MODELING AND CHARACTERIZATION OF POTATO QUALITY BY ACTIVE THERMOGRAPHY

A Thesis

by

CHIH-CHEN SUN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2008

Major Subject: Mechanical Engineering

MODELING AND CHARACTERIZATION OF POTATO

QUALITY BY ACTIVE THERMOGRAPHY

A Thesis

by

CHIH-CHEN SUN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Sheng-Jen "Tony" Hsieh
Committee Members,	Obdulia Ley
	Steve Suh
Head of Department,	Dennis O'Neal

August 2008

Major Subject: Mechanical Engineering

ABSTRACT

Modeling and Characterization of Potato Quality

by Active Thermography. (August 2008)

Chih-Chen Sun, B.En., National Taiwan University

Chair of Advisory Committee: Dr. Sheng-Jen (Tony) Hsieh

This research focuses on characterizing a potato with extra sugar content and identifying the location and depth of the extra sugar content using the active thermography imaging technique. The extra sugar content of the potato is an important problem for potato growers and potato chip manufacturers. Extra sugar content could result in diseases or wounds in the potato tuber. In general, potato tubers with low sugar content are considered as having a higher quality.

The inspection system and general methodologies characterizing extra sugar content will be presented in this study. The average heating rate obtained from the thermal image analysis is the major factor in characterization procedures. Using information on the average heating rate, the probability of achieving a potato with extra sugar content may be predicted using the logistic regression model. In addition, neural networks are also used to identify the potato with extra sugar contents. The correct rate for identifying a potato with extra sugar content in it can reach 85%. The location of extra sugar content can also be found using the logistic regression model. Results show the overall correct rate predicting the extra sugar content location with a resolution of

20 by 20 pixels is 91%. In predicting the extra sugar content depth, amounts exceeds 2/3 inches are not detectable by analyzing thermal images. The depth of extra sugar content can be discriminated in 0.3 inch increments with a high rate of accuracy (87.5%).

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Sheng-Jen “Tony” Hsieh, and my committee members, Dr. Obdulia Ley and Dr. Steve Suh, for their guidance and assistance in this research.

I would also like to thank Amy Chang, Yuting Chang, John Chiang, Damson Lai, and Jasmine Lin for their kind support over the past two years.

Finally, thanks to my parents for providing me with everything.

NOMENCLATURE

RTD	Resistance Temperature Detector
IR	Infrared Thermography
α	Absorptivity
ρ	Reflectivity
τ	Transmissivity
ε	Emissivity
E	Energy
NIR	Near Infrared Spectroscopy
PLC	Programmable Logic Controller
H	Heating Rate
H_{avg}	Average Heating Rate
P	Probability
LL	Log of the Likelihood
ΔH_{avg}	Average Heating Rate Difference
H_{avg_whole}	Average Heating Rate of the Whole Area of Interest

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

1.1 Motive

Consumers focus has traditional concentrated on the quality when purchasing food, especially for agricultural products. Consequently, the inspection of agricultural products has become an important issue in the industry. Therefore, postharvest grading of agricultural products is difficult without the aid of modern technologies. Currently, there are still many food companies using visual inspection by human to pick out the unqualified products based on the product's flavor, texture, and color. The results will be affected by many factors such as experiences of the operators, physical status of the operator, and environmental conditions. This will be costly and unreliable to obtain accurate results comparing to machine inspection. Moreover, the defects inside the food cannot be found out by human easily. So the development of a new system to characterize the quality of agricultural products is in great need.

The focus of this study is to characterize the potato with excessive sugar using active thermography imaging technique. Extra sugar content of potato is an important and difficult problem for potato growers and potato chip manufacturers. Russet potato is chosen as our experimental material due to the fact that it resembles the potato for most

---

This thesis follows the style of *Nondestructive Testing and Evaluation*.

uses. Further, it is easy to obtain and the properties of potato are well known. In addition, its shape can be approximate as an ellipsoid, which is found to be more convenient for thermal modeling comparing to irregular shapes. Consequently, the potato is a representative produce because its usage is very large comparing to other produce. The consumption of potato around the world was found to be 218.13 million tons in the year of 2006 [1]. Potato can be consumed fresh or as processed food products such as potato crisp chips, which are the most popular snack food in many countries.

The internal quality of a potato is related to its sugar contents. In general, potato tubers with low sugar contents are considered as higher quality for most uses [2]. These reasons include the idea that potato tuber produces sugar from the starch to survive by respiration, and the temperature and pressure control during the potato storage process might increase the respiration rate of potato. Excessive sugar accumulation could be a result of disease or dead tissue within the tuber. Hence monitoring the sugar concentration of potato tuber becomes an important topic for both the grower and processor.

Thermography has been a powerful evaluation method in many fields such as material defects characterizing, and electronic fault diagnosis. It is a highly potential tool when applying on the food quality inspection because it is non-destructive, easy to operate, and safe comparing to other technologies. Moreover, after taking thermal images, the food can still be consumed. However, the accuracy and reliability depend on the resolution of the infrared camera. The economy of using infrared thermography to replace human inspection is needed to be studied.

1.2 Nature of the problem

The potato quality is characterized by its external and internal quality. External qualities are related to the size, shape, and defects. The number of the buds on the tuber is also an indicator of the potato quality. Buds will start to grow if the potato is mature enough, and there will be chemical constituents that affect people's health. There might also be holes caused by disease or bacterial on the potato. Holes and cracks inside a material will emit higher rates of thermal energy than their surrounding surfaces, and the shapes will appear on the thermal images [3]. So the thermal images of the potato are examined to study these possible phenomena that have adverse effect the potato.

The internal quality is highly related to the sugar contents inside the potato tuber. When there is excessive sugar contents in potato tuber, it might cause by diseases, wounding or other unwanted effects occurring in the potato. In this study, the active thermography technique will be used to characterize the extra sugar concentration in the potato. By applying an external heat source on a potato and recording the surface temperature changes, the excessive sugar contents in the potato can be identified. The extra sugar contents regions inside the potato will have different thermal properties such as thermal conductivity and specific heat, hence might affect the heating rate of the potato. Moreover, when the heat propagates through the sugar concentration area, it will encounter a subsurface discontinuity and reduce the heat diffusion rate. Such discontinuity area causes the surface temperature to be higher than other surrounding regions, further alters the cooling rate.

1.3 Problem statement

In this research we focus on identifying potato with extra sugar contents and identifying the extra sugar concentration location and depth in potato. The development of a system and general methods to characterize the quality of the potato will be presented in this research. By using the thermography imaging analysis, the extra sugar contents of potato tubers will be identified due to the abnormalities of the thermal behavior of the specific surface area.

1.4 Scope and objective of research

The scope of this work are to understand how the heating and sugar contents may affect the heat transfer in a potato and to design and validate the algorithms for detecting potato with excessive sugar and finding the excessive sugar location and depth. The properties of each potato cannot be the same, so that some of the properties are assume to be the same if they only have minor effects to the heating rates such as skin texture, and moisture contents of potato. The environmental conditions are significant in the experiment. Hence the environment is controlled to be as identical as possible to eliminate the noise during the whole experiment. The factors that might affect the surface temperature distribution will be considered as a variable, such as the density and the surface area of the potato. The potato to be examined is also treated as a healthy and fresh one although it cannot be verified.

In constructing the potato thermal model, some assumptions are also be made. These include the shape of potato is assumed to be a perfect ellipsoid without any cavity and irregular surface change on the surface. The physical properties of the potato are considered to be uniform inside the potato although they might be varied in different areas.

The research goals are summarized as follows:

- Construct a semi-automatic prototype system for loading, heating, and scanning potato using infrared camera.
- Design and validate the algorithms to identify potato with excessive sugar, the sugar concentration location and depth within the potato
- Propose a general experimental methodology for potato quality control

1.5 Format of research

The procedures of this research start from the concept of studying the thermal images of potato surface to predict its quality, and follow by constructing experiments and thermal modeling. These procedures are described in the following chapters.

Chapter II includes the literature review on the introduction and application of infrared thermography, the food thermal modeling methods, and the current food inspecting techniques.

Chapter III talks about the design concept of the experimental setup and material used, and introduces the preliminary experiments and procedures for each experiment.

Chapter IV consists of the results obtained from experiments and the general methodologies of the potato quality characterization.

Chapter V discusses the validation of the models used in the extra sugar contents prediction.

The last chapter VI consists of the conclusion, limitation, and future work that can be done in this research.

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

The research is associated to the inspection of produce quality by using infrared thermography. The target produce to study here is russet potato, but the method can be applied to other agricultural products. The internal quality of potato highly relates to its sugar contents, especially for chip potatoes. So this chapter starts with the introduction of potato quality with possible mechanism of the sugar contents variation in potatoes. Next section will introduce the principle and the applications of infrared thermography. The last section introduces other techniques currently be applied to the produce inspection.

2.2 Quality of potato

The quality of potato can be characterized by the external quality and the internal quality. The external quality relates to the size, shape, appearance, and defects of the potato where the internal quality deals with the appearance, nutrition contents, and defects of potato. The summary of the potato quality factors are in table 1. The defects of potato might due to the physiological disorders, fungus, pest, handling or storage [4]. Common defects of potato tubers are crack, scab, greening, and rot. Although appropriate growing conditions can reduce the defects of potato, a reliable inspection method is still in demand for the potato quality control.

Table 1. Summary of potato quality factors.

	Factors of potato quality
External qualities	Size and shape Color Defect: crack, greening, scab,…
Internal qualities	Appearance: color, hardness, skin texture, … Nutrition content: starch, vitamin, sugar, … Defect: cavity, rotten, sugar end…

The internal quality of potato and the sugar contents are highly related. The lower potato sugar contents are recommended in most uses, especially for chip potatoes. After harvest, sugar content inside the potato tuber is relative to the respiration effect. Potato tuber needs to produce sugar form starch for respiration to survive. In living cells, respiration effect can release the energy and produce carbon dioxide and water from organic substances such as sugar. The respiration effect and sugar contents of potato can be affected by many factors including variety, maturity at harvest, postharvest handling stress and the storage environment [5]. The respiration effect can be represented by the equation:

$$C_6H_{12}O_6 + 6\ O_2 = 6\ CO_2 + H_2O + \text{Energy} \quad (2.1)$$

The effect of varying the sugar contents on respiration was found experimentally. As the sugar concentration of the medium increases, respiration increases until reaching very high concentrations. The addition of carbohydrates increased the amount of carbon dioxide comparing to the no sugar added plants [6]. The effect is mainly due to a higher rate of respiration.

Wounding of potato causes an increase in the rate of respiration hence can change the sugar contents in a very short period of time. The result of the experiment done by Hopkins shows clearly that an increase in wounding is accompanied by an increase in sugar contents [6].

In general, the low sugar contents of potato in every processing process are recommended. The quality of potato products such as chips and french fries is largely dependent on the its color, and the color is directly affected by the amounts of sugar contents in the potato. To have acceptable process color of chips and fries, the sugar contents should not exceed the standard range. After harvest, sucrose in the potato can transform into reducing sugar: fructose and glucose. When frying the potato tuber at high temperature, reducing sugars and amino acid will react to form the darker color on potato called browning effect, which is a serious problem when processing [2]. High sucrose contents in potato tuber can lead to reducing sugar accumulation after harvest, and increase the possibility of browning effect. The sucrose contents can be an identification of the maturity and the quality for frying potatoes.

One of the defect types of potato is sugar ends, usually happens on the stem end of the potato tuber. The accumulation of reducing sugars at the one end of potato will also cause dark color after frying the potato. Potato with sugar end will have lower starch and specific gravity than healthy potato. The sugar end on the potato appears translucent on the flesh, and might lead to tissue breakdown on the region.

In conclusion, the extra sugar contents or the sugar accumulation in the potato tuber has adverse effect to the potato quality. For process potato, it is especially important to monitor the sugar contents because the product color is directly related to it.

2.3 Principle and applications of infrared thermography

The study of the surface temperature measurement of potato can help us have the basic understanding of the effect of the extra sugar contents on the thermal behavior of potato. Temperature measurement is an important factor in many applications, especially for agricultural and food industries. Traditional ways to measure the temperature are base on the comparison of different mechanical properties such as the electrical voltage using thermal couple, resistance using temperature detector, and the expansion using liquid thermometers. However, this techniques need to touch the target well to obtain an accurate reading. The comparisons of current temperature measuring techniques are in table 2. Infrared thermography is considered to be a technique with great potentials to measure the surface temperatures of objects due to the fact that you can measure the temperature of an entire potato without coming into contacting with it. In addition, it is able to cooperates with other electronic devices to analysis the data, and easy to operate [7]. The non-contact and non-destructive properties of infrared thermography are specially desired in food temperature measuring so that few contamination and injury will occur.

Table 2. Comparisons of the temperature sensing techniques.

	Measuring Principle	Advantages	Disadvantages
Resistance temperature sensor	measure temperature by correlating the resistance with temperature	stable, reproducible, accurate, fast response	the element is fragile
Thermocouple	comparing the voltage between dissimilar metals, the difference increases with temperature	interchangeable, low cost	not precise enough
Infrared thermography	detect the radiation in the infrared range of electromagnetic spectrum	non-contact, non-destructive, large measuring area	expensive, only surface area

The integrated infrared thermography system can detect the infrared radiation from the target object. When the temperature is different between objects, heat energy will transfer from the higher temperature one to the lower temperature one until they reach the thermal equilibrium. All objects with temperatures higher than absolute zero emit energy by radiation. Thermal infrared radiation is one of the electromagnetic energy which travels at the speed of light. When viewing a surface, not only the radiation has been absorbed may be seen, but also the radiation is being transmitted through the target object or reflected by it. However, the transmitted and reflected radiation did not provide the information for the surface temperature. The radiosity which is the combined radiation reflecting from the surface of the object can be sensed by the detector of the infrared camera. The electromagnetic radiation has different traveling wavelengths, and the wavelengths of infrared radiation can be detected are between 2 and 15 μm. Once sensing the radiation, the detector will submit a signal to the electronic processing system to generate a thermal image on the display screen. After calibrating with the background temperature, the surface temperature of the object can be read directly from the personal computer.

A black body is an object that absorbs all the incident radiation, so it is treated as an ideal source of thermal radiation. However, the objects are not usually black bodies, which mean they will reflect and transmit portions of radiations. The absorptivity (α), reflectivity (ρ) and transmissivity (τ) of non-black bodies depend on the wavelength and the angle of the radiation. For thermal equilibrium, the absorptivity is equal to the emissivity (ε). The relationship between these three factors is

$$\alpha + \rho + \tau = 1 \qquad (2.2)$$

When a surface is opaque, the transmissivity equals to zero. All objects emit infrared radiation as a function of their temperatures. And only the emitted energy provides the information about the surface temperature of the object. This phenomenon allows us to see radiant surfaces with infrared sensing cameras. According to Stefan-Boltzmann law (equation 2.3), the higher the object's temperature is, the more infrared radiation it emits.

$$E = \varepsilon \sigma T^4 \qquad (2.3)$$

where ε is the emissivity of the object. When the object is assumed to be diffuse, the emissivity is equal to its absorptivity. Once the total radiation emitted and the emissivity of the object are known, the surface temperature of the object can be determined form equation 2.3.

Infrared thermography systems are made up of several common components, including lens, detector, processing electronics, controls, display, data storage, data processing and report generation software, and filter. Quantitative infrared thermography involves the use of calibrated instruments. However, calibration is expensive to achieve, which is best made for the specific temperature ranges and accuracy required [3].

Infrared thermography can be separated into two categories: the passive thermography and the active thermography. The passive thermography is when own temperature of the investigated object is naturally different from ambient temperature, while in the active thermography, an external energy source must be applied to the object and measuring its thermal response [8-10]. Furthermore, the way of data collection and analysis is different form passive thermography. The use of active thermography has grown rapidly in the past few years, especially in the aerospace industry. Fast data collection and improved spatial resolution of active thermography can detect the movement of heat transfer in the material [3]. From the sequence of thermal images taking by infrared camera, it is possible to see the relative size and depth of the flaw. In this study, the active thermography is applied. The time and spatial resolution is the key point for the infrared camera to be accurate, especially for detecting the defects inside an object.

The infrared thermography is first used in military applications during World War I. Current nonmilitary applications of infrared thermography are in the fields of non-destructive evaluations and predictive and preventive maintenance where the abnormal temperature behavior can be an early signal of future problems [11]. Some examples are the detection of electronic inspections [11-13], the non-destructive testing of materials [14-15], and medical applications [16]. In agricultural applications, infrared thermography studies have been used in estimating the surface quality of apple [17], and predicting the plant diseases [18].

2.4 Current methods and limitations for food inspection

Various methods have been applied to the inspection of the agricultural products. Traditional food inspection technologies require labor intensive works such as the sample collection and analysis in the laboratory. Moreover, because most of these methods are destructive, the sample cannot be consumed after examination. Current food inspection technologies have focused on the non destructive and real time methods to determine the quality of food based on certain parameters of the food properties. The ideal inspection method for food should be reliable, easy to operate, accurate, cost-effective and able to process large quantities in a short time.

In recent years, many inspection methods are based on machine vision to inspect the outer defects, and the digital image processing is the most essential part in machine vision. Image acquisition including charge coupled device camera, ultrasound, and magnetic resonance imaging have been used on the food inspection. For quality classification, statistical, fuzzy logic, and neural network are used [19]. Image processing technologies can compute the features of the object such as size, shape, and color. However, the data for analyze the image might be very large if every pixel is considered. Xiuqin Rao and Yibin Ying have used the line-scanned method based image description to increase the processing speed in determining the area and contour [20]. The image data are compressed with the line-scanned based digit image description during processing, so the rate is largely improved. Zhou et.al have developed a machine vision system for potato inspection. The system can process 50 images of potato per second to determine the weight, diameter, and shape with the accuracy over 80 percent

[21].

In the potato industry, specifications have been set for the size, shape, and quality. The total of unqualified potato samples must fewer than 5% by weight in industry. Some inspection methods can only sort the potato by its appearance, but not the inner defects or diseases. Muir et.al. found that the spectra will be different in defect potatoes. The relationship between optical spectral reflectance and potato tuber were studied. And the spectral performance of potato will be affected by weather conditions such as temperature and humidity [22].

The sugar contents of fruits and vegetables are usually determined by destructive methods such as Brix meter. The Brix meter is widely used in the food industry to determine the approximate sugar amounts within the product, especially in liquid food products including juices, jams, and wine. The Brix value is the comparison of the reflective index of the measuring liquid and the pure water. When sugar dissolves in the liquid, the reflective index of the liquid changes. To determine the sugar contents in the potato using Brix meter, the liquid of the potato is put on the measuring prism of the Brix meter. The refraction of light that go through the liquid and the sugar dissolved in it can show the Brix value of the liquid. The advantages of the Brix meter are portable and simple to use. However, Brix meter cannot be applied to the large quantities of potatoes because it will be time consuming and destructive.

Near infrared (NIR) spectroscopy has been a popular method for the quality control of agricultural products. The researches of NIR used in the internal quality inspection of various fruits and vegetables including apples [23-24], guava [25], and potato [26]. Take fruit quality inspection for example, the light is applied on the fruit as it passes the sorter. As the near infrared light (700-2500nm) travels through an object, it will be reflected, transmitted, or absorbed by the object. Part of the absorbed energy may transfer to different forms of radiation. The radiation leaves the fruit surface depends on the fruit's properties and the incident lights and may include emission, transmittance, and reflectance. The properties of the fruit such as carbohydrate, PH value, and moisture contents can hence be determined by the transmittance and absorption [27-29]. The sugar contents in the potato can also be known using this method. Because the bonds of organic molecules such as sugar change their vibration energy when irradiating by NIR, it will show the absorption peaks on the spectrum. The high brix potatoes can absorb more light than those low brix ones at certain wavelength.

For sugar contents determination of potato, the NIR spectroscopy can predict the sugar level with high accuracy. But for the low sugar level potatoes, the model to predict the sugar contents should be modified to have better results. Although the NIR spectroscopy is a fast and cost-effective technique, the analysis is limited to few compounds of interest, and the effective detection depth is not deep enough to inspect the whole fruit. The comparison of the sugar contents detection techniques of fruits and vegetables are in the table 3.

Table 3. Comparisons of the different sugar contents detection techniques.

	Brix Meter	NIR	IR
Advantage	portable, cost-effective	nondestructive, quantitive information of the major component, fast response	nondestructive, larger detection depth, safe, easy to operate
Disadvantage	destructive, not efficient when applying to large quantities	only few compound of interest, smaller detection depth, requiring calibration	must apply heat to the agricultural products
Applicaiton	sugar contents detection in produce	agricultural products characterization, medical application	electronics, material defect detection

2.5 Summary of the literature review

The quality of potato products such as chips and french fries is largely dependent on the its sugar contents in the potato. To have high quality chips and fries, the sugar contents should not exceed the standard range. The extra sugar contents of the potato can also indicate the defect or wounding on the potato. Hence to identify the extra sugar contents of the potato is a significant issue for both the chip manufacturer and growers.

The current sugar contents detection methods are Brix meter and near infrared spectroscopy. The Brix meter can identify the sugar contents accurately. However, it is a destructive method and not efficient when applied in large quantities. The near infrared is the popular non-destructive method for the food quality inspection. Nevertheless, the penetration depth of the near infrared of the potato is about 5 mm only [30]. The extra sugar contents exceeded 5 mm is difficult to be identified with the near infrared spectroscopy.

The infrared thermography has been a widely used tool in characterizing the defects of the materials. The advantages of the infrared thermography include non-contact, safe, and results are easy to interpret. Hence it has great potential in the sugar contents identification. With the appropriate modeling and data analyzing methodologies, identifying the extra sugar contents with infrared thermography yields to have high accuracy.

CHAPTER III

METHODOLOGY

3.1 Introduction

This chapter gives the detail description of the materials and the temperature measuring system used in this research. The features of the thermography camera and ideas to construct the automated positioning system are also presented in this chapter. The specific chamber design and the reasons to use some specific components in this research are also discussed.

In this research, the preliminary experiments should be done first to determine the parameters used in the following experiments. To obtain a better result and be efficient at the same time, the potato heating time and the sugar accumulation simulation need to be considered as well. After determining the standard experimental procedures, experiments can examine the potato quality. In this study we will establish a procedure that can effectively inspect potato quality.

3.2 Materials

The potato used in the whole experiments is the Russet potato purchased from local supermarket. The Russet potato has brown skin and white pulp, and usually been used for baking, mashing, and making fries. Detail nutrient information of a medium Russet potato is listed in the table 4. From the table, the total sugar contents of the potato are about 1 gram and compose from sucrose, glucose, and fructose. The Russet potato is used within 7 days form purchase to maintain its freshness. The potatoes are stored in controlled room temperature at 70^{o}F and sealed with plastic bags. Because the potatoes were purchased for a large quantity at a time, the desired weight and volume can be not controlled during the experiments. The weight and volume of the potato samples are form 130 to 250 grams and 95 to 185 cm^3 in this research.

The properties of the potato are assumed to be uniform. Examples of the properties include the emissivity of the surface, the density, and the thermal conductivity of the potato.

Table 4. Nutrition information of the russet potato [31].

Russet potato, serving size: 170g	
calories	134
Total Fat	0 g
sodium	9 mg
Total Carbohydrate	30.7 g
Dietary Fiber	2.2 g
Sugars	1 g
Sucrose	221 mg
Glucose	425mg
Fructose	391mg
Protein	4 g
Water	134 g

3.3 Experimental set-up

The section describes the thermal image capturing and analysis system, and operators.

The system has four principal parts: infrared camera, working chamber, control systems, and image acquisition and analysis software.

3.3.1 Infrared camera

A Compix infrared camera model PC2100, including the infrared sensor, connecting cables and software, was used to analyze the thermal images. The resolution of this product is 244 x 193 pixels. The operation temperature range is from 63 to 300 degree Fahrenheit, and the measuring temperature range is from 68 to 95 degree Fahrenheit. The infrared camera can detect small temperature changes (0.4 ^{o}F) of an object's surface. Figure 1 shows the camera and figure 2 shows the experimental setup in this study.

Figure 1. The infrared camera used in this study.

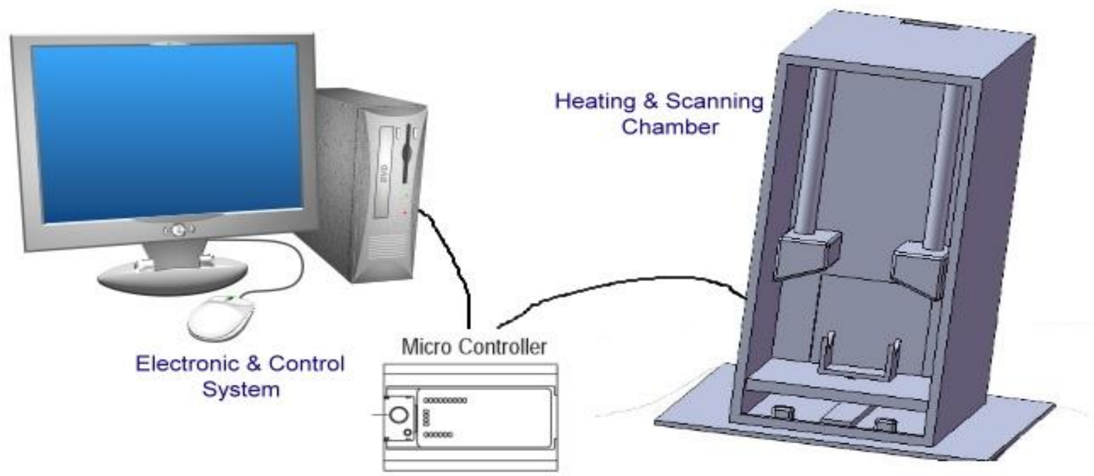

Figure 2. The experimental setup of the research.

The scanning area of the camera depends on the distance between the camera and the object. The field of view will be larger when the distance between the camera and the object is longer. To install the infrared camera, the distance between the camera and the object needs to be adjusted according to the size of the object. When the distance is set, the focus of the camera can also be fixed to have better image quality and accurate temperature readings.

When in operation, the infrared camera is a very sensitive device. Many variables might affect the accuracy of the temperature readings. Since our goal is to compare the temperature distribution of the object at different times, the operation conditions should be maintained as uniform as possible. The examples of the operation conditions are ambient temperatures, the heat sources, and the physical set-up of the system. To reduce the external infrared source and vibration during the experiment are also important. The external infrared source can affect the accurate surface temperature reading while vibration can have adverse effects of the resolution of infrared camera.

3.3.2 Experiment chamber design and layout

The experiment chamber includes the heating chamber, scanning chamber, potato transfer system, and potato fixture. To provide a stable environment for experimentation, the chamber must have special designs to meet the different requirements for both heating chamber and scanning chamber. The material to build the whole chamber is Acrylic, which has relative low thermal conductivity and easy to process. The dimension of the whole chamber is 24 x 34 x 21 inches. Figure 3 shows the chamber design of the experiment.

The heating chamber locates at the front section of the whole chamber. Two 150W halogen lamps are mounted to provide the heat source for the potato. The distance between the lamps and the potato is about two inches, depends on the height of the potato. To prevent the heat source from dissipating form the heating chamber to the scanning chamber, all of the chamber walls are cover with Styrofoam and Mylar. The thermal conductivity of Styrofoam is very low so it can prevent most of the heat transfer from conduction. Mylar is a biaxially-oriented polyethylene terephthalate film used for its chemical stability, reflectivity, and electrical insulation capabilities. The Mylar used in the chamber is metalized and can reflect more than 90% of light including much of the infrared spectrum. When the heating time is 60 seconds, the temperature from the scanning chamber only increases 0.1^{o}F. This demonstrates that the thermal insulation with Styrofoam and Mylar is effective.

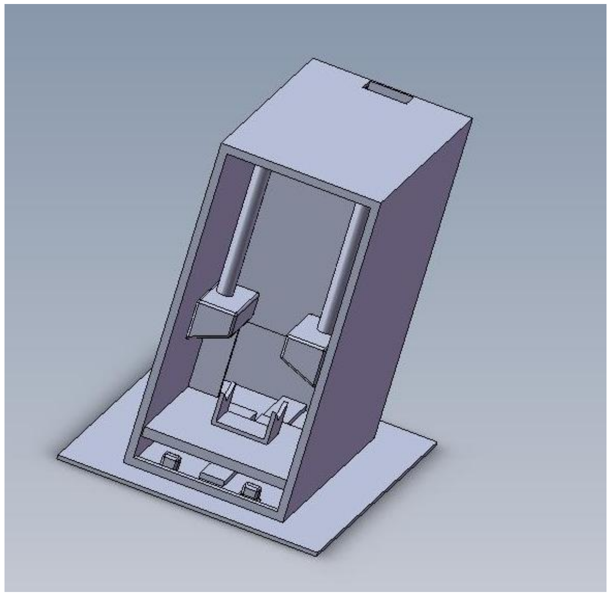

Figure 3. The chamber design of the experiment.

The infrared camera is installed in the scanning chamber to take the thermal images of the potato. The ambient temperature of the scanning chamber needs to be controlled well to eliminate the noise due to ambient temperature increasing. Three Resistive Temperature Detector (RTD) sensors are placed in the scanning chamber to monitor the ambient temperature inside the chamber.

In the scanning chamber, the inner wall is covered with Styrofoam because the good thermal insulation is also needed. The Styrofoam is painted into black color instead of Mylar to reduce the reflection energies which is a potential noise when taking thermal images. The fixture of the potato is used to hold the potato steadily for the whole experiment. The potato is in the same position when in heating and scanning process.

The material of the fixture is also Acrylics cover with Styrofoam and Mylar for their low thermal conductivity. To reduce the heat transfer from the potato to the fixture, the contact between the potato and the fixture is minimized as four points with the double V shape as shown in figure 4. Some holes are drilled on the fixture structure to reduce the heat conduction.

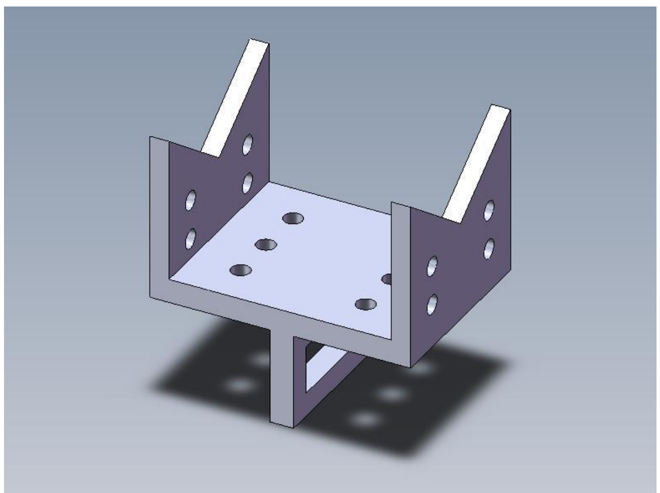

Figure 4. The fixture design to hold the potato during experiment.

After the potato is heated in the heating chamber, it will be moved to the scanning chamber immediately. The potato transfer system is constructed by a linear guide way and motors to control the motion. The fixture of the potato is mounted to the guide way directly. Two gates are placed in between the heating chamber and the scanning chamber. These two gates are closed when heating the potato. After heating, the potato is transferred to the fixed position in the scanning chamber, and the motors connected with strings will pull the gates closed.

3.3.3 Electronic and control systems

All of the motions are controlled by a programmable logic controller (PLC). The PLC programming can be modified to adjust the experiment settings such as the heating time and the gate closing times.

The PLC used in this research is MicroLogix 1000 from Allen-Bradley company. PLC plays an important role in the automated control system. With the program stored in it, PLC can monitor the status of the system. Input devices give signals to the PLC, and the PLC sends a signal to the output devices for system operation. Examples of the input devices and output devices are motors and sensors. The design concept of the control system for this experiment is in figure 5.

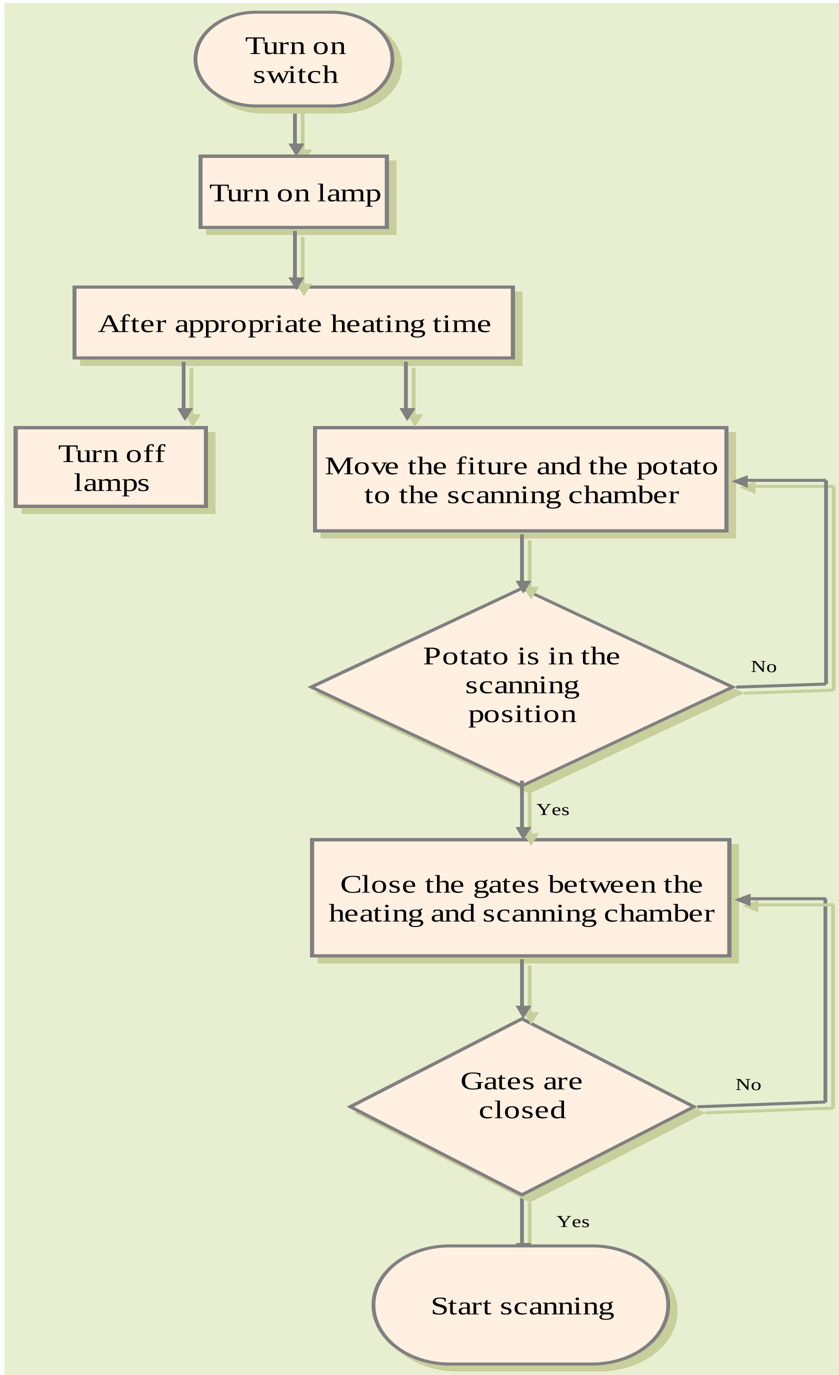

Figure 5. Design concepts of the PLC program.

3.3.4 Image processing and analysis software

To display the thermal images on the screen and to process the data are essential functions in the application of infrared thermogarphy. The softwares to analyze the thermal images are WinTES and ThermalView in this study.

The Windows Thermal Evaluation Software (WinTES) is software to control the infrared camera and show the temperature profile of the object. The variables including ambient temperature, focus, emissivity of the object, and scanning interval are set before scanning. In WinTES, the thermal images are display as gray scale as shown in figure 6. WinTES can show the real time thermal images while scanning, and save each image to the personal computer. It can also show the maximum, minimum, and average temperatures of the chosen area of the thermal image. By viewing the thermal images, we can have the basic understanding of the temperature distribution of the potato. The shape and the contour of the potato can be clearly shown in the thermal images. However, to record every temperature reading of the thermal images needs a lot of time. There are also difficulties to choose the same area of interest for every thermal image.

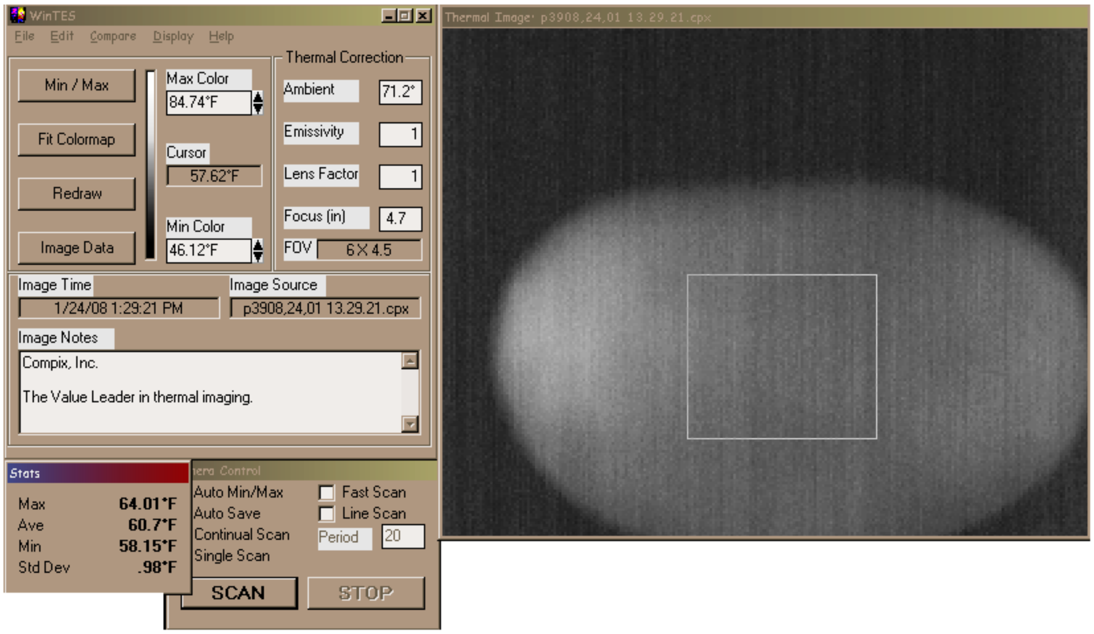

Figure 6. The WinTES program to display the thermal images.

To simplify the analyzing procedure, the software Thermal View is used in this research. The basic function of Thermal View includes showing the thermal images, drawing and saving the area interested on the thermal image, getting the temperature data from the saved area, and calculating the heating rate or cooling rate of the object. The area of interest on the potato is saved for each sample to maintain the consistency. For each area of interest, it can be divided into many columns and rows depends on the resolution needed. To know the thermal behavior of potato, the heating rate and average heating rate are calculated instead of surface temperatures [32]. The heating rate over time is calculated based on the temperature difference between each the thermal images for the area of interest. The unit of heating rate is $\frac{^{o}\mathrm{F}}{\Delta t}$, where ^{o}F and t represent degree Fahrenheit and time respectively. The average heating rate is defined as the summation of heating rates divide by the number of thermal images captured. The equations of heating rate and average heating rate are:

Heating Rate $H = \sum_{i=1}^{n} \frac{T_i - T_{i-1}}{n}$ (3.1)

Average Heating Rate $H_{avg} = \sum_{i=1}^{n} \frac{H_i}{n}$ (3.2)

where i is the number of the current thermal image, ranges from 1 to n, and n is the number of images captured for each experiment. The advantage to use heating rate and average heating rate is the signal can be easy to interpret. The noise due to the ambient temperature fluctuation can be eliminated as well.

To obtain the accurate results, the environmental data is needed when taking the images. Without this data, the results might be misinterpreted due to the large changes of the environment. The example of the environmental data includes ambient temperature and distances form camera to the object.

3.4 Design of experiments

3.4.1 Preliminary experiments

3.4.1.1 Determine the appropriate method to simulate the extra sugar contents area

Extra sugar contents in the potato is a serious problem in the potato industry. However, there are difficulties related to obtaining the potato with extra sugar contents before analyzing it. In this experiment, the appropriate method to simulate the extra sugar contents will be decided.

Sucrose accumulation in the potato is a serious problem in the potato industry. Sucrose can also transform into fructose and glucose that can cause dark coloration after frying the potato. In this experiment, we use cane sugar as the material to simulate extra sugar contents accumulation. In this research, we assume the excessive sugar contents is concentrated in a small region.

The sugar accumulation areas inside the potato need to be appropriately simulated in order to meet the real situation. The first method is to cut a corner of the potato and soak it in the sugar solution. After soaking, the potato corner will be put back in the original position. However, the sugar solution will not diffuse into the potato. The potato corner shrinks because the water contents balances the osmotic pressure.

Since over 80% of the inside of the potato is water, the sugar solution will be used directly to simulate the sugar accumulation area inside the potato as possible. The simulation occurs by injecting the sugar solution inside the potato tuber after drilling a hole at specific locations. The location of the sugar accumulation is set as 45 degrees from the centerline as shown in figure 7. After marking the center of the potato and measuring the depth of the potato, it is easy to drill a hole of 45 degrees. The size of the drill bit is 1/8 inch. The volume of the sugar solution can be calculated as $\pi(\frac{1}{16})^2 \text{ x } \sqrt{2} \text{ x } \frac{1}{2}$ potato height. This configuration can let us study the depth and area of sugar concentration by analyzing the thermal images taking from the top of the potato. The disadvantage is that the sugar solution will not be inside the potato; instead, it will extend to the potato's surface. The sugar solution area on the thermal image will be 0.012 inch^2.

The concentration of the sugar solution is set as 37.5% (20 ml water with 12 grams of cane sugar). When injecting the sugar solution into the potato, we should make sure that there is no air in the solution. Because the thermal properties of air are greatly different from the potato and the sugar solution, the appearance of air will affect the result.

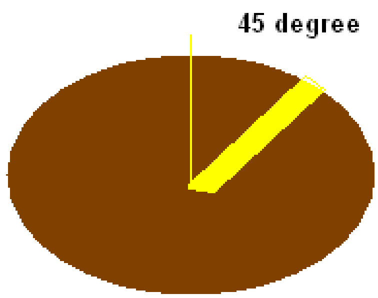

Figure 7. The sugar solution injection position.

3.4.1.2 Determine the appropriate heating and cooling time

Different heating times for the potato are applied in this experiment with the best time to be selected. The appropriate heating time should not be too short so that the heat cannot propagate to the center of the potato, and it should not be too long, or it may affect the potato quality. In real application, the shorter heating time is preferred considering the efficiency. The infrared camera scanning time must be long enough to record the surface temperature change of the potato. However, it is not efficient to keep scanning until the surface temperature of the potato decreases to the original temperature. So the appropriate heating and scanning time should be determined in this preliminary experiment.

Considering 30, 60, and 90 seconds of heating time, the temperature is measured from the RTD sensors first. After heating, the potato is put in room temperature for 300 seconds, and the temperature changes of the potato's subsurface are recorded to determine the proper heating and cooling time.

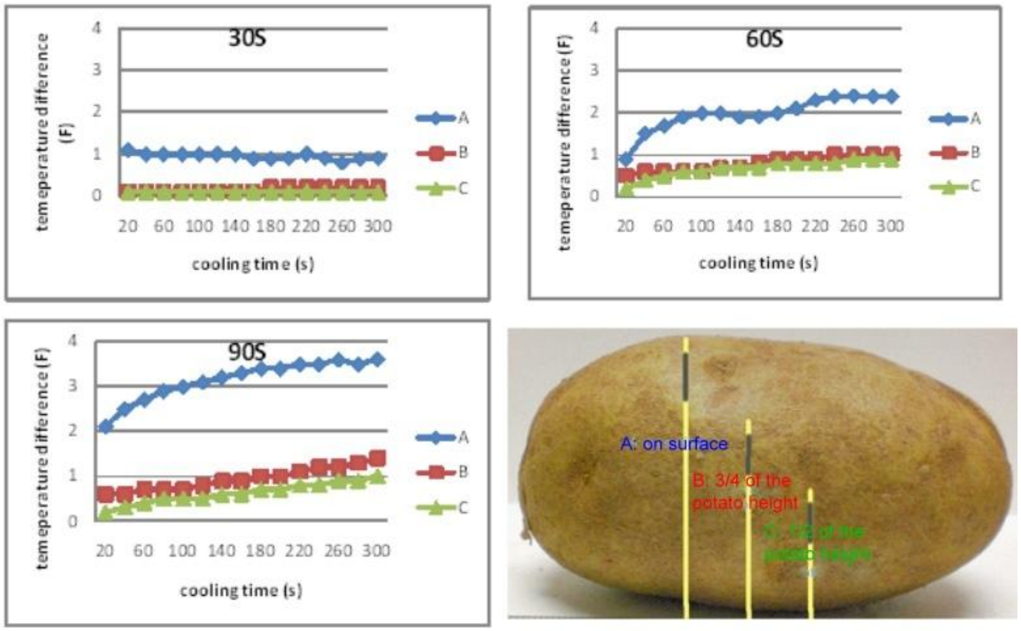

Figure 8. Temperature difference of 30, 60, and 90 seconds heating time and the RTD sensor positions.

From the above figures, we can see that as the cooling time increases, the temperature also increases. The reason for this may be that the heat propagates from the potato surface into the inner potato and dissipates from the inner potato to the surface when the potato's temperature is higher than the ambient temperature. The figure 8 shows that when heating for 30 seconds, the heat propagates to the middle of the potato is very small, and the signal is not large enough to discriminate the potato quality. For heating 60 and 90 seconds, the trends are similar except that the curves in for heating for 90 seconds are smoother than for 60 seconds. In this preliminary experiment, we can exclude the 30 seconds of heating time in order to have more apparent results. Each data point is an average of 3 samples.

In another experiment to determine the appropriate heating time, the RTD sensor is used to measure the center temperature of the potato. A hole is drilled in the potato with the depth half of the potato height. And the RTD sensors are inserted into the hole to measure the center temperature of the potato. The center temperature of the potato is recorded from the time the potato is heated until it returns to the initial value. This experiment's intent is to determine the cooling time needed between the first experiment and the second experiment of the same potato with sugar. Here, the ambient temperature is assumed to be constant. The time needed to cool for the 30, 60, and 90 seconds of heating are 140, 160, and 180 minutes, respectively.

In this experiment, the heating time of 60 seconds is applied. Extending the heating time might affect the quality of the potato, even if it has a larger signal in thermal images. After taking the thermal images, the potato is placed in room temperature for at least 3 hours to reach its original temperature. Then the potato is injected with the sugar solution immediately before taking the thermal images for the second time. If the shorter heating time is needed, the lamps can be replaced with higher watts ones. But the watts of the lamps should not be too high that might affect the properties of the potato.

3.4.1.3 Determine the appropriate scanning time

In the grayscale thermal images, the darker colors indicate lower temperatures. We can observe and compare the temperature of each thermal image taken to determine the appropriate scanning time. The scanning time must be long enough to record the heat transfer behavior of the potato. However, it is not efficient if the scanning time is too long for real application. The surface temperature of the potato after heating decreases until it equals the original temperature. From the thermal images of the potato before heating, the surface temperature of the potato is lower than the background. Figure 9 shows the thermal images taken every 60 seconds beginning from A. In the figure 9 (F), after cooling for 300 seconds, the thermal image becomes very blurry. Figure 10 shows the average heating rates within 300 seconds. The slope is very small in the end of the curve, which means the average heating rate does not change greatly after scanning for 300 seconds. The slope of the curve will be even smaller after scanning for 300 seconds. So in this research, the scanning for 300 seconds is enough for the heat transfer analysis. The thermal images are taken every 20 seconds for a total of 300 seconds. For each potato there will be 16 images in each experiment, including the first snapshot taken after heating immediately.

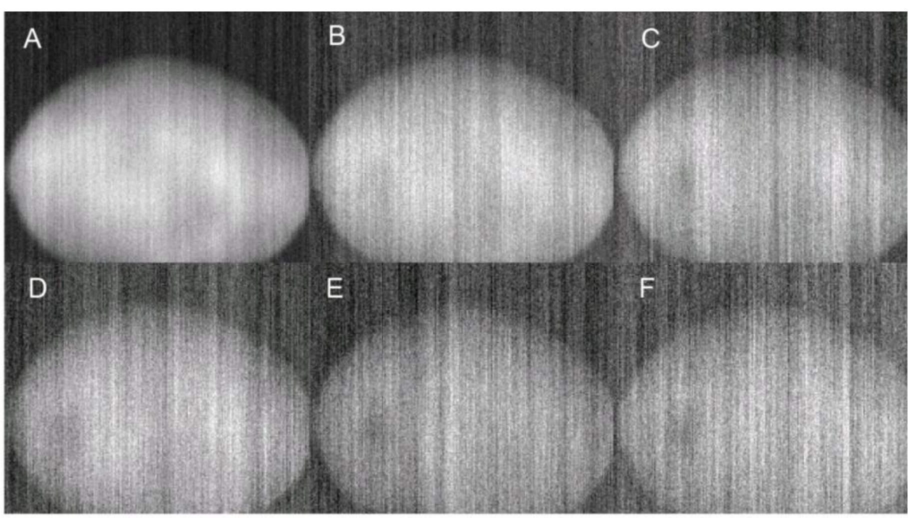

Figure 9. Thermal images of the potato taken every 60 seconds.

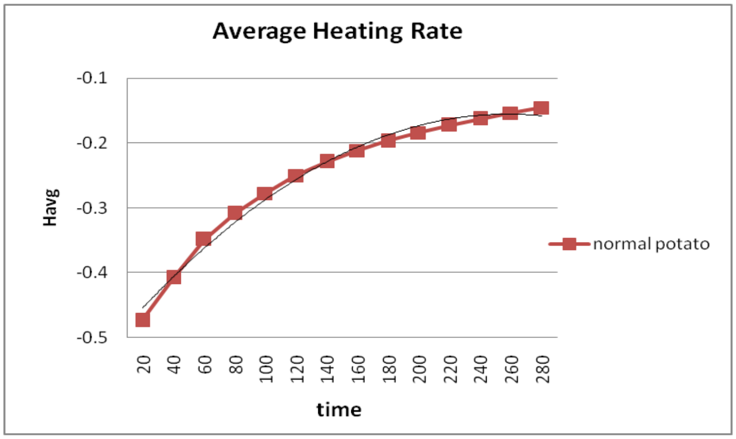

Figure 10. The average heating rates within 300 seconds.

3.4.1.4 *Effect of ambient temperature change*

The ambient temperature can vary greatly during the whole experiment hence affect the accuracy of the result. In this research, the ambient temperature is controlled between 70 and 74 degree Fahrenheit. To make sure the ambient within this range will not affect the result significantly, the summation of the average heating rate of the same potato is examined at different temperature levels. Three samples are selected and take the average values for the statistical test. H_{avg} values of the same potato obtained from different ambient temperature are shown in table 5.

Table 5. H_{avg} values of the same potato sample obtained from different ambient temperature.

Temperature Range (F)	70-71	71-72	72-73	73-74
Sample Number	3	3	3	3
Havg	-4.406	-4.303	-4.454	-4.443
Standard deviation	0.124	0.140	0.156	0.161

The procedures to test the means of these four groups are:

1. Ho: $\overline{x_1} = \overline{x_2} = \overline{x_3} = \overline{x_4}$

   Ha: At least one of the sample means differ from the rest.

   α =0.05

2. Total sample number $n_T = 12$, $n_1 = n_2 = n_3 = n_4 = 3$

   The overall mean = (3(-4.406)+3(-4.303)+3(-4.454)+3(-4.443))/12

   = - 4.402

   Sum of squares between samples =

   $3(-4.406+4.402)^2+3(-4.303+4.402)^2+3(-4.454+4.402)^2 +3(-4.443+4.402)^2 =0.043$

   Sum of squares within samples =

   (3-1)(0.124)+(3-1)(0.140)+(3-1)(0.156)+(3-1)(0.161) = 1.162

   Total sum of squares = 0.043+1.162=1.205

3. The AOV table can be generated in table 6.

Table 6. The ANOVA table of the experiments done with different ambient temperatures.

Souce	SS	df	Mean Square	F
Between samples	0.043	3	0.0143	0.0143/0.145 = 0.984
Within samples	1.162	8	0.145	
Totals	3.434	11		

The critical value of F is obtained from the percentage points of the F distribution table. For α=0.05, df_1=3, and df_2=8, the value is 4.07.

4. Because the computed F value is 0.984, which is smaller than the critical value, we fail to reject the null hypothesis of equality of the means of the H_{avg} obtained from different ambient temperatures. So the ambient temperature varies in this research does not affect the Havg significantly.

3.4.2 Experiments to identify if there is extra sugar contents in the potato

After the standard experimental procedures are developed, potato samples are selected to compare the thermal behavior with and without the extra sugar contents within the same potato. The heating rates and average heating rates of potatoes with and without the extra sugar contents will be discussed. The prediction of the extra sugar contents in the potato is accomplished by using artificial neural networks and the statistic model. The correct rates of the prediction by these two methods will be discussed.

To determine whether the potato has extra sugar contents, 60 potato samples are selected to compare the thermal behavior before and after injecting sugar solution. Before doing the experiment, the weight and dimension of the potato is recorded. The experiment procedures are as follows:

1. Place the potato on the fixture, mark the position of the potato, and fully close the chamber to ensure stable environmental conditions.
2. Start heating the potato for 60 seconds by 2 halogen lamps. After heating, the potato will move automatically to a predetermined position for taking the thermal images. All of the gates are closed to prevent the heat transfer from the heating chamber to the scanning chamber.

3. Take one thermal image every 20 seconds, for a total of total of 300 seconds per potato.
4. Cool the potato at room temperature for 3 hours. Inject a sugar solution into the potato to simulate a area of sugar concentration. After injection, scan the potato again following steps 1 to 3 above.
5. Analyze and compare the thermal images of the potato with and without excessive sugar contents.

3.4.3 Experiments to characterize the extra sugar contents locations

The experimental procedures are similar to the procedures in 4.3. The analysis procedures are divided into two steps. First, the average heating rate differences of different areas on the same potato are compared. One of the selected areas on the potato is the predetermined sugar injection region, and the other is the region far from the sugar injection region. Because the properties of each potato are not identical and the properties of the same potato are not uniform, the average heating rates of the same potato needs to be compared first. The average heating rate difference is calculated by subtracting the average heating rate of the potato with extra sugar contents from the average heating rate of the normal potato, illustrated as follows:

$$\Delta Havg = Havg_{,\ normal} - Havg,_{\ extra\ sugar\ contents} \quad (3.3)$$

After studying the average heating rate difference of the potato with and without the extra sugar contents, the next step is to find the location of the sugar contents without comparing the original data. The whole thermal image of the potato is divided into small squares (20 x 20 pixels), and the average heating rates of these small squares are calculated. The area of interest on the potato is set to the center of the potato as shown in figure 11 to eliminate the noise due to the edges' uneven cooling rates. The length and width are 160 and 100 pixels, so there will be a total of 40 small squares on the area of interest. The average heating rate of each small area is plotted into two dimensional figures. Statistical models will be applied in this procedure and the correct rate of identification of the extra sugar contents location is discussed.

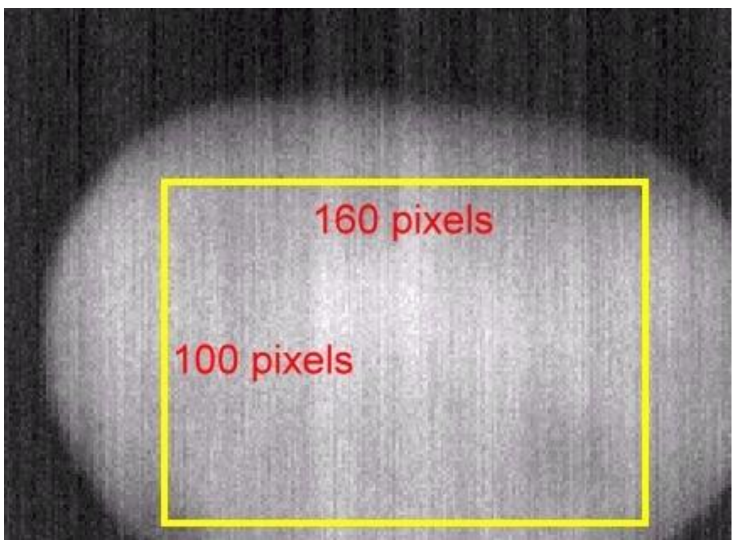

Figure 11. The whole area of interest of the extra sugar contents location prediction.

3.4.4 Experiments to characterize the extra sugar contents depth

When the location of the extra sugar contents is identified, the next goal is to predict the depth of the extra sugar content. In this experiment, the average heating rate difference of different depths of the extra sugar content will be compared. For the extra sugar content potatoes, the relationship between the average heating rate and the depth of the extra sugar content will be predicted using a statistical model. The correct rate of the depth of the extra sugar content prediction will be discussed. Table 7 summarizes the experiments done in this research.

Table 7. Summarized of the experiments needed to be done.

Experiment	Purpose
Sugar contents simulation	To determine the appropriate sugar simulation method
Heating time	To determine the appropriate heating time for potato
Scanning time	To determine the appropriate scanning time for potato
Ambient temperature effect	To see how different ambient temperature may affect the surface temperature distribution of potato
Extra sugar contents identification	Use different analysis approaches to identity if the extra sugar contents is in the potato
Extra sugar contents location identification	Use different analysis approaches to identify the extra sugar content location
Extra sugar content depth identification	Use different analysis approaches to identify the extra sugar content depth

CHAPTER IV

RESULTS

4.1 Introduction

This chapter presents the results of the extra sugar contents detection, the location of the sugar concentration prediction, and the depth of the sugar concentration prediction. For the extra sugar contents detection, a statistical model and Neural Network model are applied. For the location and depth of the extra sugar concentration prediction, statistical model is applied. The correct rate of each experiment is discussed in this chapter. The general methodologies for characterizing the extra sugar contents potato and identifying the location and depth of the extra sugar contents are proposed.

4.2 Experimental results of the extra sugar contents detection

4.2.1 Comparison of the heating rate (H) and average heating rate (H_{avg}) of potato with and without the extra sugar contents

The first objective is to determine if there is an average heating rate difference between a normal potato and a potato with extra sugar contents. Sixty potato samples are used in this experiment. The specific potato sample is scanned for the first time without adding any sugar solution. After the appropriate cooling time, the same potato is injected with the sugar solution for the second scan. The scanned surface of the potato is maintained in both of the scanning processes. The summation of heating rate (H) and average heating rate (H_{avg}) of the 60 potato samples are listed in Appendix A. The values are calculated from the summation of the heating rate and average heating rate of each snapshot for each potato sample. The area of interest on the thermal image is the largest available area of the potato. Since the shape of the potato is approximately an ellipse viewed from the top; and the area of interest is a rectangle, the edges will not be included in the heating rate analysis.

As Appendix A demonstrates, most the H and H_{avg} of normal potatoes appear higher than those with extra sugar contents. We can then conclude that a potato with extra sugar contents cools faster than a normal one. The average value of H and H_{avg} of each snapshot for each potato is listed in table 8. Figure 12 and 13 are generated from these values for comparing H and H_{avg} easily. The heating rates and average heating rates decrease with time. The curves of H and Havg fit well into second-order polynomial equations, which means the H and Havg decrease faster in the preceding snapshots. In addition, the average values of H overlap except in the first few snapshots, while the average values of H_{avg} always separate. Although the relationships of the H and H_{avg} between potato with and without extra sugar contents can be seen from the figures, conclusions cannot be made before examining the statistical procedures. Figure 14 shows the boxplots of the H_{avg} for each snapshot.

Table 8. The average value of H and H_{avg} of each snapshot for each potato.

	H of normal potatoes	H of extra sugar contents potates	H_{avg} of normal potatoes	H_{avg} of extra sugar contents potatoes
snapshot 1	-0.47133	-0.53043	-0.47285	-0.53039
snapshot 2	-0.32791	-0.36105	-0.40134	-0.44554
snapshot 3	-0.23796	-0.2371	-0.34638	-0.37588

Table 8. Continued.

	H of normal potatoes	H of extra sugar contents potates	H_{avg} of normal potatoes	H_{avg} of extra sugar contents potatoes
snapshot 4	-0.18668	-0.18704	-0.30634	-0.32865
snapshot 5	-0.14648	-0.15014	-0.2743	-0.29335
snapshot 6	-0.12019	-0.12927	-0.24874	-0.26572
snapshot 7	-0.09582	-0.1059	-0.22699	-0.24298
snapshot 8	-0.09513	-0.09289	-0.21046	-0.22422
snapshot 9	-0.07722	-0.08526	-0.19571	-0.20873
snapshot 10	-0.07171	-0.06578	-0.18314	-0.19437
snapshot 11	-0.0532	-0.06818	-0.17157	-0.18302
snapshot 12	-0.04632	-0.05913	-0.16105	-0.17272
snapshot 13	-0.05793	-0.0502	-0.15299	-0.16328
snapshot 14	-0.0448	-0.03387	-0.14529	-0.15404

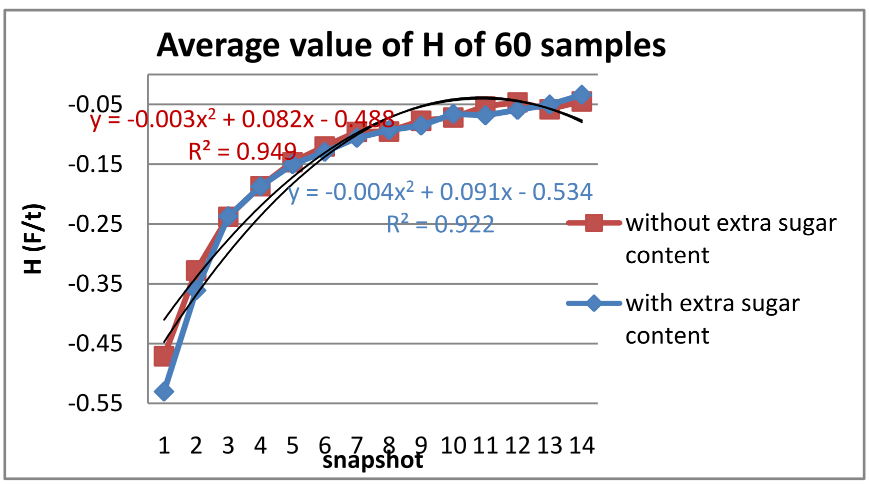

Figure 12. The average value of heating rate of 60 samples.

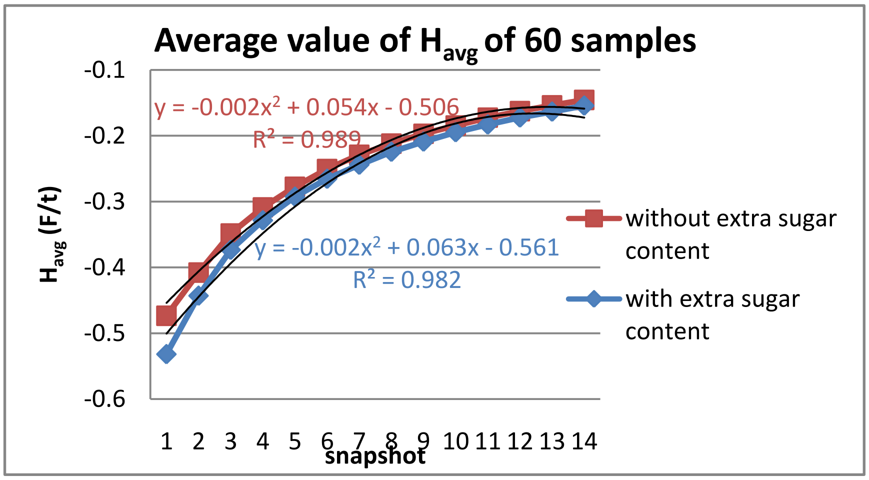

Figure 13. The average value of average heating rate of 60 samples.

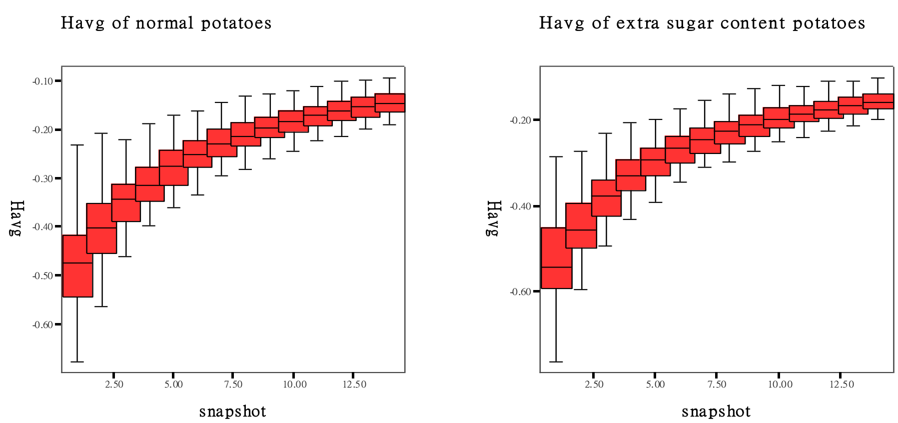

Figure 14. Boxplots of H_{avg} of normal and extra sugar contents potatoes.

Table 9. Comparison of the H and H_{avg} of the normal and the extra sugar contents potatoes.

	H, normal	H, sugar	Difference of H	H_{avg}, normal	H_{avg}, sugar	Difference of H_{avg}
average	-2.033	-2.156	0.123	-3.497	-3.787	0.289
standard deviation	0.356	0.337	0.146	0.607	0.607	0.263

The paired t-test is used to check if there is a significant difference between the summations of average heating rates. The paired t-test can be used to compare two population means when each measurement in one sample is paired with a particular measurement in the other sample [33]. In our experiment, the data generated before and after the sugar injection are from the same potato for each test. Thus, the paired t-test was applied in this study with the data in table 9.

The procedures of paried t-test to compare the summation of the heating rates are:

1. H_o: $\mu_d = \mu_1 - \mu_2 \leqq 0$

   H_a: $\mu_d > 0$

   $\alpha = 0.05$

2. T.S: $t = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{0.123}{0.146/\sqrt{10}} = 2.66$

3. R.R: For degree of freedom (df)=n-1= , reject Ho if $t \geqq t_\alpha$

   From the table of the t distribution, $t_{\alpha,\,9} = 1.833$

4. Because $t > t_{\alpha,\,9}$, we reject the null hypothesis that the mean of the two group is the same. There is a difference of heating rates of potato with and without extra sugar contents.

The procedures of paired t-test to compare the summation of the average heating rates are:

1. H_o: $\mu d = \mu 1 - \mu 2 \leqq 0$

   H_a: $\mu d > 0$

   $\alpha = 0.05$

2. T.S: $t = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{0.289}{0.263/\sqrt{10}} = 3.475$

3. R.R: For degree of freedom (df)=n-1= , reject Ho if $t \geqq t_{\alpha}$

   From the table of the t distribution, $t_{\alpha, 9} = 1.833$

4. Because $t > t_{\alpha, 9}$, we reject the null hypothesis that the mean of the two group is the same. There is a difference of average heating rates of potato with and without extra sugar contents.

From the hypothesis testing, we conclude that there is difference of H and H_{avg} between potato with and without extra sugar contents. However, the difference is larger for the H_{avg}, so the following discussions focus on the H_{avg} exclusively.

4.2.2 Introduction of the logistic regression model

After examining the difference of H and H_{avg}, the next step is to use the statistical model to predict the probability of the appearance of extra sugar contents in the potato. Because the outcome of the prediction of extra sugar contents appearance is simply “yes” or “no”, the logistic regression model is applied here.

Logistic regression is a model used to predict the probability of an event occurrences, especially when the dependent variable is binary. The independent variables can be continuous or categorical. For binary outcomes, if the independent variables are continuous, the ordinary linear regression model might not fit the data very well [34]. To largely improve the fit, using the logistic model is possible, as shown in figure 15. The outcomes are set to 0 and 1in the logistic regression model, and the mean of the binary distribution is the proportion of 1s, as shown in figure 16.

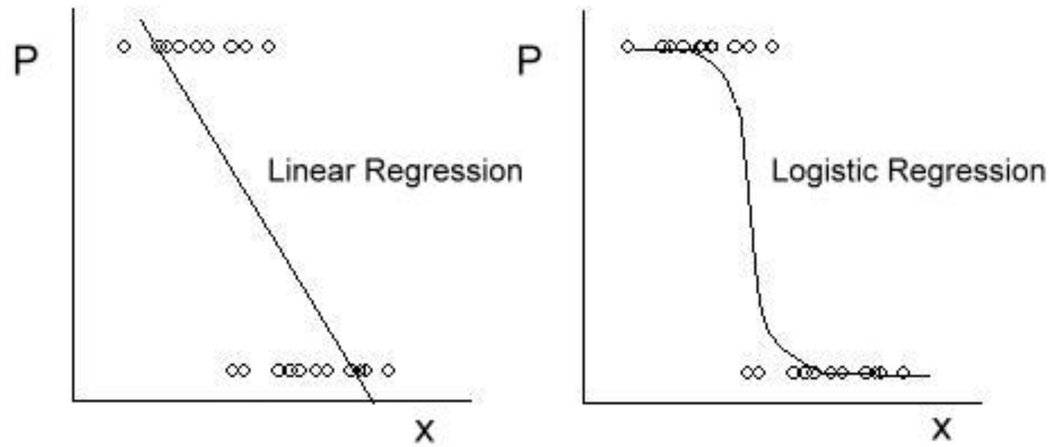

Figure 15. The comparison of fitting in linear and logistic regression model.

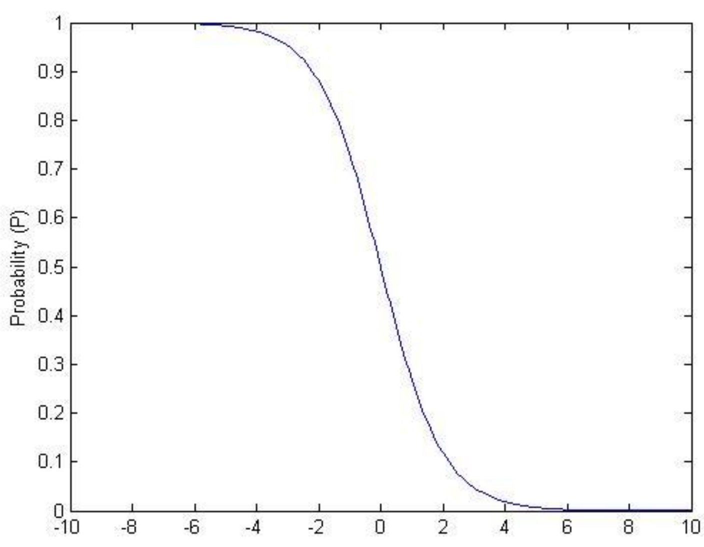

Figure 16. The general figure of the logistic function.

The logistic equation is in the following form:

$$P = \frac{e^x}{1+e^x} \tag{4.1}$$

By this equation, the estimate value of P can only be in the range of 0 to 1for any input x, as shown in the figure x. Here P represents the probability of the outcome. The variable x is the factor affecting the outcome, can be in the form

$$x = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{4.2}$$

In equation (4.2), β_0 is the intercept which will be P when x equals zero, and $\beta_1$to β_n are the regression coefficients of x_1 to x_n, respectively. Each of the regression coefficients can show the influence of the specific factor.

From equation

$$\frac{P}{1\text{-}P} = \frac{e^x/(1+e^x)}{1/(1+e^x)} = e^x \tag{4.3}$$

In statistics, odds is defined as the likelihood of a given event occurring compared to the likelihood of the same event not occurring, which can be represented as $\frac{P}{1-P}$. The logit of P is the natural logarithm of the odds, which can be obtained from

$$\ln\left(\frac{P}{1-P}\right) = x \tag{4.4}$$

Hence, to model P with the logistic function is equivalent to fitting the logit into the linear regression model. The logistic regression model is to calculate changes in the natural logarithm of odds of the dependents, not changes of the dependents as in the ordinary linear regression models. Thus, the relationship between x and P is nonlinear, and the regression coefficient does not have direct interpretation as in the ordinary linear regression.

The assumptions of logistic regression model are not as restrictive as in the ordinary linear regression model. For example, the dependent variable does not need to have the same variance for each level of the independents. In addition, the logistic regression model does not require that the independents to be intervals or unbounded. However, when using logistic regression, the following basic assumptions should be met:

1. The model is specified correctly. The dichotomous dependent variable is assumed for binary logistic regression.
2. The logistic regression model should include all relevant variables and exclude all irrelevant variables. The independent variables should not be linear functions of the other independent variables, or the standard error of the regression coefficient will be too high.
3. No outliers. Outliers can affect the results substantially, hence should be removed from the model. The values exceed more than 2 standard deviation are considered as outliers.

4.2.3 Extra sugar contents identification results from a logistic regression model

The dependent variable in this model is set to be 0 or 1, which represents the normal or the potato with extra sugar contents, respectively. The possible independent variables are H_{avg}, density of the potato, weight of the potato, and the ambient temperature. The density and weight of these two data in the same potato will not change, while the H_{avg} and ambient temperature might vary. The H_{avg} of all potato samples are obtained after analyzing the thermal images. The summation of H_{avg} in the previous section is applied.

Other factors that might affect the outcome, such as density and weight are also consider as variables. Density can affect the thermal diffusivity of the potato and weight is the most convenient property to measure when sorting the potato. Although the ambient temperature is controlled within 4 degree of Fahrenheit when doing experiment, it is still considered as a potential factor that could affect the results.

The independent variables are selected from the likelihood ratio test of the model. The likelihood ratio is the probability of the observed results of the dependent be predicted form the independents, varies from 0 to 1. It is customary to use -2 times the log of the likelihood (-2LL) to test the significance of the logistic regression model because it is approximately a chi-square distribution. A good logistic regression model will have high likelihood and results in very low -2LL. Analogous to the chi-square test in the linear regression model, the likelihood ratio test is applied to test the null hypothesis that all the coefficient of the independent variables are zero. The test is done by comparing the difference in -2LL of the whole model with a nested model, which drop one of the independents.

A nonsignificant likelihood ratio test means there is no difference between the full and the nested model, hence the coefficient of the dropped independent variable can be considered as 0. The likelihood ratio test to select the appropriate independent variables can be done by the software SPSS as shown in the table 10 below. From the table we can see only the H_{avg} is significant with the 0.05 significance, which means we can reject the null hypothesis that the coefficient of H_{avg} is 0. However, models without the other independent variables are not significantly different from the full model. Hence the only variable to be considered in predicting the occurrence of extra sugar contents in the potato is H_{avg}.

Table 10. Statistical results focus on selecting the independent variables.

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	106.546	.201	1	.654
weight	106.349	.003	1	.954
amb_temp	106.505	.160	1	.689
density	106.926	.581	1	.446
Havg	110.432	4.087	1	.043

After deciding the variable to be used in the logistic regression model, the model is established and the accuracy of prediction is discussed. There are total sixty samples and the samples are separated into modeling and testing groups. Forty samples are randomly selected for the modeling group, there are two data sets in each sample, one form the normal potato and another from the potato with the sugar solution injection. The other twenty samples formed the testing group. For each sample in the testing group, data from normal and potato with extra sugar contents cannot be used at the same time since when an unknown potato is to inspected, there will not be data of the normal and the potato with extra sugar contents available simultaneously. As a result, there are 80 data sets and 20 data sets to be used in the modeling and testing group, respectively.

The logistic regression model is constructed by SPSS using the data from the modeling group. The raw data used to build the logistic regression model in the modeling group and the testing group appear in the Appendix B. The coefficient of the H_{avg} and the constant are listed in the following table 11:

Table 11. Coefficients of the logistic model in extra sugar contents prediction.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step	Havg	-.776	.414	3.511	1	.061	.460
1(a)	Constant	-2.814	1.522	3.417	1	.065	.060

The equation of the logistic regression model is now:

$$P = \frac{e^{-2.814-0.776*Havg}}{1+e^{-2.814-0.776*Havg}} \tag{4.5}$$

The next step requires using this model to fit the data of the testing group to determine the accuracy of classification. It is an accurate prediction if the predicted probability of the potato being an extra sugar one is higher than 0.5 and the inputs are from the potato with extra sugar contents, or the predicted probability of potato being an extra sugar one is lower than 0.5 and the inputs are from the normal potato. The mathematical interpretation to calculate the correct rate of the prediction is list as follows:

Actual value A(x):

$A(x) = 0$ potato without extra sugar contents

1 potato with extra sugar contents

Predicted value: P(x): the probability of having the extra sugar contents

$x=1\sim m$ m: total sample number

If A(x) =0 and P(x)<0.5 , C(x)=1

If A(x) =1 and P(x)≧0.5, C(x)=1

Else C(x)=0

Correct rate= (∑ C(x)) / m

For the twenty data in the testing group, the classification table of the correct and incorrect estimates is in table 12. From the table we can also have the information about the hit rate, sensitivity, and specificity, which are concluded in the table 13.

Table 12. The classification table in extra sugar contents prediction, the cut value is 0.5.

Observed (Actual)		Predicted		
		Has extra sugar contents		Percentage correct
		Yes	No	
Has extra sugar contents	Yes	9	1	90
	No	2	8	80
Overall Percentage				85

Table 13. The hit rate, sensitivity, and specificity of the extra sugar contents prediction.

	Implicaiton	Value
Hit Rate	Number of correct prediction divided by total sample size	85%
Sensitivity	Fraction of correct predictions in the having extra sugar contents potato case	90%
Specificity	Fraction of correct predictions in the normal potato case	80%

The results show that the logistic regression model can predict if there is extra sugar contents in the potato at an overall 85% correct rate. The false positive might due to the uneven surface roughness or sprouts on the potato. And the false negative might due to the close H_{avg} values to the normal potatoes. While the probability of the extra sugar contents occurrence does not match the target, the value is not very far from the correct range.

4.2.4 Development of the neural network

The neural network model is used to obtain the correct rate of the sugar concentration prediction. Neural network is a useful tool for research and industry in many applications, such as electronics, financial, and manufacturing. Neural networks are composed of interconnecting neurons inspired by the biological nervous systems. Neural networks are trained to perform a specific target output when entering particular input and bias parameters. The values of the weights between the elements are adjusted based on the comparison of the output and the target until they are match. Neural networks can be used to solve complex functions, and have been used successfully for applications such as data processing, identification, classification, and function approximation [35].

In this research, backpropagation networks were constructed to detect if the potato has extra sugar contents. The gradient in backpropagation is computed for nonlinear multilayer networks, and the weights moved along the negative of the gradient of the performance function. The inputs and the targets are used to train the network until it can associate the inputs with specific outputs. The advantage of a backpropagation network is that it can approximate any function with some discontinuities. With a properly trained backpropagation networks, a new input leads to an output similar to the correct output for input in training. So there is no need to train all possible input and output pairs to obtain good results.

The steps of using backpropagation functions to train neural networks to predict the sugar contents are as follows:

1. Assemble the training data.

   All of the inputs and targets are normalized and arranged. The target is set to be 0 and 1 for potato without and with extra sugar contents, respectively. The original data of possible inputs such as H, Havg, and density are normalized to a value between 0 and 1.

2. Create a network.

   A feedforward network is created in Matlab program. The variables of the network structure including the number of inputs, the number of hidden nodes, the number of hidden layers, and the transfer and training functions are decided in this step.

3. Train the Simulate the network.

   The weights and biases of the network are initialized before training the network. The network is trained for classification in this research. When training, the inputs and the target outputs are entered in the neural network model. The weights and biases are changing iteratively to minimize the performance function – mean square error, which is the average squared error between the target outputs and the network output. The gradient of the performance function is determined by the backpropagation algorithm.

4. Testing the network with new inputs when the training of the network is finished.

New inputs and target outputs are entered to test the feasibility of the network. The correct rate is computed as the target output and the network output matches.

5. Modify the network function if necessary.

   If the correct rate of the network design is not high enough, the parameters or the functions in network are modified to have a higher correct rate.

4.2.5 Results of neural network model using MATLAB

The same data used in the logistic model are applied in the inputs of the neural network model. There are also 80 data sets and 20 data sets in the modeling and testing group, respectively. The data are randomly selected in the testing group as in the logistic regression modeling. Before entering the raw data as inputs, they are adjusted to the values between zero to one. There are four input nodes and one output node in the neural network model. The neural network structure used in this research is plotted in figure 17.

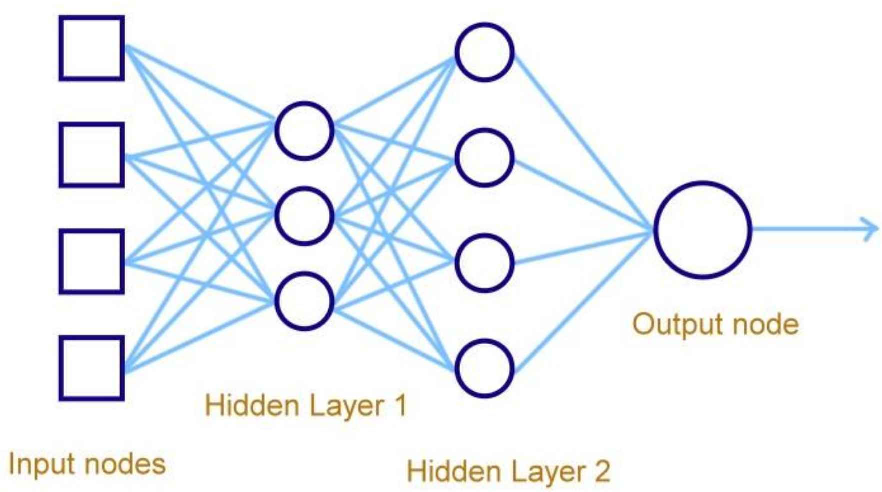

Figure 17. The neural network structure used in extra sugar contents prediction.

There are 4 input nodes: weight, ambient temperature, density, and H_{avg}. There are 2 hidden layers with 3 and 4 nodes in each layer. The transfer functions of the hidden layers are tansig, which can calculate the layer's output between -1 and 1. The transfer function in the output is logsig, which can calculate the layer's output between 0 and 1 since our targets are between 0 and 1.

Various training functions of the neural network can be used under backpropagation algorithm. The correct rates of the extra sugar contents identification are listed in the table 14.

Table 14. Backpropagation training functions with the correct rates.

Training functions	Learning Algorithm	Correct Rate(%)
Trainrp	Resilient back propagation	80
Trainoss	One step secant method	55
Trainscg	Scaled conjugate gradient algorithm	60
Traingda	Gradient descent with adaptive learning rate	55
Trainlm	Levenberg-Marquardt algorithm	70
Traingd	Basic gradient decent	50
Traingdm	Gradient descent with momentum	55
Traincgb	Powell-Beale conjugate gradient algorithm	55
Traincgp	Polak-Ribiere conjugate gradient algorithm	50
Traincgf	Fletcher-Reeves conjugate gradient algorithm	60
Traingdx	Adaptive learning rate algorithm	65

The highest correct rate of neural network model is 80% using Trainrp to predict the extra sugar contents. The cut value is set to 0.5 as well. Figure 18. shows the actual and predicted values for the 20 testing samples. Circles stand for the targets and stars stand for the predicted values.

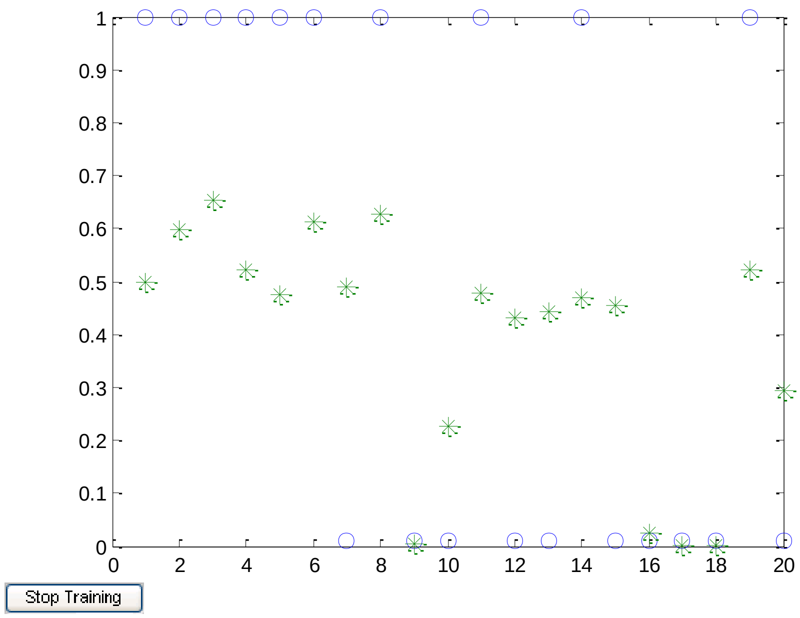

Figure 18. Results of the extra sugar contents prediction using neural network model.

4.2.6 General methodology for the extra sugar contents detection

The general methodology to identify the extra sugar contents potato is summarized as follows:

1. Heat the potato sample with appropriate time and heat flux.
2. Use infrared camera to take thermal images of the potato sample.
3. From the sequences of thermal images, obtain the average heating rate of the sample area.
4. Randomly select the data to be in the training group or the testing group.

5. Use data in the training group to construct the logistic regression model

   $P = \frac{e^x}{1+e^x}$. Select variables that is significant to the model.

6. With the existing logistic regression model, the probability of the extra sugar contents potato can be calculated by the average heating rate.

7. If the probability of extra sugar contents potato exceeds 0.5, the specific potato sample is characterized as the extra sugar contents one.

   If $P \geqq 0.5$, the potato has extra sugar contents.

   If $P < 0.5$, the potato is a normal one.

4.3 Experimental results of identifying the extra sugar contents locations

4.3.1 Comparison of the H_{avg} of locations with and without the extra sugar contents

First, we want to know if there is a Havg difference between extra sugar contents area and normal area on the same potato. The areas where sugar solution was injected were compared with areas without injecting sugar solution on the same potato. Area 1 is where the sugar solution injected, and area 2 is the symmetric area that far from the sugar injection, as shown in figure 19. The area 2 should be far from the sugar injection region because the average heating rate will not be affect by the sugar solution easily. Table 15 shows ΔH_{avg} of the areas with and without the extra sugar contents for each snapshot. The average of ΔH_{avg} of the sixty potatoes from area 1 and area 2 are in the table 16. The ΔH_{avg} is calculated by taking the H_{avg} of normal potato subtracting the H_{avg} of extra sugar contents potato for each area. The reason to apply the ΔH_{avg} in comparing the H_{avg} in different areas is due to the non-uniform thermal properties of the potato in reality. In other words, each potato might have different skin texture and non flat surface hence has non-uniform emissivity. So it is needed to take into consideration of the natural of the potato properties.

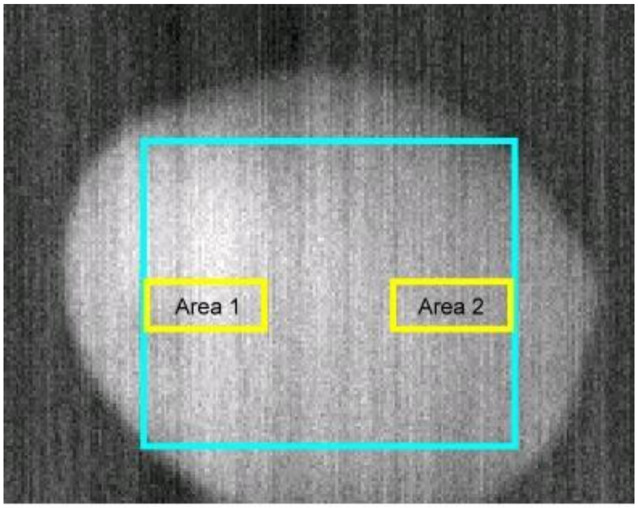

Figure 19. ΔH_{avg} comparison of the locations with and without the extra sugar contents.

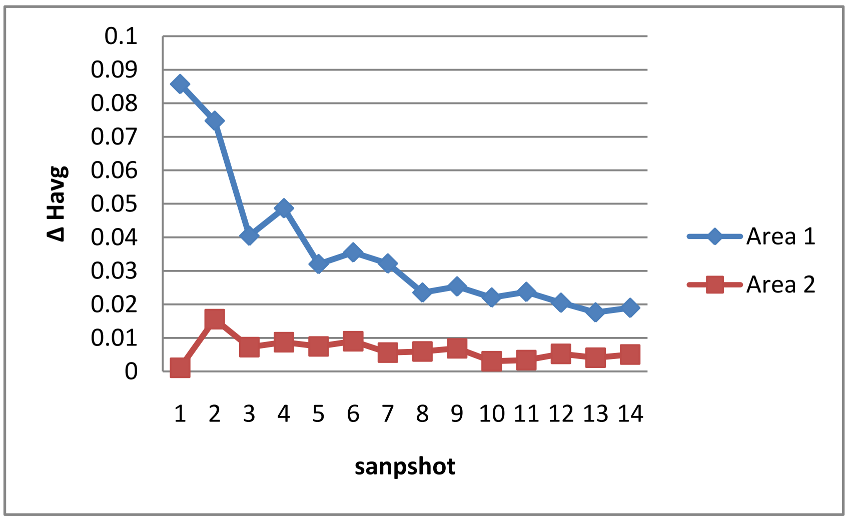

Figure 20. ΔH_{avg} comparison of the areas with and without the extra sugar contents for each snapshot.

Table 15. ΔH_{avg} of the areas with and without the extra sugar contents for each snapshot.

	ΔH_{avg} of Area 1	ΔH_{avg} of Area 2
snapshot 1	0.085719	0.001077
snapshot 2	0.074766	0.015584
snapshot 3	0.040469	0.00722
snapshot 4	0.048692	0.008715
snapshot 5	0.032021	0.007448
snapshot 6	0.035524	0.009002
snapshot 7	0.03221	0.005594
snapshot 8	0.023532	0.005952
snapshot 9	0.025364	0.006879
snapshot 10	0.022053	0.003044
snapshot 11	0.023742	0.00337
snapshot 12	0.020509	0.005235
snapshot 13	0.017612	0.004097
snapshot 14	0.018982	0.005056

Table 16. Average ΔH_{avg} values of area 1 and area 2.

Sample	Summation of ΔH_{avg}, Area l	Summation of ΔH_{avg}, Area 2	Difference of summation of ΔH_{avg}
average	0.501	0.088	0.413
standard deviation	0.647	0.586	0.101

The average value of ΔH_{avg} on different areas of each snapshot is plot in figure 20. From the figure we can assume that Area1- the sugar solution area- induces larger ΔH_{avg} than Area 2. ΔH_{avg} of area 2 is very close to zero, hence the t-test is used to see if ΔH_{avg} is significant different from zero.

1. H_o: $\bar{x} = 0$

   H_a: $\bar{x} \neq 0$

   $\alpha = 0.05$

2. T.S: $t = \frac{\bar{y}}{S/\sqrt{n}} = \frac{0.088}{0.586/\sqrt{60}} = 1.163$

3. R.R: For degree of freedom (df)=n-1= , reject Ho if $|t| \geqq t_{\frac{\alpha}{2}}$

   From the table of the t distribution, $t_{\frac{\alpha}{2}, 59} = 2.001$

4. Because $t < t_{\frac{\alpha}{2}, 59}$, we fail to reject the null hypothesis that the mean of ΔH_{avg} is equal to zero. The ΔH_{avg} of area 2 is not significant different from zero.

To see if there is significant difference of ΔH_{avg} of area 1 and area 2, paired t-test is used here. The procedures of paired t-test to compare the summation of the average heating rates are:

1. Ho: $\mu_d = \mu 1 - \mu 2 \leqq 0$

   Ha: $\mu_d > 0$

   $\alpha = 0.05$

2. T.S: $t = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{0.412}{0.101/\sqrt{60}} = 31.692$

3. R.R: For degree of freedom (df)=n-1= , reject Ho if $t \geqq t_\alpha$

   From the table of the t distribution, $t_{\alpha, 59} = 1.671$

4. Because $t > t_{\alpha, 59}$, we reject the null hypothesis that the mean of the two group is the same. There is a significance difference of the ΔH_{avg} between area 1 and 2. The ΔH_{avg} of area 1 is larger than the ΔH_{avg} of area 2.

From the results of the hypothesis testing, we can also verify that the extra sugar contents location on the potato cools faster than the normal areas on the same potato.

4.3.2 Results of identifying the extra sugar contents location using the logistic regression model

After understanding how the ΔH_{avg} behave on the extra sugar contents location, the next objective is to identify the extra sugar contents location of the potato without comparing to the original one. Data of potatoes with extra sugar contents only are used here because we cannot have the identical potato without extra sugar contents in a real situation. The whole area of interest on the thermal image is limited in the rectangular inside the potato. And the thermal images of each potato are divided into 40 small squares with 20 x 20 pixels of each square. The H_{avg} of each square and the whole area of interest are calculated for each potato. With the logistic regression introduced in the previous section, the location of the extra sugar contents could be predicted.

First, the independent variables are selected in the logistic regression model. In addition to the independent variables in the previous section, the H_{avg} of the whole area of interest on the specific potato is considered as a variable. The H_{avg} might vary due to different properties of each potato, such as surface roughness and skin color. The likelihood ratio test is used to select the appropriate independent variables. In table 17, H_{avg} means the average heating rate of the small square and H_{avg_whole} means the average heating rate of the whole potato.

Table 17. Statistical results of selecting the independent variables.

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	1851.358	173.961	1	.000
Havg_whole	1764.187	86.790	1	.000
Havg	2546.983	869.586	1	.000

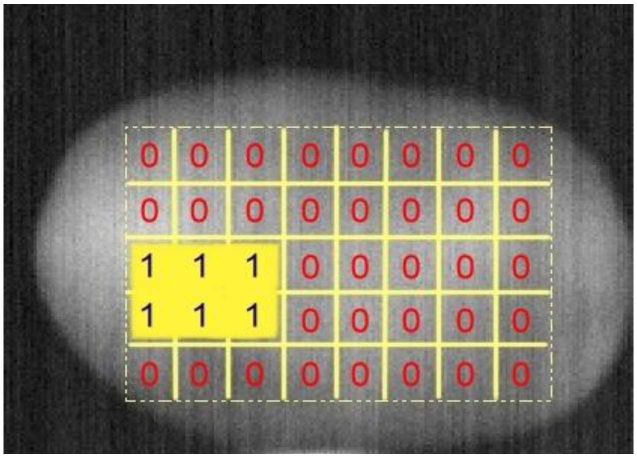

Figure 21. The target settings of the potato with sugar injection region.

Since other factors such as density of potato and ambient temperature are tested in the last section, we only consider the H_{avg} and H_{avg_whole} for this discussion. From the table 17, it is apparent that H_{avg} and Ha_{vg_whole} are significant with the significance level 0.05, which means we can reject the null hypothesis that the coefficient of H_{avg} and H_{avg_whole} are 0. Hence H_{avg} and H_{avg_whol}e are the independent variables used in the logistic regression model to predict the extra sugar contents locations.

In constructing the logistic regression model, the dependent variables are set to 0 or 1. For each potato with extra sugar contents, there are 6 squares close to the extra sugar contents location, which are set to be 1. Other squares far from the extra sugar contents are set to be 0. An example to set the targets is shown in figure 21.

The total potato samples are randomly separate as modeling and testing group, which contains 40 and 20 samples, respectively. There are 40 squares on each potato, so the total data sets of the modeling and testing group will be 1600 and 800 in each group. The data sets in the modeling group are used to construct the logistic regression model. The coefficients of the independent variables are calculated from SPSS as shown in the table 18.

Table 18. The coefficients of the logistic regression model in the extra sugar contents location prediction.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	Havg	-2.055	.092	493.546	1	.000	.128
	Havg_whole	1.484	.148	100.013	1	.000	4.411
	Constant	-5.451	.521	109.569	1	.000	.004

The equation of the logistic regression model is now:

$$P = \frac{e^{-5.451-2.055*\mathrm{Havg}\ +1.484*\mathrm{Havg\_whole}}}{1+e^{-5.451-2.055*\mathrm{Havg}\ +1.484*\mathrm{Havg\_whole}}} \qquad (4.5)$$

The next step is using this model to fit the data of the testing group to see the accuracy of classification, as in the previous section. The cut line is set to 0.5 as well. For the 800 data sets in the testing group, the classification table of the correct and incorrect estimates appears in table 19. From the table we can also have the information about the hit rate, sensitivity, and specificity, which are included in the table 20.

Table 19. The classification table in extra sugar contents prediction, the cut value is 0.5.

Observed (Actual)		Predicted		
		Has extra sugar contents on the square area		Percentage correct
		Yes	No	
Extra sugar contents on the square area	Yes	70	50	58.33%
	No	22	658	96.76%
Overall Percentage				91.00%

Table 20. The hit rate, sensitivity, and specificity of the extra sugar contents location prediction.

	Implication	Value
Hit Rate	Number of correct prediction divided by total sample size	91.00%
Sensitivity	Fraction of correct predictions in the location with extra sugar contents potato case	58.33%
Specificity	Fraction of correct predictions in the location without extra sugar contents potato case	96.76%

The result shows that the logistic regression model can predict if the specific small square is the location of extra sugar contents on the potato as an overall 91 % correct rate. The false positive might due to the uneven skin texture in the location. And the false negative might due to the depth of the extra sugar contents is too deep that are not detectable. While the probability of false negative location prediction of the extra sugar contents doesn’t reach 0.5, the values are higher than other normal locations.

In addition, the correct rate is also calculated for each potato sample, listed in table 21. The overall correct rate is 91% used this method. The results of the prediction of the extra sugar contents location for the 20 samples is in figure 22. The best correct rate is 100% and the worst is 85% for individual potato (6 errors out of 40 squares).

Table 21. The correct rate of each sample of the extra sugar contents location prediction.

Sample	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	
Correct Rate(%)	97.5	85	100	87.5	95	90	87.5	87.5	90	90	
Sample	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	Overall
Correct Rate(%)	92.5	90	90	87.5	90	100	90	100	85	85	91.0

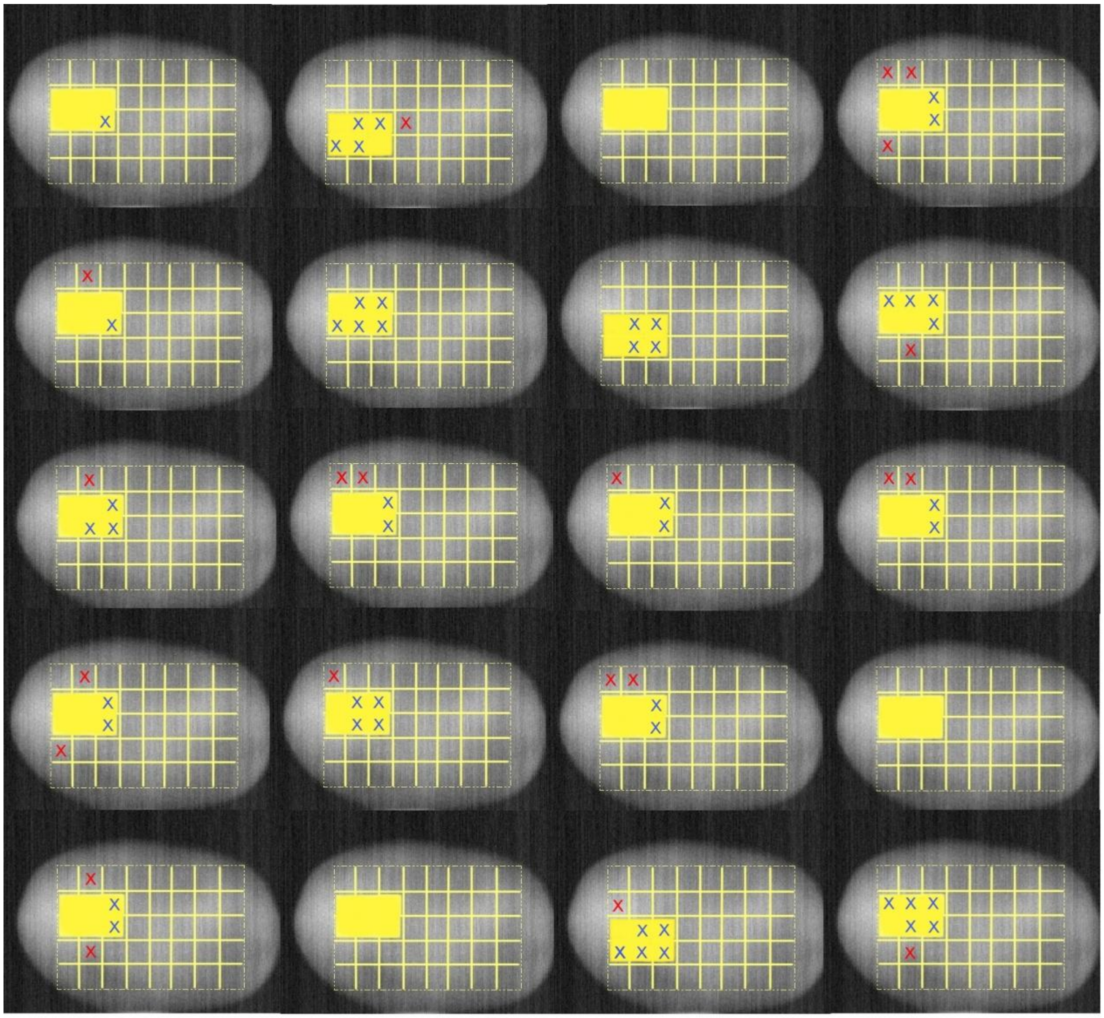

Figure 22. The visualization results of the extra sugar contents location prediction.

4.3.3 General methodology for identifying extra sugar contents location

The general methodology to identify the location of the extra sugar contents on the potato is summarized as follows:

1. Heat the potato sample with appropriate time and heat flux.
2. Use infrared camera to take thermal images of the potato sample.
3. From the sequences of thermal images, obtain the average heating rate of the sample

area.

4. Randomly select the data to be in the training group or the testing group.
5. Identify the potato with extra sugar contents by the methodology in the previous section.
6. For potatoes with extra sugar contents, divide the thermal image into small squares. Calculate the average heating rate of each small square.
7. Use data in the training group to construct the logistic regression model

   $P = \frac{e^x}{1+e^x}$. Select variables that is significant to the model.
8. With the existing logistic regression model, the probability of the specific location with the extra sugar contents can be calculated by its average heating rate and the average heating of the whole potato.

   If $P \geqq 0.5$, the location on the potato has extra sugar contents.

   If $P < 0.5$, the location on the potato does not have extra sugar contents.

4.4 Experimental results of identifying the extra sugar contents depth

4.4.1 Comparison of H_{avg} of different extra sugar contents depths

In section 4.3 we determine that the location of the extra sugar contents could be predicted with high accuracy. However, if the depth of the extra sugar contents is too deep, it might not easily be identified form that procedure. The objective here is to focus on the relationships between the depth of the extra sugar contents and the H_{avg} of the potato. Only the location of the sugar solution injection is selected on the thermal image. To simplify the analysis, the area of interest is set to 60 x 20 pixels for each potato on the extra sugar contents location with different depth. The depth of the extra sugar contents is from 0 to 1 inch for larger potatoes, and 0 to 0.875 inch for smaller potatoes. The area of interest is divided into 3 columns, so the area for each column is 20 x 20 pixels, as shown in the figure 23.

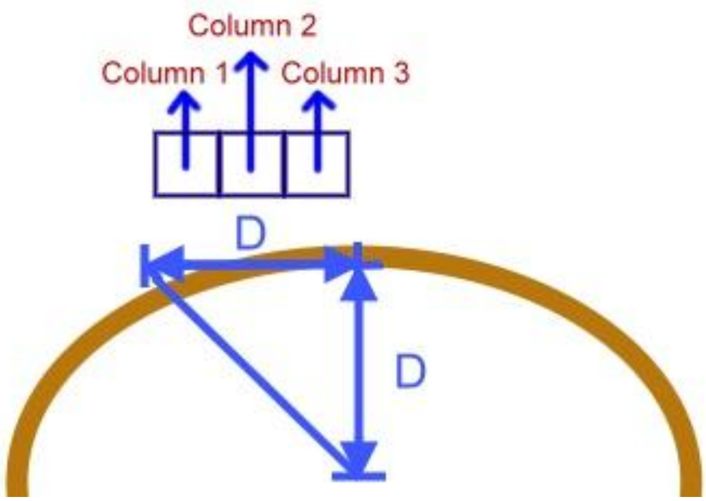

Figure 23. The extra sugar contents location is divided into 3 columns.

Table 22. Values of ΔH_{avg} of the different extra sugar contents depths of each snapshot.

	ΔH_{avg} of Column 1	ΔH_{avg} of Column 2	ΔH_{avg} of Column 3
snapshot 1	0.109844	0.035828	-0.00495
snapshot 2	0.097396	0.010018	0.002033
snapshot 3	0.078867	0.014094	-0.00152
snapshot 4	0.072162	0.022084	-0.00518
snapshot 5	0.054371	0.009428	-0.00661
snapshot 6	0.051177	0.020478	-0.00178
snapshot 7	0.046261	0.014533	-0.00292

Table 22. Continued.

	ΔH_{avg} of Column 1	ΔH_{avg} of Column 2	ΔH_{avg} of Column 3
snapshot 8	0.036767	0.010122	-0.00259
snapshot 9	0.036499	0.010779	0.002199
snapshot 10	0.031802	0.008343	9.25E-05
snapshot 11	0.031269	0.014199	0.00302
snapshot 12	0.030162	0.009023	0.00034
snapshot 13	0.026368	0.008799	-0.00196
snapshot 14	0.026138	0.010492	0.003567

Table 23. The summation of ΔH_{avg} of different extra sugar contents depths.

Sample	Summation of ΔH_{avg}, Column l	Summation of ΔH_{avg}, Column 2	Summation of ΔHavg, Column 3	Difference of ΔH_{avg}, column1-2
average	0.729	0.198	-0.016	0.053
standard deviation	0.843	0.663	0.609	0.029

The relationship between ΔH_{avg} and depth is studied first. The ΔH_{avg} is calculated for the three columns of different sugar injection depths as shown in figure 23. The total sample number is sixty. Column 1 to column 3 represents the sugar solution injection region form shallow to deep, relative to potato surface. Table 22 shows the values of ΔH_{avg} of the different extra sugar contents depths of each snapshot. And table 23 lists the summation of ΔH_{avg} of different extra sugar contents depths. Figure 24 is plotted from the data in table 22, and shows that the deeper the extra sugar contents in the potato, the smaller the ΔH_{avg} on the selected area. The ΔH_{avg} are very close to zero in column 3, where the depth of the extra sugar contents is form $\frac{2}{3}$ to 1 inch.

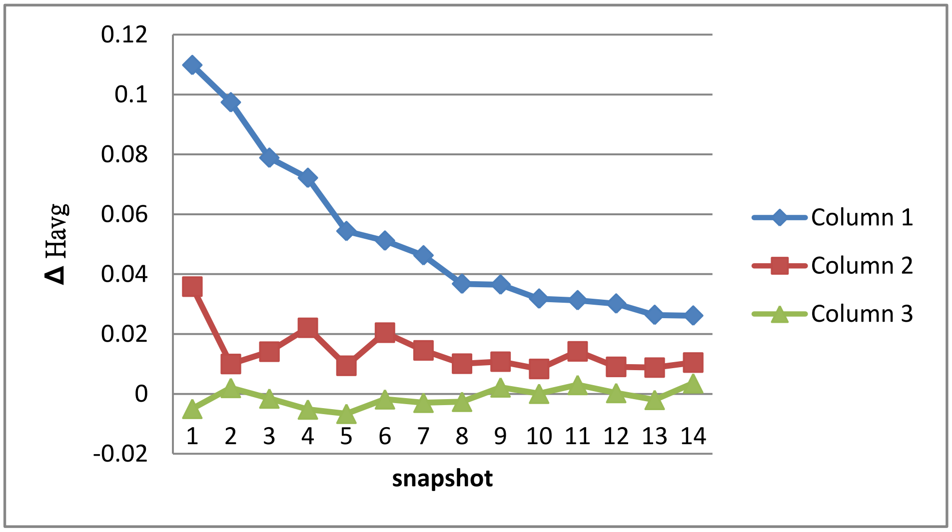

Figure 24. ΔH_{avg} comparison of the different extra sugar contents depths.

To test if the ΔH_{avg} in column 2 are significance different from zero, the t-test is used.

1. H_o: $\mu = 0$

   H_a: $\mu \neq 0$

   $\alpha = 0.05$

2. T.S: $t = \frac{\bar{y}}{S/\sqrt{n}} = \frac{-0.016}{0.609/\sqrt{60}} = -0.204$

3. R.R: For degree of freedom (df)=n-1= , reject Ho if $|t| \geqq t_{\frac{\alpha}{2}}$

   From the table of the t distribution, $t_{\frac{\alpha}{2},\,59} = 2.001$

4. Because $t < t_{\frac{\alpha}{2},\,59}$, we cannot reject the null hypothesis that the mean of ΔH_{avg} is equal to zero. The ΔH_{avg} of column 3 is not significant different from zero. Since the ΔH_{avg} of column 3 can be considered as zero, we may assume that the depth of the

extra sugar contents larger than $\frac{2}{3}$ inch is not detectable in this research.

The paired t-test is also used here to test if there is significant difference of ΔH_{avg} between column 1 and column 2.

1. H_o: $\mu d = \mu 1 - \mu 2 \leqq 0$

   H_a: $\mu d > 0$

   $\alpha = 0.05$

2. T.S: $t = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{0.053}{0.082/\sqrt{60}} = 5.006$

3. R.R: For degree of freedom (df)=n-1= , reject Ho if $t \geqq t_{\alpha}$

   From the table of the t distribution, $t_{\alpha, 59} = 1.671$

4. Because $t > t_{\alpha, 59}$, we reject the null hypothesis that the mean of the two group is the same. There is a significance difference of the ΔH_{avg} between column 1 and 2. The ΔH_{avg} of column 1 is larger than the ΔH_{avg} of column 2.

4.4.2 Results of identifying the extra sugar contents depth using linear regression model

Since the ΔH_{avg} of different depths of extra sugar contents varies, we may predict the depth of the extra sugar contents by the value of H_{avg}. The potatoes with extra sugar contents are used without comparing the originals. The H_{avg} of the different depth of extra sugar contents region are calculated for the areas of interest. The area of interest is now reset to 40 x 20 pixels neglecting the column 3, which the H_{avg} is no significant difference comparing to the normal potatoes.

Unlike the prediction of the extra sugar contents occurrence and the location of the extra sugar contents, the linear regression model is used to predict the depth of the extra sugar contents. The dependent variable is now the depth of the extra sugar contents, and the independent variables are H_{avg} and H_{avg_whole}. The potato samples are also randomly separated as modeling and testing group with 40 and 20 samples in each group.

For consistency, the total area of interest is divided into 2 columns, so the area of each column will be 20 x 20 pixels. The value of depth is set according to the middle value of each column, as shown in Figure 25. For example, for large potatoes with the maximum detectable depth of $\frac{2}{3}$ inch, if the area of interest is divided into 2 columns, the value of depth used in the modeling group will be $\frac{1}{6}$ and $\frac{1}{2}$ inch.

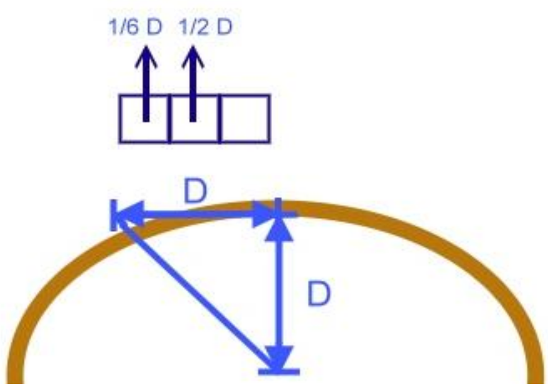

Figure 25. The setting of target depths in the linear regression model.

The linear regression model used to predict the depth of the extra sugar contents is calculated using SPSS in table 24.

Table 24. Coefficients of the linear regression model.

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta	B	Std. Error
1	(Constant)	.537	.108		4.969	.000
	Havg	.125	.019	.701	6.469	.000
	Havg_whole	-.107	.033	-.354	-3.270	.002

From the coefficient table we can see that the coefficients of Havg and Havg_whole are both significant. So these two independent variables can be used to predict the depth of extra sugar contents. The linear regression model will now be

Predicted depth = 0.125*Havg – 0.107*Havg_whole + 0.537 (4.6)

for the forty data sets in the modeling group. Use equation this to fit the data sets in the testing group we can have the predicted depth of the extra sugar contents. The predicted depth and the actual depth of the 20 samples in the testing group are listed in the Table 25. The cut value is set to 0.3 inch, which means the predicted depth in column 1 and column 2 should be 0- 0.3 and 0.3-0.66 inch, respectively. It is recorded as an error prediction if the values are not in the predicted ranges. There are 5 errors out of the 40 data, so the correct rate of the extra sugar contents depth prediction is 87.5% if we set the resolution to 0.3 inch.

Table 25. The predicted depth and the error prediction of the testing samples.

Sample	Column	Actual depth	Havg	Havg_whole	Predicted depth	Error prediction
T1	column 1	0.165	-6.411	-4.4	0.21	
T1	column 2	0.495	-5.615	-4.4	0.31	
T2	column 1	0.165	-6.827	-4.62	0.18	
T2	column 2	0.495	-5.870	-4.62	0.30	
T3	column 1	0.165	-7.046	-4.25	0.11	
T3	column 2	0.495	-5.963	-4.25	0.25	x
T4	column 1	0.165	-5.95462	-3.72	0.19	
T4	column 2	0.495	-4.35632	-3.72	0.39	
T5	column 1	0.165	-5.944	-4.67	0.29	
T5	column 2	0.495	-5.619	-4.67	0.33	
T6	column 1	0.165	-6.116	-3.88	0.19	
T6	column 2	0.495	-5.663	-3.88	0.24	x
T7	column 1	0.165	-5.14568	-3.66	0.29	
T7	column 2	0.495	-4.9422	-3.66	0.31	
T8	column 1	0.165	-5.359	-3.33	0.22	
T8	column 2	0.495	-3.958	-3.33	0.40	
T9	column 1	0.165	-5.753	-3.58	0.20	
T9	column 2	0.495	-4.261	-3.58	0.39	
T10	column 1	0.145	-5.092	-3.68	0.29	
T10	column 2	0.438	-4.502	-3.68	0.37	
T11	column 1	0.165	-5.818	-3.46	0.18	
T11	column 2	0.495	-4.150	-3.46	0.39	
T12	column 1	0.145	-5.557	-3.2	0.18	
T12	column 2	0.438	-3.595	-3.2	0.43	
T13	column 1	0.145	-4.666	-3.69	0.35	x
T13	column 2	0.438	-3.704	-3.69	0.47	
T14	column 1	0.165	-4.349	-2.91	0.30	
T14	column 2	0.495	-3.992	-2.91	0.35	
T15	column 1	0.165	-4.377	-3.05	0.32	x

Table 25. Continued

Sample	Column	Actual depth	Havg	Havg_whole	Predicted depth	Error prediction
T15	column 2	0.495	-3.215	-3.05	0.46	
T16	column 1	0.165	-4.835	-2.85	0.24	
T16	column 2	0.495	-3.007	-2.85	0.47	
T17	column 1	0.165	-6.617	-4.3	0.17	
T17	column 2	0.495	-5.591	-4.3	0.30	
T18	column 1	0.145	-5.964	-4.29	0.25	
T18	column 2	0.438	-5.539	-4.29	0.30	
T19	column 1	0.145	-5.278	-3.86	0.29	
T19	column 2	0.438	-4.434	-3.86	0.40	
T20	column 1	0.145	-4.93467	-3.69	0.31	x
T20	column 2	0.438	-3.62901	-3.69	0.48	

4.4.3 General methodology for identifying extra sugar contents depth

The general methodology to identify the depth of the extra sugar contents on the potato is summarized as follows:

1. Heat the potato sample with appropriate time and heat flux.
2. Use infrared camera to take thermal images of the potato sample.
3. From the sequences of thermal images, obtain the average heating rate of the whole sample area.
4. Randomly select the data to be in the training group or the testing group.
5. Identify the potato with extra sugar contents and locate the extra sugar contents area by the methodology in the previous section.
6. Focus on the thermal image of the extra sugar contents location and divided the location into small squares. Calculate the average heating rates of the small squares

on the extra sugar contents location.

7. Use data in the training group to construct the linear regression model. Select variables that is significant to the model.
8. With the existing linear regression model, the extra sugar contents depth can be predicted with a 0.3 inch resolution.

4.5 Summary of the experimental results

Table 26 shows the experimental results of this research with the correct rate of each experiment. From the table we can see that the logistic regression model predicts the occurrence of the extra sugar contents and the location of the extra sugar contents with high accuracy. The depth of the extra sugar contents can be predicted with the linear regression model, but the resolution is 0.3 inch with the current experimental settings. For the neural network model in the extra sugar contents occurrence prediction, the reasons of the lower correct rate than the logistic regression model might due to the small sample numbers and inadequate parameter settings.

Table 26. Summary of the experimental results.

Experiment	Identify if the potato has extra sugar contents		Identify the location of the extra sugar contents on the potato	Identify the depth of the extra sugar contents on the potato
Analysis Methods	Logistic regression model	Neural network model	Logistic regression model	Linear regression model
Variables	Havg	Havg, weight, ambient temp., density	Havg, Havg of the whole area	Havg, Havg of the whole area
Correct rate (%)	85	80	91	87.5 (0.3in resolution)

CHAPTER V

MODEL VALIDATION

5.1 Introduction

The models constructed in the previous chapters can predicted the occurrence, location, and depth of the extra sugar contents of the potato in considerably high accuracy. However, how the model fit into the data was not discussed. Moreover, the models might be misleading if the correct rate varies greatly using data from different modeling/testing combinations. To validate the models used to predict the occurrence, location, and depth of the extra sugar contents in the potato, the goodness of fit test is discussed in this chapter. In addition, the correct rates of the predictions are recalculated using different data sets form the original one.

5.2 Effects of the experiments done in different seasons

The whole experiment is conducted within 12 months. However, for different seasons, the potato samples might have different properties that can affect the H_{avg}. In this section, the mean and the variance of the H_{avg} for two batches of samples in different seasons are compared. The normal samples and the samples with extra sugar contents are compared respectively. Ten potato samples of similar weight are selected in each batch.

- For normal potato samples

Table 27. Data of the normal potatoes generated from different seasons.

Experiment done in April 2008				
	weight(g)	temp(F)	density(g/cm^3)	Havg
average	159.7	70.797	1.333	-3.728
std	16.261	0.417	0.088	0.345
Experiment done in December 2007				
	weight(g)	temp	density(g/cm^3)	Havg
average	161.2	71.623	1.371	-3.896
std	18.371	0.896	0.162	0.487

Compare the mean of Havg of experiments done in spring and winter using a 95% confidence interval. The data in table 27 is used.

1. Ho: $\overline{y_1} - \overline{y_2} = 0$

   H1: $\overline{y_1} - \overline{y_2} \neq 0$

2. T.S: $t = \frac{\bar{y}_1 - \overline{y_2}}{S_p / \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{-3.723 + 3.896}{0.422 / \sqrt{\frac{1}{10} + \frac{1}{10}}} = 0.183$

   Sp = $S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(10-1)0.345^2 + (10-1)0.487^2}{10+10-2}} = 0.422$

3. R.R: For degree of freedom df = $n_1+n_2-2= 18$, reject Ho if $|t| \geqq t_{\frac{\alpha}{2}}$

   From the t table, $t_{0.025,\,18} = 2.101$

4. Because $t < t_{\alpha/2,\,9}$, we fail to reject the null hypothesis that the mean of the two group is the same. There is no difference of Havg of experiment done in different season for normal potatoes.

Comparing the variance of Havg of experiments done in spring and winter using a

95% confidence interval.

1. Ho: $\sigma_1^2 = \sigma_2^2$

   H1: $\sigma_1^2 \neq \sigma_2^2$

2. T.S: $F = \frac{s_1^2}{s_2^2} = 0.503$

3. R.R: For $\alpha = 0.05$ and $df_1 = n1-1$, $df_2 = n2-1$, reject Ho if $F \leqq F_{1-\alpha/2, df1, df2}$ or

   $F \geqq F_{\alpha/2, df1, df2}$

   From the F distribution table, $F_{0.025, 9, 9} = 4.03$, $F_{0.975, 9, 9} = 0.248$

4. Because F is not in the rejection region, we fail to reject the null hypothesis that the variances of the two groups are the same. There is no significant difference of variance of H_{avg} of experiment done in different season for normal potatoes.

- For extra sugar contents potato samples

Table 28. Data of the extra sugar contents potatoes generated from different seasons.

Experiment done in April 2007				
potato	weight(g)	temp(F)	density(g/cm^3)	Havg
average	160.8889	70.75922	1.337	-3.95665
std	16.72287	0.423908	0.089	0.259787
Experiment done in December 2008				
potato	weight(g)	temp(F)	density(g/cm^3)	Havg
average	164.2222	71.63704	1.379	-4.11368
std	16.64165	0.68383	0.170	0.486039

Compare the H_{avg} of experiments done in spring and winter using a 95% confidence interval. The data in table 28 is used.

1. Ho: $\overline{y_1} - \overline{y_2} = 0$

   H1: $\overline{y_1} - \overline{y_2} \neq 0$

2. T.S: t = $\frac{\bar{y}_1 - \bar{y}_2}{S_p / \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{-3.957 + 4.114}{0.390 / \sqrt{\frac{1}{10} + \frac{1}{10}}} = 0.180$

   Sp = $S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(10-1)0.260^2 + (10-1)0.486^2}{10+10-2}} = 0.390$

3. R.R: For degree of freedom df = n_1+n_2-2= 18, reject Ho if $|t| \geqq t_{\frac{\alpha}{2}}$

   From the t table, $t_{0.025, 18} = 2.101$

4. Because $t < t_{\alpha/2, 9}$, we fail to reject the null hypothesis that the mean of the two group is the same. There is no difference of Havg of experiment done in different season for extra sugar contents potatoes.

Comparing the variance of Havg of experiments done in spring and winter using a 95% confidence interval.

1. Ho: $\sigma_1^2 = \sigma_2^2$

   H1: $\sigma_1^2 \neq \sigma_2^2$

2. T.S: F = $\frac{s_1^2}{s_2^2} = 0.286$

3. R.R: For α = 0.05 and df_1 =n1-1, df_2=n2-1, reject Ho if $F \leqq F_{1-\alpha/2, df1, df2}$ or

   $F \geqq F_{\alpha/2, df1, df2}$

   From the F distribution table, $F_{0.025, 9, 9} = 4.03$, $F_{0.975, 9, 9} = 0.248$

4. Because F is not in the rejection region, we fail to reject the null hypothesis that the variances of the two groups are the same. There is no significant difference of variance of H_{avg} of experiment done in different season for the extra sugar contents

potatoes.

5.3 Model validation of the extra sugar contents occurrence prediction

5.3.1 Goodness of fit of the model

The recommended test of the overall fit of the logistic model is the Hosmer and Lemeshow test [36]. This test is more robust than the traditional chi-square test especially for small size samples. The Hosmer and Lemeshow's goodness of fit test divides the observations into 10 groups based on the predicted probabilities, and calculates the chi-square form the observed and expected values. Where a significance is larger than α (0.05 in this research), we fail to reject the null hypothesis that there is no difference between the actual and predicted values. Thus, the model prediction is not significantly different from the actual values, indicating a well-fit model [37].

The Hosmer and Lemeshow test is completed using SPSS. From table 29, the significance is 0.941, which is very large comparing to 0.05. We can conclude that the logistic regression model fits the data well in the extra sugar contents prediction.

Table 29. Results of the Hosmer and Lemeshow goodness of fit test of the extra sugar contents occurrence predictions.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	2.886	8	.941

Contingency Table for Hosmer and Lemeshow Test

		Target = .00		Target = 1.00		Total
		Observed	Expected	Observed	Expected	Observed
Step 1	1	5	5.647	3	2.353	8
	2	5	4.903	3	3.097	8
	3	5	4.611	3	3.389	8
	4	5	4.352	3	3.648	8
	5	3	3.969	5	4.031	8
	6	4	3.715	4	4.285	8
	7	5	3.573	3	4.427	8
	8	3	3.445	5	4.555	8
	9	2	3.104	6	4.896	8
	10	3	2.681	5	5.319	8

5.3.2 Correct rates discussion using different data sets

The original logistic regression model is generated from the randomly selected 40 samples. And the other 20 samples in the testing group are used to examine the correct rate of the prediction. The correct rate of the original model is 85% with 10 normal and 10 extra sugar contents samples. To test the validity of the model, 20% to 100% of the samples in the testing group are exchanged with the samples in the modeling group. When exchanging the data, the ratio of the normal and extra sugar contents sample are remained the same to reduce the noise from selecting larger sample numbers than the other.

When the data in modeling group is moved to testing group, only one data as extra sugar contents or normal sample can be used. The data being replaced from the testing group to the testing group should retrieve the original data for both extra sugar contents and normal ones. The sensitivity, specificity, and overall correct rates of the replaced data are listed in table 30.

Table 30. Correct rates comparison of the different data sets of the extra sugar contents occurrence predictions.

Replaced sample percent	0%	20%	40%	60%	80%	100%	Overall	SD
Sensitivity (%)	90	90	100	100	90	90	93.333	5.164
Specificity (%)	80	70	70	80	90	90	80.000	8.944
Correct rate (%)	85	80	85	90	90	90	86.667	4.082

Table 30 shows that the overall sensitivity is higher than the overall specificity. It might due to the outliers are not removed or the potatoes might have potential extra sugar contents. The overall correct rate 86.667% is fairly high for the prediction of the occurrence of extra sugar contents. The logistic regression model is valid in predicting the occurrence of extra sugar contents in the potato.

5.4 Model validation of the extra sugar contents location prediction

5.4.1 Goodness of fit of the model

The prediction of the extra sugar contents location is also done by the regression model. To test the goodness of fit, the Hosmer and Lemeshow test is applied as well. The

Hosmer and Lemeshow test is done by SPSS as shown in table 31. From table 31, the significance is 0.206, which is larger than 0.05. We can conclude that the logistic regression model fits the data well in the extra sugar contents prediction.

Comparing to the significance in the last section which is 0.941, the significance in extra sugar contents location prediction is closer to the α. This might due to the relative smaller samples of extra sugar contents areas.

Table 31. Results of the Hosmer and Lemeshow goodness of fit test of the extra sugar contents location predictions.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	10.919	8	.206

Contingency Table for Hosmer and Lemeshow Test

		Target = .00		Target = 1.00		Total
		Observed	Expected	Observed	Expected	Observed
Step 1	1	160	159.817	0	.183	160
	2	160	159.424	0	.576	160
	3	160	158.925	0	1.075	160
	4	160	158.135	0	1.865	160
	5	158	156.665	2	3.335	160
	6	155	153.728	5	6.272	160
	7	151	148.511	9	11.489	160
	8	127	135.883	33	24.117	160
	9	95	100.240	65	59.760	160
	10	34	28.672	126	131.328	160

5.4.2 Correct rates discussion using different data sets

The original logistic regression model is generated from the randomly selected 40 samples with extra sugar contents only. And the other 20 samples in the testing group are used to examine the correct rate of the prediction. There are 40 squares for each sample, so the number of the total data sets is 800 in the testing group. To test the validity of the model, 20% to 100% of the data in the testing group are exchanged with the data in the modeling group. The exchanged data are based on different samples but not number of data only.

The sensitivity, specificity, and overall correct rates of the replaced data are listed in table 32. Table 32. shows when the data in the modeling and testing groups are swapped, the overall correct rate only varies within 2%. The prediction of the extra sugar contents location is considerably reliable. The sensitivity is always lower than the specificity. This might due to the signal of the deeper extra sugar contents is not detectable. However, for relative shallow extra sugar contents, the location of the extra sugar contents can be predicted with high accuracy. So the logistic regression model is valid in predicting the location of extra sugar contents in the potato.

Table 32. Correct rates comparison of the different data sets of the extra sugar contents location predictions.

Replaced sample percent	0%	20%	40%	60%	80%	100%	Overall	SD
Sensitivity (%)	58.33	59.17	60.00	58.33	66.67	63.33	60.97	3.060
Specificity (%)	96.76	95.00	95.00	95.29	94.67	94.56	95.05	0.734
Correct rate (%)	91	89.63	89.75	89.75	89.63	89.88	89.94	0.482

5.5 Model validation of the extra sugar contents depth prediction

5.5.1 Goodness of fit of the model

The linear regression model is applied in this research to predict the extra sugar contents depth. The linear regression model is obtained base on the ordinary least squares method that the sum of the residuals is minimized. From the model summary generated from SPSS, as shown in table 33, the significance of the whole model can be tested by the F test. The R square is 0.486 in this model, meaning that 48.6% of the variance in depth can be predicted from the independent variables. In the ANOVA table, the F value can indicate the significance of the model. F is the ratio of the mean square for the model divided by the mean square for the error, and the mean squares are the sums of squares divided by the df. The probability of F in the table is 0.000, which is smaller than α (0.05). We can reject the null hypothesis that there is no linear relationship of the dependent and the independent variables. We can say the linear regression model with independent variables H_{avg} and H_{avg_whole} is significant.

Table 33. Results of the goodness of fit test of the extra sugar contents depth prediction model.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.697(a)	.486	.480	.40209

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.836	2	.418	20.943	.000(a)
	Residual	1.537	77	.020		
	Total	2.373	79			

5.5.2 Correct rates discussion using different data sets

The original linear regression model is generated from the randomly selected 40 samples with extra sugar contents only. And the other 20 samples in the testing group are used to examine the correct rate of the prediction. For each sample, the area of interest is divided into 2 squares. So the numbers of the data sets are 80 and 40 in the modeling and testing group, respectively. To test the validity of the model, 20% to 100% of the data in the testing group are exchanged with the data in the modeling group. The cut value is set to 0.3 inch as well.

The correct rates of the replaced data are listed in table 34. Table 34 shows when the data in the modeling and testing groups are swapped, the correct rates are between 85 to 90%. The prediction of the extra sugar contents depth using linear regression model is considerably reliable.

Table 34. Correct rate comparison of the different data sets of the extra sugar contents depth predictions.

Replaced sample percent	0%	20%	40%	60%	80%	100%	Overall	SD
Correct rate (%)	87.5	85	87.5	85	85	85	85.83	1.29

CHAPTER VI

CONCLUSION AND FUTURE WORK

6.1 Introduction

In this study, the quality of a potato relating to its sugar contents is successfully characterized using infrared thermal images. Instead of the surface temperature, the average heating rates of potatoes with and without extra sugar contents are studied. Based on the different properties of extra sugar contents regions, abnormal heating rates can vary between normal and extra sugar contents potatoes. Statistical models and neural networks models are used in this study to identify the occurrence, location, and depth of the extra sugar contents region of a potato. After selecting appropriate variables and testing the validity of the model, general methodologies for potato quality characterization are suggested.

Detailed conclusions for experiments and statistical models are presented in this chapter. Current limitations of using infrared thermography to identify potato quality are also discussed. The last section proposes future works for improving the results of this research.

6.2 Conclusions

This research outlines using infrared thermography to take thermal images of potatoes with further analysis can determine the quality of a potato. Here, the quality of the potato is associated with its sugar contents. By simulating extra sugar contents in the potato and comparing the average heating rates to the normal one, the quality of the potato can be predicted.

To identify extra sugar contents in potatoes, the logistic regression model is recommended. The logistic regression model can predict the probability of whether or not the potato has extra sugar contents. By implementing the Hosmer and Lemeshow goodness of fit test, the model provides a good fit for the data. The correct rate for identifying extra sugar contents in a potato is 85% for this study. Although there are some false predictions, the values of probabilities are not far from the range designate as correct. For the neural networks model, the highest correct rate is only 80%.

The area of interest is the largest available area of the potato image. However, the signal is not strong enough because normal areas are included. If potential extra sugar contents areas can be located in advance, the correct rate might be higher due to stronger signals.

In addition to identifying potatoes with extra sugar contents, the location can be found though analyzing thermal images. By comparing the average heating rate differences of the sugar injection location and locations far from the sugar injection position of the same potato, one can conclude that the extra sugar contents location will have a larger average heating rate difference. To find the exact extra sugar contents location, the entire thermal image is divided into small squares, for whichthe average heating rates are calculated. The logistic regression model is also applied to predict the probability of having extra sugar contents in a specific area. The sensitivity of this model is not very high (58.33%), due to extra sugar contents being too deep to detect. However, the overall correct rate (91%) is still high enough for real applications.

The depth of extra sugar contents is researched in this research as well. The average heating rate difference of different depths is compared using statistical testing. Results show the maximum detectable depth of extra sugar contents is $\frac{2}{3}$ inches in this research. The deeper the extra sugar contents, the smaller the average heating rate difference. To predict the depth of the extra sugar contents, linear regression model is used. This area of interest is limited to the sugar injection location. Results show the increment of 0.3 inches of the extra sugar contents depth can be identified with 87.5% correct rate.

By analyzing infrared thermal images, the extra sugar contents in a potato is characterized with a high level of accuracy. Developed methodologies can apply to other agricultural products as well.

6.3 Limitations

Potato quality characterization using infrared thermal images and statistical models has the following limitations:

1. The scanning rate of the infrared camera is 20 second per frame in this research. If the scanning rate of the infrared camera was higher, the more detailed thermal behavior of the potato could be studied. The time needed to inspect the potato is expected to be shorter.
2. The experimental setting can only be applied to potatoes of certain sizes. The approximate volume ranges of the potato are 5.5 to 11 $inch^3$ in this research. For extra large potatoes, the whole potato might exceed the imaging area. The area of

the smallest potato viewed from top should be larger than the fixture.

6.4 Future Work

The work done in this research can be further extended to improve the characterization procedures. Recommend future works are summarized as follows:

1. Develop the statistical model for different breeds of potatoes. The variables in the statistical model must be reconsidered.
2. Use alternative heating methods, such as a high power flash lamp, to improve efficiency.
3. Build the thermal model of the potato using a finite element analysis. Use the Finite Element Model (FEA) to validate the findings from statistical and neural network models.
4. Increase the sample sizes of potato samples for more accurate and reliable results.
5. Use an infrared thermal camera with a higher speed for better results in determining the location and depth of the extra sugar contents predictions.
6. Cooperate with optical inspection techniques to inspect the external quality of the potato.
7. Compare the results with other produce inspection methods.

REFERENCES

[1] Caracalla, V.T., International year of the potato: hidden treasure, 2008, Available at: http://www.potato2008.org.

[2] Pritchard, M.K. and Adam, L.R., Relationships between fry color and sugar concentration in stored Russet Burbank and Shepody potatoes. *American Journal of Potato Research*, 1993, **71**(1), 59-68.

[3] Hellier, C., *Handbook of Nondestructive Evaluation*, McGraw-Hill, New York, 2003.

[4] Topoleski, L.D. and Hogen, C., Vegetable identification, 2008, Available at: http://instruct1.cit.cornell.edu/courses/hort220/.

[5] Brook, R.C., CRIS project, 2001, Available at: http://www.msu.edu/~brook/cris/cris.htm.

[6] Hopkins, E.F., Variation in sugar contents in potato tuber caused by wounding and its possible relation to respiration. *Botanical Gazette*, 1927, **84**(1), 75-88.

[7] Manickavaasagan, A., Jayas, D.S, White, D.G., and Paliwal, J., Applications of thermal imaging in agriculture- A review. *CSAE Meeting*, No. 05-00, 2005.

[8] Wiecek, B., Review on thermal image processing for passive and active thermography. *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005, pp. 686-689.

[9] Shepard, S.M., Introduction to active thermography for non-destructive evaluation. *Anti-Corrosion Methods and Materials*, 1997, **44**(4), 236-239

[10] Maldague, X., *Theory and Practice of Infrared Technology for Non Destructive*

Testing, John Wiley and Sons, New York, 2005.

[11] Finney, P. J., Applications of infrared thermography to electric inspection and measurement. *Optical Sensing and Measurement Symposium*, 1992, **73**, 80-88.

[12] Allred, L.G. and Howard, T.R., Application of thermal imaging to electronic fault diagnosis. *SPIE Thermosense*, 1994, **2245**, 224-230.

[13] Nishino, S. and Ohshima, K., Fault detection for IC board using histogram of thermography. *Systems and Computers in Japan*, 2001, **32**, 38-48.

[14] Tretout, H., Infrared thermography development for composite material evaluation. *Proceedings of SPIE--The International Society for Optical Engineering*, 1988, **918**, 12-22.

[15] Connolly, M.P., Review of factors influencing defect detection in infrared thermography: Applications to coated materials. *Journal of Nondestructive Evaluation*, 1991, **10**(3), 89-96.

[16] Jiang, L.J., Ng, E.Y.K., Yeo, A.C.B., Wu,S., Pan, F., Yau, W.Y., Chen, J.H., and Yang, Y., A perspective on medical infrared imaging. *Journal of Medical Engineering & Technology*, 2005, **29**(6), 257-267.

[17] Veraverbeke, E.A, Verboven, P., Lammertyn, J., Cronje, P., De Baerdemaeker, J. and Nicolai, B.M., Thermographic surface quality evaluation of apple. *Journal of Food Engineering*, 2006, **77**, 162-168.

[18] Chaerle, L. and Straeten, D.V.D., Seeing is believing: Imaging techniques to monitor plant health. *Biochimica et Biophysical Acta*, 2001, **1519**, 153-166.

[19] Du, C.J. and Sun, D.W., Recent developments in the applications of image processing techniques for food quality evaluation. *Trends in Food Science & Technology*, 2004, **15**, 230-249.

[20] Rao, X., A line-scanned based digit image description method and its application in fruit quality inspection. *Proceedings of SPIE-the International Society for Optical Engineering*, 2004, **5587**(1), 63-70.

[21] Zhou, L., Chalana, V., and Kim, Y., PC-based machine vision system for real-time computer-aided potato inspection. *International Journal of Imaging Systems and Technology*, 1998, **9**(6), 423-33.

[22] Muir, A.Y., Ross, D.W., Dewar, C.J., and Kennedy, D., Defect and disease detection in potato tubers. *Proceedings of SPIE-The International Society for Optical Engineering*, 1999, **3543**, 199-207.

[23] Liu, Y., Ying, Y., Fu, X., and Lu, H., Experiments on predicting sugar content in apples by FT-NIR technique. *Journal of Food Engineering*, 2007, **80**, 986-989.

[24] McGlone, A.V., Jordan, B., and Paul, J., Vis/NIR estimation at harvest of pre- and post-storage quality indices for 'Royal Gala' apple. *Postharvest Biology and Technology*, 2002, **25**, 135–144.

[25] Hsieh, C. and Lee, Y., Applied visible/near-infrared spectroscopy on detecting the sugar content and hardness of pearl Guava. *Applied Engineering in Agriculture*, 2005, **21**(6), 1039-1046.

[26] Hasse, N., Rapid estimation of potato tuber quality by near-infrared spectroscopy. *Starch/Staerke*, 2006, **58**(6), 268-273.

[27] Kang S., Lee K., Choi W., Son J., Choi D., and Kim G., A near-infrared sensing technique for measuring the quality of potatoes. *ASAE Annual Meeting*, 2003, Paper number 033137.

[28] Scanlon M.G., Pritchard M.K., and Adam L.R., Quality evaluation of processing potatoes by near infrared reflectance. *Journal of the Science of Food and Agriculture*, 1999, **79**(5), 763-771.

[29] Hartmann R., and Büning-Pfaue H., NIR determination of potato constituents. *Potato Research*, 1998, **41**(4), 327-334.

[30] Dagerskog, M., & Österström, L., Infra-red radiation for food processing I. A study of the fundamental properties of infra-red radiation. *Lebensmittel Wissenschaft Technologie*, 1979, **12**(4), 237-242.

[31] CondeNet, Inc., Nutrition fact: potatoes, russet, flesh, and skin, raw [includes USDA commodity food A214]., 2008, Available at: http://www.nutritiondata.com/facts-C00001-01c20f5.html.

[32] Hsieh, S.J. and Huang, S.L., A methodology for microcontroller signal frequency stress prediction. *Microelectronics Reliability*, 2005, **45**, 1243-1251.

[33] Ott. R.L. and Longnecker M., *An Introduction to Statistical Methods and Data Analysis. 5th Edition*, Thompson Learning, Connecticut, 2001.

[34] Garson G.D., Logistic Regression from *Statnotes: Topics in Multivariate Analysis.*, 2008, Available at http://www2.chass.ncsu.edu/garson/pa765/statnote.htm.

[35] Demuth, H., Beale, M., and Hagan, M., *Neural Network ToolboxTM 6 User's Guide*, The Math Works, Inc., Massachusetts, 2008.

[36] Bruin, J., Newtest: command to compute new test. UCLA: Academic Technology Services, Statistical Consulting Group., 2006, Available at http://www.ats.ucla.edu/stat/stata/ado/analysis/.

[37] Hosmer, D.W. and Lemeshow, S., *Applied Logistic Regression. 2nd Edition*, John Wiley and Sons, New York, 2000.

APPENDIX A

Sample	H, normal	H, sugar	Difference of H	Havg, normal	Havg, sugar	Difference of Havg
P1	-2.347	-2.405	0.058	-3.929	-4.251	0.322
P2	-2.376	-2.39	0.014	-4.321	-4.201	-0.12
P3	-2.504	-2.486	-0.018	-4.284	-4.626	0.342
P4	-2.499	-2.627	0.128	-4.137	-4.469	0.332
P5	-2.627	-2.555	-0.072	-4.509	-4.671	0.162
P6	-2.652	-2.752	0.1	-4.559	-4.597	0.038
P7	-2.16	-2.358	0.198	-3.601	-3.955	0.354
P8	-2.392	-2.463	0.071	-4.235	-4.355	0.12
P9	-2.358	-2.406	0.048	-3.99	-4.214	0.224
P10	-2.06	-2.211	0.151	-3.458	-3.966	0.508
P11	-2.322	-2.61	0.288	-3.789	-4.251	0.462
P12	-2.174	-2.186	0.012	-3.693	-3.811	0.118
P13	-1.852	-2.102	0.25	-3.117	-3.636	0.519
P14	-2.21	-2.251	0.041	-3.847	-4.017	0.17
P15	-1.937	-2.276	0.339	-3.413	-4.218	0.805
P16	-2.572	-2.577	0.005	-4.357	-4.734	0.377
P17	-2.616	-2.645	0.029	-4.388	-4.738	0.35
P18	-2.166	-2.311	0.145	-3.887	-3.992	0.105
P19	-2.19	-2.284	0.094	-3.81	-3.879	0.069
P20	-2.293	-2.277	-0.016	-3.866	-3.954	0.088
P21	-2.26	-2.282	0.022	-3.9	-4.267	0.367
P22	-2.273	-2.768	0.495	-3.927	-4.692	0.765
P23	-2.308	-2.298	-0.01	-3.976	-3.847	-0.129
P24	-1.798	-2.21	0.412	-3.134	-3.672	0.538
P25	-2.266	-2.408	0.142	-3.757	-4.242	0.485
P26	-1.642	-2.025	0.383	-2.785	-3.731	0.946
P27	-1.448	-1.94	0.492	-2.227	-3.37	1.143
P28	-1.968	-1.921	-0.047	-3.286	-3.273	-0.013
P29	-1.354	-1.39	0.036	-2.241	-2.442	0.201

Continued.						
Sample	H, normal	H, sugar	Difference of H	Havg, normal	Havg, sugar	Difference of Havg
P30	-1.942	-2.061	0.119	-3.528	-3.722	0.194
P31	-1.896	-1.824	-0.072	-3.142	-3.251	0.109
P32	-1.766	-1.959	0.193	-3.173	-3.41	0.237
P33	-2.076	-2.092	0.016	-3.504	-3.749	0.245
P34	-1.679	-1.706	0.027	-2.967	-2.918	-0.049
P35	-1.997	-2.04	0.043	-3.491	-3.897	0.406
P36	-1.916	-1.957	0.041	-3.306	-3.633	0.327
P37	-2.162	-2.353	0.191	-3.83	-4.085	0.255
P38	-1.769	-1.957	0.188	-3.255	-3.462	0.207
P39	-1.812	-1.864	0.052	-3.073	-3.197	0.124
P40	-2.235	-2.209	-0.026	-3.839	-4.024	0.185
P41	-1.957	-1.94	-0.017	-3.462	-3.453	-0.009
P42	-1.788	-1.924	0.136	-3.245	-3.394	0.149
P43	-1.36	-1.556	0.196	-2.394	-2.703	0.309
P44	-1.774	-1.898	0.124	-3.182	-3.354	0.172
P45	-1.678	-1.695	0.017	-2.973	-2.964	-0.009
P46	-1.594	-1.717	0.123	-2.862	-2.919	0.057
P47	-1.498	-1.778	0.28	-2.461	-3.057	0.596
P48	-1.317	-1.498	0.181	-2.274	-2.403	0.129
P49	-1.805	-1.808	0.003	-3.013	-3.118	0.105
P50	-1.437	-1.55	0.113	-2.615	-2.839	0.224
P51	-2.281	-2.674	0.393	-3.857	-4.544	0.687
P52	-2.027	-2.131	0.104	-3.484	-3.61	0.126
P53	-2.422	-2.538	0.116	-4.213	-4.408	0.195
P54	-2.378	-2.404	0.026	-4.039	-4.287	0.248
P55	-2.111	-2.293	0.182	-3.572	-4.211	0.639
P56	-2.554	-2.757	0.203	-4.483	-4.995	0.512
P57	-1.797	-1.848	0.051	-2.988	-3.211	0.223
P58	-1.996	-2.013	0.017	-3.367	-3.459	0.092
P59	-1.998	-1.979	-0.019	-3.411	-3.553	0.142

Continued.						
Sample	H, normal	H, sugar	Difference of H	Havg, normal	Havg, sugar	Difference of Havg
P60	-1.351	-1.922	0.571	-2.395	-3.289	0.894
average	-2.033	-2.156	0.123	-3.497	-3.787	0.289
standard deviation	0.356	0.337	0.146	0.607	0.607	0.263

APPENDIX B

Modeling Group					
potato	weight(g)	temp(F)	density(g/cm^3)	H_{avg}	Target
P1	219	71.43333	1.418705	-3.92864	0
P1s	219	71.73333	1.418705	-4.25103	1
P5	175	71.8	1.457574	-4.50951	0
P5s	175	71.83333	1.457574	-4.67089	1
P6	154	71.76667	1.381331	-4.55935	0
P6s	154	71.96667	1.381331	-4.59765	1
P7	223	71.46667	1.529595	-3.60123	0
P7s	223	72	1.529595	-3.95509	1
P8	147	71.73333	1.52365	-4.23526	0
P8s	147	72.06667	1.52365	-4.35524	1
P10	137	71.5	1.331251	-3.45756	0
P10s	137	72.63333	1.331251	-3.96583	1
P11	186	72.26667	1.37706	-3.78877	0
P11s	186	73.43333	1.37706	-4.25132	1
P12	199	73.23333	1.160228	-3.69368	0
P12s	199	73.03333	1.160228	-3.81119	1
P13	213	72.33333	1.379836	-3.11749	0
P13s	213	73.66667	1.379836	-3.63609	1
P15	167	72.23333	1.153966	-3.41324	0
P15s	167	72.33333	1.153966	-4.21772	1
P16	194	72.13333	1.175145	-4.35723	0
P16s	194	72.76667	1.175145	-4.73431	1
P18	251	72.33333	1.463404	-3.88675	0
P18s	251	72.63333	1.463404	-3.992	1
P20	247	72.76667	1.355372	-3.8663	0
P20s	247	73.03333	1.355372	-3.95351	1
P21	183	72.66667	1.264526	-3.90021	0
P21s	183	72.8	1.264526	-4.26691	1
P23	162	73.13333	1.476154	-3.9761	0

Continued.					
potato	weight(g)	temp(F)	density(g/cm^3)	H_{avg}	Target
P23s	162	73.4	1.476154	-3.8468	1
P24	168	73.03333	1.339473	-3.13413	0
P24s	168	73.16667	1.339473	-3.67177	1
P27	175	71.56667	1.133668	-2.22677	0
P27s	175	71.6	1.133668	-3.36953	1
P28	194	71.7	1.256752	-3.28592	0
P28s	194	71.5	1.256752	-3.2728	1
P29	204	71.46667	1.243796	-2.24122	0
P29s	204	71.66667	1.243796	-2.44168	1
P31	134	71.9	1.302099	-3.14241	0
P31s	134	71.63333	1.302099	-3.25074	1
P33	188	71.56667	1.217884	-3.50391	0
P33s	188	71.7	1.217884	-3.74865	1
P34	211	71.53333	1.157828	-2.96718	0
P34s	211	71.53333	1.157828	-2.91794	1
P35	128	71.53333	1.263539	-3.49099	0
P35s	128	71.33333	1.263539	-3.89698	1
P37	129	71.36667	1.227931	-3.83042	0
P37s	129	70.6	1.227931	-4.08482	1
P38	208	71.1	1.347446	-3.25497	0
P38s	208	71.83333	1.347446	-3.46211	1
P39	168	70.96667	1.492555	-3.0733	0
P39s	168	70.83333	1.492555	-3.19716	1
P40	143	70.3	1.361195	-3.83851	0
P40s	143	70.5	1.361195	-4.03297	1
P44	135	71.5	1.199375	-3.18247	0
P44s	135	71.53333	1.199375	-3.35405	1
P46	150	71.66667	1.110532	-2.8618	0
P46s	150	71.63333	1.110532	-2.91927	1
p48	183	71.8	1.264526	-2.27494	0
P48s	183	71.33333	1.264526	-2.403	1

Continued.					
potato	weight(g)	temp(F)	density(g/cm^3)	H_{avg}	Target
P49	152	71.96667	1.266007	-3.01328	0
P49s	152	71.33333	1.266007	-3.11829	1
P50	173	71.86667	1.260801	-2.61487	0
P50s	173	71.33333	1.260801	-2.83871	1
P51	181	70.8	1.172537	-4.10334	0
P51s	181	70.8	1.172537	-4.30094	1
P52	148	70.867	1.232698	-3.2896	0
P52s	148	70.867	1.232698	-3.61639	1
P53	159	70.933	1.412597	-3.95844	0
P53s	159	70.933	1.412597	-3.76034	1
P54	165	70.633	1.465903	-3.84636	0
P54s	165	70.633	1.465903	-4.28787	1
P55	158	70.3	1.403713	-3.91839	0
P55s	158	70.3	1.403713	-3.85829	1
P56	182	70.2	1.326385	-3.96901	0
P56s	182	70.2	1.326385	-4.03053	1
P57	135	70.3	1.285036	-3.34062	0
P57s	135	70.3	1.285036	-3.68931	1
P58	149	71.133	1.323754	-3.23625	0
P58s	149	71.133	1.323754	-3.74549	1

Testing Group						
potato	weight(g)	temp	density(g/cm^3)	Havg	Target	Predicted probability
P2s	182	71.76667	1.515877	-4.20072	1	0.609629
P3s	203	71.8	1.352628	-4.626	1	0.684768
P4s	172	72.1	1.337081	-4.46905	1	0.657907
P9s	191	71.9	1.391983	-4.21398	1	0.612075
P14s	172	70.46667	1.637242	-4.01659	1	0.575144

Continued.						
potato	weight(g)	temp	density(g/cm^3)	Havg	Target	Predicted probability
P17s	246	73.13333	1.434252	-4.73824	1	0.70326
P19	195	72.36667	1.299322	-3.80969	0	0.53552
P22s	216	72.8	1.43925	-4.69238	1	0.69578
P25	134	73.43333	1.302099	-3.75682	0	0.525301
P26	183	71.6	1.33368	-2.78463	0	0.342284
P30s	168	71.6	1.160876	-3.72233	1	0.518622
P32	196	71.56667	1.354356	-3.17346	0	0.413044
P36	179	71.4	1.49089	-3.30614	0	0.438208
P41s	179	70.43333	1.40319	-3.45341	1	0.466511
P42	172	70.4	1.337081	-3.2453	0	0.426621
P43	192	71.7	1.243796	-2.39398	0	0.277625
P45	132	71.66667	1.256488	-2.9739	0	0.376071
p47	177	71.8	1.146624	-2.46103	0	0.28818
P59s	180	71.367	1.399271	-4.28465	1	0.625015
P60	140	71.433	1.33263	-3.42524	0	0.461075

VITA

Name: Chih-Chen Sun

Address: Department of Mechanical Engineering, Texas A&M University, College Station, Texas 77843-3123

Email Address: smallmo@gmail.com

Education: B.En., Mechanical Engineering, National Taiwan University, 2006
M.S., Mechanical Engineering, Texas A&M University, 2008