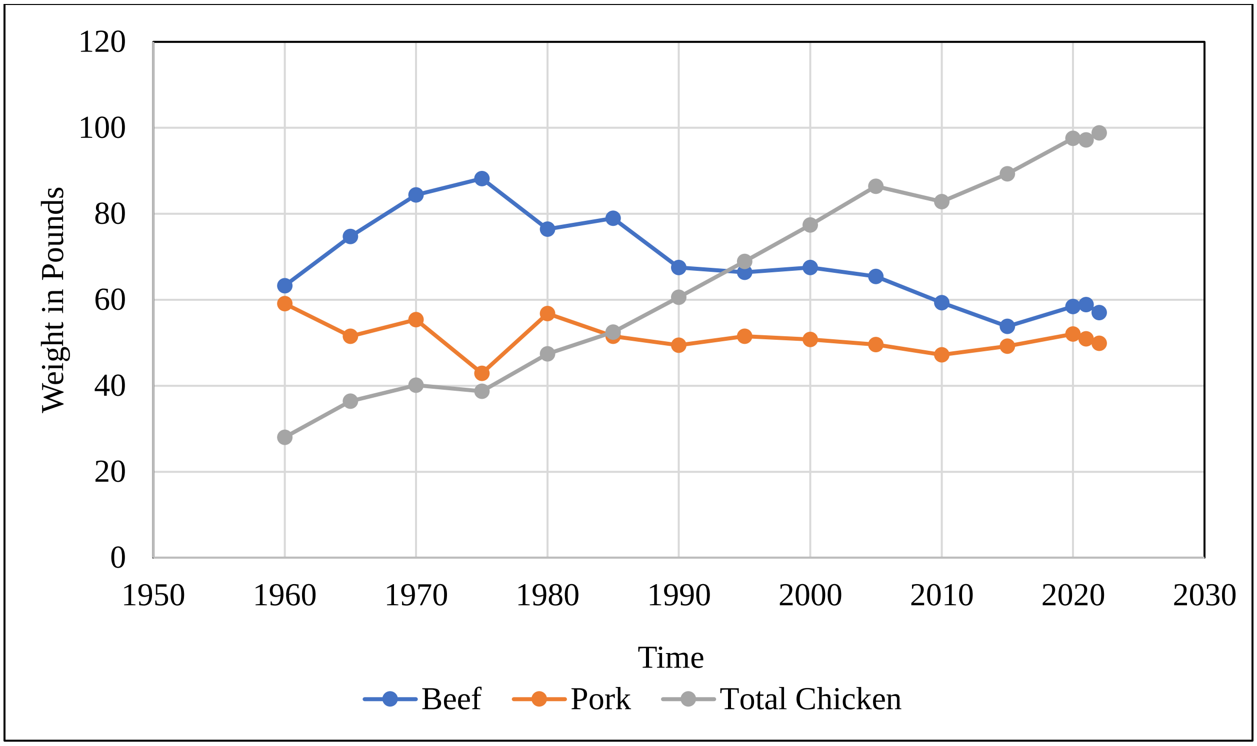


Figure 5-1: Consumption trend of chicken compared to beef and pork (Modified from NCC, 2021)

Figure 5-2: Identification of myopathic fillets using Hand-palpation technique based on 4 point scale (Tijare et al. 2016)

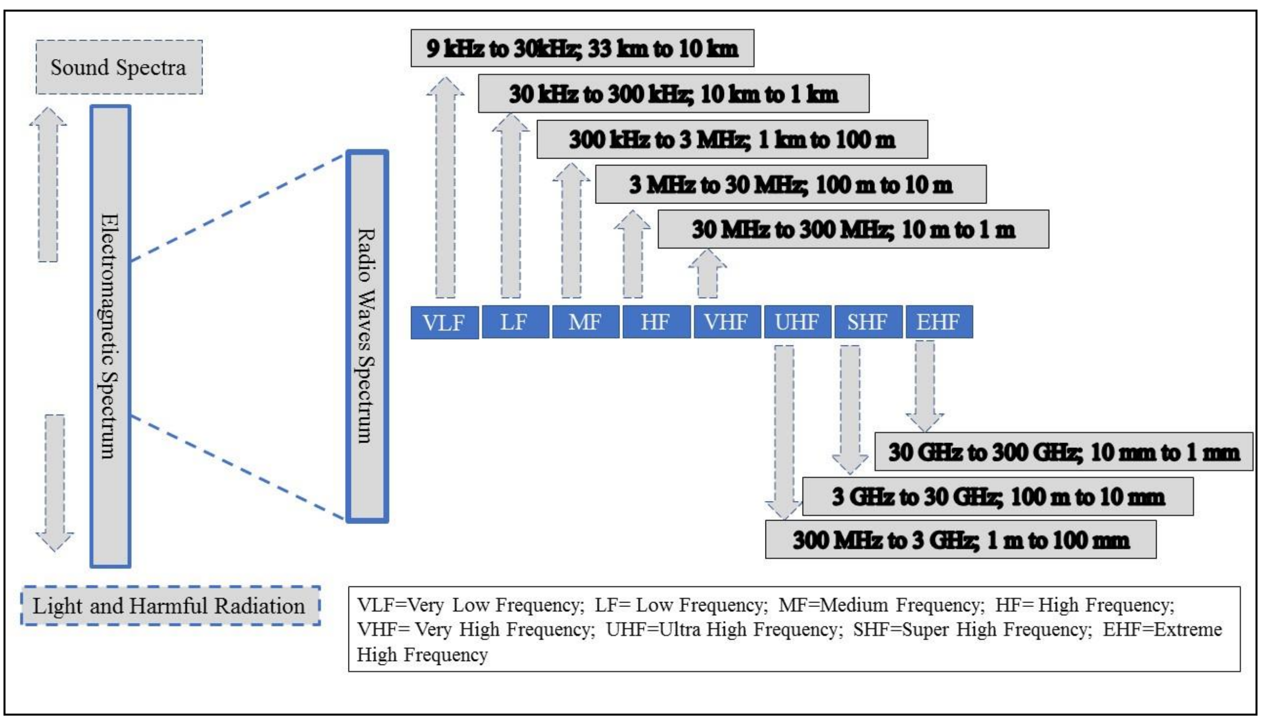


Figure 5-3: Descriptive Radio frequency spectrum bands with travelling distances

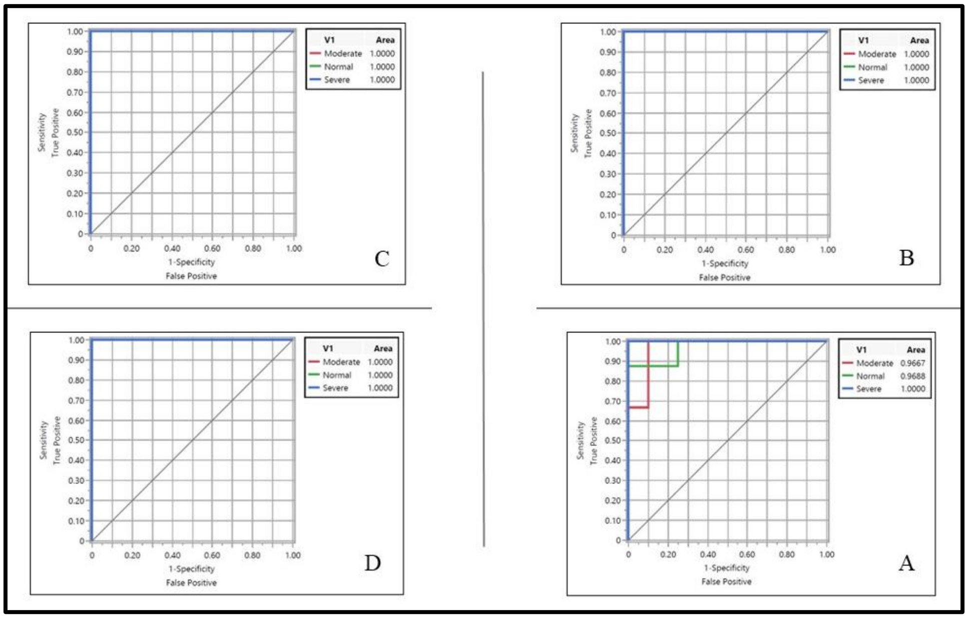


Figure 5-4: Receiver Operating Characteristics for different processing steps

Table 5.1:Different subjective scales used for the classification of woody breast meat.

Woody Breast Subjective Classification Scale[1]	**Condition**	**Description**
2 Point Scale	**Normal**	No toughness or Hardness
	Severe	Tough fillets
3 Point Scale	**Normal**	No toughness or Hardness
	Moderate	Medium toughness up to 50%
	Severe	More than 50% toughness
4 Point Scale	**Normal**	No toughness or hardness
	Mild	Hardness at cranial region
	Moderate	Filets extremely hard and rigid through from cranial region of caudal tip filets that were hard throughout but flexible in mid-to caudal region
	Severe	More than 50% of fillet area is woody

[1]2 point scale (Sihvo et al. 2014), and 3 point scale (Sihvo et al. 2014), 4 point scale (Tijare et al. 2016)

Table 5.2: Summary table for specific radio-wave frequency from top 100 frequencies for different processing steps

Step	Condition	Frequency Ranges
Live Birds	**Normal**	2.14 GHz, 2.33 GHz, 6.06 GHz, 8.73 GHz, 10.21 GHz, 12.60 GHz, 16.95 GHz
	Moderate	2.16 GHz, 6.06 GHz, 8.41 GHz, 9.27 GHz, 10.36 GHz, 12.61GHz, 16.95 GHz
	Severe	2.16 GHz, 6.06 GHz, 8.77 GHz, 10.21 GHz, 12.61 GHz, 16.95 GHz
Pre-Chill WOGs	**Normal**	2.18GHz, 3.21 GHz, 4.60 GHz, 5.71 GHz, 7.50 GHz, 8.90 GHz, 10.06 GHz, 11.16 GHz, 16.09 GHz
	Moderate	2.18 GHz, 3.15 GHz, 3.92 GHz, 4.85 GHz, 5.71 GHz, 6.93 GHz, 7.69 GHz, 8.90 GHz, 9.97 GHz, 10.93 GHz, 16.07 GHz
	Severe	2.18 GHz, 3.33 GHz, 4.75 GHz, 5.90 GHz, 8.52 GHz, 10.00 GHz, 10.93 GHz, 15.99 GHz
Post-Chill WOGs	**Normal**	9.40 GHz, 9.90 GHz, 11.67 GHz, 12.48 GHz, 13.01 GHz, 14.79 GHz
	Moderate	9.40 GHz, 9.90 GHz, 11.98 GHz, 12.48 GHz, 13.01 GHz, 14.79 GHz
	Severe	9.40 GHz, 9.90 GHz, 12.24 GHz, 12.86 GHz, 14.86 GHz
Deboned Fillets	**Normal**	9.54 GHz, 9.93 GHz, 10.37 GHz, 14.86 GHz, 17.92 GHz
	Moderate	9.41 GHz, 9.71 GHz, 10.12 GHz, 14.96 GHz, 17.14 GHz, 17.91 GHz
	Severe	9.52 GHz, 10.03 GHz, 14.99 GHz, 17.63 GHz

Table 5.3: Classification accuracy summary table for Live Birds, pre-chill WOGs, post-chill WOGs and Deboned fillets using BPNN algorithm.

Processing Steps	Model Measurement		Testing Accuracy		
			Normal	Moderate	Severe
Live Birds	Generalized R^2	0.83	87.5	83.3	100
	Entropy R^2	0.64			
	RASE	0.33			
	Misclassification Rate	0.12			
Pre-Chill WOG	Generalized R^2	0.78	100	87.5	100
	Entropy R^2	0.56			
	RASE	0.37			
	Misclassification Rate	0.06			
Post-Chill WOG	Generalized R^2	0.91	69.7	100	100
	Entropy R^2	0.78			
	RASE	0.26			
	Misclassification Rate	0.10			
Deboned Fillets	Generalized R^2	0.85	100	66.7	100
	Entropy R^2	0.66			
	RASE	0.33			
	Misclassification Rate	0.26			

Table 5.4: Support vector machines classification accuracy summary table for Live Birds, pre-chill WOGs, post-chill WOGs and Deboned fillets

Condition	Type	Sensitivity	Specificity	Testing
Live Birds	**Normal**	68.4	55.5	**61.9**
	Moderate	45.5	73.0	**59.2**
	Severe	42.8	96.6	**69.7**
Pre-Chill WOG's	**Normal**	34.6	75.0	**54.8**
	Moderate	50.0	77.7	**63.8**
	Severe	70.0	67.5	**68.7**
Post-Chill WOG's	**Normal**	38.4	84.0	**61.2**
	Moderate	53.3	77.7	**65.5**
	Severe	60.0	63.4	**61.7**
Deboned Fillets	**Normal**	46.1	92.3	**69.2**
	Moderate	75.0	72.2	**73.6**
	Severe	80.0	80.9	**80.4**