CS CAPSTONE FINAL REPORT

JUNE 1, 2018

EBOLA PREDICTION MODEL

PREPARED FOR

PROFESSOR BILL SMART

PREPARED BY

GROUP 34
EBOLA TEAM
CLAUDE MAIMON
BRIAN LEE HUANG
BIANCA BEAUCHAMP

Abstract

There are four major parts to this project, image processing, machine learning, production and evaluation. The image processing portion will import the thermal images from the camera, format the images, select necessary pixels, and summarize those pixels into a single temperature value. The machine learning portion of the project will create a mathematical model to represent the relationship between skin temperature and core body temperature and then test this model on different sets of data. The production portion will use the model that has been created to produce a estimated core body temperature from the estimated skin temperature and then produce a true or false output as to weather or not thatThere are four major parts to this project, image processing, machine learning, production and evaluation. The image processing portion will import the thermal images from the camera, format the images, select necessary pixels, and summarize those pixels into a single temperature value. The machine learning portion of the project will create a mathematical model to represent the relationship between skin temperature and core body temperature and then test this model on different sets of data. The production portion will use the model that has been created to produce a estimated core body temperature from the estimated skin temperature and then produce a true or false output as to weather or not that

CONTENTS

1 Introduction 2

2 Requirements Document 2

2.1 old requirement . . . 2

2.2 new requirements . . . 7

2.3 final Gantt Chart . . . 7

3 Design Document 7

3.1 original design document . . . 7

3.2 new design changes . . . 11

4 Tech Reviews 12

4.1 Brian's Tech Review . . . 12

4.2 Claude's Tech Review . . . 15

4.3 Bianca's Tech Review . . . 16

5 Weekly Blog Posts 19

5.1 Brian . . . 19

5.2 Claude . . . 21

5.3 Bianca . . . 22

6 Final Poster 23

7 Project Documentation 25

8 Recommended Technical Resources 27

9 Conclusions and Reflections 27

10 Appendix 1 27

11 Appendix 2 27

12 Conclusion 27

References 27

1 INTRODUCTION

There are four major parts to this project, image processing, machine learning, production and evaluation. The image processing portion will import the thermal images from the camera, format the images, select necessary pixels, and summarize those pixels into a single temperature value. The machine learning portion of the project will create a mathematical model to represent the relationship between skin temperature and core body temperature and then test this model on different sets of data. The production portion will use the model that has been created to produce a estimated core body temperature from the estimated skin temperature and then produce a true or false output as to weather or not that subject has a fever. The evaluation portion will evaluate how accurate the true or false output of the production portion is. The machine learning, production and evaluation portions will have user interfaces. The first piece of the project that will be developed is be the image processing portion since both the learning, and production code rely on processing the thermal image to get a skin temperature. The second piece that will be worked on is the machine learning portion since the model needs to be created and tested before it goes into production. The third piece that will be done is the production portion and the forth piece that will be done is the evaluation portion.

Who requested it?

Why was it requested?

What is its importance?

Who was/were your client(s)?

Who are the members of your team?

What were their roles?

What was the role of the client(s)? $I.e., did they supervise only, or did they participate in doing development$

2 REQUIREMENTS DOCUMENT

2.1 old requirement

Abstract

The end goal of this project is to end up with a research paper. The paper should outline the problem that the project is trying to solve, the steps taken to solve the problem and how successfully we were at solving the problem. The paper should also allow the project to be continued if someone chooses to. The whole process should be explained in detail allowing whoever wants to continue the project to continue without any problem. The main body of this research paper will be about the program that we develop to predict a person's core body temperature. The program should first be able to extract data from a thermal image. The data of the image should come from the top half, focusing on the head. It will then interpret the data to create a mathematical model that uses the temperature of a person's skin as data and analyzes that information and predicts what their core body temperature is. A high accuracy rate is not strictly required as that is not the point of the project, the goal of the is to determine if this method will be effective to detect whether a person is symptomatic with Ebola. A high accuracy rate is a good indicator that a mathematical model is a good way to predict, where a low accuracy rate indicates that we should look for an alternative method.

Introduction

Purpose

In the medical industry it is known that one of the first signs of illness is an elevated core body temperature. Currently, the only way to get this temperature is with contact sensors which put health care workers at risk of infection if a patient is ill. This project will aim to solve this problem by creating a device that will be able to quickly take a person's core body temperature from a distance using only stand-off sensors.

Scope

In Africa throughout the years there have been many different Ebola outbreaks. This project is part of a larger project that is trying to use robotics and automation to fight against Ebola. The National Institutes of Health are funding this project and we are also working with Medecins Sans Frontieres (Doctors Without Borders) in Brussels, Belgium who lead the response to the 2015-2017 Ebola outbreak. This project, if successful, will be part of an automated system that will sort people into two groups (symptomatic and asymptomatic) as they come in for care. This will lower the risk of health care workers and asymptomatic patients being exposed to Ebola.

Definitions, Acronyms and Abbreviations

OpenCV: an open source computer vision library.

Overview

We will be creating this automated system by using pictures from thermal imaging camera and analyzing them to estimate core body temperature. We are going to attempt to do this by taking an average temperature of the head of the person in the thermal image, taking their core body temperature with an ear thermometer and comparing the average temperature from the image to their actual core body temperature to build a mathematical model that will estimate core temperature from the thermal image temperature. To get the average temperature of the person's head from the thermal image we plan to isolate the head from the thermal image, make the pixels of the image into data points on a graph, get rid of the outliers, and take the average of the rest of the data points. We will also be collaborating with the mechanical engineering team to create this automated system. The mechanical team will be providing us with a thermal imaging camera and the physical structure that will hold the camera.

Overall Description

Product Perspective

Our program will be working with a thermal camera for the data collection. Our program will be working with a thermal camera for the data collection. We will also be using OpenCV for the image processing. Our program will also have a user interface that allows for easy data input. This user interface should be friendly enough where someone without an engineering background can easily use our program.

Product Functions

Processing Thermal Images

We will create a program that will process data taken from a thermal camera image. The program will then isolate the person in the picture and only take data from the upper body of the person. In order to isolate the person, we will use a specific background or a door frame to cut only the person person from the image. Once the image is fully processed the pixels will be used as data points. The temperature value will depend on the color of the pixels and the program will analyze the data from all the pixels and process it.

Processing Data from Pixels

After isolating the head of a person from the image the pixels will be used as data points for our program. The temperature of each pixel will graphed, to create a histogram which we can analyze. The extreme outliers in the histogram will not be counted towards the analysis as those data points could come from the background. We will then analyze the data to get an estimation of the person's temperature.

Collecting Data

To collect data we will use a thermal camera to take the temperature of their skin, mainly focusing on their head and neck. Every person that has their picture taken will also have their real core temperature checked through the ear. The two pieces of data will then be stored for the program to analyze.

Analysing Data from Thermal Image

Once we've collected enough data, we will statistically analyze it different ways. We will then look for the best statistical analysis that best connects the collected ear temperature to the predicted processed temperature from the thermal image. The statistical analysis will then be used as a baseline for a mathematical model to predict a person's core body temperature.

Mathematical Model

The data collected from the analysis will then be used to feed into a the mathematical model that predicts core body temperature. The accuracy of the model will improve over time as more data is fed into it.

User Characteristics

Our program should require no background knowledge in engineering to use. The usage of our program should be as simple as walking through a doorway and having a simple yes or no output to a screen. The hope of this project is to have it be deployed in areas where there is an Ebola outbreak. For this to be effectively deployed an expert should not be required to use it.

Constraints

Working with the Mechanical Team

For this project, we will be partnering up with a mechanical engineering team. Our project progress would be highly affected by the team's input. The mechanical team will be responsible for working with the camera and building the sensors set up. Our progress might be affected by the other team's progress.

Getting People to be Checked

Getting data will involve participants. We will need to check people for their temperatures. Participants can be people from our class but not other people. This constraint might make it harder to get sufficient data collection to create a comprehensive model.

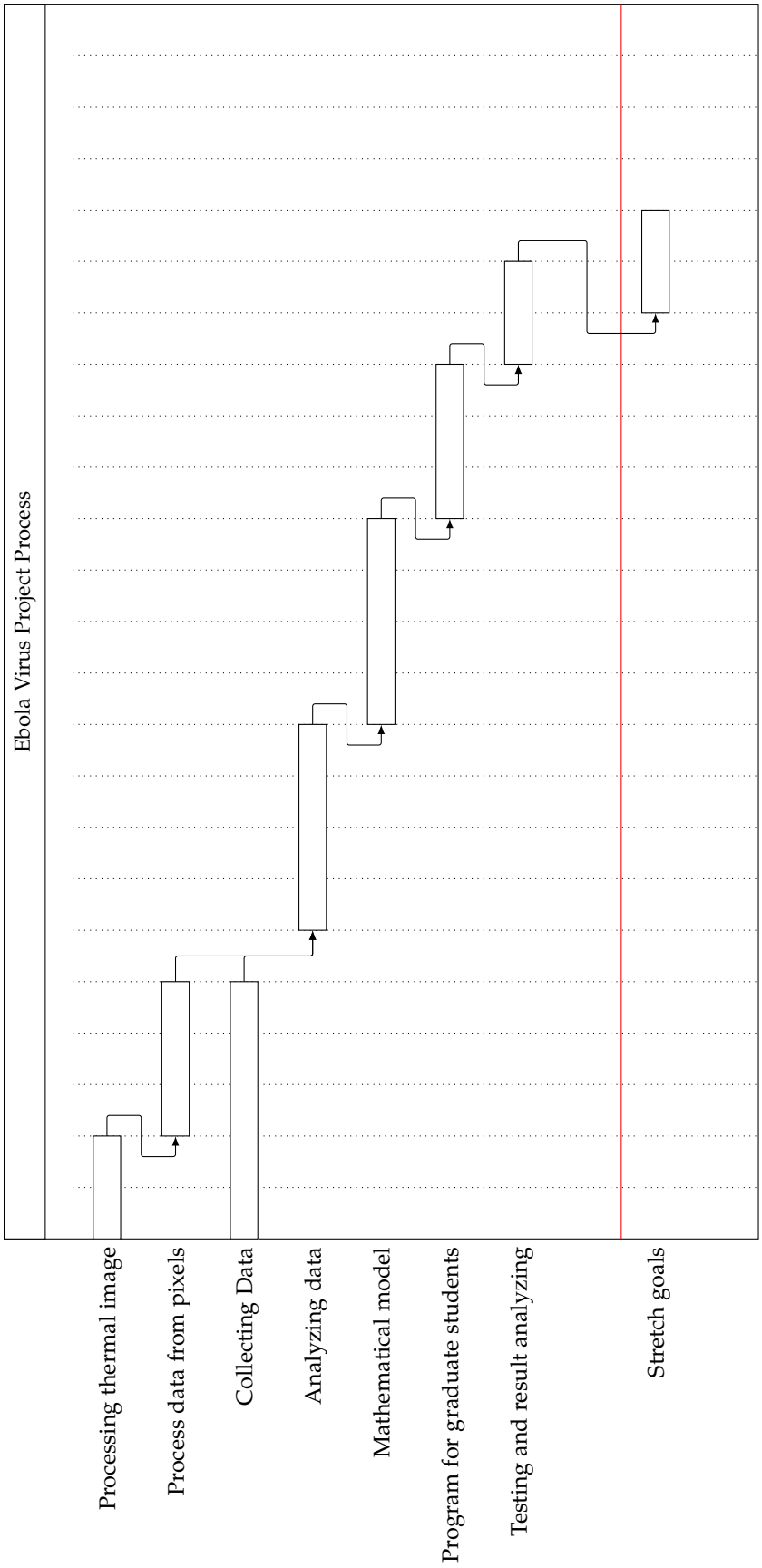
Specific Data Needed

In order for this project to work, we need to collect data from people. The data would be a set of ear temperature and a thermal image of a person.For the project to work, we need enough data from all spectrums. This means that we need to get data from people with different core body temperatures. We need to collect data from people with high and low core body temperatures. This mean people will either have to have a fevers or they would have to elevate their body temperature by exercise. This constraint will make it harder to collect large sets of data.

Limitations of the Camera

The accuracy of the camera will reflect on the research's results. If the camera is not accurate, the model will not be accurate. We might end up spending a lot of time creating an insufficient model.

Assumptions and Dependencies


Ebola Virus Project Process
Processing thermal image
Process data from pixels
Collecting Data
Analyzing data
Mathematical model
Program for graduate students
Testing and result analyzing
Stretch goals

Specific requirements

Our program will be able to take a thermal image and perform pixels selection. After that, the program will analyze the pixels in the image and produce an average temperature of the skin.The program will then run the observed skin temperature calculated from the camera through a mathematical model that will produce an estimated core body temperature.

The whole process will take less than one minute and will no have more than more than 40% of false negative results. In the end of the process, the program should provide a simple yes or no to whether or not the person's core body temperature is elevated.

Stretch Goals

We will go out and collect data to feed into the model, and make it able to predict more accurately.If we will achieve our baseline of 40% accuracy[1] early, we will work to lower the percentage of the false negative.

2.2 new requirements

add new changes

2.3 final Gantt Chart

add new chart

3 DESIGN DOCUMENT

3.1 original design document

Abstract

The end goal of this project is a research paper as well as a prototype of the software. The paper should outline the problem that the project is trying to solve, the steps taken to solve the problem and how successful the implemented solution was. The prototype of the software as well as the information in the paper should allow the project to be continued by someone else. The whole process should be explained in detail allowing whoever wants to continue the project to continue without any problem. The main body of the research paper will be about the program that is developed to predict a person's core body temperature using a thermal image. The program will first be able to extract data from a thermal image. The data from the image will come from the top half, and then be narrowed down to the head of the person. It will then use the data to create a mathematical model that will predict core body temperature. A high accuracy rate is not strictly required as this is an exploratory project. The main goal is to determine if this method will be effective to detect a elevated core body temperature. If a high accuracy rate is achieved then there this method should continue to be perfected. If the outcome is a low accuracy rate then this method may not be best and a new method should be considered.

Introduction

There are four major parts to this project, image processing, machine learning, production and evaluation. The image processing portion will import the thermal images from the camera, format the images, select necessary pixels, and summarize those pixels into a single temperature value. The machine learning portion of the project will create a mathematical model to represent the relationship between skin temperature and core body temperature and then test this model on different sets of data. The production portion will use the model that has been created to produce a estimated core body temperature from the estimated skin temperature and then produce a true or false output as to weather or not that subject has a fever. The evaluation portion will evaluate how accurate the true or false output of

the production portion is. The machine learning, production and evaluation portions will have user interfaces. The first piece of the project that will be developed is be the image processing portion since both the learning, and production code rely on processing the thermal image to get a skin temperature. The second piece that will be worked on is the machine learning portion since the model needs to be created and tested before it goes into production. The third piece that will be done is the production portion and the forth piece that will be done is the evaluation portion.

Importing Images and Image data

The first step for image processing is importing the image from the camera. It's important that the image will stay in the right format to keep the pixel data. This means that the image needs to be in a format which maintains the temperature data. The FLIR Tools software will be used to achieve this.

FLIR Tools is a software for importing and analyzing images from FLIR cameras. The software can be used to import images to a personal computer, search the image library using various filters, store search criteria and manipulate images. The software is free to use and has a simple installation process. FLIR Tools is suitable for the FLIR A315 camera. Other than importing images the tool can be used to create PDF reports, add header logos to images and sort the image folder by specific variables.[2]

This software is the best option because it's free and it offers all the features needed for this process. Even though FLIR Tools+ offers more features, it is not necessary and this option will make the project cheaper.

Other than importing images, the core body temperature will need to be collected from many people. These temperatures are needed to compare to the estimated skin temperature from the camera. People's temperature will be collected and stored in a file. Every temperature will be paired with an image, so it is clear which temperature belongs to which person.

The thermal camera may not arrive in time for data collection. If this occurs, black and white images taken from a web camera will be used as mock thermal images. The results from these images won't be accurate but they will make a good place holder so that work can be done.

Formatting Images

In order to process the thermal images, they have to be in the right format. After importing the images from the camera to the computer, the program will transfer the images into a two-dimensional array format. This format will allow for easy image processing. Every element in the two-dimensional array will hold a value of a pixel in the image. This structure will allow for easy processing of the pixel values.

The program will use OpenCV with Python to manipulate the image. OpenCV (Open Source Computer Vision Library) is a free open source computer vision software. It has interfaces for C++, C, Python, and Java and it supports various operating systems. OpenCV's library has more than 2500 algorithms which offer many features. Some of those features are facial recognition, gesture recognition, motion understanding, bio-medical analysis and more. The software is used all around the world, has a strong user community and offers technical support. OpenCV is a good option for manipulating the image because it is free and has a large user community. It has shown to have the best performance in comparison to other image processing libraries and there are many easy to understand tutorials available.[2]

Pixel selection

In order to get a good temperature estimation from the image, only the pixels from the warm parts of the subject's head are needed. The program will go through the pixel values in a two-dimensional array and select specific pixels. It will

only select values that are higher than 98.8 and lower than 105. The pixel values might not be represented as degrees. If they are not represented as degrees the pixel values that represent 98.8 and 105 degrees will be found and those values will be used instead. This approach will work because by only selecting these values it will, in theory, isolate only the warm parts of the subjects's head and it will leave out the cooler parts of the head as well as the background behind the subject.

Since this is a primarily a research project this method will tried first, but if is fails a different approach to selecting the pixels may be taken. Another approach may be to isolate the subject's head from the picture before getting the pixel values. This way the background would be removed prior to selecting the warmest parts of the face which would in theory eliminate the chance of the background being included in the data.

Summary Statistics

Once the pixels of skin have been isolated from the image, they need to be simplified into a single value that is a good representation of the entire set of data. This single value would represent the temperature of the subjects skin. This is important because the value be used in the model to relate to core body temperature.

The mean will be used for the summary statistics because it is the most accurate representation of all the data points and it is simple to find. In this case the mean used would be the arithmetic mean which is the average of the entire data set. The arithmetic mean is found by adding up all of the values in a data set and then dividing by the number of data points in the data set. The mean is the best used as a representation of the entire data set because it minimizes the sum of squared deviations from the typical value, meaning that it is a value that is the closest to all values in the set. This will be important because the value from the summary statistics will be what is used to relate the data from the camera to the measured value in the model.[3]

Model

The condensed data needs to be related to the measured value in order to produce an equation that represents the relationship between skin temperature and core body temperature. The value produced by the summary statistics represents the skin temperature of a subject and the measured core body temperature of the same subject will be provided. These two pieces of data will be taken from many subjects and then used to create a model. The model will then be able to take just the summary statistics, which is the subjects estimated skin temperature, and produce the core body temperature.

The model must be able to represent the relationship as accurately as possible. This means that the type of model used will need to be determined by trial and error since there is no known relationship. The order in which the different models are tried should go from least complex to most complex.

The simplest model is a linear regression model. This model is used for modeling the relationship between one dependant variable and one or more explanatory variables. This type of model is typically used for predictions, forecasting and error reduction. The predictive model is made by finding the equation of the line of best fit for a set of dependant and explanatory variables. Then the model can be given an explanatory variable and predict the dependant variable.

A slightly more complex model is a polynomial regression model. This model is used for modeling the relationship between dependant and explanatory variables as an nth degree polynomial. The model is made by fitting a nonlinear

relationship between explanatory variables and the corresponding conditional mean of dependant variables. Polynomial regression models are used for nonlinear phenomena.

A even more complex model is a ridge regression model. This is a type of regression model that is used when there are explanatory variables that have a high correlation. It is very similar to the linear regression but it is more complex since in order to account for the explanatory variables with a high correlation it uses the prediction errors.

The linear regression will bet the best model to try and test first because it is the simplest of all the models. If the linear regression does not produce a high enough accuracy during testing, the next model tried and tested will be the polynomial regression since it is slightly more complex. If the polynomial regression is not accurate enough, then the ridge regression will be tried and tested. If all three do not produce the accuracy that is desired then other models besides these three may need to be considered or it may not be possible to produce an accurate relationship between skin and core body temperature. [3]

Model Testing

Once the model has trained on the set of data that has both the subjects estimated skin temperature and their measured core temperature, it needs to be tested on a new set of data to determine how accurate the model is. This will be done by providing the model with only the estimated skin temperature and comparing the core temperature the model calculates to the measured core temperature. To quantify this comparison it is best to calculate the error between the calculated core temperature and the measured core temperature. This will need to be done for a large set of data to get a good idea of how well the model is working. The error calculation will provide valuable information that will allow for an accurate evaluation of any error in the relationship between the calculated core temperature and the measured core temperature.

Absolute error will be used for finding the error of this model. Any error at all is detrimental in this case and absolute error provides a simple evaluation of if the model is accurate or not. Absolute error averages the size of the error and it weights each error the same. This is done by subtracting a measured quantity from a calculated quantity and taking the absolute value of the result, repeating this for all sets of measured and calculated values, adding all of these values together and dividing by the number of sets. [3]

Production Mode

The production mode is the third piece of the overall project. The production mode will use the model produced by the learning portion of the project to analyze the image. The output of the production mode will be the estimated core body temperature produced by the model as well as a true (fever) or false (no fever) statement. This portion of code will be written in python like the other pieces of the project to eliminate any need for cross communication of languages. Although this is an important piece of the project most of it can only be done after the image analysis portion is complete. The production code also requires the model created by learning portion of the project. The production mode will use the image processing portion of the project to get the data from the image and then it will use the model created by the learning portion of the project to estimate the core body temperature of the subject.

The production mode should be accurate as possible since it will be used to predict a patient's core body temperature. This would mean keeping the false negative rate as low as possible. However, since this is primarily a research project it is not realistic to have a perfect program in such a small time frame. So the goal that has been established is a 40 percent false negative rate. [4]

Evaluation

After the production model is created it needs to be tested. The accuracy can be evaluated by examining either false positives or false negatives. In this case false negatives are critical to keep as low as possible since placing a sick person in a group of healthy people could result in spreading disease to the masses. False positives are still very important but not as important since having a healthy patient among unhealthy patients will not spread disease to the masses.

To gather data for the false negative and false positive rates, the binary output of the program will be collected into a text file. In another text file there will be the corresponding known binary values. A program will be created to parse the text files, count the number of false negatives and false positives and output the percentage of both false negatives and false positives.

User Interface

Once the project gets started a user interface is going to be one of the first things constructed. The initial user interface will have a command line user input and print out the results to the screen. The user interface will be created using python's file input and print statements. It will be up to the user to properly parse the input files and properly input the relevant information. This simple interface will be used for most of the project, if not the whole project. This will depend on how accurate the model becomes and how much time is left. If the model is very accurate and there is time left a nicer user interface will be created. As the project becomes more and more complete the user interface should also improve with the project. The user interface will slowly develop as the project continues to progress since more inputs and outputs may become necessary.

If there is time to improve the user interface a GUI (graphical user interface) will be created. Instead of using command line for input and output the user interface will allow the user to select the input file using a mouse and keyboard and then allow the user to choose where the output file should go. The output of the machine learning portion should output a text file that contains the model. The output of the production mode should output a temperature that is predicted from the model as well as a true or false statement to the screen.[4]

Conclusion

The goal of this project is to increase the safety of health care workers as well as the people coming to those workers for help. By taking a thermal image of the patient and using that data to predict their core body temperature, health care workers and healthy patients will not have to be within a close proximity to sick patients. To accomplish this task, thermal images will be taken, the images will be processed, a mathematical model will be made, the model will be tested, a production mode will be created and the production mode will be evaluated.

If successful, this project will create something that will be usable in the field for doctors without borders. If the accuracy is not high enough for use in the field, it will be handed over to graduate students to continue the project. Ultimately, if this project is successful and has accurate prediction rates it could reduce the spread of illnesses.

3.2 new design changes

add changes

4 Tech Reviews

4.1 Brian's Tech Review

Abstract

The goal of this project is to be able to determine a person's core body temperature just by taking a thermal image of them, and analyzing the data take from that picture. This requires many parts, but the three I am tasked with is the Production code, Statistical Analysis, and the User Interface. To create these piece of the project I need to choose a pieces of technology to create them. For the production code I chose to use python, and analysis I chose to use R, a language specifically designed for statistical computing and analysis. For the user interface I am choosing to use command line as it is something easy to implement through print statements and command line arguments. However the choice for the user interface may change in the future depending on how far the project gets. The most important feature about all my choices is that each option is open source meaning there will be no licensing issues.

Introduction

The goal of this project is to create a program that uses a thermal camera to measure a person's skin temperature, then use that data to predict their core temperature. An elevated core temperature is an indicator that the person is symptomatic with the Ebola virus. The model and the analysis will be what predicts the core body temperature given a skin temperature.

For this project I am tasked with creating the Production code, User Interface and the Statistical Analysis. The production code and the analysis are mostly done towards the end of the project, while the user interface can be done at almost anytime during the project. The user interface should be very simple and not require much experience to use it. For most of the project a command line user interface will be sufficient.

Production Model

Overview

The production code is the second step, and what would be the final product of this project. As inputs it will take in a model, then take in a picture to then analyze the picture using the model provided. The output of this will just output a core body temperature.

Critera

The production code should try and be as accurate as possible, but should try to keep the false negative rate as low as possible. False negative should kept as low as possible as it can be extremely detrimental to misdiagnose someone that is symptomatic and putting them in the healthy group, as opposed to misdiagnosing someone that doesn't have it and putting them the unhealthy group.

We have found other projects that are similar to our project. They took a thermal image of a person's ear and head to derive a person's core body temperature. In this paper they achieved a positive predicted value of 40

Potential Choices

For this aspect of the project we have chosen three different possibilities to create a model. They are python, MATLAB and R.

First Choice

Python is a versatile language that we planned to use from the very beginning. It is easy to use, and has very many useful functions that makes data manipulation easier. It also has many plugins and libraries that can be used for model creation. We also plan to use python for our image analysis, so it will be easier to use the same language for both of these portions to make it easy to incorporate it.

Second Choice

MATLAB is a computing language created by MathWorks which is primarily used for numerical computing. This makes it an obvious potential choice to analyze the data from the image. MATLAB has many functions built into it so we mostly do not write any of the functions ourself for the analysis. The main difference between Python and R is that MATLAB requires a licence.

Third Choice

R is language used for statistical analysis and graphing. It can easily analyze the data taken from the image. R also includes many packages, so we can simply add in any feature that we need in the future. One small potential problem that we may run into with R is to have it run with our image processing program. We would need to set up a properly formatted text document as the output of our image processing and then run R with the document as the input.

Discussion

The first and most important distinction to make between the three choices is that Python and R are both open source, while MATLAB is not. This in itself makes MATLAB not a good option for us. Python is more of a general purpose language, while R has a specific use. For inputting data R is much better than Python because it is simply a few simple commands, while in python we will most likely have to write our own parser. Analyzing the data will most likely be easier in R as its intended use is to analyze data, while python is just general purpose.

Conclusion

Matlab requires a licence while the other two choices are free, so we most likely will not use MATLAB. We will most likely use python as our image handling program will also most likely be written in python. This allows us to just combine the program rather than handling the input and output of two different languages. However R will also be very useful for analyzing the data from the image, but incorporating the pre made model could be difficult.

Statistical Analysis

Overview

The statistical analysis is the intermediate step between the mathematical model and the data processing of the thermal images. We will use different statistical analysis to determine which is best for the creation of the mathematical model.

Critera

The analysis should show some kind of trend in the data. It should be easy to tell what different factors affect a person’s core body temperature.

Potential Choices

For this aspect of our project we would use similar tools, if not the same tool for the mathematical model. Our choices for this piece are MATLAB, R, and Microsoft Excel.

First Choice

MATLAB can easily process huge amounts of data and create multiple different analysis of the data. It has many built in functions so the different analysis could simply be changed by changing a name, or a number of a function.

Second Choice

R is an obvious choice for as the whole point of R is statistical analysis. Just like MATLAB it can easily process huge amounts of data and create many different analysis. R also has many built in functions that do different statistical analysis to make the whole process easier. Data input and output would also be extremely simple with R as it simply uses a function call or two for data input and then another to process. Data output can also be setup to output to either a text file or an excel document.

Third Choice

Excel is another choice we had as it is a tool to create spreadsheets and analyze data. Data input is simple, and most of the analysis is simple. Output is also simple as it is all done through the Excel environment. No code has to be written and it is all done through the user interface of the program. The statistical analysis most likely would need to be done by hand, where we would need to set up equations using the cells as the data.

Discussion

In terms of power, Excel is lacking when compared to MATLAB and R. MATLAB and R are both stronger in than excel in the analysis portion as they both can import different libraries and packages if the built in tools are not enough. However both MATLAB and Excel require licences while R is open source. The versatility and the whole purpose of R make it the a clear choice for the statistical analysis. With easy input and output, being open source and its whole purpose being to analyze data.

Conclusion

R is the clear choice for this piece of the project as it is the most powerful and it is also free. Both MATLAB and Excel require licences that could become an issue in the future. Using an open source software makes it easier for us, and easier for anyone that intends to work on this in the future.

User Interface

Overview

The user interface is what we will be using to input our data and display the output. In the end we need to create something simple that can be used by someone without engineering or computer experience. However for most of the project a extremely simple UI can be used. A nicer UI is only required if our project gets into its end stage, where it is almost ready to be deployed in the field. If we do not get to that stage, a graduate student may take over the project and for them I command line UI will be sufficient.

Critera

For our end product we need something simple and clear to use. The output of the program should be displayed with a simple yes or no. For the graduate student version, it should be simple to input data and have a clear output showing what it is doing. The usage of should be quick and clean so data input can be fast.

Potential Choices

For this piece of the project depending on where we are, we could use command line. However for the simpler version we chose to use either Java or Python. Command line is a good option because it can be written in almost any language and does not require any extra libraries or packages to install. Java and Python are also good options for a nicer looking UI because they have so many libraries and packages available to them.

First Choice

Command line is an easy choice for someone with a background in computer science. If a graduate student were to take over this project in the end a simple command line UI should be sufficient. A command line UI is easy to create, compatible with almost any language and easy to use, with a slight learning curve. Inputting data can be done with some command line arguments, while outputting can just be done with a couple of print statements.

Second Choice

Python is a good choice for UI design as it is easy and simple to use and there are many different libraries and toolkits available for python User Interface design. Python is also a good option for us in particular as we have planned to create the rest of the project in Python. This means that there should be no miscommunication between the languages when putting all the pieces together.

Third Choice

Java, just like Python has many libraries and toolkits built just for creating an user interface. One problem we may run into with using Java is the communication between the other parts of our project. We intend to create the other pieces using Python so creating a Java UI requires us have the two languages communicate which could cause issues.

Discussion

The command line interface will be easy to create as it can be created inside the main program.The other choices will have some learning required as the we would need research which toolkit we would want to use, and how to use the toolkit. However if we were to use one of the tool kits the user interface would be more robust and friendlier to the input and output of data.

Conclusion

Depending on where the project ends we can choose what kind of UI we want to make. For most of the project we will most likely be using a command line UI, where all the input and output is done from the program itself. In the end we will most likely create the UI using Python, as it will most likely be the easiest because we have already planned on using python.

4.2 Claude's Tech Review

add my tech review

4.3 Bianca's Tech Review

Abstract

There is currently no way to take a persons temperature without being within a close proximity to their bodies, putting health care workers at a great risk of infection. The purpose of this project is to reduce this risk by creating a device that will be able to quickly take a persons core body temperature from a distance using a thermal camera. In order to do this the thermal image must be processed and a model needs to be created. To process the image the pixel data must first be extracted and sorted, then the outliers will be removed. To create the model, summary statistics will be preformed to consolidate the data into a single value. That value will be used to model the relationship between skin temperature and core body temperature. Then the model will be tested by calculating the error of the model. This document will discuss some of the possible options and come to a conclusion best option for implementing the summary statistics, model, and model test. The best option for the summary statistics was determined to be the mean because it is the most accurate option and it is the simplest to find. The best option for the model was determined to be the linear regression because it is the simplest type of model and using it is a good starting place to see if it works of if a more complex model is needed. The best option for testing the model is absolute error because it provides a simple evaluation of if the model is accurate or not.

Summary Statistics

Overview

Once the pixels of skin have been isolated from the image, they need to be simplified into a single value that is a good representation of the entire set of data. This single value would represent the temperature of the subjects skin. There are three good options to consider for this purpose. The first is taking the median of the data, the second is to find the variance of the data and the third is to find the mean of the data.

Criteria

The option chosen has to be the most accurate representation of the data. This is important because this value will represent the skin temperature and will be used in relation to core body temperature. The model that creates the relationship is the most important part of this project. Therefore, the value produced by this step must be a good representation of the skin temperature.

Potential Choices

Median

The median is the value that separates the upper and lower halves of the data. The median is found by sorting all of the data points from least to greatest or from greatest to least and then finding the value that is in the middle of this set of data. If the number of data points is odd then a value will exist in the middle of the data set but if the number of data points is even the two data points that are closest to the middle are averaged. One advantage of using the median is that it minimizes the impact of outliers in its representation of the data set.[5]

Variance

The variance measures how far a set of random values from the data set are from the average value of the data set. Variance is calculated by squaring the standard deviation of the set of data. The variance is typically used to identify the causes of variability in a data set.[6]

Mean

In this case the mean used would be the arithmetic mean which is the average of the entire data set. The arithmetic mean is found by adding up all of the values in a data set and then dividing by the number of data points in the data set. The mean is the best used as a representation of the entire data set because it minimizes the sum of squared deviations from the typical value, meaning that it is a value that is the closest to all values in the set. [7]

Discussion

The key characteristic of the median is that it minimizes the impact of the outliers in that data set. This is not very important in this case because by the time the data is ready have the summary statistics taken, the outliers have already been removed. The key characteristic of the mean is that it is the best representation of every value in the data set. This is very important because the value produced by the summary statistics will be used to represent the the data set in the model. The key characteristic of the variance is identifying the causes of variability. This is not very useful in this case because the value from the summary statistics will be used to represent the data and not analyze it.

Conclusion

The mean is the best choice for the summary statistics because it is the most accurate representation of all the data points and it is simple to find. This will be important because the value from the summary statistics will be what is used to relate the data from the camera to the measured value in the model.

Modeling

Overview

The condensed data needs to be related to the measured value in order to produce an equation that represents the relationship between skin temperature and core body temperature. The value produced by the summary statistics represents the skin temperature of a subject and the measured core body temperature of the same subject will be provided. These two pieces of data will be taken from many subjects and then used to create a model. The model will then be able to take just the summary statistics, which is the subjects estimated skin temperature, and produce the core body temperature.

Criteria

The model must be able to represent the relationship as accurately as possible. This means that the best option for the model will need to be determined by trial and error since there is no known relationship. The order in which the different models are tried should go from least complex to most complex.

Potential Choices

Linear Regression

A linear regression model is used for modeling the relationship between one dependant variable and one or more explanatory variables. This type of model is typically used for predictions, forecasting and error reduction. The predictive model is made by finding the equation of the line of best fit for a set of dependant and explanatory variables. Then the model can be given an explanatory variable and predict the dependant variable.[8]

Polynomial Regression

A polynomial regression model is used for modeling the relationship between dependant and explanatory variables as an nth degree polynomial. The model is made by fitting a nonlinear relationship between explanatory variables and the corresponding conditional mean of dependant variables. Polynomial regression models are used for nonlinear phenomena.[9]

Ridge Regression

Ridge regression is a type of regression model that is used when there are explanatory variables that have a high correlation. It is very similar to the linear regression but it is more complex since in order to account for the explanatory variables with a high correlation it uses the prediction errors.[10]

Discussion

The linear regression is the least complex of the models since it is finding a simple line to represent the relationship between core and skin temperatures. The polynomial regression is mildly complex since it is used for modeling nonlinear relationships. The ridge regression is more complex since it is similar to the linear regression but considers independent variables that are correlated and takes the error into account.

Conclusion

The linear regression will bet the best model to try and test first because it is the simplest of all the models. If the linear regression does not produce a high enough accuracy during testing, the next model tried and tested will be the polynomial regression since it is slightly more complex. If the polynomial regression is not accurate enough, then the ridge regression will be tried and tested. If all three do not produce the accuracy that is desired then other models besides these three may need to be considered or it may not be possible to produce an accurate relationship between skin and core body temperature.

Model Testing

Overview

Once the model has trained on the set of data that has both the subjects estimated skin temperature and their measured core temperature, it needs to be tested on a new set of data to determine how accurate the model is. This will be done by providing the model with only the estimated skin temperature and comparing the core temperature the model calculates to the measured core temperature. To quantify this comparison it is best to calculate the error between the calculated core temperature and the measured core temperature. This will need to be done for a large set of data to get a good idea of how well the model is working.

Criteria

The calculation used to find the error needs to be the best representation of the accuracy of the model. To do this it needs to provide valuable information that will allow for an accurate evaluation of any error in the relationship between the calculated core temperature and the measured core temperature.

Potential Choices

Absolute Error

Absolute error averages the size of the error and it weights each error the same. This is done by subtracting a measured quantity from a calculated quantity and taking the absolute value of the result, repeating this for all sets of measured and calculated values, adding all of these values together and dividing by the number of sets.[11]

Squared Error

Squared error measures the differences between the calculated and measured values by putting a larger emphasis on values that are more consistent throughout the data set. This is done by subtracting a measured quantity from a calculated quantity and squaring the result, repeating this for all sets of measured and calculated values, adding all of these values together and dividing by the number of sets.[11]

Root Mean Square Error

Root mean square error represents the spread of the calculated and measured values from each other. This is done by subtracting a measured quantity from a calculated quantity and squaring the result, repeating this for all sets of measured and calculated values, adding all of these values together, dividing the result by the number of sets, and then taking the square root of the result.[12][13]

Discussion

The absolute error represents the size of the error of all the data sets and puts the same emphasis on all data sets. Unlike the absolute error, the squared error puts more emphasis on the errors that occur more often in the data sets. The root mean square error measures the magnitude of the error much like the absolute error but it puts a larger emphasis on larger errors.

Conclusion

The absolute error is the best error calculation for finding the error of this particular model. Any error at all is detrimental in this case. Putting emphasis on larger errors or on errors that occur more often won't offer any extra benefit overall. They do offer benefits for fine tuning the model because they offer extra information but are not needed to know the overall error.

5 Weekly Blog Posts

5.1 Brian

Fall

Week1

test

Week2

test

Week3

test

Week4

test

Week5

test

Week6

Week7

Week8

Week9

Week10

Winter

Week1

Week2

Week3

Week4

test

Week5

Week6

Week7

test

Week8

Week9

Week10

Spring

Week1

Week2

Week3

test

Week4

Week5

Week6

test

Week7

Week8

Week9

Week10

test

5.2 Claude

Fall

Week1

test

Week2

test

Week3

test

Week4

test

Week5

test

Week6

Week7

Week8

Week9

Week10

Winter

Week1

Week2

Week3

Week4

test

Week5

Week6

Week7

test

Week8

Week9

Week10

Spring

Week1

Week2

Week3

test

Week4

Week5

Week6

test

Week7

Week8

Week9

Week10

test

5.3 Bianca

Fall

Week1

test

Week2

test

Week3

test

Week4

test

Week5

test

Week6

Week7

Week8

Week9

Week10

Winter

Week1

Week2

Week3

Week4

test

Week5

Week6

Week7

test

Week8

Week9

Week10

Spring

Week1

Week2

Week3

test

Week4

Week5

Week6

test

Week7

Week8

Week9

Week10

test

6 Final Poster

COLLEGE OF ENGINEERING | Electrical Engineering and Computer Science | CS34

Screening for Ebola

Saving Lives Through Thermal Imaging

The Ebola Threat

- 11,325 people died in the most recent outbreak that lasted from 2014 to 2016
- People in contact with Ebola patients are at the highest risk of infection
- Symptoms include: Fever, headache, muscle pain, weakness, fatigue, diarrhea, vomiting, abdominal pain and unexplained bleeding
- It is transmitted through bodily fluids or objects that came in contact with an infected person, bat or primate

Figure 1: Doctors training for deployment in the 2014 outbreak of Ebola. http://news.trust.org//item/20141008090039-aydvf?view=print

Thermometer

- Better accuracy
- Cheaper
- Easier to access
- Requires close proximity to patients
- Takes longer
- Requires a physician

Thermal Camera

- Faster
- No need for an operator
- Safer for both doctors and patients
- Less accurate
- Expensive ($11,000)
- Difficult to purchase

Oregon State University

Current Problem

Currently, to check for Ebola, doctors must use thermometers to check core body temperature in patients. This process is slow and can potentially lead to infection of staff members and patients. With a large volume of patients, as there was in the recent Ebola outbreak, this method is extremely inefficient and can contribute to the spread of the Ebola Virus.

Our project aims to create a device that would be able to quickly estimate a person's core body temperature. The device would work from a distance using a thermal camera.

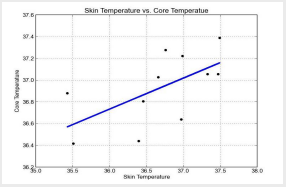


Figure 3: Mathematical Model trying to fit a line to the data points.

Future Use

The program as well as the structure is a proof of concept. If we'll evaluate that thermal image analysis is a reasonable method to detect Ebola, it will then be improved by graduate students and possibly used in the field.

In the future, the device will be able to direct patients to a specific direction. If a patient's core body temperature is elevated an arrow will light up, indicating them to go one way, otherwise an arrow would light up telling them to go the other way. This essentially quarantines the patients, reducing the spread of the disease.

Results

The model is highly dependent on data so our current model is not very accurate. The current model is a prototype so that when there is more data available in the future it will be able to generate better results.

Figure 2: Thermal Image of a person next to a coffee cup

Thermal Imaging

The program connects to a thermal camera, take an image, and convert that image into a CSV file. The CSV file holds the temperatures of each pixel in the image. Then, the program finds the head of the person and produces the average temperature of those pixels. This average temperature is then passed to the model.

Mathematical Model

Data was collected and used to create a mathematical model that can predict core body temperatures. Once the model was created, it was then used to predict core body temperature based on skin temperature from thermal images.

Production Mode

The production mode is the final piece of this project. It combines the thermal imaging and the model parts. This creates a program that is meant to be used in the field which does all of the tasks at once. It takes the picture, extracts the important data from it, and then passes it through the model to output a predicted core body temperature.

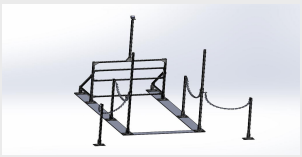


Figure 4: The Mechanical Team's structure that will hold the thermal camera.

Research Team

In the picture from left to right: Brian Huang, Bianca Beauchamp and Claude Maimon

Team Members

- Bianca Beauchamp beauchbi@oregonstate.edu
- Claude Maimon maimonc@oregonstate.edu
- Brian Huang huangbr@oregonstate.edu

Supporting Client

- Bill Smart smartw@oregonstate.edu

Advisors

- Christopher Bollinger bollingc@oregonstate.edu
- Austin Whitesell whitesea@Oregonstate.edu

Affiliation

- Work partially supported by NIH award EB024330-03
- Doctors without borders

NIH

MEDECINS SANS FRONTIERES

7 Project Documentation

Code

Overview Of The Code

Our code has two main modes. One is a learning mode and one is a production mode. In the learning mode, the images processing code and the model code are used separately. In the production mode, the separate pieces of code are used together to output an estimated temperature.

In the learning mode, the image processing code processes a folder of CSV files that were taken in advance. This mode is to be used when trying to collect data to train the model. While using this mode, the user should collect the actual temperature of the people that are being photographed. After running the image processing code on the CSV folder, the thermometer temperatures need to be added to the file that the code produces. Then, that file (that holds both the actual and estimated temperature) needs to be sent into the model to create a new model.

In the production mode, the whole process happens at the same time. The code can only run while the camera is connected to the computer and can only run on Linux environment. When running the code, the camera will take a few pictures in a row, our code will process them and produce an estimated core body temperature.

Image Processing

The image processing code is used to process an image CSV file and produce a temperature mean value. That value can later be sent to the model code to create a model. The code finds the head of the person in the image, calculates the mean of the head pixels and returns that value. This code doesn't include the process of connecting to the camera and producing a CSV from the image. That code is available under the production mode.

Running the Program

Correcting Values

Finding the Person's Head

Getting the Mean

Output

Mode

Our Model

Other possibility

Production Mode

Command Line

Testing

PROBLEMS AND SOLUTIONS

Camera

Data

RESULTS

RECOMMENDATIONS FOR FUTURE

Model

We are currently using a least squares regression line for our model creation. This creates a simple linear line which produces a slope and intercept to correlate a person's skin temperature to their estimated core body temperature. This simple method could be the best method, but we did not much data to create a more robust model. Multiple linear regression is the next logical step to test for the model. It can account for more variables that could possibly change someone's core body temperature. Some of these variables could be the weather, the ambient temperature, and the humidity.

Camera

The camera must be calibrated correctly in order for this project to work. Currently, the code uses the default calibration. We were stuck on this problem through the whole project. We couldn't find a way to get the camera to work correctly. It also wasn't in our requirements to calibrate it. We think that a possible fix to the problem might be to take more pictures and videos with the camera, add to them their temperatures, and use those as calibration files. We weren't able to do this, but it might help increasing the accuracy of the camera.

Data collection

Code Modifications

CONCLUSIONS

This project relies heavily on data. As mentioned in this report, we weren't able to collect good data. Both the camera problems and the nature of the needed data made it hard to achieve that goal. We believe that once the camera is fixed

and enough data is collected, our code can be used to estimate a core body temperature of a person using a thermal image. There are some code modifications that need to be done. Also, new prediction models need to be tested. But more importantly, a lot of data needs to be collected for this project to succeed.

8 Recommended Technical Resources

9 Conclusions and Reflections

10 Appendix 1

11 Appendix 2

12 Conclusion

The goal of this project is to increase the safety of health care workers as well as the people coming to those workers for help. By taking a thermal image of the patient and using that data to predict their core body temperature, health care workers and healthy patients will not have to be within a close proximity to sick patients. To accomplish this task, thermal images will be taken, the images will be processed, a mathematical model will be made, the model will be tested, a production mode will be created and the production mode will be evaluated.

If successful, this project will create something that will be usable in the field for doctors without borders. If the accuracy is not high enough for use in the field, it will be handed over to graduate students to continue the project. Ultimately, if this project is successful and has accurate prediction rates it could reduce the spread of illnesses.

References

[1] C. M., L. P., L. L., and C. H., "Mass screening of suspected febrile patients with remote-sensing infrared thermography: Alarm temperature and optimal distance," http://www.sciencedirect.com/science/article/pii/S0929664609600176, 2008, (Accessed on 11/03/2017).

[2] C. Maimon, "Claude's technical review," https://github.com/maimonc/Ebola-Virus-Project/blob/master/Documents/TechReview/Claude_Maimon_Tech_Review/Tech_Review_Claude_Maimon.pdf, 2017.

[3] B. Beauchamp, "Bianca's technical review," https://github.com/maimonc/Ebola-Virus-Project/blob/master/Documents/TechReview/Bianca_Beauchamp_Tech_Review/Technology_Review.pdf, 2018.

[4] B. Haung, "Brian's technical review," https://github.com/maimonc/Ebola-Virus-Project/blob/master/Documents/TechReview/Brian_Huang_Tech_Review/Brian_Huang_Tech_Review.pdf, 2017.

[5] "Median," https://en.wikipedia.org/wiki/Median, (Accessed on 1/10/2018).

[6] "Variance," https://en.wikipedia.org/wiki/Variance, (Accessed on 1/10/2018).

[7] "Mean," https://en.wikipedia.org/wiki/Mean, (Accessed on 1/10/2018).

[8] "Linear regression," https://en.wikipedia.org/wiki/Linear_regression#Applications_of_linear_regression, (Accessed on 1/10/2018).

[9] "Polynomial regression," https://en.wikipedia.org/wiki/Polynomial_regression#Interpretation, (Accessed on 1/10/2018).

[10] S. Ray, "7 types of regression techniques you should know!" https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/, August 2015, (Accessed on 1/10/2018).

[11] "Squared or absolute? how different error can be." http://archive.is/0hHpF, September 2013, (Accessed on 1/10/2018).

[12] S. Holmes, "Rms error," http://statweb.stanford.edu/~susan/courses/s60/split/node60.html, November 2000, (Accessed on 1/10/2018).

[13] "Mae and rmse-which metric is better?" https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d, March 2016, (Accessed on 1/10/2018).