AMERICAN UNIVERSITY OF BEIRUT

AI SYSTEM IN THE REAL WORLD: CAPTURE CHILDREN FOOD EXPOSURE USING WEARABLE CAMERAS

by
Yorgo Toni Zoughby

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Science
to the Department of Computer Science
of the Faculty of Arts and Science
at the American University of Beirut

Beirut, Lebanon
June 2020

AMERICAN UNIVERSITY OF BEIRUT

AI SYSTEM IN THE REAL WORLD: CAPTURE CHILDREN FOOD EXPOSURE USING WEARABLE CAMERAS

by

Yorgo Toni Zoughby

Approved by:

Dr.Shady Elbassuoni, Assistant Professor — Advisor
Computer Science

Dr. Fatima Abu Salem, Associate Professor — Member of Committee
Computer Science

Dr. Hala Ghattas, Associate Research Professor — Member of Committee
Epidemiology and Population Health

Dr. Hazem Hajj, Associate Professor — Member of Committee
Electrical and Computer Engineering

Date of thesis defense: June 19, 2020

From: Hala Ghattas
Sent: Thursday, July 2, 2020 10:08 AM
To: Shady Elbassuoni
Cc: Yorgo Zoughby (Student)
Subject: RE: My Thesis signature delegated to Prof. Shady Elbassuoni

Dear Shady,
This email is to authorize you to sign the Thesis Approval Page for Yorgo Zoughby's thesis on my behalf.
Please let me know if anything else is needed from my side.
Very best wishes and congratulations,
Hala

Hala Ghattas, PhD
Associate Research Professor &
Interim Director, Center for Research on Population and Health
Faculty of Health Sciences
American University of Beirut
Tel: +961 1350000 Ext 4679

From: Hazem Hajj <hh63@aub.edu.lb>
Sent: Wednesday, July 1, 2020 11:48 PM
To: Shady Elbassuoni <se58@aub.edu.lb>
Cc: Yorgo Zoughby (Student) <ytz00@mail.aub.edu>; Hala Ghattas <hg15@aub.edu.lb>; Fatima Abu Salem <fa21@aub.edu.lb>; Hazem Hajj <hh63@aub.edu.lb>
Subject: My Thesis signature delegated to Prof. Shady Elbassuoni

To whom it may concern,

I authorize you (Prof. Shady Elbassuoni) to sign on my behalf the Thesis Approval Page of Yorgo Zoughby.
Congratulations!
Please let me know if you need any additional information from my side.

Hazem

From: "Yorgo Zoughby (Student)" <ytz00@mail.aub.edu>
Date: Tuesday, June 30, 2020 at 1:14 PM
To: Shady Elbassuoni <se58@aub.edu.lb>, Hala Ghattas <hg15@aub.edu.lb>, Hazem

From: Fatima Abu Salem
Sent: Tuesday, June 30, 2020 4:47 PM
To: Shady Elbassuoni
Cc: Yorgo Zoughby (Student)
Subject:

Hello,

I authorize Dr. Shady Elbassuoni to sign on my behalf the Thesis Approval Page of Yorgo Zoghbi

Best,
Fatima

AMERICAN UNIVERSITY OF BEIRUT

THESIS, DISSERTATION, PROJECT
RELEASE FORM

Student Name: Zoughby Yorgo Tomi
Last First Middle

☒ Master's Thesis ○ Master's Project ○ Doctoral Dissertation

☐ I authorize the American University of Beirut to: (a) reproduce hard or electronic copies of my thesis, dissertation, or project; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

☒ I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of it; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes after: **One ___ year from the date of submission of my thesis, dissertation or project.**
Two ___ years from the date of submission of my thesis , dissertation or project.
Three X years from the date of submission of my thesis , dissertation or project.

Yorgo Zo
Signature

2/7/2020
Date

This form is signed when submitting the thesis, dissertation, or project to the University Libraries

Acknowledgements

Special thanks are for Dr. Shady Elbassuoni for his great help and support in the technical work, for Dr. Hala Ghattas for her assistance and support in the practical field work and last but not least for Dr. Fatima Abu Salem and Dr. Hazem Hajj for their supportive and constructive feedback.

My recognition and gratitude are addressed to the American University of Beirut, the department of computer science and The CRPH (the Center for Research on Population and Health) for its financial support and for providing all the hardware tools needed.

An Abstract of the Thesis of

Yorgo Toni Zoughby for Master of Science
Major: Computer Science

Title: AI System in the Real World: Capture Children Food Exposure Using Wearable Cameras

Children's dietary habits are influenced by complex factors. Identifying community-level influencers and measuring their effect is traditionally based on self-reported data and prone to measurement error. In addition, researchers are usually unable to accurately capture chiddren's dietary habits and food exposure using traditional surveying techniques. We propose to develop a culturally acceptable AI based data collection system that objectively captures school-children's exposure to food items, food ads, markets, etc... We engaged students, parents and school staff in a user-centered study for the food exposure design of school children through a feasible model in Beirut and Tunis. Findings suggest that wearable cameras are suitable when used for a limited period (24 hours max). Nonetheless, some ethical challenges were raised related to privacy, confidentiality and anonymity along with suggestions and solutions on how to address them with technology in a scientific objective manner and how to make such an AI system acceptable so that it can be used in similar research studies. We also survey a list of the most popular wearable cameras, and present their pros and cons. We train a machine learning model for automatically detecting food-exposure images and blurring faces that will be automatically captured by wearable cameras. This report also discuss how we automatically collected from the Web the training dataset for training such an ML model and how we managed to overcome any possible bias in our model and report on its performance. Furthermore, we explain and expose the software tool that packs the AI system that we developed and we showcase its simple-to-use interface and we report on its throughput, efficiency and hardware requirement. Finally, we discuss how we deployed this AI based system in the real world in a real study and we provide a high level of analysis on the collected data from the real world.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The last two decades have been marked by substantial dietary and lifestyle changes in the Arab region, where the prevalence of overweight and obesity among children is estimated to have doubled during this period [15, 16]. Greater Beirut area, in Lebanon and Greater Tunis, in Tunisia are two urban agglomerations located in middle-income countries where estimates of childhood overweight match those of high-income countries, reaching about 30% in Lebanon and 20% in Tunisia [16].

Assessment of factors influencing food choices among children and adolescents is challenging, and is usually based on self-reported data, which are prone to measurement error [36–38]. Considering the multiplicity of factors influencing food choices, innovative tools are needed to better assess these among children. A review highlights the gap in studies assessing the impact of children's usual trajectories to and from schools on their actual exposure to obesogenic environments [39]. These trajectories include exposures within a child's immediate environment, most commonly; home, school, and the neighborhoods around them. The multifactorial nature of food choice drivers necessitates the use of innovative tools that foster a 'people-based' approach to measuring these exposures [39]. Technology-based tools enable an objective and comprehensive measurement of children's nutrition-related behaviors and experiences around food [37, 40, 41]. These include factors in the child's environment such as setting, social interactions and media exposure [40]. Digital technologies are also favored over traditional data collection tools since they focus on children's individual lived experience and engage younger participants in documenting their own behavior, which may also lead to more accurate and representative data on their food experiences [41, 43]. These technologies have been positively received by younger participants, and are considered to be feasible and acceptable [38, 40, 41, 44].

Evidence suggests that children's dietary habits are established early in life and continue into adolescence and beyond [21, 22]. Children's food choices are influenced by school, family and community environments but food exposures in these environments are difficult to capture [3–6]. In order to quantify children's exposures to such factors, innovative tools such as wearable cameras can be used. These cameras can continuously capture videos for long periods or capture images every few seconds. The captured footage acts as a recorded diary of what the the child wearing the camera is being exposed to. Since the amount of footage can be arbitrarily large, and only a small percentage corresponds to food exposure related footage, machine learning can be employed to identify only images (or video frames) corresponding to food exposure. These include images of food items but also food advertisement, food markets, etc. Once such food exposure related images are identified, they can be further studied to assess how healthy the child's diet is, what type of food advertisement the child is exposed to, what food items the child is exposed to in markets or at home, etc..., as well as the measurement of the association with children's nutritional status and health outcomes.

This study includes multiple components and employs a mix of research methods, including standard surveys, checklists, a choice experiment in the form of a game on a tablet, and micronutrient assessment from a finger-stick. Additionally, an informative phase has guided the development of an acceptable data collection model that documents children's immediate surroundings, objectively and comprehensively. Children will wear cameras that automatically capture images of their exposure to the obesogenic environment on their trajectories to and from school, with various technological safeguards in place to minimize exposure to risks.

To assess the acceptability of using wearable cameras to capture food exposure among school children, we conduct a user study involving school children (aged 10-12y), parents and school staff in twelve interactive workshops in two different urban cities, namely Beirut and Tunis. These workshops included discussions, mind-mapping and storyboarding to identify challenges associated with wearable cameras and to inform the design of a device that meets the cultural and ethical requirements specified. Discussions were audio-recorded, transcribed and analyzed thematically along with the mind-maps and storyboards. Commonly reported challenges were: invading children's and third-parties' privacy; distracting school-students; and obtaining biased data. Safety issues related to photo-capture in public (third-party questioning/aggression) only emerged in Tunis since Lebanese parents stated that children were rarely unsupervised in public. To overcome these challenges, participants suggested wearable cameras capturing exposures automatically, for a short period. Two rounds of image-filtering were proposed to safeguard privacy: automated selection of images with food exposures, followed by parental manual screening. To protect anonymity, participants suggested automated blurring of faces.

We also survey a list of wearable cameras currently available online and highlight their pros and cons. To do this, we started by identifying which cameras have been used in similar other studies and found out that most of these cameras are currently out of stock. We then surveyed the most recommended wearable cameras that are nowadays used for life logging purposes and we were able to identify a set of these cameras. We ordered a sample of these cameras and experimented with them. We will be providing an in depth survey of all the cameras that were tested by the research team and we will highlight the pros and cons of each one.

In addition, we build a machine learning model for automatic detection of images related to food exposure. To train such a model, we crawled a dataset consisting of 700,000+ images from the Web. We used an open source web crawler that uses Google and Bing search queries to collect the images. The images in our dataset are automatically labeled based on the search query that returned each image. We then applied some filtering and pre-processing on the constructed dataset such as removing any duplicates, or watermarked or low resolution images. This in turn reduced the number of data instances in the dataset to approximately 500,000 images. Also, we did some dataset analysis and made sure that the dataset is representative and will not lead to a biased food exposure detection model. After building the dataset, we trained a series of deep learning models based on the constructed dataset to automatically detect images related to food exposure and blur faces in these images if they do exist.

Finally, We developed a AI based software tool which basically contains the food exposure and the face blurring models and handle the data transfer from the cameras to a computer used by the researchers where the tool will automatically filter the data and delete the images without food exposure and blur all the faces that appears in the food exposure images. This AI based software tool will be used in the real world by scientist and was designed with a set of specifications and requirements that makes the usage of a wearable camera for 1 day widely acceptable by schools children and their parents since the time frame is short and the data is guaranteed to remain private and anonymous. This research tool is one of the major contributions of this thesis project along side all the work that has been conducted to be able to design and develop this tool as well as the real field study that was conducted in the real world.

The contributions made in this thesis can be summarized as follows:

1. Design a qualitative Human-Centered study for acceptable AI among school children to capture food exposure and highlight the associated ethical challenges and solutions to such challenges and report on the following:
    - Approach of the study
    - findings of the study and the acceptable system settings
2. Design and develop a system for data collection in real-world taking into consideration all of the below:
    - Incorporation of acceptable AI requirements
    - Guarantee a safe usage of technology

- Secure data collection and data transfer
- Assure reliable prediction by the ML model
- Minimize system cost and optimize system performance and throughput

3. Train an AI Model for recognition of food exposure

4. Build an annotated dataset that can be used to train a binary classifier to detect images related to food exposure. The dataset has the following info:

    - 2 main classes of images: food exposure and non food exposure
    - More than 500,000 images in total equally split among the 2 classes
    - All images have a resolution of 224x224 suitable for most state of the art convolutional neural networks
    - This dataset will be made publicly available in the future

5. Survey most popular wearable cameras and detail their pros and cons and propose a list of requirements that a wearable camera should have to be used in similar studies.

6. Investigate the presence of socio-economic, gender, food types and food healthiness biases in our dataset and trained ML models.

7. Conduct a real user study and deploy the AI based tool in the real world and collect a big amount of data (photos) from schools in Tunisia.

This report is organized as follows. In section 2, we review related work. Section 3 describes our user study for assessing the acceptability of using wearable cameras and deploying an AI based system among school children. Section 4 describes our survey of currently available wearable cameras and what are the specifications that makes using a wearable camera in a similar study acceptable and feasible. Section 5 describes the dataset construction, preprocessing and labeling process as well as how we managed to elimiate socio-economic and food types and healthiness biases while building the dataset. In section 6, we describe the machine learning model we built to capture food exposure as well as how we blurred the faces and we report their performance, accuracy, throughput, efficiency and other metrics. In section 7, we describe the data collection process and the AI based software tool that was developed to make this process effective, confidential, safe and private. In section 8, we describe how we deployed the AI system in the real world and we report on the real world acceptability of this system. We also provide some analysis and insights on the data that we managed to collect in this study using this AI system. Finally, we conclude and provide future directions in section 9.

Chapter 2

Literature Review

In this section, we review related work on using technology to capture children's food exposure. In addition, we review related work on using machine learning in the context of food applications.

2.1 Usage of Technology to Capture Food Exposure

Assessment of factors influencing food choices among children and adolescents is challenging, and is usually based on self-reported data, which are prone to measurement error [38]. Considering the multiplicity of factors influencing food choices, innovative tools are needed to better assess these among children. Technology-based interventions show promise in engaging schoolchildren in collecting data, which may also lead to more accurate and representative data [38]. As the popularity and use of mobile technologies and video gaming platforms increases, opportunities arise to use these resources to collect data on variables that affect food choice, dietary intake and associated outcomes.

A recent review highlights the gap in studies assessing the impact of children's usual trajectories to and from schools on their actual exposure to obesogenic environments [41]. These trajectories include exposures within a child's immediate environment, most commonly; home, school, and the neighborhoods around them. The multifactorial nature of food choice drivers necessitates the use of innovative tools that foster a people-based approach to measuring these exposures [41]. Technology-based tools enable an objective and comprehensive measurement of children's nutrition-related behaviors and experiences around food [42–44]. These include factors in the child's environment such as setting, social interactions and media exposure [43]. Digital technologies are also favored over traditional data collection tools since they focus on children's individual lived experience and engage younger participants in documenting their own behavior, which may also lead to more accurate and representative data on their food experiences [38, 44, 45]. These technologies have been positively received by younger participants, and are considered to be feasible and acceptable [43, 44, 46, 47].

2.2 Machine Learning in the Context of Food

There exists already in the literature many approaches for automatically detecting or classifying food items in images, recognizing their types, estimating their calorie count, estimating their weight or recommending them. However, these approaches have their limitations and they are not suitable for detecting food exposure at schools, home or the neighborhood. For example, In [2–5], the authors assume that every image only contains one food item. This assumption is suitable for an image containing one large dish, but is not suitable at all in case the image contains multiple small dishes. In [8], the researchers used a model called "Deformable Part Model" (DPM). The DPM is a two-layered hierarchical model, which consists of a global "root" filter and several part models. Each part model specifies a spatial model and a part filter. The spatial model defines a set of allowed placements for a part relative to a detection window, and a deformation cost for each placement. Both root and part filters are scored by computing the dot product between a set of weights and HoG features [9] within a given window. To detect object regions, a sliding window approach is adopted in the DPM. In addition, the DPM is defined at a fixed scale, and the food items are detected by searching over an image pyramid. This approach is not efficient and very computationally expensive and has a high failure rate in detecting small food items that are very close to each other.

In [6], the authors proposed a two-step method to recognize multiple-food items trying to detect candidate regions using several methods and classifying them using various kinds of features. In the first step, they detect several candidate regions by fusing outputs of several region detectors including Felzenszwalb's deformable part model (DPM) [8], a circle detector and the JSEG region segmentation. In the second step, they apply a feature-fusion-based food recognition method for bounding boxes of the candidate regions with various kinds of visual features including bag-of-features of SIFT and CSIFT with spatial pyramid (SP-BoF), histogram of oriented gradient (HoG), and Gabor texture features. This approach seems to outperform the other approaches.

In the computer vision domain, automatically predicting the amount of calories in food items based on their images has received some attention. For example, Pouladzadeh et al. [10] proposed an approach to automatically predict the amount of calories in food items by dividing an image of a food item into multiple segments, such that all the pixels represented by one segment have the same characteristics in terms of color, texture, size and shape. After segmenting the image, an SVM classifier is used to predict the amount of calories in the food item. However, their experiments were done only on images of single ingredient food items, which limits the applicability of their approach for specific types of food. In addition, their experimental results showed inconsistent performance ranging between 58.13% and 98.34% in accuracy of the prediction based on the type of the food item.

Sudo et al. [11] address the problem of predicting calories and other nutritional values, by doing some segmentation and some regression analysis directly on image features. The dataset they used contains 2,500 recipes where each recipe is represented by an image, an ingredient list and cooking instructions. The first step proposed is to extract a label histogram where the picture is divided into regions, called segments, each corresponding supposedly to an ingredient and then every segment is tagged, and based on different tags obtained in an image, labels are assigned. They also compute the color histograms in order to compare the results for different sets of features. Finally, they use a Support Vector Regression model and reported an average error of 33.6% and 31.8% for calories prediction, using color histograms and label histograms, respectively.

Meyers et al. [12] published a paper tackling the problem of calories prediction as well. They used previous results from Beijbom et al. [13] and Bettadapura et al. [14] in this field.They built a deep-learning approach to predict the amount of calories in a food item based on its image using convolutional neural networks. Their dataset was based on 23 restaurants with 2,517 food items in total. They also proposed predicting the size of the food items and other labels. Moreover, they used a GPS service to identify the geographical location from which the image was taken and hence map it to a certain restaurant. They obtained an error in prediction of about 20% on average.

The authors in [1] proposed a system that does not require any input from the user, except from an image of the food item. First, they identified the type of the food item in the image. Second, they estimated the size of the food item in grams. Finally, by taking into consideration the output of the first two phases, they predicted the amount of calories in the photographed food item. All these three phases were based purely on supervised machine learning. They showed that this pipelined approach is very effective in predicting the amount of calories in a food item as compared to baseline approaches which directly predicts the amount of calories from the image.

In [53], they did a literature review of all the work and research done on food recommender systems. The authors provide a summary of the state-of-the-art in food recommender systems, highlighting both seminal and most recent approaches to the problem, as well as important specializations, such as food recommendation systems for groups of users or systems which promote healthy eating. They examine which algorithms have been used in the food domain, how systems are typically evaluated, and the resources available to those interested in building or studying recommender systems in practice.

In [54], researchers investigate the contribution of common recipe features to the estimation of nutritional properties – and therewith the healthiness – of online recipes and associated meals in a data driven manner. They build regression models using different kinds of features that are derived from a recipe's title, ingredient list and cooking directions, from popularity indicators such as the number of ratings and the user comments, as well as from state of the art image analysis methods. Recent research in the area of online recipes has shown that these are important indicators and cues when it comes to online recipes food choices [60]. They also investigate which of these features contribute most to the quality of these estimations. Further, they directly compare the performance of automated methods with human performance (using crowdsourcing) to find out to what extent automatic estimates would now already improve people's insight in the healthiness of their meal choices. Finally, they investigate which recipe features people take into account for their nutrition estimations.

In [55], researchers proposed a Hierarchical Attention based Food Recommendation (HAFR) system to infer users' preference over recipes for food recommendation. The HAFR aims to learn more comprehensive recipe representation via jointly leveraging user-recipe interaction history, food image, and food ingredients with a hierarchical attention. In this study, researchers have collected a large-scale dataset for food recommendation and conducted extensive experiments, demonstrating that HAFR consistently outperforms the state-of-the-art models.

In [56], they proposed a method of design and implementation of an ingredient-based food calorie estimation system using nutrition knowledge and the fusion of brightness and heat information. This proposed design and implementation method aims to construct a system that measures food calories in each dish, which should calculate calories with more accurate results than the average statistical data of food calories. In the method, ingredients that are components of food are recognized by the fusion of brightness and heat information processing, whose features are boundaries and temperature-range order. The calories of the food are obtained by the sum of all ingredient calories using calories per weight of all ingredients, based upon nutrition knowledge. In performance evaluation of the proposed method, the experiments were performed on prototypes in software and hardware with 10 types of popular Thai curries, which are difficult to segment into ingredients. The results from each dish significantly showed more accurate calories than the average statistical data of food calories in comparison with the standard destructive-evaluation method of food calories.

In [57], they proposed a deep learning based approach to perform calories estimation from food images. They propose a deep learning-based food calorie estimating approach. The user takes a photograph of the food, and the object detection model identifies the location and category of the food in the picture. The weight prediction model predicts the weight of the food and finally calculates the calorie according to the category of the food.

In [58], they proposed a new approach based on deep learning that precisely estimates the composition of carbohydrates, proteins, and fats from hyperspectral signals of foods obtained by using low-cost spectrometers. Specifically, they developed a system consisting of multiple deep neural networks for estimating food nutrients followed by detecting and discarding estimation anomalies. Their performance evaluation demonstrates that the proposed system can maximize estimation accuracy by automatically identifying wrong estimations. As such, if consolidated with the capability of reinforcement learning, this proposed system will likely be positioned as a promising means for personalized health-care in terms of food safety.

In [59], the researchers proposed the classification of food nutrients composition utilizing deep learning techniques. The proposed framework uses convolutional neural networks (CNN) as a basis of recognizing images of food and classifying the food into their corresponding nutrients composition such as fats, carbohydrates, proteins and more.

It is worth mentioning that we were not able to find and review any work related to building a AI based food exposure detection system to be used as a major component in a research study in the domain of food, health and/or nutrition. Also, None of the papers that we managed to review discuss the usage of Machine learning and computer vision to train a food exposure detector. Also, None of the widely known and publicly available datasets exists to train a food exposure detector and none of the food datasets are suited for detecting lebanese and tunisian type of foods.

Chapter 3

System Acceptability and Feasibility

3.1 Informative research

Based on recommendations to ensure ethical conduct in visual research [48, 49], we conducted an informative participatory research study to assess the feasibility and acceptability of using cameras to capture the food environment and food exposures of school children. We employed a user-centered design approach to engage with school students (10-12 years), their parents and school staff (directors and teachers) in the development of a suitable, culturally-relevant and acceptable technology that is AI based.

We relied on multiple ethical frameworks that guide the use of wearable technology in research in order to ensure that we address the major ethical issues during the implementation of the current phase [48–50].

We mapped the issues related to wearable camera use (privacy, anonymity, confidentiality, data protection, social acceptability, etc.) to features of the study design and the technology itself, which could be integrated in the final data collection model. These features were then translated into an "inspiration pack" (cards), which parents and school staff used during the engagements to design the most feasible and acceptable combination of features **(Figure 3.1)**. A summary of all the models suggested by participants is available in **Appendix A**. The data from the formative phase was used to design a wearable camera with culturally-acceptable features, that takes into consideration the ethical concerns that arose during the formative research.

This mixed-methods study includes multiple components at different levels:

1- The major component involves the collection of data using traditional survey and interview methods from children, their parents/caregivers, and schools

Figure 3.1: Example of a combination of features suggested by parents during the workshops

2- An assessment of children's hemoglobin and micro-nutrient status by analyzing finger-pick blood samples

3- A mapping of children's trajectories to and from schools using wearable cameras

The study population included all students enrolled in grades 4-5-6 (aged between 9-12 years) attending schools located in Greater Beirut and Greater Tunisia, and their parents/caregivers. The study also included school directors, teachers, and nutrition/health educators working at selected schools.

Only a sub-sample of students enrolled in the overall study was included. We invited all the students to participate, and of those who agreed to participate, we chose a random sample of 10 students per schools, to reach a total of 500 students.

Parents/caregivers of children enrolled in the selected schools were invited to participate in this study by sending them consent forms with their children. The school administration/teachers distributed the consent forms, and made sure they are brought back by students within 2 days.

Based on the ethical frameworks and the findings of our informative study, we conceptualized and are developing a final camera model and an implementation protocol. The final model will be a wearable camera, attached with a lanyard around the neck, and pinned to the child's clothes.

The camera will be wrapped in a pocket of neutral color (black), which only shows the lens and the on/off switch. We will not use vests to hide the cameras, since some parent groups perceived that concealing the reality from others is against the values of openness and integrity that they teach their kids. The camera has a long battery life ($9 \sim 10$ hours) and can take high resolution pictures from the children's perception of the food environment each 5-10 seconds intervals. The camera is password protected, hence the captured images will only be accessible to the primary investigator and the research team member who will retrieve the pictures from students.

The camera will be worn by included children for one day only, outside the confines of school, mainly on the trajectory to and from school, and inside the house. On the day of data collection, children who agree to participate in this study component, are given the camera kit including the camera, the lanyard with the pin, the charger and an instruction manual. Trained data collectors will demonstrate camera advise students to wear the device as soon as they leave school in the afternoon and until they go to bed at night. Students will also be asked to wear the camera the next morning on their way to school. Students will be informed that they can turn the camera off anytime they do not feel comfortable wearing it, they will also be instructed to remove the camera in these conditions:

1- before going to the bathroom or changing clothes

2- in pubic bathrooms

3- in public places where there is a notice/sign that prohibits taking pictures

4- before engaging in physical activity or sports

Parents will be urged to remind their kids of turning the camera back on/wearing it again whenever they turn it off. Children will be provided with a written note that they can show to those who ask about the camera. The note will inform third parties that this is part of a research study conducted by the American University of Beirut, and that the objective is to capture the food environment and not the individuals in it. The note will also state that all faces will be blurred in the retrieved pictures, and that these pictures will not be shared nor disseminated. Students will also be advised to turn the camera off/remove it if requested by third parties.

The research team will visit the school the next day to pick up the cameras from children at a set time. Parents of these children will be invited to schools on that date and time and will be allowed to screen and delete any images they do not want the research team to view.

Before having access to the images, we will transfer the pictures from the cameras to a protected computer using a USB cable, in a private room and in the presence of the parents only. On the computer, and before viewing the pictures, we will run an AI Based Application/Tool on the captured pictures, which will:

1- Delete all pictures that do not contain a food item, food marketing, etc.

2- Blur all the faces that appear in pictures.

The collected set of pictures will then be stored on a password protected computer maintained by the primary investigator and will be destroyed within 3 years of study completion.

3.2 Analysis

Data collected from the multiple levels will be merged and analyzed using standard statistical methods. Descriptive statistics will be used to describe child diets and socio-demographic variables at the child and household level. School-level variables will be generated in regard to policy, nutrition education and food availability patterns. The determinants of children's food choices, diet quality, and nutritional status will be examined using logistic and linear regression. Econometric analysis of data from the food choice experiment will enable the estimation of preference intensities for each food item as inflected by its characteristics and attributes. We will use deep learning models to conduct the photo analysis and quantify the intensity, frequency and types of food exposures, stratified by gender, socio-economic status, geographic areas, and nutritional status. Multilevel models will be used to analyze the various levels of influence on child diets and nutritional status, allowing for the analysis of hierarchical data. School environmental factors, individual socio-economic characteristics and contextual effects in choice tasks, will be accommodated by the econometric analysis as preference intensity shifters. The multiple layers of analysis included in this project will ensure the triangulation of data from different sources and will allow us to fill knowledge gaps that exist at each of the levels of analysis.

3.3 Ethical considerations

The research team will conduct the study following the approval of local authorities and communities; meetings have been conducted with representatives of MEHE, and their official approval of the study is imminent. This study will be conducted following ethical approval by the American University of Beirut Institutional Review Board.

All research team members obtained their CITI certificates. Data collectors attended an extensive training session on the ethical conduct of research. The training was conducted by AUB (PowerPoint presentation attached) and aims to ensure voluntary enrollment of participants and protection of their privacy and confidentiality. Before data collection, informed consent will be sought from school directors, school staff, and parents/caregivers and assent will be sought from children. The consent will clearly explain the purpose of the study, the activities involved, the risks and benefits of participating, while detailing how confidentiality will be handled and indicating the right to refuse or withdraw from participating without any penalty. We will take special precautions to protect participant confidentiality, safety, and autonomy.

In order to address the ethical concerns raised by the use of wearable cameras in research, the study protocol was developed based on:

1- the findings of the participatory research on the acceptability of the use of technology in research.

2- ethical guidelines developed specifically for wearable camera research 48,49,50.

A mapping of the guidelines that ensure ethical research conduct to the implementation protocol is available in **Appendix B**.

3.4 Risks

The use of wearable camera in research could impose the potential risk of invading participants' privacy and confidentiality. Multiple measures were included in the protocol that address this issue:

1- Participants are advised to remove/turn off the camera anytime they don't feel comfortable wearing it or anytime they don't want their surroundings to be recorded.

2- The cameras will be password protected and access to parent-approved captured images will be restricted to the primary investigator and to Citi certified research staff.

3- In case any faces appear in the captured images, they will be blurred directly after the transfer to a computer, and the original unblurred pictures will be permanently deleted.

4- Two rounds of image selection will be applied: the first one will be conducted automatically by a machine learning model that will delete the majority of the picture not containing food or food related items, without anyone viewing the original dataset; and the second one will be conducted by children's parents in a private space in the school, where they will delete unwanted pictures – parents will review the images after faces have been blurred.

5- Data analysis will be conducted using a machine learning model, which reduces the need for research team members to view the pictures.

Additionally, parents might consider this research as an opportunity to monitor their children's activities during that day, which is considered as an intrusion to children's privacy. Therefore:

1- Children will be informed that their parents will view captured pictures at the end of data collection.

2- It will be emphasized to children that they have the freedom to remove/turn off the camera anytime they don't feel comfortable using it, or when they don't want something to be recorded.

3- Parents will only be able to view the images after the first round of automated filtering, so they will only view a sub-sample of the data.

4- The research team will only be able to view the images after two rounds of filtering (automated filtering and parent-filtering).

Participants might also be exposed to the risk of being questioned by third parties about the device or the image capture activity. Although findings from the participatory research study indicate that students of this age-group are almost always accompanied by a responsible adult who can protect them in such a case, we included in the protocol the following measures in the aim of reducing this risk, and that of invading third party privacy:

1- Children are advised to inform their close relatives and household members about the camera and the purpose of its use.

2- Children will be provided with a note that explains to third parties that they are part of a research study about food, and that the aim is to capture the food environment and not third parties.

3- The note will explain that all faces will be blurred, and that pictures not including food-related exposures will be automatically deleted.

4- Students will also be advised to turn the camera off/remove it if requested by third parties.

3.5 Benefits

This study will generate contextual knowledge on environmental influences on child dietary behaviors, which will inform multi-level interventions and policies that address childhood obesity. Findings can also be used to derive lessons-learned for similar middle-income countries in the region undergoing a nutrition transition. Participants in the micronutrient assessment component will be directly informed about their hemoglobin level status and referred to the nearest primary healthcare center in the case anemia. We will also generate profiles for individual schools enrolled in the study using the school assessment module which will be shared with MEHE and participating schools to inform improvements to the school food environment.

3.6 Confidentiality of data and records

For this study component, we will not record any direct identifiers such as names, addresses or telephone numbers. On the day of data collection, the cameras will be labeled with each child's unique ID, and the retrieved food-image data will be given the same unique ID. Moreover, no identifying images will be retrieved as the faces will be blurred in all the images, nor will these be shared in publications or conferences.

The cameras are password protected and the password is only known by the PI and Citi certified members of the research team. Therefore, only members of the research team can have access to the data. Research team meetings will be conducted regularly to ensure adherence to study protocol by all team members.

The collected dataset will be stored on a password-protected computer maintained by the PI. Within 3 years of study completion, the PI will destroy the datasets, erasing them from the encrypted hard drive completely. Hard copies of the signed consent forms will be stored in a locked cabinet at the PI's office and destroyed/shredded within 3 years of study completion.

3.7 Findings of the participatory research study on Acceptable AI in the real world

Based on all of the above, and after conducting several qualitative user centered workshops with parents, children and schools' directors and staff, we reached a consensus that wearing a camera is acceptable in general by most parents and children if it is for a limited time (1 day) and if we can develop an AI based system that that guarantee the user privacy and minimize the risk invading the user private life. Most of the participants in these workshops, especially the parents' workshops, proposed the following:

- The camera should is acceptable to be worn for a limited time only (1 day).
- The camera should not have any internet connection capabilities or Bluetooth capabilities to make sure that the data cannot be transferred to another device wirelessly or uploaded and stored on a cloud server.
- The camera should ave a password protection.
- The camera should not have a removable SD card for storage to avoid any attempt to steal the data from the camera in case the camera was stolen.
- There should be a system on the camera or that support the camera that automatically detect the images that contain food exposure and delete all the other images.
- There should be a system on the camera or that support the camera that automatically blur any face in the food exposure photos.
- The parents would like to screen the filtered data by the AI system before the research team can use it.
- The recording should be automated to minimize the chance of distracting the kid at school, on the road and at home for his own safety and to guarantee a better capturing of food exposure.
- The camera should be attached in a way that is simple and suitable for the kid while still allowing him to carry his bag or wear his jacket without interfering with the camera. Also, the camera should be attached in a way that keeps it firmly tied to the child's chest to avoid dropping or losing it while running or doing some normal activities during the day.
- The camera can be turned off at any time or any place where cameras are not allowed to be used or where using it might provoke a risk or danger on the child wearing it or when asked by someone to turn it off.

Under these constraints and specifications, we concluded that the majority of the parents belonging to different socio-economic backgrounds would accept that their kids wear a camera for a day to participate in a study that aims to assess the factors that influence their food choices. Also, the parents were behind the idea of using technological means to only detect images with food exposure and to blur the faces of people in these images. Also, as scientist, we made sure to provide a high level explanation on these technologies and how we develop them as well as how we are planning to design a system that meets all these constraints and specifications using technological means at our disposal nowadays.

Based on those findings, we will discuss in the next chapters how we proceeded and managed to meet all these requirements and specifications to develop an entire AI based wearable camera system that capture food exposure and guarantee the anonymity, privacy and confidentiality of the data that this AI based system will be collecting from the real world. This AI system will is a Major component as stated previously in a mutli-level research project it will be used by the researcher to help them capture children's exposure to food on a typical school day. Also, we would like to mention that the use case of such a system won't necessarily be limited to this single research project and it might be used in other similar field research studies.

Chapter 4

Camera Survey

In this study, we tried to test and benchmark most of the available wearable camera is the market at the moment. During our assessment,we took into consideration the following general aspects:

1- Overall image quality

2- Battery life

3- Security features

4- Premium quality of the product

5- Reliability of the camera while operating under different scenarios

4.1 Required Camera's Hardware Specifications

The following are the hardware specs that we took into consideration, detailing the benefits and effects that every single hardware specification has on image quality, security and usability:

1- The actual resolution of the image sensor: the resolution is quite important since higher resolution usually results in better image/video quality. Image/video quality is a key factor for getting accurate predictions from the ML models at a later stage. Also, higher image resolution will enable the research team to extract details of items that are far or tiny in the captured image.

2- The battery capacity and active battery life time: a long battery life is crucial in this study because the cameras need to record for at least 10 hours without the need of recharging it.

3- The form factor: the camera should be as simple as possible to make it less intrusive and easier to attach to the clothes.

4- The weight: the camera should be as light as possible. A lighter camera is always a bonus and makes the process of attaching the camera to the clothes and making sure it is as stable as possible way easier than a camera that is heavy.

5- The camera dimensions: a camera will smaller dimensions is always a better option since the smaller the camera is, the less intrusive and less noticeable it will become when seen from a distance.

6- The Price: price is an important factor. Usually higher price indicates a better product quality while a lower price indicates the opposite, but this is not always the case. We tried in our camera testing and review process to pick the cameras that have the best price to features ratio instead of aiming to minimize or maximize the cost of the camera. In other words, we tried to purchase and test the cameras that offer the best set of features and specs in its price range.

7- The storage capacity and storage type: a storage capacity of 32/64 GB is optimal for a full day of video recording. Also, built-in memory is preferable over micro-SD cards for security and performance reasons. Built-in storage usually delivers higher transfer speed of data and ensures that the data on the camera is well secured in case the camera was lost or stolen. A micro-SD card is usually slower at transferring data and the data can be easily stolen from the camera in case the camera is lost or stolen.

8- The hardware protection features: since the cameras will be used by kids, the chances of running into situations where the camera might get damaged is higher. For that reason, cameras that are resistant to water, dust and shocks are highly recommended. Usually an IP66 rating for water and dust resistance is recommended.

9- Optical image stabilization (OIS): the optical image stabilization feature allows the camera to compensate the vibration and motion of the camera lens when shooting a photo or video. which results in a better quality recording with less motion blur.

10- Wide angle lens enabling wide angle shots and footage: a Wide angle lens (greater than 120 degrees) allow the camera to capture more of the surroundings. This is a very important feature since it would allow the research team to capture a wider scene which will provide more information.

11- Infrared capable camera: an infrared capable camera is an excellent choice for capturing images and video in low light and dark places. it is also great since the light emitted by the infrared LEDs are not visible to the human eye which helps in keeping the camera unnoticed at night as well.

4.2 Required Camera's Software Features

The following are the software related features that we took into consideration, detailing the benefits and effects that every single software feature has on image quality, security and usability.

1- Password lock: password Protection is a very crucial and needed feature to make sure that the data on the camera is secured. Also, we made sure that no data can be accessed via USB unless the camera is unlocked.

2- Silent mode: the camera should be able to operate in silent mode i.e. without making any beep, sound, noise or light. This is very important because it makes the camera less noticeable and allows the user to wear it without disturbing people that are sitting nearby.

3- Record split: splitting the recorded videos into splits of 5/10 minutes to avoid losing data in case of a power interrupt or system failure or unexpected camera damage.

4- Photo time-lapse: photo time-lapse is a very important recording mode. It captures an image every 5/10 seconds instead of continuous video recording. This allows the camera to save storage and reduce data size by a great margin while making sure that most of the information that could have been extracted from video recording is represented in these time-lapse images.

4.3 Findings of the Camera Survey

The specification table in **Appendix C** contains all the cameras that we have surveyed for reference. This table contains all the cameras that we have tested and/or reviewed and that we do not recommend for a similar study because they do not comply with our list of requirements. From all the tested cameras, only 1 camera that is still in production and available in the market matched our requirements list and will be used in this study. This camera is the MIUFLY 2K Body Pro. This camera has a very reasonable price of 180 USD. It has good quality, wide angle image sensor, a long battery life (9 hours of continuous video recording at 720p), a GPS, infrared sensor, password protection, water and dust resistance, good shock resistance, an average and acceptable form factor and weight, a silent mode that disables all LED lights and sound, a vibration motor for operation feedback and a 32 GB of built-in flash storage.

Chapter 5

Dataset

In this section, we detail how we collected and built the food dataset that we used in this study. Also, we will explain why we needed to build our own dataset instead of using an open source dataset available online.

5.1 Review of Publicly Available Datasets

Initially, we looked for all the available food datasets available online. We discovered many limitations in the open source datasets that made them unpractical to be used for training a neural network that can perform food exposure detection. These limitations are as follows:

- A low number of images
- A low number of food categories
- The absence of Lebanese and Tunisian food items from the datasets
- The absence of images that reflect food exposure

5.2 Dataset Collection and Labeling Procedure

For all of the above reasons, we needed to collect and build our own dataset. The dataset was automatically collected and labeled using an open source web crawler that crawl images from Google and Bing search engines [61]. We crawled data for around a month. More than 1 million images (1,037,459 images to be exact) were crawled in total during this process. Also, around 85% of the photos were crawled from the Google search engine and the remaining 15% were crawled from Microsoft Bing search engine. This 85-15 split happened because crawling using Google search engine was way faster than using Bing and returned way more images for every query and the crawler API provided more filtering and

Figure 5.1: Percentage of Noise in 4 Random Data Samples from the Dataset

	Sample Size	Correctly Labeled Images	Incorrectly Labeled Images	Noise Percentage
Sample 1	1000	907	93	9%
Sample 2	1000	894	106	11%
Sample 3	1000	916	84	8%
Sample 4	2000	1762	238	12%

control options using the Google search engine. In order to automatically label the data during the crawling process, every image was saved in a folder named similar to the search query that returned the image. For example, if 'burger' was the query used to crawl photos, all the photos returned during that process are stored in a folder labeled 'burger'. We relied on the accuracy of the web search engines to label our dataset in this work. We are completely aware that relying on search engines to label the data will result in having some images incorrectly labeled, which will add some noise to the dataset. However, since we have a very large dataset and the percentage of noise in the data is low (less than 12%), this noise was not a problem for training and might actually help the machine learning models to generalize better. In order to validate the noise ratio in the dataset, we randomly selected 3 subsets of the data each containing a 1000 randomly selected images and 1 subset of the data containing 2000 randomly selected images. We manually validated the labeling of each image in these 4 random samples. The percentage of the noise was around 11% as shown in **figure 5.1**.

5.3 Dataset Classes

The dataset has 2 main classes: images that contains food exposure and images that do not contain food exposure. What we mean by food exposure is anything related to the food context. These are some examples of food exposure images:

- actual food items
- food ads
- restaurants
- food shops
- vehicles that are dedicated for food transportation and covered with ads
- food brands
- vending machines
- markets
- refrigerator
- kitchen
- someone eating

The second class of images that does not contain food exposure can be any image where all of the above is not present in the image i.e. mountains, bedrooms, trees, animals, etc...

Appendix D contains the list of all queries that were used to crawl the food exposure class of the dataset and **Appendix E** contains the list of all queries that were used to crawl the non food exposure class of the dataset for reference.

For the food exposure class, we used around 290 queries to get images from the search engine. The queries were a list of the most popular food items available in Lebanon and Tunisia, the most popular restaurants and food brands in both countries, as well as the most common places that might contain food during an ideal day of a kids' school day. For the non-food exposure images, we crawled the results of around 150 queries of locations, places, things, animals, objects, etc... that the kids might see during a classical school day while making sure that the images does not contain any food-related exposure.

5.4 Dataset Balacing and Pre-Processing

After running the crawler for about a month, 1,037,459 images were collected. While crawling we made sure to crawl at least 2000 images per food exposure related query and 4500 images per non food exposure related query in order to keep the data in every class balanced. In order to make sure that the dataset is ready for training, we performed some pre-processing and filtering on the collected images. First, we used a duplicate remover to eliminate all the duplicate images from the dataset. We also made sure to eliminate duplicate images that have different resolutions. After eliminating all the duplicate images, only 702,096 images remained from the 1 million images that we crawled. This high number of duplicates was anticipated from the beginning of the crawling process because some different queries usually return similar images. For example, if we take 2 queries: "fruits" and "apple", it is very likely that some images returned by the "fruits" query will be also returned by the "apple" query since apple is a type of fruit. Also, we eliminated all the images that have a resolution lower than 224 x 224 since they will be irrelevant and useless for training. In addition, we deleted all the images with transparent background and most of the images that have watermarks. At the end, we made sure to re-balance both classes by randomly deleting images from the class with higher image count until both classes became even. After all these filtering processes, we ended up having approx. 510,000 images (255,000 images per class).

Finally, the dataset was split as follows: 80% for training (approx. 195,000 images per class), 10% testing (approx. 25,000 images per class) and 10% validation (approx. 25,000 images per class).

Also, in order to accelerate the training task, we uploaded a around 50% of the dataset to amazon S3 (240,000 images). Then, we resized the entire dataset to 4 different resolutions and 3 different aspect ratio in order to accelerate the training jobs later on and avoid the need of resizing every image to match the input shape of the neural network before using it. These are the resolutions and aspect ratios to which the uploaded subset of the dataset was entirely resized to:

- resolution: 224x224; aspect ratio: 1:1
- resolution: 360x240; aspect ratio: 4:3
- resolution: 256x144; aspect ratio: 16:9
- resolution: 480x360; aspect ratio: 4:3

All the resized datasets were also stored on amazon S3. This subset of the dataset was used to run some experiments and training jobs using all the models that we planned on testing their performance on our dataset in order to check how these different neural networks will perform and then take the best 3 network architectures and resume the training on the full dataset and the fine tuning process using a local workstation which stores the entire dataset (510,000 images).

5.5 Dataset Bias

In order to assure that our dataset is well representative of the real world data and is not biased toward some food categories or has any socio-economic and gender based bias, we performed a set of measures to make sure to minimize any risks of having a biased ML model when training on our dataset. We present the list of measures and checks that we applied to try to avoid or minimize any chances of Data Bias:

1- We made sure to balance the number of images for both dataset classes.

2- For every class, we made sure that the number of images that belongs to every sub class item is almost equal. for example, the non food exposure class contains 150 categories of images and we made sure to have around the same number of images for every category

3- We removed all duplicate photos even the ones with different resolutions to make sure the dataset is wide, diverse and representative.

4- We made sure that the queries we used to crawl the images does not have and gender, age, race, social or economical terms or meaning and we made sure to not have any adjectives in the queries we used to crawl data (see **Appendix D** and **F**). However, we cannot fully guarantee that the Google search engines won't return at all any images that might be subjective or contain any gender, age, race, social or economical bias. This factor

Figure 5.2: healthiness Distribution of Food Items Present in the Dataset

Caterogie	Number of Food Items
Healthy	42
Medium Healthy	37
Unheathly	41

is beyond our control and our scope of research and we expect that the world leading and most used search engine on the web to return fair and unbiased images in general. Also, it is beyond our ability to manually check thousands of images.

5- Regarding the non food exposure class, we inspired most of the list of queries from the imagenet datatset classes and we picked all the locations, places, timing (day shots and night shots), objects, animals, etc... that we expect a 10-12 years old kid to see during a day at school.

6- Regarding the food exposure class:

- The queries where equally split between queries of food items and queries of restaurants, markets, food ads, etc... (around 130 food items and around 160 food related exposures).
- The Food items related were mostly inspired from a food recommender system (Zomato) which is one of the ost popular food recommender systems in the arab region especially in Lebanon. Also, we had a team of Tunisians that helped us in including the most famous types of tunisian homemade food as well as the popular and most consumed types of food from restaurants.
- The food items were also equally splitted among Lebanese and Tunisian ones to make sure that our model won't be biased towards one of them.
- We also had a team of nutritionists who managed to label and classify all the food items that we have in our dataset based on their healthiness. We did this classification to check and verify that we have managed to fairly represent the healthy, medium healthy and unhealthy food items in our dataset(check **Figure 5.2** for reference).

Based on all of the above, We can safely assume that our dataset won't lead to a biased model at least in terms of socio-economic bias, gender based bias, and any bias related to favoring some types of food over others. However, this dataset will lead to a biased model if it will be used for example to train a food exposure detector outside of our study scope and especially in countries that doesn't share the majority of food items that are popular in Lebanon and Tunisia. For example, using this dataset by a chinese research team to build a food classifier or food exposure detector or any sort of food research will lead to a biased model in that case since our dataset does not include chinese food items.

In brief, We made sure to follow all the possible scientific measures that we are aware of to make sure to minimize any chances of introducing bias to our ML model in the context and scope of our research and study (we will provide further evidence on that below). Also, we will not recommend the usage of our dataset as it is for similar studies in countries doesn't share at least most of the food types that are similar to what we have in Lebanon and Tunisia.

In order to back up our claim that our dataset is bias free, we will provide some evidence based on the data collected from the real world after conducting this study in around 30 schools in Tunisia prior the spread of the covid-19 pandemic worldwide and the global lock-down. We managed to get 265 children to participate in our study from 29 Schools in Tunisia (9 children on average per school). We managed to collect 567506 photos in total, 62745 of which were filtered by our Ml model to be containing food exposure (around 11% of the total captured photos). Unfortunately, Due to the global pandemic, we were unable to start field study in Lebanon and unable to finish the entire data collection from all the schools in Tunisia (we aim for 50 schools in every country, and for a total of 1000 kids (500 kids from every country)). Based on that, all the statistics that will be presented in this thesis are based on the subset of real world data that we managed to collect prior the pandemic global spread and despite having an acceptable amount of data to base our analysis on, it is important to mention that all the reported numbers based on the real wold data are prone to change when we manage to complete and finish the data collection later this year hopefully since as of now, we just have 26% of the total data that we are planning to collect by the end of the study.

Based on the above, we will base our evidence that the trained models on our dataset are bias-free on the statistical tables generated from the real world data found in **Appendix F**. In **Table F.1**, as mentioned above, we managed to collect 567506 images in total. only 11% (62745 images) were images that contain food exposure. The average recorded hours per child was 3.27 hours (including the empty cameras) and the average recorded food exposure time was 22 minutes (including the empty cameras). Please note that an empty camera means that the child and his parents/guardians accepted to participate in the study and took the camera but returned it empty or with less than 1 min of recordings. These children were included in the computed values stated above. Based on **Table F.1**, if we look and compare the values of Tables **F.2**, **F.3**, **F.4**, **F.5**, **F.6** which respectively represent the same statistical values based on gender, age, family income, parents educational level and the availability of internet access at home, we can notice a minor and slight variation in these statistical values. Since no major difference is present across all these metrics, this provide a scientific evidence that the ML models trained on our dataset did not have any major and noticeable gender bias, age bias or socio-economic bias.

Chapter 6

Machine Learning

In this section, we detail the machine learning models that were trained and used to detect images of food exposure. Then, we report on the best model performance, throughput, precision, recall and other metrics. We also detail how we added a face blurring feature to the pipeline. Finally, we report some qualitative results.

6.1 Models and Requirements

We used Amazon SageMaker cloud services and a very powerful local workstation to run all the training experiments. we used Keras framework with tensorflow as a backend. we used 5 different models for training:

- VGG16
- VGG19
- MobileNet V1
- 2 custom convolutional neural network architecture.

VGG16 and VGG19 were loaded with their pretrained ImageNet weights [61], then, the last layer was modified to have 2 classes instead of 1000 classes. Then, we trained both models 3 times, to retrain respectively the last 1, then the last 2 and then last 3 layers while freezing all the other layers (transfer learning).

MobileNet V1 was also a pretrained model with weights trained on ImageNet. We chose this model for several reasons. First of all, it is one of the most recent models that won the ImageNet competition. Second, MobileNet V1 has an excellent top-1 and top-5 accuracies that challenge the accuracy of VGG16 and VGG19. Finally, and the most important among all, MobileNet was designed to run efficiently on Mobile phones and requires very minimal computational power. The whole network has 4.8 million weights when compared to VGG16 (134 million) and VGG19(144 million).

This small number of weights that MobileNet V1 has allowed us not only to run transfer learning jobs, but to also train the entire model from scratch in a considerably short time when compared to the time needed to train networks like VGG16 and VVG19 from scratch. We used two approaches to train the MobileNet V1 Network. In the first approach, similar to what we did in the case of VGG16 and VGG19, we loaded MobileNet V1 with its pretrained ImageNet weights, then, the last layer was modified to have 2 classes instead of 1000 classes. Then, we trained the model 3 times, to retrain respectively the last 2 layers (transfer learning), then the last 12 layers (transfer learning) and then all the layers (full network training from scratch). In the second approach, we loaded MobileNet V1 with its pretrained ImageNet weights, then, the last layer with 1000 classes was removed and replaced with 2 fully connected layers. The first fully connected layer had 2000 neurons and a final layer with 2 neurons (2 classes). The purpose behind that customization is to increase the number of parameters of the model by around 2 million extra parameters (from 4.8 million to 6.9 million parameters) since the number of parameters in the original MobileNet network is considerably low. We did that to see if this increase in the number of neurons in the last 2 fully connected layer of the MobileNet V1 will improve the accuracy and precision of the model. Then, we trained the model 3 times, to retrain respectively the last 3 layers (transfer learning), then the last 12 layers (transfer learning) and then all the layers (full network training from scratch). It is worth mentioning that over-fitting is not a problem when training the MobileNet V1 from scratch since we have a very large amount of training data.

In addition to these 3 predefined models for object classification, we created 2 custom convolutional neural networks with 5 millions parameters and 20 millions parameters, respectively. Both of these models were also trained from scratch. These 2 custom models were used for preliminary testing of the dataset and in the development stage because they are faster to train and easier to deal with but they were not intended to be used as the final best models in this study.

As for the hardware specifications that were needed to train these networks, we needed to rent an amazon SageMaker instance with the following specifications detailed below for 3 days (74 hours in total) in order to run a subset of these training jobs just to test how the different network architectures will perform on a subset of the dataset (around 250,000 images). The specifications are as follows:

- CPU: 32 cores
- GPU: 8 x NVIDIA Tesla K80 with 96 GB of dedicated GPU memory
- RAM: 488 GB
- Network bandwidth: 10 Gigabit used to load and store data on Amazon S3

- Storage: 1 TB of Amazon EBS storage

After running these experiments and getting some primary insights on how the different neural networks performed, we decided to eliminate VGG16, VGG19, and resume the training jobs using only both variations of the MobileNet V1 (both version with 4.8 millions and 6.9 millions parameters respectively as explained in the paragraph above). Experiments that we did run on AWS showed that both MobileNet models outperform the other models in terms of training and validation accuracy and have a shorter training time per epoch due to their lower number of parameters. Also, using AWS was so expensive, so we decided to resume training of the selected neural networks on larger datasets and fine tune them on a local workstation that has some powerful specifications (detailed below). Since training time was not a limitation for us, we decided to use this workstation and save on budget even if training and fine tuning these networks would require up to few months (2-3 months). The specifications of the workstation are as follows:

- CPU: Dual intel Xeon Gold with 28 physical cores , 56 logical cores (threads).
- GPU: NVIDIA Quadro P2000 with 5 GB of dedicated GPU memory
- RAM: 64 GB DDR3
- Network bandwidth: 6 Mbps used to download the dataset using a Google search engine based crawler
- Storage: 1 TB NVME SSD

6.2 Model Training and Metrics

On average, every training job took between 48 and 240 hours to finish. Also, every training job had its own set of optimizers and fine tuned parameters. In brief, we used Adam and RMSprop as optimizers with a learning rate raging between 0.0001 and 0.05. Also, learning decay was set to either 0.0001 or kept as the default decay value of the optimizer. The dropout rate ranged between 0.25 and 0.5. Also, the batch size was set to 32 for MobileNet V1 for its both variants while using a batch size of 32 and 128 for VGG 16 and VGG 19. We kept on training the models until the training accuracy flattened and stopped increasing for 5 successive epochs. After all the training experiment were done, MobileNet V1 with both its variations was the network to perform the best when trained from scratch. The training accuracy reached 98.93% and 98.73% while the validation accuracy reached 93.24% and 92,75% for both its variations respectively (see **Figure 6.1**). Image Augmentation was used in most training jobs but not all since the dataset is large and diverse and for the sake of accelerating the training

Model Name	Training Loss	Traning Accuracy	Validation Loss	Validation Accuracy
MobileNet V1 (4.6M parameters)	0.0293	0.9893	0.2353	0.9324
Customized MobileNet V1 (6.9M parameters)	0.035	0.9873	0.2504	0.9275

Model Name	Test Accuracy	Precision	Recall	F1 Score	Cohens Kappa	ROC AUC
MobileNet V1 (4.6M parameters)	0.9253	0.9541	0.8967	0.9204	0.9138	0.9594
Customized MobileNet V1 (6.9M parameters)	0.9189	0.9407	0.8897	0.9123	0.9105	0.9473

Figure 6.1: Best MobileNet V1 Model Performance

time as well. As for the other models, the training accuracy ranged between 84% and 94% while the validation accuracy ranged between 79% and 84%. The reasoning behind training from scratch instead of using transfer learning is as follows:

1. Training from scratch yielded higher accuracy and better recall and precision.

2. There is a very high chance that the pre-trained weights on ImageNet dataset will introduce a bias into the trained models. The majority of the ImageNet dataset classes are non food exposure classes and using the pre-trained weights of these models will much likely result in models that are biased toward detecting non food exposure images.

3. we have a very large dataset that allows us to train from scratch.

6.3 Face Blurring

Face Blurring was needed to hide the identity of the people that are present in the collected photos with food exposure in order to keep the data private and confidential. To achieve face blurring, we tested 3 different approaches to perform this task. Two of these approaches performed extremely well while the third approach had an acceptable performance with some errors. In the first approach, we used OPENCV's open source library [51] to detect faces and draw virtual bounding boxes around them. This approach was not used because it did not perform well and sometimes it blurred objects that are not human faces in the picture. In the second approach, we used an open source CNN named MTCNN [52] that also detect faces and draw a virtual bounding box around them. This CNN had a very good accuracy (up to 90% of the faces gets detected and blurred). Only cropped faces or images that have very far people in the background are usually not detected and blurred. In the third approach, we used a python library called CVLIB that also perform face detection. Also, similar to MTCNN, This library had more or less the same accuracy and delivered almost the same results when tested on the same data sample. In order to maximize the detection accuracy, we used both MTCNN followed by CVLIB to detect faces in every image. In order to blur the faces after being detected, we took the x-y coordinates of the virtual bounding boxes that contains faces and we applied Gaussian Blurring to the bounding box area. The accuracy of face blurring is around 91% when both methods are used in a pipeline.

6.4 Performance and Throughput

We tested our best classifier (MobileNet V1) followed by MTCNN and CVLIB (for face blurring) on a subset of the dataset (around 16,000 images) that was not used in training nor in validation. We obtained an accuracy of 94.86% and an f1-measure of 0.9482. Also, we sampled some real world data and checked for the precision of food exposure detection on real data. The model precision ranged between 87.23% and 94.57% depending on the samples. The throughput of the MobileNet V1 that we achieved using a core i7 (6-cores CPU) with an Nvidia GTX 1050Ti is around 1100 images per minute for food exposure detection. Having such a high throughput is very recommended in a real world AI system and would minimize the time needed to filter large amount of data from the wearable cameras. The throughput of MTCNN is close to 600 images per minute and the throughput of CVLIB is around 200 images per minute on the same hardware. The combined throughput of both approached when used as a pipeline is around 600 images every 4 minutes (around 150 images per minute). This is an acceptable throughput since we will only be blurring faces in the food exposure images. Some of the resulted filtered images are shown in **Figure 6.2**.

Figure 6.2: Sample of Detected Food Exposure Images With Blurred Faces

Chapter 7

AI System

In this section, we detail how the data collection of images happened. Also, we will explain how the camera synchronization application works and we will report on accuracy of the tool and show some filtered data samples from the study.

7.1 Image Data Collection Protocol

We used the MIUFLY 2K Pro Body Camera that we discussed in the camera survey section to collect video footage from the students. The camera is fixed to the center of a black shoulder strap that the students wear when they leave the school until it is bed time and then from the morning when they wake up until they reach the school. We asked the students to keep the camera on as long as it does not put them at risk, in an awkward situation or disrupt their privacy. The cameras will be picked up by a member of the data collection team when the students arrive to school on the next day. In order to keep the data safe and confidential, the data collection will happen at the director's office directly after collecting the cameras from the students by using a desktop application. The interface of the application is very simple and straight forward. It just requires from the data collector to connect the cameras (up to 10 cameras at once) to the computer via a USB cable, to enter the password on every camera to enable the access to the camera storage and the automated filtering process will be carried automatically without seeing the data. When the filtering process is terminated, the results will be saved on tablets for another round of screening by parents and all the data from the cameras will be deleted automatically. The parents of the children will be asked to show up to the school in order to review the filtered data that will be loaded into tablets. The parents will have the freedom to delete any filtered image that they do not want the research team to have and will return the tablets to the data collector. By using this data collection protocol, we make sure that we are keeping the collected data private and confidential.

7.2 Data collection and Filtering Tool

In order to simplify the data collection of images, we developed and built a desktop application that will use our food exposure and face blurring models to filter the food exposure images. The front-end of this software tool is written in Java and the back-end uses Python code and Python libraries to perform most of the tasks and computations. We used Java in the front-end for its portability and ease of debugging and its easy-to-build and sophisticated graphical user interface (GUI) and for its better multi-threading support and better performance when compared to Python. Also, we used Python functions and libraries in the back-end to run the functionality and tasks that the application handles because all our ML models, testing, development and prototypes were coded in Python so it was easier and more reasonable to import these models and functions into a Java-based GUI Application that execute the python functions and run the Machine Learning Models Using Python frameworks and libraries. This AI based software tool just requires from the user to specify the number of cameras he would like to synchronize and then asks the user to connect the specified number of cameras and to enter the password of the cameras to enable the USB access to the cameras. Then, it will check if all the cameras have footage and/or images and will notify the user if any camera was found empty. Then, the application will ask the user to enter the school ID, Student ID and student name for every connected camera and will validate the entries and then start the automated filtering process. This process is fail safe and the data protection is always guaranteed (no data loss will occur in case of an unexpected failure caused by a system error or a hardware error at any point during the process).

The synchronization process is divided into 5 main phases/tasks:

- Task 1: initializing all parameters and variables, creating required directories, verifying the data on the connected cameras, copying the data from the cameras to the computer for faster and safer processing.
- Task 2: processing all the videos. the cameras are set to record videos so the application will extract a frame every 5-10 seconds from the recorded videos and delete the video files once all the needed frames are extracted.
- Task 3: running the food exposure detection models on all the images that were extracted in task 2. All images with no food exposure are deleted on the spot.
- Task 4: running face blurring on the remaining images with food exposure to blur any face that might be present in the image.
- Task 5: cleaning all the data from the cameras and generating a small report showing the total number of food exposure images being filtered, the percentage of filtered images, etc.. In case of errors that were handled

during the process, a warning message would also be generated at the end of the process stating what the error was and how it was handled.

This software tool is very computationally aggressive and requires a computer with very high specifications to run it at its optimal performance, especially when the number of cameras is 5 or higher. The application rely intensively on multi-threading and run most of the tasks in parallel for all the connected cameras. We provide the recommended specifications needed to run this application without problems:

- CPU: Intel Core i-7 with 6 cores or higher, clocked at 3.2 GHZ or higher
- GPU: NVIDIA GeForce GTX 1050 Ti or higher with 4 GB GDDR5 VRAM or more
- RAM: 16 GB DDR4
- Network bandwidth: No internet Connection is required for this application
- Storage: The application is nearly 100 MB in size and requires around 5 GB of space worth of libraries to run properly.

However, since the application is dealing with large amount of data (10 cameras maximum with an average of 20 GB of data per camera to be processed), it is recommended to have at least 500 GB of free storage available on the computer. In addition, since the application performs a lot of I/O operations simultaneously, it is recommended to have an SSD, NVME or PCIe based storage for faster I/O speed and better I/O Bandwidth to avoid I/O bottleneck caused by traditional HDD.

Figure 7.1 shows some screen-shots showing the user interface of the synchronization application.

7.3 Results and Samples

We started data collection in Tunisia on February 1st, 2020. So far the study has been done in more than 25 schools and the entire data collection process went flawless (no bugs, no technical problems). The accuracy of the machine learning models is excellent on real data from the outside world. So far, the lowest accuracy achieved by the food exposure model was 88% and most of the times, it is around 93%. As for the face detection and blurring accuracy, it is around 89%. **Figure 7.2** shows some samples from the data collected from the field study.

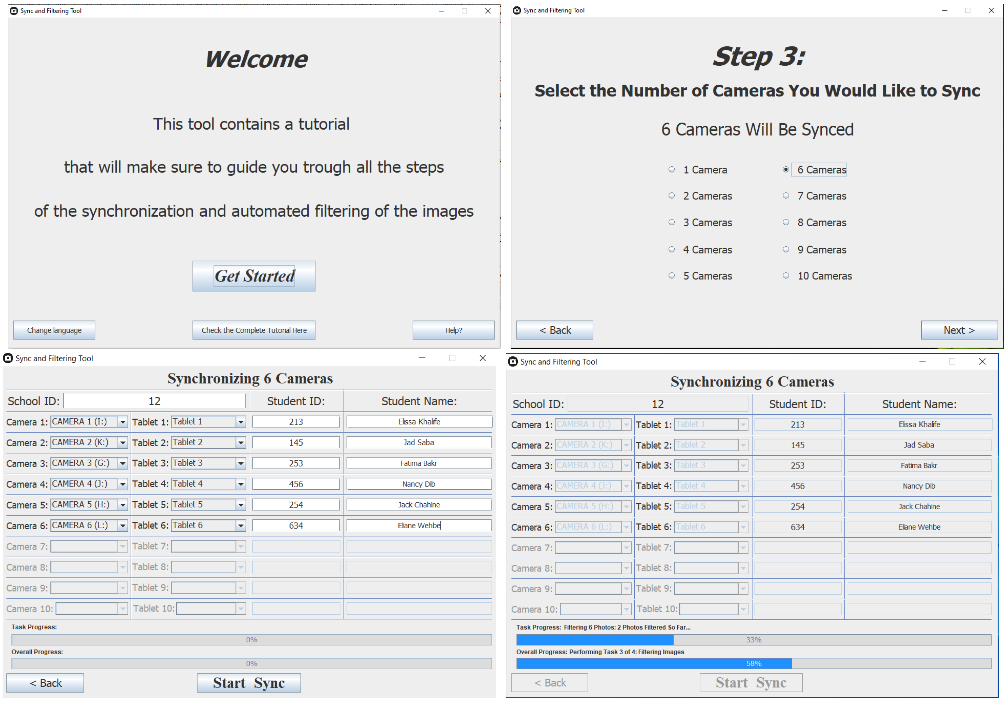


Figure 7.1: GUI of the AI Filtering Tool

Figure 7.2: Some Random Samples from the Collected Data After Automated Filtering

Chapter 8

System Deployment

In this section, we will discuss the deployment protocol and provide some high level analysis on the collected data from the real world.

8.1 Deployment Protocol and Status

This Study aims to deploy this AI based wearable cameras system in 100 schools in total (50 schools in Lebanon in the area of greater Beirut and 50 schools in Tunisia in the area of greater Tunis). We aim to have 10 participants from every school (a total of 1000 (500 tunisian and 500 lebanese) participants aged 11-12 years old from grade 5-6). The child will wear the camera after he leave the school and will return it the next day in the morning. A member of the research team will be waiting in the director office and in presence of parents to connect the cameras to the software tool that will handle the filtering and face blurring of the recordings found on the cameras. The parents will be allowed to screen the data after the AI system and delete any photos they feel uncomfortable to share with the research team using tablets.

In mid January 2020, we, a team of researchers, traveled to Tunisia where we trained a Tunisian data collection team on the deployment protocol. We spent around 3 weeks there and went to schools and had the chance to test everything in the real world and to experience and feel the responsibility of using technology and AI in the real world. Data collection started in Tunisia at the end of January 2020 and was paused by mid March because of the spread of the COVID-19 pandemic while being delayed in Lebanon because of the current lock-down situation. Data collection has been conducted in around 30 schools so far in Tunisia and the overall data collection progress is 26%. We still have around 20 schools in Tunisia and all 50 schools in Lebanon and we are waiting for the pandemic spread to minimize in order to resume the collection process.

8.2 Real world data Analysis

We conducted some statistical analysis on the data that we collected from the real world. First, we will start by discussing some of these statistics on the school level. As we can see in **Table G.1** (**Appendix G**), the majority of the schools that we managed to collect data from are public Tunisian schools .This is because the Tunisian data collection team decided to finish all the public schools first than start with the private schools later. For that reason, We won't be able to compare the acceptability of the technology in private vs public schools for now. However, we would like to note that our school selection protocol makes sure that the school selection is fair and reflect all socio-economic levels. The participation at the school level was very good as shown in **Figure G.1**; where the majority of schools (86%) had 8 or more participants in the study and 52% of schools had 10 or more participants. **Figure G.2** represent the school participation quality. The participation quality is based on the number of empty cameras returned by children over the total number cameras distributed to the children who accepted to participate in the study. If this ratio is between:

- $0 - 0.25 \Rightarrow$ Excellent participation
- $0.25 - 0.5 \Rightarrow$ Good participation
- $0.5 - 1 \Rightarrow$ Shy participation

Also, the majority of schools had excellent participation (52%) and good participation (17%). **Figure G.2** represent the usage awareness and/or confidence of using wearable cameras at the school level. As we can see, the majority of the schools were not aware or confident using the camera. The usage awareness and confidence index is based on the collection ratio index. The collection ratio index is measured as follows:

- We track the number of video files recorded on every camera
- The camera is set to automatically split the recordings in 3 minutes long videos.
- Based on the number of files available on the cameras, we are able to compute the theoretical recorded time and compared it against the actual recorded time screened by AI.
- A low collection ratio index means that the camera was turned on and off very frequently and for short amount of time (less than 3 continuous minutes). This usage pattern reflect that the kid was not familiar and aware on how to operate the camera in a correct manner or that he was not comfortable using it for a long time and he was just trying to capture in an active manner the food exposure.

- If the collection ratio is between:
  - 0-60 ⇒ not aware/confident
  - 60-100 ⇒ aware confident

Moving to Child statistics, as we can see in all tables and figures of **Appendix H**, around 55% to 65% of the participants in the study were confident and aware of using the cameras and had a good or excellent participation quality in the study. When looking at Age based statistics in **Appendix I**, while age difference didn't show any major difference regarding the awareness and / or confidence of using technology, however, 12 years old kids seems to have used the cameras to record more time of the day in general and this explain the slight advantage in the recorded food exposure time as well. Moving to Gender based statistics in **Appendix J**, as we can see, male participants were more engaged in the study than the female participants and male participants where more aware and confident in using a wearable camera as well with 67% of male participants being aware and confident compared to 52% for female participants. As for Income based statistics shown in **Appendix K**, participants with lower family income have better participation than participants with higher family income level in terms of recorded food exposure time and the acceptability of using wearable cameras. Also, we can notice that participants where their parents have a higher educational level were not involved in the study as much as the participants that their parents have a lower educational level (**Appnedix L**). This seems to be correlated with parents income level that we already checked since higher educational level usually lead to higher family income and if we compare the income based stats (**Appendix K**) and the educational level based stats (**Appendix L**), we can see a close similarity in the reported numbers based on those 2 factors.

In brief, based on the amount of data that we managed to collect from the real world so far and the statistical analysis that we conducted on this data, we can conclude that the specifications and requirements defined in the chapter 3 of this thesis that would make the usage of an AI based wearable camera system acceptable for a majority of children and their parents succeeded to a large extent in making the usage of AI in the real world acceptable by around 60% to 75% of participants who were asked if the would like to take part of our study. All these numbers are still prone to change since the just collected around 26% of the estimated total amount of data that this study aims to collect by its end, however, those percentages are promising and reflect a good acceptability rate in general for such a new and advanced technology to be used in the real world.

Chapter 9

Conclusion and Outlook

We conclude that it is feasible to use wearable cameras and AI to capture children food exposure while keeping the data confidential and secure. We also can conclude that using AI in the real world is acceptable under several requirements and specifications and it is not considered as a threat to personal privacy. In addition, we explored the possibility to generate a dataset using a web crawler and to label it by relying on a web search engine which turned out to be a reliable way to label very large data sets and eliminate the need of crowd-sourcing to label the data. Moreover, we surveyed most of the available wearable cameras in the market and highlighted the features, pros and cons of every camera and described a way to evaluate cameras in general for similar studies. In addition, we trained a machine learning model that would work as a classifier for images that contain food exposure and we added on top of it a face blurring layer that would blur faces in the food exposure images if available. We used this pipeline of models to create an AI tool for researchers that will filter the images / video frames captured by up to 10 wearable cameras, delete the images without food exposure on the spot and keep the images with food exposure while blurring all the faces to preserve the privacy and confidentiality of the study and the involved 3rd party. Finally, we deployed and tested this tool in the real world and collected a large amount of data and provided some statistics and analysis on the collected data.

In future work, we will try to improve on the accuracy of the model by collecting a few thousands of images from the real world to improve the real word accuracy of the model, which will improve the results we obtain from the entire procedure. These images will be reflective of the real-life scenarios and should help improve the accuracy of our model in some tricky cases that were not reflected in the dataset that we crawled from the Internet. Also, we will resume the data collection process in Tunisia and Lebanon when the Covid-19 pandemic stabilize and schools return to normal. Furthermore, we are looking into expanding the pipeline of neural networks to provide us with more information about the food exposure present in the image like predicting its type (if it is a food item, a restaurant, an ad, a brand, etc...). Also, we are interested in exploiting scientific ways to annotate and label the images collected from the study by nutritionists in order to build another machine learning model that will provide a healthiness score (or health index) for every food item present in a given image or in other words, to try to build a machine learning model that can play the role of a nutritionist by inferring from an image containing a food item how much this food is healthy.

Appendix A

Findings of the participatory research study

Table A.1: Findings of The User Centered Acceptability Workshop In School 1, Beirut

	Lebanon		
	School 1		
	Team 1	Team 2	Team 3
Device	Camera	Camera	Camera
Image capture	Wearable	Wearable	Wearable
		Vest	Vest
	Passive [15s]	Passive	Passive
	Active	Active	
Image filtering	Manual selection	Manual selection	
		Automated (keeps only food)	Automated (keeps only food)
		Filter by time	
		Filter by location	
Privacy	Blurring faces	Blurring faces	Blurring faces
		Encrypt	Encrypt
	Password	Password	Password
		Turn off	
		Reminder to turn back on	
Transfer	Cable transfer	Cable transfer	Cable transfer
	Bluetooth	Uploading	

Table A.2: Findings of The User Centered Acceptability Workshop In School 2, Beirut

	Lebanon		
	School 2		
	Team 1	Team 2	Team 3
Device	Phone	Camera	Phone
Image capture	Wearable	Wearable	Wearable
	Vest	Vest	Vest
		Passive	Passive
		Active	
Image filtering	Manual selection		Manual selection
	Automated (keeps only food)	Automated (keeps only food)	Automated (keeps only food)
		Filter by time	
	Filter by location	Filter by location	
Privacy	Blurring faces	Blurring faces	Blurring faces
	Encrypt	Encrypt	Encrypt
		Password	Password
	Turn off		Turn off
	Reminder to turn back on		Reminder to turn back on
Transfer	Cable transfer	Cable transfer	Cable transfer
	Messages	Uploading	Uploading
	Bluetooth		Bluetooth

Table A.3: Findings of The User Centered Acceptability Workshop In School 1, Tunisia

	Tunisia		
	School 1		
	Team 1	Team 2	Team 3
Device		Camera or Phone	Phone
Image capture	Vest		Vest
	Passive		Passive
	Active	Active	
Image filtering		Manual selection	Manual selection
	Automated (keeps only food)	Automated (keeps only food)	Automated (keeps only food)
	Filter by time		Filter by time
	Filter by location		Filter by location
Privacy	Blurring faces		Blurring faces
	Encrypt	Encrypt	Encrypt
		Password	Password
	Turn off		Turn off
	Reminder to turn back on	Reminder to turn back on	Reminder to turn back on
Transfer	Cable transfer	Cable transfer	Cable transfer
	Uploading		Uploading
	Bluetooth		

Table A.4: Findings of The User Centered Acceptability Workshop In School 2, Tunisia

	Tunisia		
	School 2		
	Team 1	Team 2	Team 3
Image capture	Wearable	Wearable	Wearable
	Vest	Vest	Vest
	Passive		Passive
	Active	Active	
Image filtering	Manual selection	Manual selection	Manual selection
	Automated (keeps only food)	Automated (keeps only food)	Automated (keeps only food)
			Filter by time
			Filter by location
Privacy			Blurring faces
	Encrypt	Encrypt	Encrypt
			Password
	Turn off		Turn off
	Reminder to turn back on		Reminder to turn back on
Transfer		Cable transfer	Any mode
	Uploading	Uploading	
		Messages	

Table A.5: Findings of The User Centered Acceptability Workshop In School 3 and 4, Tunisia

	Tunisia		
	School 3	School 4	
	Team 1	Team 1	Team 2
Device	Phone	Camera	
Image capture	Wearable		
		Vest	Vest
		Passive	Passive
			Active
Image filtering	Manual selection	Manual selection	Manual selection
	Automated (keeps only food)	Automated (keeps only food)	Automated (keeps only food)
	Filter by time		
Privacy	Blurring faces		
	Encrypt	Encrypt	Encrypt
		Password	
	Reminder to turn back on		
Transfer	Cable transfer	Cable transfer	Any mode
	Uploading		

Appendix B

Mapping of ethical frameworks to implementation protocol

Table B.1: Mapping of ethical frameworks to implementation protocol

Ethical principal	Guidelines	Measures for implementation
Privacy and confidentiality	Devices should be configured so that data can only be retrieved by the research team. It should be impossible for participants or third parties who find devices to access images.	The device is not equipped with a screen to avoid participants viewing the captured images. The cameras are password protected and the password is only known by the PI and Citi certified members of the research team. Therefore, only members of the research team can have access to the data.
	Data should be stored according to national data protection regulations	The collected dataset will be stored on a password-protected computer maintained by the PI. Within 3 years of study completion, the PI will destroy the datasets, erasing them from the encrypted hard drive completely.
	Identifying images should not be used without express consent of those individuals who are depicted	No identifying images will be retrieved as the faces will be blurred in all the images, nor will these be shared in publications or conferences.
	Devices should be configured to allow participants to cease recording for short periods. Participants should be allowed to remove the device of any time, with examples of where this might be appropriate (e.g. airport security).	Students will be informed that they can turn the camera off anytime they don't feel comfortable wearing it, they will also be instructed to remove the camera (1) before going to the bathroom or changing clothes, (2) in pubic bathrooms, (3) in public places where there is a notice/sign that prohibits taking pictures, and (4) before engaging in physical activity or sports.
	Appropriate training should be provided for all those in the research team who have contact with the image data	Only Citi certified personnel will have access to the collected data from this study component. Analysis will be mostly conducted using the automatic machine learning model, therefore researchers will not directly view the data.
	Narrowing the scope of captured data	Data collection will only be for the duration of 1 afternoon/evening and the following morning. Images will be automatically filtered to keep only those related to food.
Non-maleficence	Participants should be prepared for questions by the public with a short sentence that explains the device and concludes with an offer to remove if they are feeling uncomfortable.	Children will be provided with a written note that they can show to those who ask about the camera. The note will inform third parties that this is part of a research study conducted by the American University of Beirut, and that the objective is to capture the food environment and not the individuals in it. The note will also state that all faces will be blurred in the retrieved pictures, and that these pictures will not be shared nor disseminated. The note will include the PI's and IRB's contact information for reference. Students will also be advised to turn the camera off/remove it if requested by third parties.
	Participants should be instructed to remove device in any situation where it is attracting unwanted attention, or they feel threatened or uneasy wearing the device.	As mentioned above.
Third party consent	Taking informed consent from people familiar to the research participant who are likely to be recorded	1- Children are advised to inform their close relatives and household members about the camera and the purpose of its use. 2- They will be provided with a card that explains that they are part of research study about food, and that the aim is to capture the food environment and not third parties. 3- The card will explain that the pictures are only accessible to the research team and that all faces will be blurred, and that pictures will not be shared.
	Minimizing the recording of third parties by reducing the scope of what is being recorded	The scope of collected images will be reduced through a first round of automated filtering which will delete all pictures not containing food items.

Ethical principal	Guidelines	Measures for implementation
Autonomy of third parties	Participants should seek verbal permission from family members and cohabitants before study commencement.	To begin with, parental permission is required before giving students the cameras. Then they will be instructed to inform people who are familiar to them about the camera and its purpose.
	Participants should seek verbal permission of workplace mangers or supervisor. If possible, this should be prior to study commencement, but in reality, may be a rolling process. Appropriateness of device to work setting should be assessed by researcher	Not applicable. Schools will be aware of this study component, and the cameras will not be worn at schools.
	Participants should inform friends and acquaintances of device when encountered and offer to remove device if they are uncomfortable.	Participants will be instructed to inform people who are familiar to them about the camera and its purpose.
	Participants should be told to inform third parties that they also can request image deletion by asking the participant to inform the research team or contacting them directly.	Participants will be told to inform third parties that all images not containing food items will be deleted automatically by the machine learning model, without anyone viewing them. Third parties will also have access to the PIs contact information to place any clarifications or requests.
	The privacy and anonymity of third parties must be protected; no image that identifies them should be published without their consent.	No identifying images will be retrieved as the faces will be blurred in all the images, nor will these be shared in publications or conferences. A second round of image filtering will be conducted by parents, who will only have access to images with blurred faces.
	Photography is inappropriate in some cultural settings, and automated, wearable cameras should not be used in these instances.	Children will be advised to turn the cameras off/remove them if they feel it is inappropriate to take pictures.
Anonymity, lifelogging, and increasing data protection	Automated image analysis (researchers do not even see third parties)	The final dataset will be analyzed by a machine learning model, with all the faces being be blurred in the final dataset.
	Physical separation between data capture, data storage and analysis	Data will be captured on the cameras and transferred via a USB cable to a password protected computer. Data will be analyzed through the machine learning model which is conducted automatically, without accessing/viewing the data.
	Protected data storage	The collected dataset will be stored on a password-protected computer maintained by the PI. Within 3 years of study completion, the PI will destroy the datasets, erasing them from the encrypted hard drive completely.
Maleficence/social acceptability	Protecting participants from third party suspicion	1- Children are advised to inform their close relatives and household members about the camera and the purpose of its use. 2- They will be provided with a card that explains to third parties that they are part of research study about food, and that the aim is to capture the food environment and not third parties. 3- The card will explain that the pictures are only accessible to the research team and that all faces will be blurred, and that pictures will not be shared.
Surveillance, intrusion and discipline	Ensure that the device is not used by parents to monitor their children	1- Parents will only be able to view the images after the first round of automated filtering, so they will only view a small amount of data. 2- Children will be informed that their parents will view captured pictures at the end of data collection. 3- Children will be advised that they have the freedom to remove/turn off the camera anytime they don't feel comfortable using it, or when they don't want something to be recorded.

Ethical principal	Guidelines	Measures for implementation
Informed written consent should state:	How much information is collected	The consent form states the duration and time intervals of image capture
	The nature and type of data that can be collected by wearing an automated wearable camera (images will depict where you go, what you do, and for how long) with examples	The parental permission and child assent forms state: "As long as the camera is on, it will capture images of the child's surroundings, including his/her neighborhood environment, home environment, activities that they do (e.g. eating, studying, playing with friends). The research team is only interested in pictures that describe the food environment (e.g. meals, food advertisements, food shops, food in the kitchen, etc.). Therefore, we will protect participants' and third-party privacy and confidentiality by **automatically** deleting all irrelevant pictures that do not contain food, without viewing the pictures."
	Participants can forget they are wearing device and record unwanted and unflattering images with examples provided (e.g., bathroom visits, online banking)	The parental permission and child assent forms state: "Your child is advised to remove or turn off the camera when they feel uncomfortable wearing it; specifically: 1- Before going to the bathroom or changing clothes 2- In pubic bathrooms 3- In public places where it is prohibited to take pictures 4- Before engaging in physical activity or sports Your child might forget to remove the camera/turn it off in such places. In this case, the captured images will be **automatically** deleted without anyone viewing them. You will also screen the obtained set of pictures in a private room and delete unwanted images."
	Data of illegal activities may not be protected by confidentiality and may be passed to law enforcement depending on the national law and nature if the activity	The parental permission and child assent forms state: "Data of illegal activities may not be protected by confidentiality and may be passed to law enforcement depending on the national law and nature if the activity".
	No individual will be identifiable in any research dissemination without their consent	The parental permission and child assent forms state: "We will only share pictures containing food-related items, and all the faces will be blurred in the final set of pictures. Your and your child's anonymity will be protected as you will not be identified in any published data."
	Participants will have the opportunity to view (and delete if necessary) their images in privacy	Stated in the parental permission and child assent forms
	Participants are able to remove the device or temporarily pause image capture whenever they wish	Stated in the parental permission and child assent forms
	Participants will not get copies of their images	The parental permission and child assent forms state: "You and your child will not have a copy of any of the captured images."
	A team of specifically trained researchers will have access to the image data	The parental permission and child assent forms state: "Only trained researchers will handle and analyze the data. The majority of the analysis will be conducted by an automated program (machine learning model), which does not require researchers to view any of the captured images."

Appendix C

Findings of the Camera Survey

Table C.1: List of Non-recommended Cameras that were Tested and/or Reviewed

Camera Name	Brand	Resolution	Battery Life
EE Capture Cam	EE	8MP	2 hours continious runtime
FrontRow	FrontRow	5MP & 8MP	16 hours continious runtime
HP lc100w	HP	8MP	2 hours continious runtime
iON 1046 Lite SnapCam	iON	5MP	2.5 hours continious runtime
iON SnapCam	iON	8MP	3000 photos/charge
MeCam Classic	Mecam	5MP	80 minutes continious runtime
MeCam HD	Mecam	8MP	2 hours
MeCam NeoMe	Mecam	8MP, wide angle	75 minutes continious runtime
Narrative Clip	Narrative	5MP	all day long
Narrative Clip 2	Narrative	8 MP , 90 degree	all day long
Qlippie	Qlippie	8MP, wide angle	2 hours continious runtime
SereneLife	SereneLife	12MP	2 hours continious runtime
YoCam	YoCam	N/A, 30FPS	2 hours
Hero 7 Black	GoPro	12 MP	2.5 hours continious runtime

Camera Name	Storage	Protection	Features
EE Capture Cam	up to 32 GB microSD	-	photo and video recording
FrontRow	up to 32 GB microSD	-	timelapse and video recording
HP lc100w	up to 32 GB microSD	IP66	video recording
iON 1046 Lite SnapCam	up to 32 GB microSD	-	photo and video recording
iON SnapCam	up to 32 GB microSD	-	video streaming, magnetic clip
MeCam Classic	B built-in, expandable to 32	-	photo and video recording
MeCam HD	up to 32 GB microSD	-	melapse every 2, 30, or 60 second
MeCam NeoMe	B built-in, expandable to 32	30M waterproof	5 timelapse options
Narrative Clip	8GB built-in, 4000 Photos	-	photo and video recording
Narrative Clip 2	8GB built-in, 4000 Photos	-	elapse every 10, 30, 60, 120 seco
Qlippie	32 GB built-in	-	timelapse and video recording
SereneLife	up to 32 GB microSD	-	photo and video recording
YoCam	up to 32 GB microSD	IP68	video streaming
Hero 7 Black	32 GB microSD card	IP67	timelapse and video recording

Camera Name	Form Factor	Weight	Dimensions	Price
EE Capture Cam	attachable square	-	-	180$
FrontRow	attachable necklace	-	-	400$
HP lc100w	attachable rectangle	240 g	180x100x50 mm	150$
iON 1046 Lite SnapCam	attachable square	-	-	50$
iON SnapCam	attachable square	30 g	38x38x10 mm	150$
MeCam Classic	attachable necklace	30 g	50 mm diameter	40$
MeCam HD	attachable square	80 g	50x50x10 mm	200$
MeCam NeoMe	attachable square	40 g	42x42x20 mm	150$
Narrative Clip	attachable square	-	36x36x9 mm	200$
Narrative Clip 2	attachable square	20 g	36x36x9 mm	200$
Qlippie	attachable square	45 g	48x48x20 mm	300$
SereneLife	attachable square	-	-	50$
YoCam	attachable rectangle	-	-	300$
Hero 7 Black	attachable rectangle	-	-	450$

Appendix D

List of Food Exposure Dataset Queries

- tunisian food
- tunisian food popular
- traditional tunisian food
- tunisian cuisine
- arabic food
- real scenario tunisian food images
- people eating tunisian food
- people in tunisian restaurant
- very far images tunisian food
- eating tunisian
- best tunisian restaurants
- lebanese food
- lebanese food popular
- traditional lebanese food
- lebanese cuisine
- real scenario lebanese food images
- people eating lebanese food
- people in lebanese restaurant
- very far images lebanese food
- eating lebanese
- best lebanese restaurants
- food images low light
- food ads
- food billboards
- food billboards lebanon
- food billboards Tunisia
- food transportation
- food vehicules
- food packages
- food packaging
- food menus
- real scenario food images
- people eating
- supermarket
- market
- food market
- real scenario images containing food
- very far images food
- eating
- pizza
- burger
- kitchen
- water
- water bottles
- water galloons
- refrigerator
- vending machine
- chocolat
- candy
- cream
- juice bottles
- biscuit
- jello
- custard
- soft drinks

- fries
- potato
- milk
- croissant
- marshmallow
- nuts
- popcorn
- steak
- chips
- corn flakes
- barbeque
- hot dog
- tartine
- mortadella
- mortadella sandwish
- nutella
- chocomax
- laban
- labneh
- cake
- lebanese food lebanon
- lebanese food pinterest
- lebanese food images in restaurants
- lebanese restaurants
- lebanese food ads
- lebanese food breakfast
- lebanese food lunch
- lebanese food dinner
- lebanese food dessert
- lebanese food fruits
- lebanese food vegetables
- lebanese cocktail
- lebanese drinks
- lebanese coffee
- lebanese shisha
- lebanese cigarettes
- lebanese smoking
- lebanese caffe
- lebanese bakery
- lebanese manakish
- lebanese fast food
- lebanese pizza
- lebanese burger
- lebanese sandwich
- lebanese salad
- lebanese ice cream
- lebanese saj
- lebanese seafood
- lebanese kebab
- lebanese steak
- lebanese vegetarian food
- lebanese tea
- lebanese crepes
- lebanese foul
- lebanese juices
- lebanese pub food
- lebanese shawarma
- lebanese falafel
- lebanese oriental food
- lebanese kaak
- lebanese healthy food
- lebanese sushi
- lebanese food kiosks
- lebanese bakeries
- lebanese furn
- lebanese food courts
- lebanese pubs
- lebanese food locations
- lebanese kibbeh
- lebanese hummus
- lebanese tabouli
- lebanese mujadara
- lebanese homemade food
- lebanese food pastry
- tunisian food tunisia
- tunisian food pinterest
- tunisian food images in restaurants
- tunisian restaurants
- tunisian food ads
- tunisian food breakfast
- tunisian food lunch
- tunisian food dinner
- tunisian food dessert
- tunisian food fruits

- tunisian food vegetables
- tunisian cocktail
- tunisian drinks
- tunisian coffee
- tunisian shisha
- tunisian cigarettes
- tunisian smoking
- tunisian caffe
- tunisian bakery
- tunisian manakish
- tunisian fast food
- tunisian pizza
- tunisian burger
- tunisian sandwich
- tunisian salad
- tunisian ice cream
- tunisian saj
- tunisian seafood
- tunisian kebab
- tunisian steak
- tunisian vegetarian food
- tunisian tea
- tunisian crepes
- tunisian foul
- tunisian juices
- tunisian pub food
- tunisian shawarma
- tunisian falafel
- tunisian oriental food
- tunisian kaak
- tunisian healthy food
- tunisian sushi
- tunisian food kiosks
- tunisian bakeries
- tunisian furn
- tunisian food courts
- tunisian pubs
- tunisian food locations
- tunisian kibbeh
- tunisian hummus
- tunisian tabouli
- tunisian mujadara
- tunisian homemade food
- tunisian food pastry
- mcdonalds lebanon
- burger king lebanon
- kfc lebanon
- roadster diner lebanon
- zaatar w zeit lebanon
- tonino lebanon
- malak al tawouk lebanon
- hardees lebanon
- le sam lebanon
- cafeteria lebanon
- cafeteria
- school catereia lebanon
- deek duke lebanon
- classic burger lebanon
- starbucks lebanon
- lody's lebanon
- al saniour lebanon
- go tango lebanon
- bliss house lebanon
- kb doner lebanon
- domino's pizza lebanon
- pizzahut lebanon
- wooden bakery lebanon
- breakfast to breakfast lebanon
- sea sweet lebanon
- dip and dip lebanon
- subway restaurant lebanon
- mcdonalds tunisia
- burger king tunisia
- kfc tunisia
- roadster diner tunisia
- zaatar w zeit tunisia
- hardees tunisia
- le sam tunisia
- cafeteria tunisia
- school cafeteria tunisia
- school kiosks tunisia
- deek duke tunisia
- classic burger tunisia
- starbucks tunisia
- kb doner tunisia
- domino's pizza tunisia

- pizzahut tunisia
- dip and dip tunisia
- Biscuits industriels
- Cake industriel
- Casse-croûte au Fromage
- Céréales Grain d'Or
- Couscous agneau
- Couscous poulet
- Danette cream
- Dinde grillée
- Grilled turkey
- Fromage fondu triangle
- Spread cheese
- Fruits
- Industrial wafers
- wafers
- Jus frais
- Fresh juice
- Industrial juice
- Lait
- Légumes
- Macaroni agneau
- Macaroni poulet
- spagetti
- Nouacer poulet
- oeuf frit
- Fried egg
- Pizza fromage thon
- Pizza cheese tuna
- Pomme de terre frite
- French fries
- Ragoût haricot blanc poulet
- Ragoût petit pois poulet
- Ragoût pomme de terre poulet
- Riz poulet
- Rice with chicken
- Salade méchouia
- Salade verte
- lettuce salad
- Sodas
- Tagine poulet
- Viennoiseries
- Yaourt
- Yogurt
- Mloukhia boeuf
- Strew with Corète and beef
- Chawarma dinde
- Chawarma
- Casse-croûte thon
- Sandwich with tuna
- Purée pomme de terre
- Mashed potatoes
- pepsi
- coca cola
- 7 up
- mirinda
- fanta
- srite
- wings
- fasoulia
- yakhne
- Chocolat
- Brick potato cheese tuna
- Crepe au chocolat
- Omelette
- oeuf
- beurre
- fromage
- tomate
- concombre

Appendix E

List of Non Food Exposure Dataset Queries

- birds
- animals
- insects
- animals macro shots
- reptiles
- fish
- cat
- dog
- bear
- panda
- hamster
- souris
- moskitoes
- pets
- home animals
- night shots
- landscape
- sunset
- low light landscape
- night shots landscape
- beach
- sand
- sunrise
- sky
- snow
- wonders
- river
- hills
- oceans
- sea
- forest
- desert
- space
- tree
- mountain
- flowers
- cloud
- sun
- valley
- rocks
- wood
- plants
- planets
- stars
- roads
- lights
- light bulbs
- storms
- thunder
- rain
- t shirt
- tshirt
- t-shirt
- socks
- shoes
- pants

- jeans
- arms
- hands
- legs
- bag
- school bag
- jacket
- keys
- sofa
- curtains
- fingers
- cars
- busses
- airplanes
- jet
- tablets
- metro
- train
- computers
- technology
- smart phones
- iPhone
- iPad
- bikes
- ships
- television
- screen
- house
- garden
- bedroom
- living room
- fourniture
- bathroom
- door
- window
- music instruments
- music
- painting
- football
- basketball
- tennis
- volley
- baseball
- hokey
- bawling
- ball
- casino
- swimming
- running
- walking
- walk
- chat
- text
- exam
- score
- laboratory
- lab
- people talk
- social media
- tv
- street
- libraries
- university
- campus
- stage
- arena
- festival
- highway
- asphalt
- walls
- rooms
- classes
- tables
- chairs
- labs
- city
- village
- subways
- garbage
- clothes
- shops
- people night shot
- people face
- people

Appendix F

Real Data Statistics Reflecting a Bias Free ML Model

Table F.1: General Average Metrics

General Metrics	Values
Total Number of Captured Images	567506
Total Number of Detected Food Exposure Images	62745
Average Exposure Ratio	11.21%
Average Deletion Ratio	88.79%
Average Recorded Hours	3.27
Average Recorded Minutes	196
Average Exposure Minutes	22

Table F.2: Gender Based Metrics

Gender Based Metrics	Both Genders	Male	Female	Female to Male Ratio
Average Exposure Ratio	11.21%	11.06%	11.49%	1.04
Average Deletion Ratio	88.79%	88.94%	88.51%	1.00
Average Recorded Hours	3.27	3.63	2.91	0.80
Average Recorded Minutes	196	217	174	0.80
Average Exposure Minutes	22	24	20	0.83

Table F.3: Age Based Metrics

Age Based Metrics	All Age Groups	11 Years Old	12 Years Old	12 Y.O. to 11 Y.O. Ratio
Average Exposure Ratio	11.21%	10.47%	12.12%	1.16
Average Deletion Ratio	88.79%	89.53%	87.88%	0.98
Average Recorded Hours	3.27	3.19	3.30	1.03
Average Recorded Minutes	196	191	198	1.04
Average Exposure Minutes	22	20	24	1.20

Table F.4: Income Based Metrics

Income Based Metrics	All Salary Ranges	Less Than 1500 TD	More Than 1500 TD	> 1500 TD to < 1500 TD Ratio
Average Exposure Ratio	9.03%	9.03%	8.39%	0.93
Average Deletion Ratio	90.97%	90.97%	91.61%	1.01
Average Recorded Hours	2.77	2.77	2.78	1.00
Average Recorded Minutes	166	166	167	1.00
Average Exposure Minutes	15	15	14	0.93

Table F.5: Parents' Education Level Based Metrics

Parents' Education Level Based Metrics	All Education Levels	Less than Baccalaureate	Baccalaureate or Higher	Ratio
Average Exposure Ratio	9.77%	8.48%	11.54%	1.36
Average Deletion Ratio	90.23%	91.52%	88.46%	0.97
Average Recorded Hours	2.9	3.34	2.31	0.69
Average Recorded Minutes	174	200	139	0.69
Average Exposure Minutes	17	17	16	0.94

Table F.6: Internet Access Based Metrics

Internet Access Based Metrics	Total	Internet Access	No Internet Access	Ratio
Average Exposure Ratio	9.84%	9.54%	10.46%	1.10
Average Deletion Ratio	90.16%	90.46%	89.54%	0.99
Average Recorded Hours	2.88	2.97	2.71	0.91
Average Recorded Minutes	173	178	163	0.91
Average Exposure Minutes	17	17	17	1.00

Appendix G

School Level Statistics

Table G.1: School Distribution Between Public and Private Sectors

Schools	Count	Percentage
Total Schools	29	100%
Public Schools	27	93%
Private Schools	2	7%

Figure G.1: Schools Participation Distribution

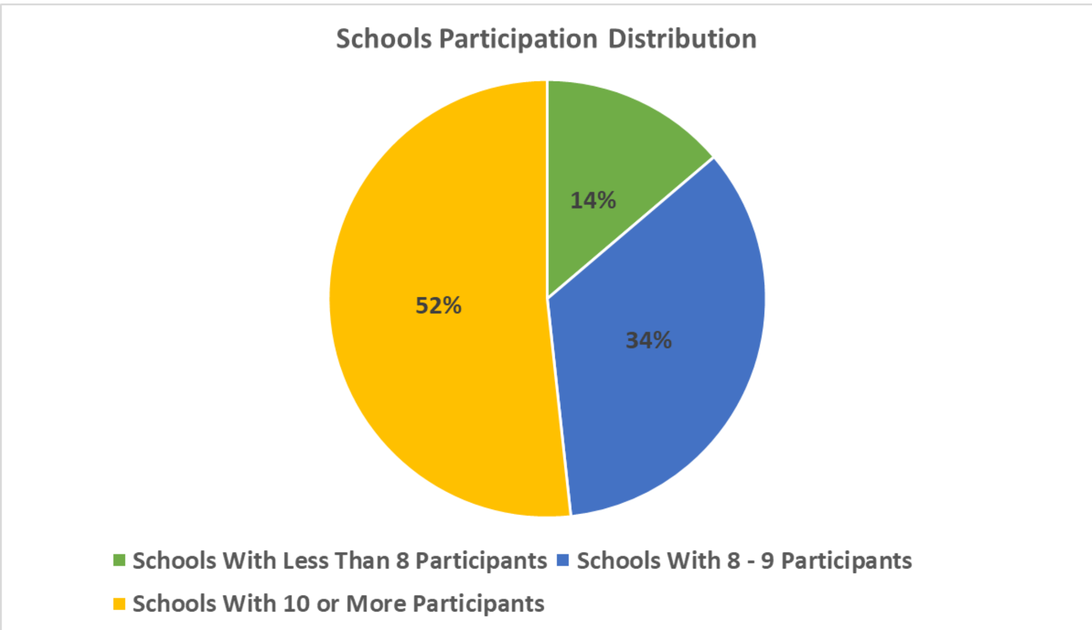


Figure G.2: Schools Participation Quality

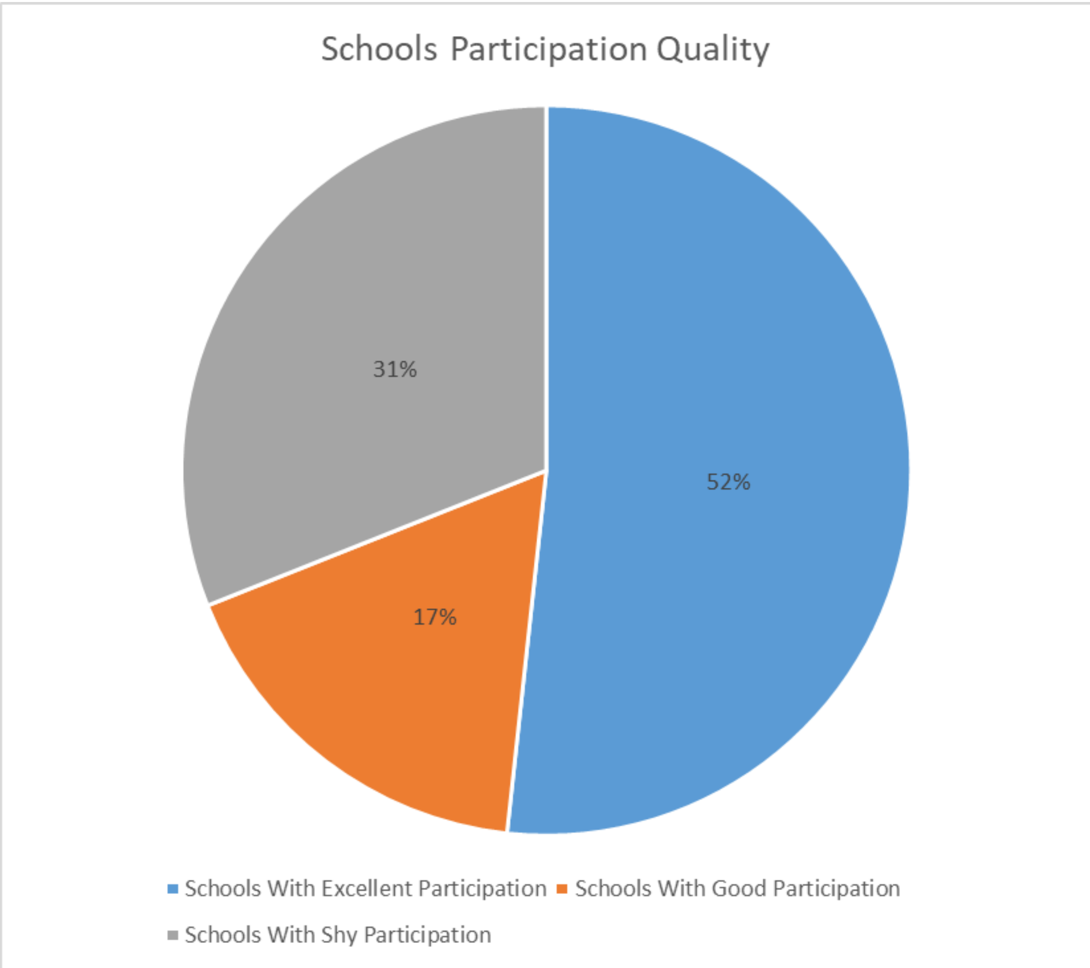
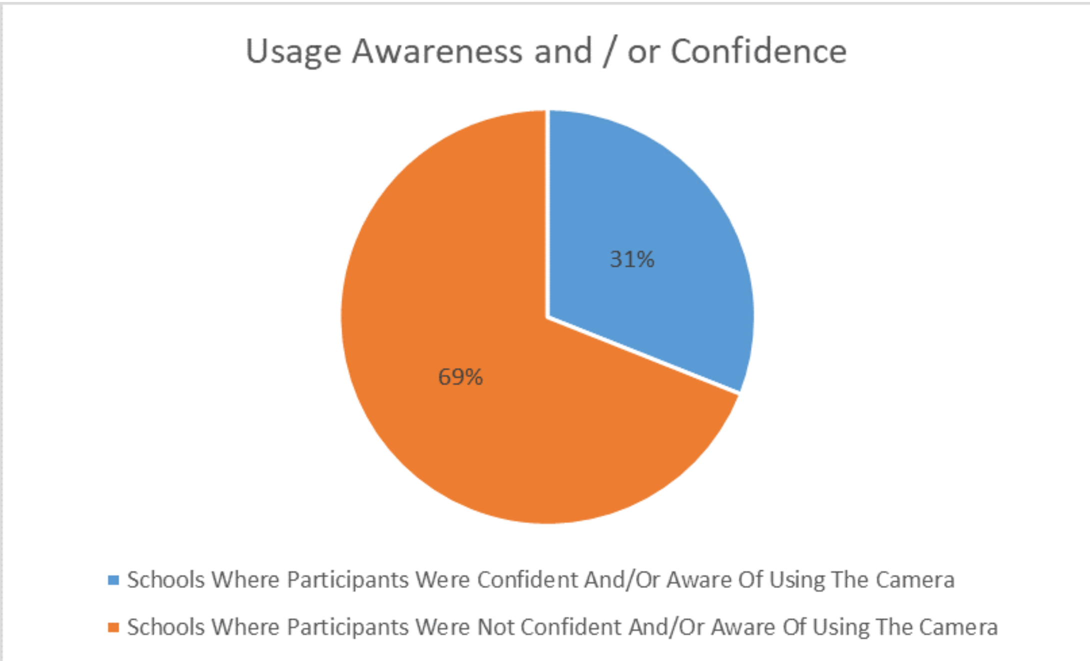


Figure G.3: Usage Awarness and/or Condifence Distribution at Schools



Appendix H

Child Level Statistics

Table H.1: Distribution of Recording Time

Recorded Time	Count	Percentage
Total Number Of Participants	265	100.00%
Participants With Empty Cameras	22	8.30%
Participants With Less Than 10 Minutes Of Recordings (Including Empty Cameras)	74	27.92%
Participants With Less Than 10 Minutes Of Recordings (Excluding Empty Cameras)	52	19.62%
Participants With 10 Min To 1 Hour Of Recordings	50	18.87%
Participants With 1 To 2 Hours Of Recordings	30	11.32%
Partricipants With 2 To 4 Hours Of Recordings	38	14.34%
Participants With 4 To 8 Hours Of Recordings	45	16.98%
Participants With 8 To 12 Hours Of Recordings	18	6.79%
Participants With 12 To 21 Hours Of Recordings	10	3.77%

Figure H.1: Children Participation Quality

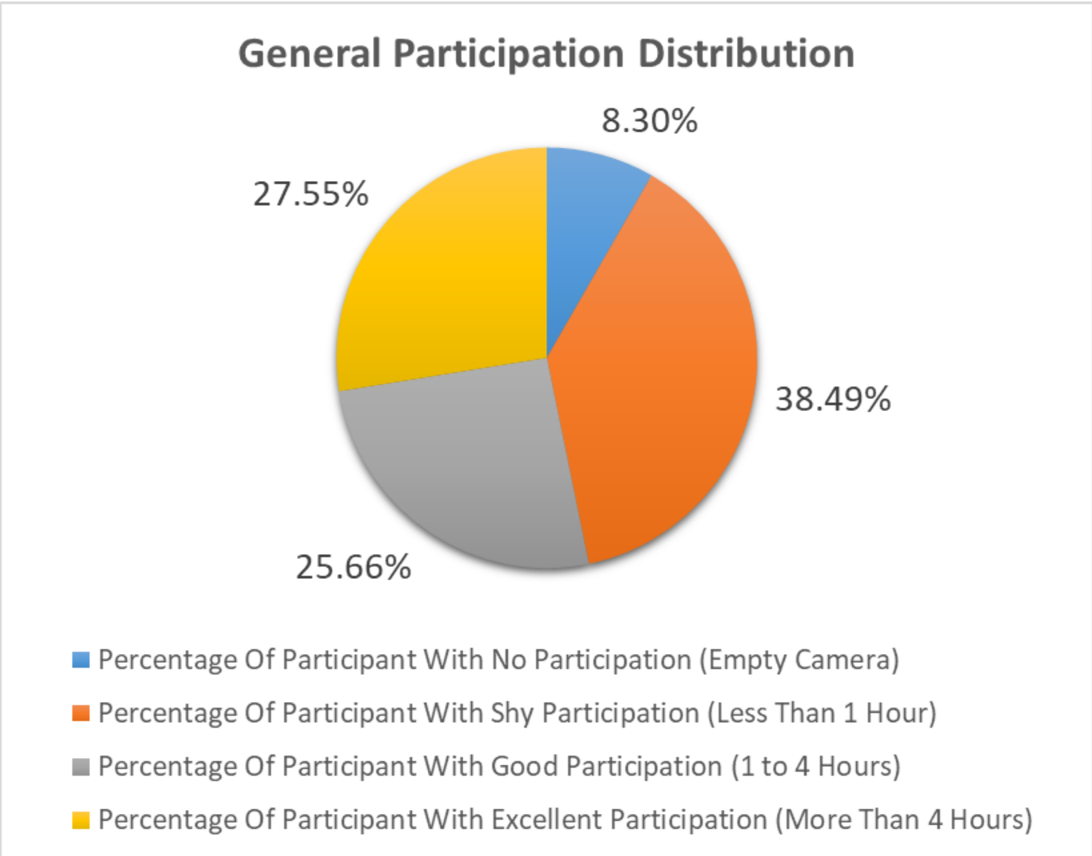
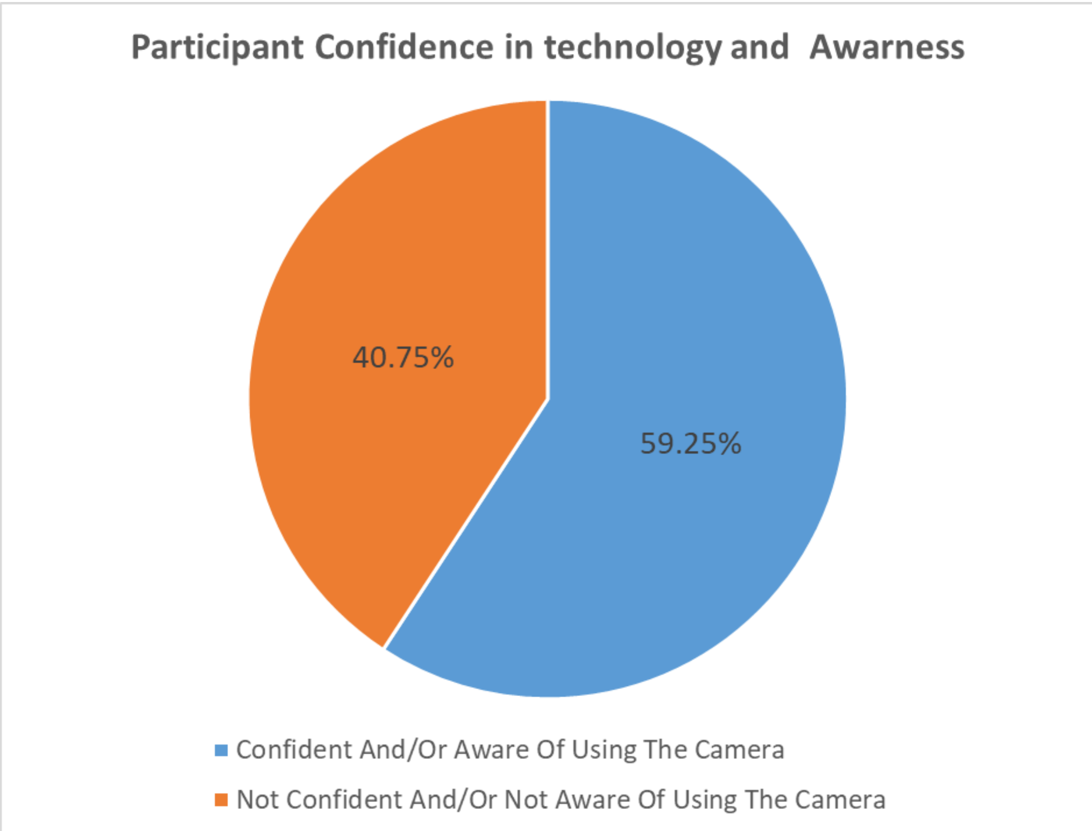


Figure H.2: Awareness and Confidence Index



Appendix I

Age Based Statistics

Figure I.1: Age Distribution

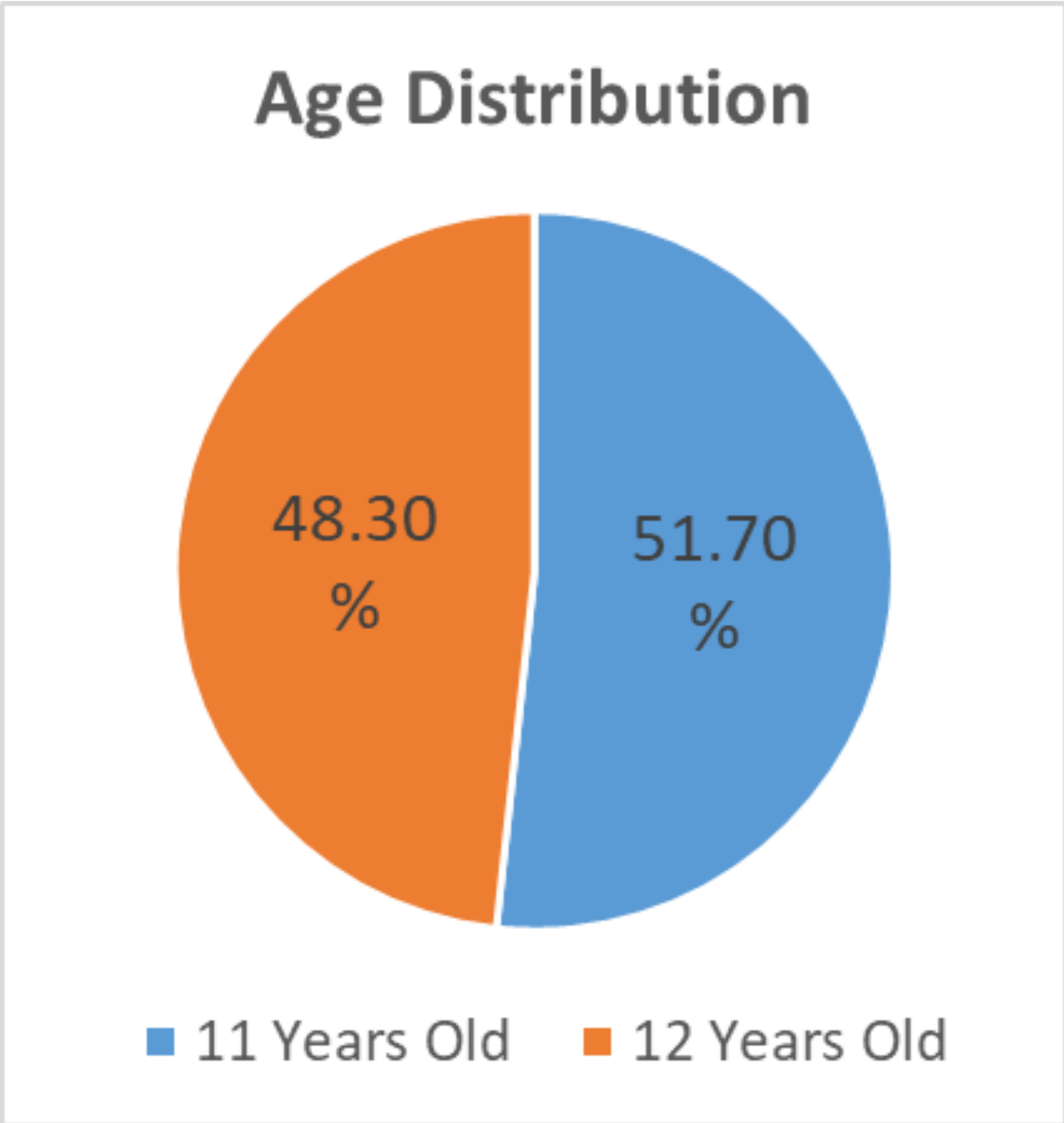


Figure I.2: Age based Participation Quality

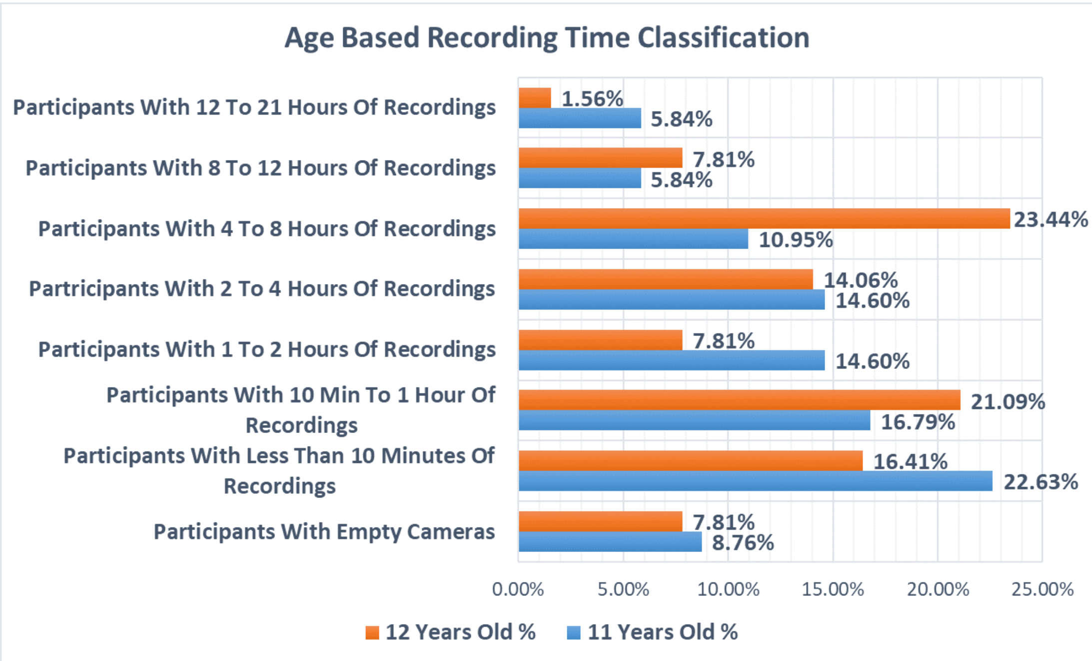


Figure I.3: Age Based Recorded Food Exposure Time

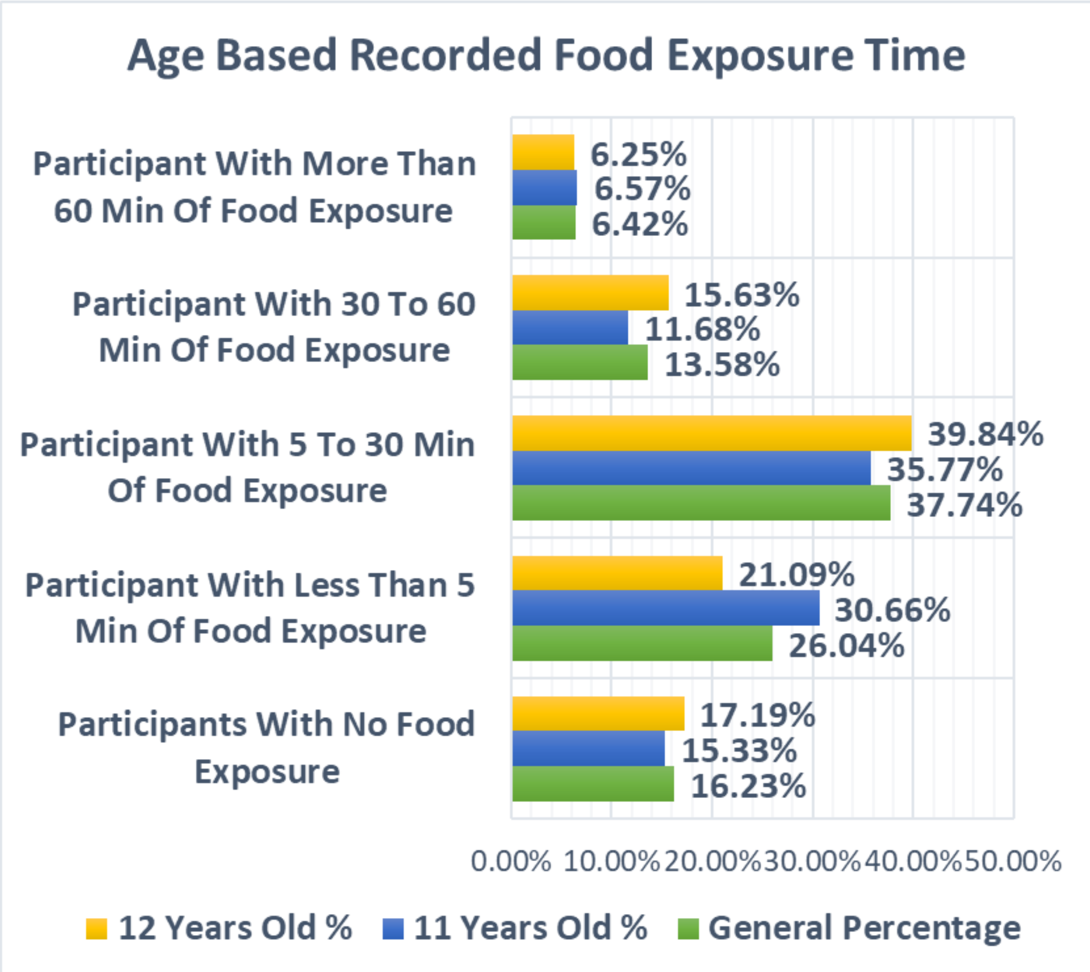
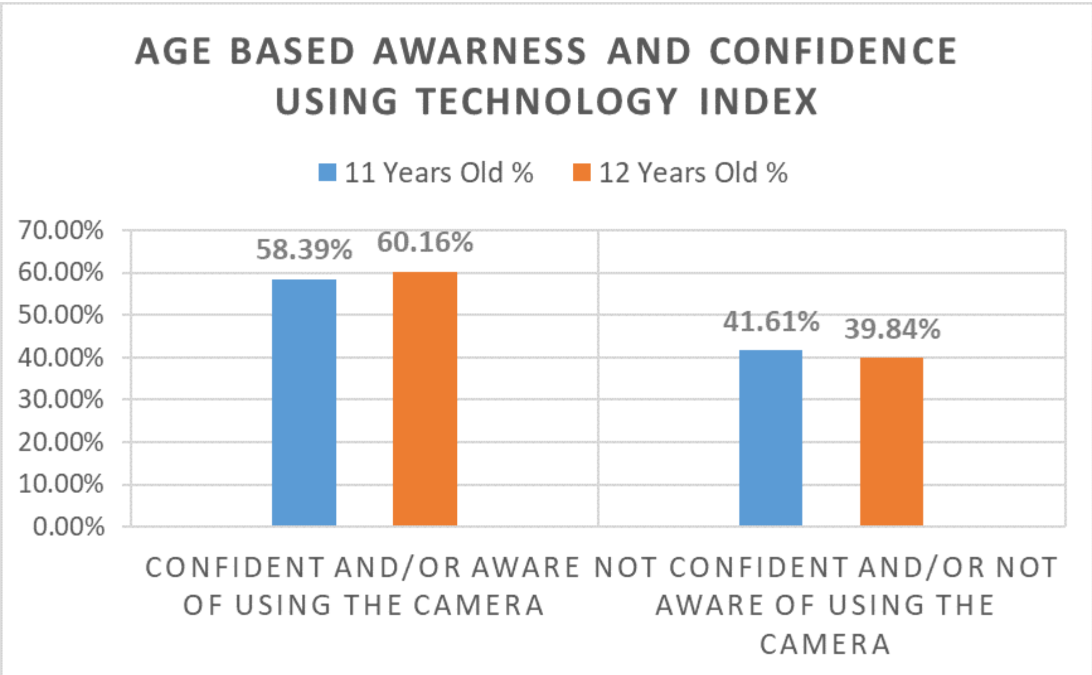


Figure I.4: Age Based Awareness and Confidence Index



Appendix J

Gender Based Statistics

Figure J.1: Gender Distribution

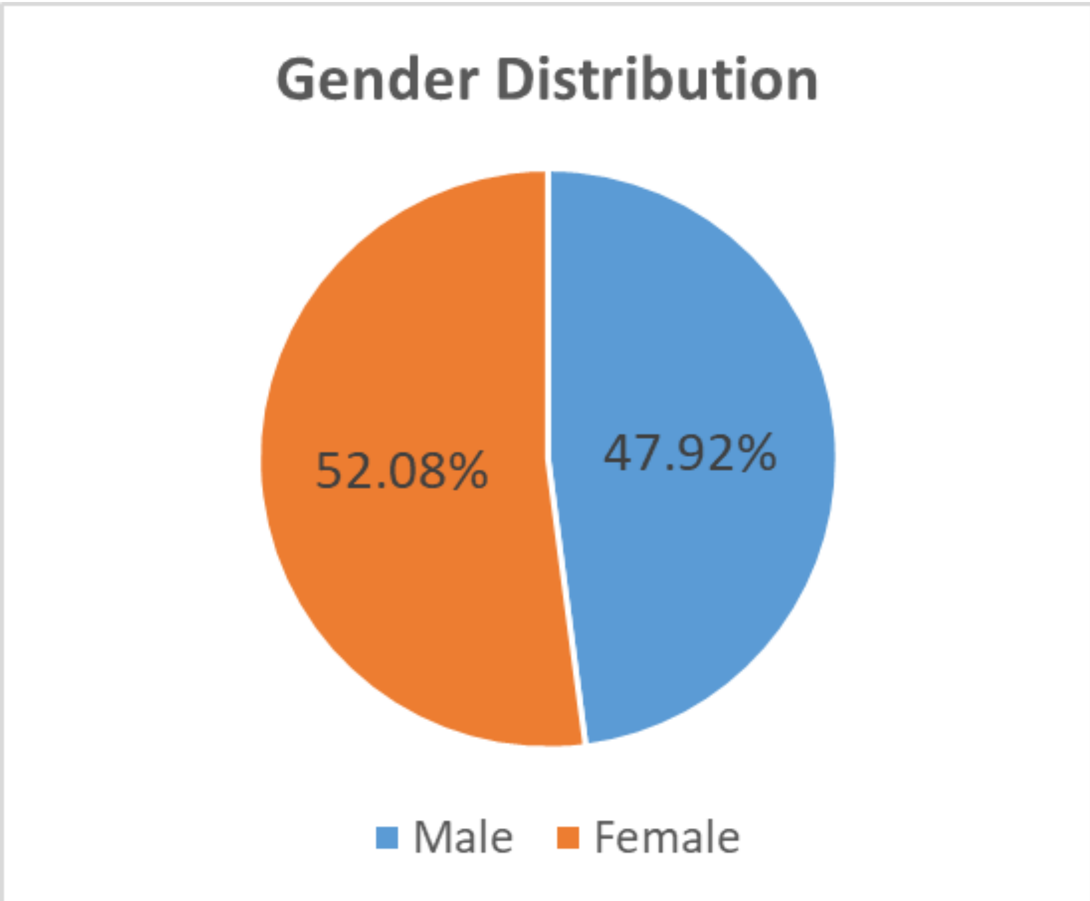


Figure J.2: Gender Based Recorded Food Exposure Time

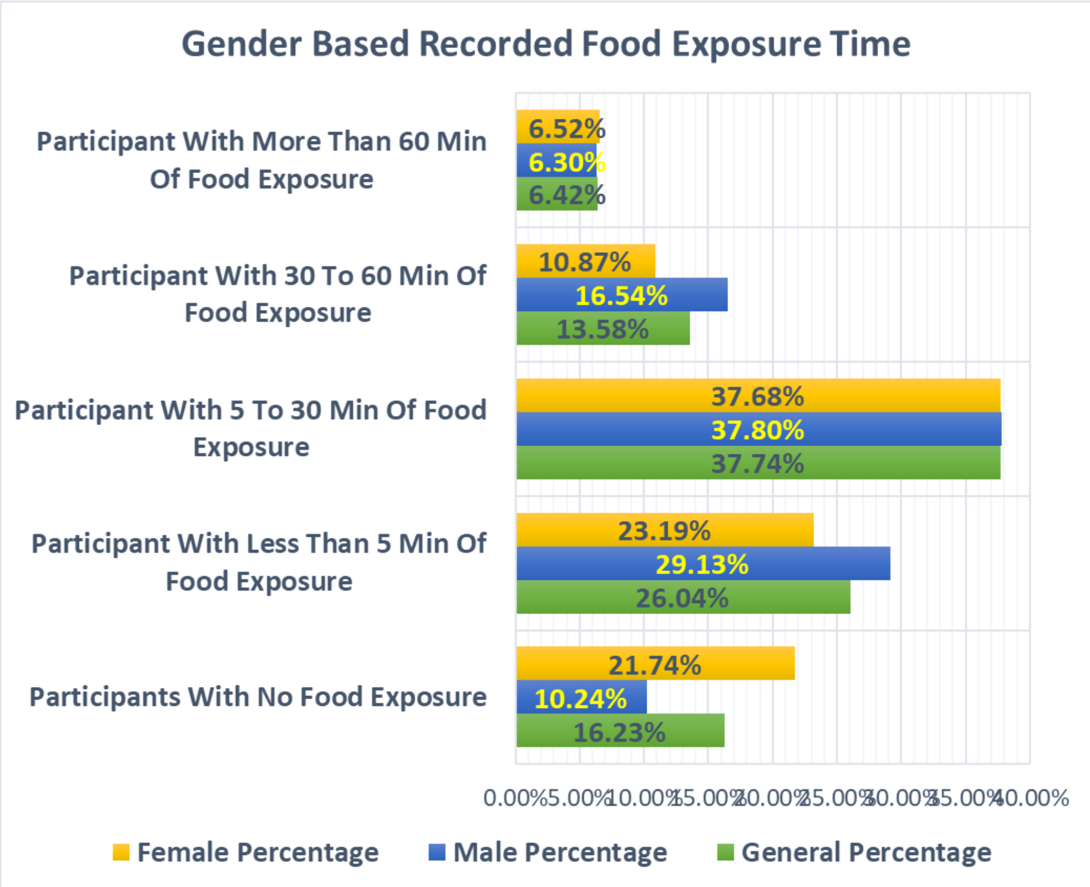
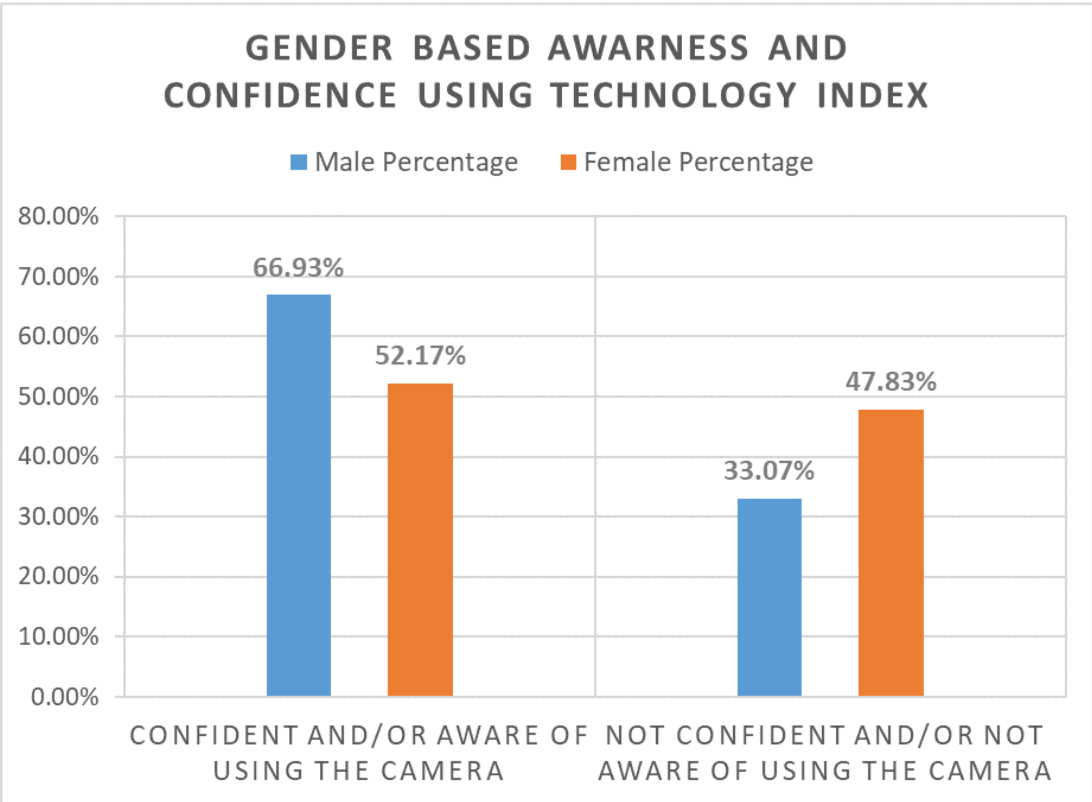


Figure J.3: Gender Based Awareness and Confidence Index



Appendix K

Income Based Statistics

Figure K.1: Income Distribution

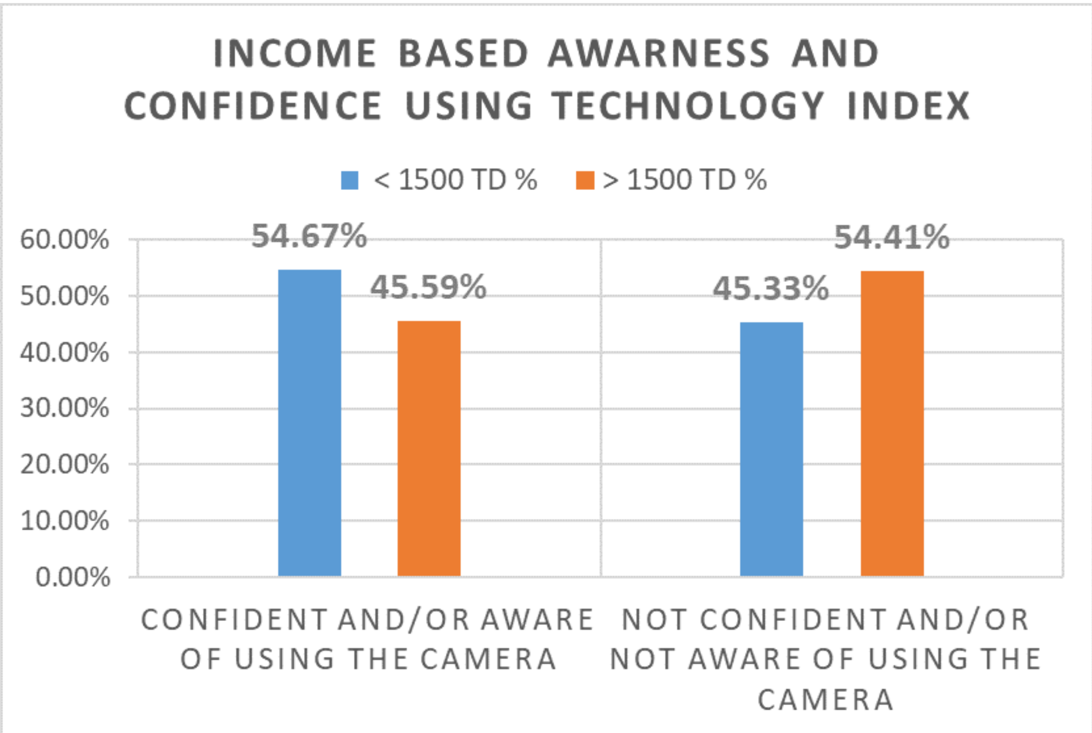


Figure K.2: Income Based Recorded Food Exposure Time

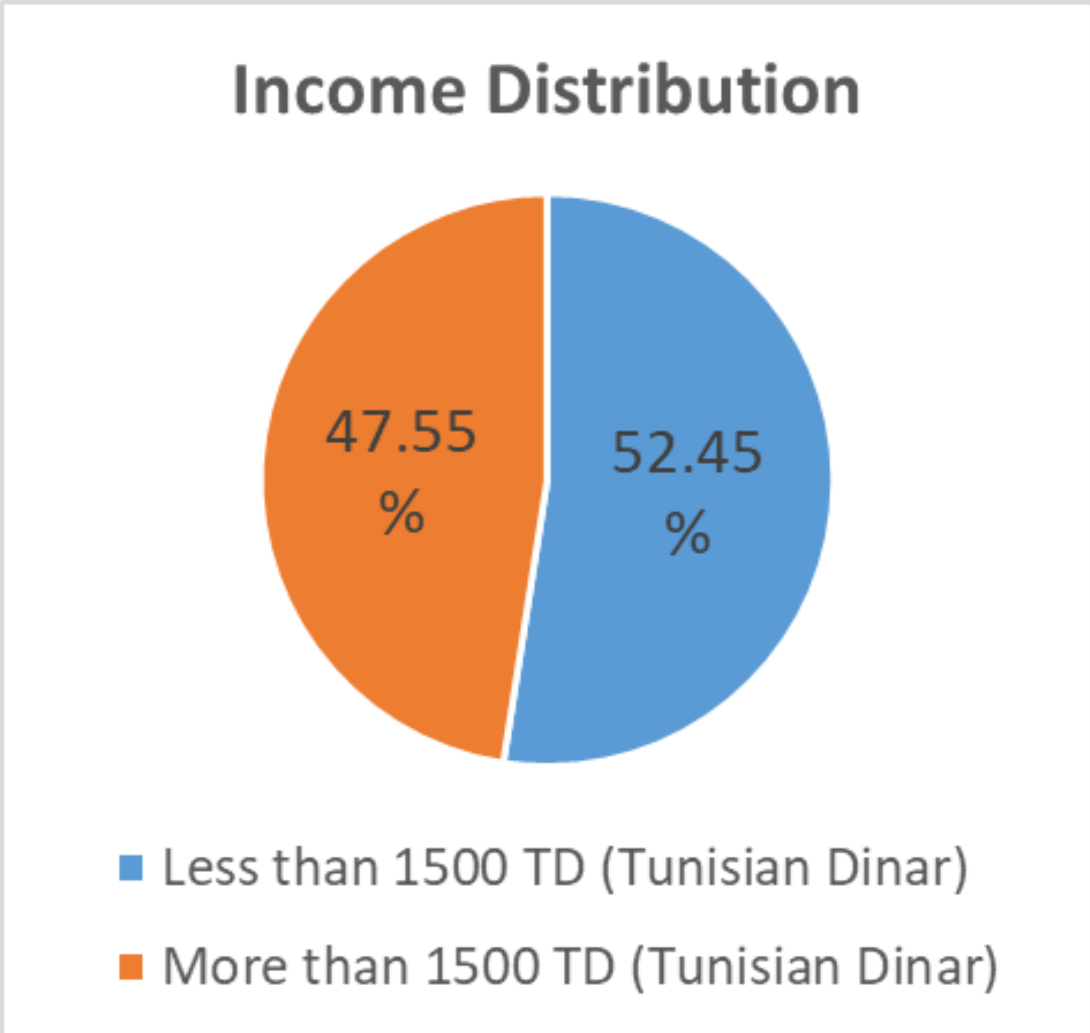
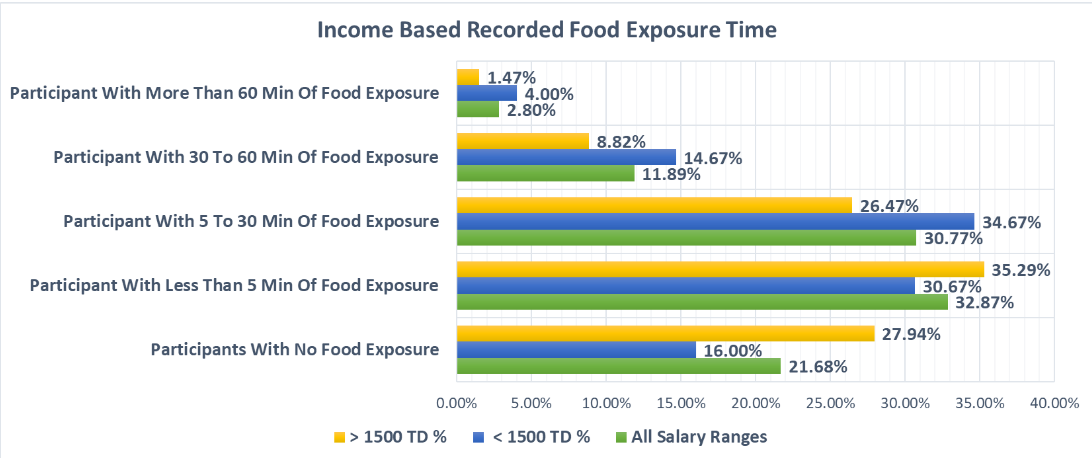


Figure K.3: Income Based Awareness and Confidence Index



Appendix L

Parents Educational Level Based Statistics

Figure L.1: Parents Educational Level Distribution

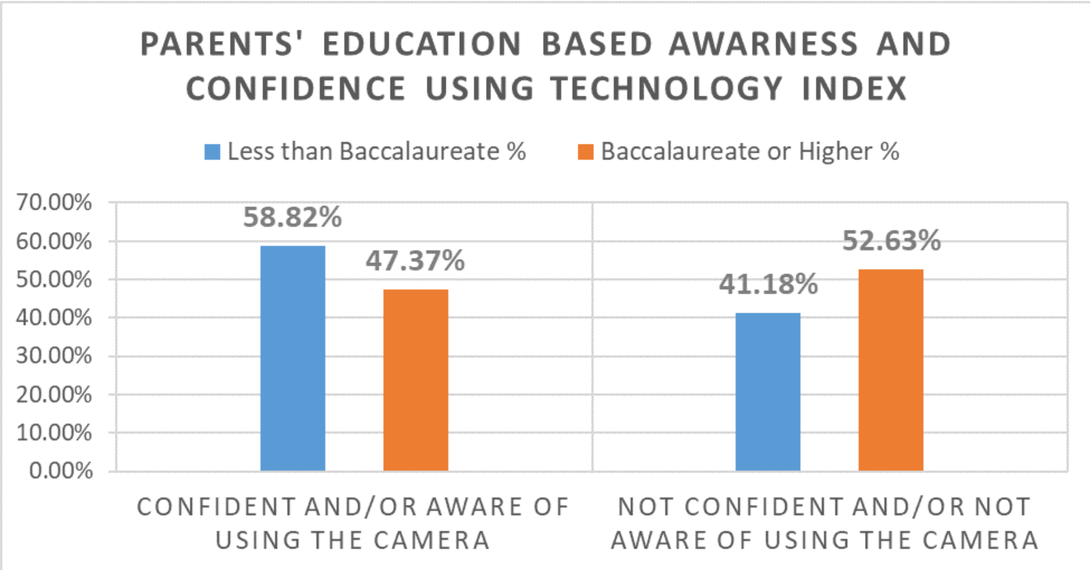


Figure L.2: Parents Educational Level Based Recorded Food Exposure Time

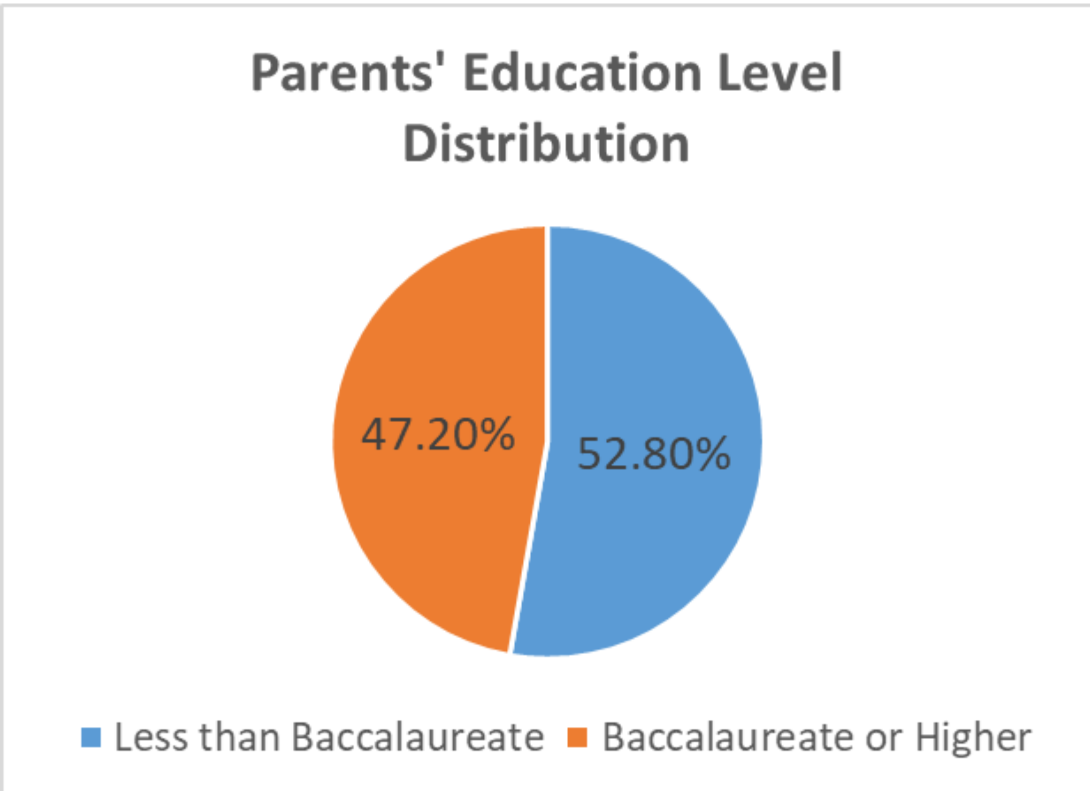
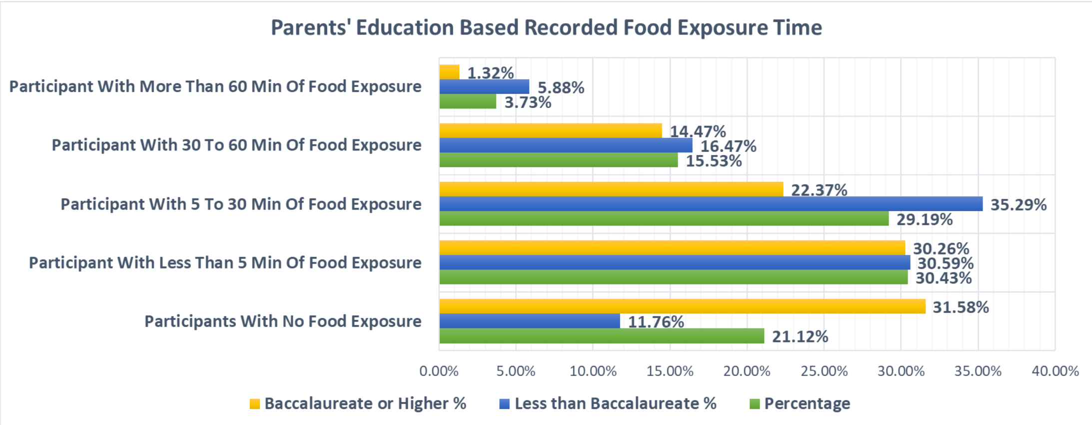


Figure L.3: Parents Educational Level Based Awareness and Confidence Index



Appendix M

Abbreviations

AI	Artificial Intelligence
ML	Machine Learning
DPM	Deformable Part Model
HoG	Histogram of Oriented Gradients
JSEG	A method for unsupervised segmentation of color-texture regions in images
SIFT	scale-invariant feature transform
CSIFT	color scale-invariant feature transform
PI	Primary Investigator
SP-BoF	Spatial Pyramid Bag Of Features
SVM	Support Vector Machine
GPS	Global Positioning System
HARF	Hierarical Attention based Food Recommendation system
CNN	Convolution Neural Network
CRPH	Center of Research on Population and Health
IRB	Institutional Review Board
Citi	Collaborative Institutional Training Initiative
USB	Universal Serial Bus
AUB	American University of Beirut
SD Card	Secure Digital Card
IP66	Ingress Protection 66
OIS	Optical Image Stabilization
LED	Light Emitting Diode
USD	United States Dollar
GB	Gega Byte
720p	Video resolution corresponding to 1280x720 pixels
API	Application Programming Interface
S3	the name of amazon cloud storage service
VGG16	name of a neural network architecture
VGG19	name of a neural network architecture
CPU	Central Proccesing Unit

GPU	Graphical Processing Unit
RAM	Random Access Memory
SSD	Solid State Drive
HDD	Hard Disk Drive
EBS	amazon Elastic Block Store
AWS	Amazon Web Services
NVME	Non-Volatile Memory Express
TB	Tera Byte
Mbps	Mega Bits Per Second
MTCNN	Multi-task Cascaded Convolutional Neural Networks
OpenCV	Open Computer Vision library
CVLIB	Computer Vision library
GUI	Graphical User Interface
ID	Identity Number
GHZ	Gega Hertz
DDR4	Double Data Rate 4
GDDR5	Graphics Double Data Rate 5
PCIe	Peripheral Component Interconnect express
I/O	Input/Output

Bibliography

[1] Chokr, Manal, and Shady Elbassuoni. "Calories Prediction from Food Images." In AAAI, pp. 4664-4669. 2017.

[2] T. Joutou and K. Yanai, "A food image recognition system with multiple kernel learning", in Proc. of IEEE International Conference on Image Processing, pp. 285–288, 2009.

[3] H. Hoashi, T. Joutou, and K. Yanai, "Image recognition of 85 food categories by feature fusion", in Proc. of International Symposium on Multimedia, pp. 296–301, 2010.

[4] S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar, "Food recognition using statistics of pairwise local features", in Proc. of IEEE Computer Vision and Pattern Recognition, 2010.

[5] Z. Zong, D.T. Nguyen, P. Ogunbona, and W. Li, "On the combination of local texture and global structure for food classification", in Proc. of International Symposium on Multimedia, pp. 204–211, 2010.

[6] Matsuda, Yuji, Hajime Hoashi, and Keiji Yanai. "Recognition of multiple-food images by detecting candidate regions." In Multimedia and Expo (ICME), 2012 IEEE International Conference on, pp. 25-30. IEEE, 2012.

[7] Turmchokksam, Sirichai, and Kosin Chamnongthai. "The Design and Implementation of an Ingredient-Based Food Calorie Measurement System Using Nutrition Knowledge and Fusion of Brightness and Heat Information." IEEE Access(2018).

[8] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627–1645, 2010.

[9] N. Dalal, B. Triggs, I. Rhone-Alps, and F. Montbonnot, "Histograms of oriented gradients for human detection", in Proc. of IEEE Computer Vision and Pattern Recognition, pp. 886–893, 2005.

[10] Pouladzadeh, P.; Villalobos, G.; Almaghrabi, R.; and Shirmohammadi, S. 2012. A novel svm based food recognition method for calorie measurement applications. In ICME Workshops, 495–498.

[11] Sudo, K.; Murasaki, K.; Shimamura, J.; and Taniguchi, Y. 2014. Estimating nutritional value from food images based on semantic segmentation. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, 571–576. ACM.

[12] Meyers, A.; Johnston, N.; Rathod, V.; Korattikara, A.; Gorban, A.; Silberman, N.; Guadarrama, S.; Papandreou, G.; Huang, J.; and Murphy, K. P. 2015. Im2calories: towards an automated mobile vision food diary. In Proceedings of the IEEE International Conference on Computer Vision, 1233–1241.

[13] Beijbom, O.; Joshi, N.; Morris, D.; Saponas, S.; and Khullar, S. 2015. Menu-match: restaurant-specific food logging from images. In 2015 IEEE Winter Conference on Applications of Computer Vision, 844–851. IEEE.

[14] Bettadapura, V.; Thomaz, E.; Parnami, A.; Abowd, G. D.; and Essa, I. 2015. Leveraging context to support automated food recognition in restaurants. In 2015 IEEE Winter Conference on Applications of Computer Vision, 580–587. IEEE.

[15] Nasreddine L, Naja F, Chamieh MC, Adra N, Sibai A-M, Hwalla N. Trends in overweight and obesity in Lebanon: evidence from two national cross-sectional surveys (1997 and 2009). BMC Public Health. 2012;12(1):798.

[16] Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet 2014;384(9945):766-781.

[17] Obermeyer CM, Bott S, Sassine AJ. Arab adolescents: health, gender, and social context. Journal of Adolescent Health. 2015;57(3):252-262.

[18] Mokhtar N, Elati J, Chabir R, et al. Diet culture and obesity in northern Africa. The Journal of nutrition. 2001;131(3):887S-892S.

[19] Aounallah-Skhiri H, Traissac P, El Ati J, et al. Nutrition transition among adolescents of a south-Mediterranean country: dietary patterns, association with socio-economic factors, overweight and blood pressure. A cross-sectional study in Tunisia. Nutrition journal. 2011;10(1):38.

[20] Hawkes C, Harris J, Gillespie S. Changing diets: Urbanization and the nutrition transition Global Food Policy Report. Washington, DC: International Food Policy Research Institute (IFPRI); 2017:34-41.

[21] Lytle LA, Seifert S, Greenstein J, McGovern P. How Do Children's Eating Patterns and Food Choices Change over Time? Results from a Cohort Study. American Journal of Health Promotion. 2000;14(4):222-228.

[22] Freedman DS, Khan LK, Dietz WH, Srinivasan SR, Berenson GS. Relationship of childhood obesity to coronary heart disease risk factors in adulthood: the Bogalusa Heart Study. Pediatrics. 2001;108(3):712-718.

[23] Swinburn BA, Sacks G, Hall KD, et al. The global obesity pandemic: shaped by global drivers and local environments. The Lancet. 2011;378(9793):804-814.

[24] Gortmaker SL, Swinburn BA, Levy D, et al. Changing the future of obesity: science, policy, and action. The Lancet. 2011;378(9793):838-847.

[25] Davison KK, Birch LL. Childhood overweight: a contextual model and recommendations for future research. Obesity reviews. 2001;2(3):159-171.

[26] Stok FM, Hoffmann S, Volkert D, et al. The DONE framework: Creation, evaluation, and updating of an interdisciplinary, dynamic framework 2.0 of determinants of nutrition and eating. PloS one. 2017;12(2):e0171077.

[27] Briefel RR, Crepinsek MK, Cabili C, Wilson A, Gleason PM. School food environments and practices affect dietary behaviors of US public school children. Journal of the American Dietetic Association. 2009;109(2):S91-S107.

[28] Fox MK, Dodd AH, Wilson A, Gleason PM. Association between school food environment and practices and body mass index of US public school children. Journal of the American Dietetic Association. 2009;109(2):S108-S117.

[29] Harrison F, Jones AP. A framework for understanding school based physical environmental influences on childhood obesity. Health & place. 2012;18(3):639-648.

[30] Neumark-Sztainer D, French SA, Hannan PJ, Story M, Fulkerson JA. School lunch and snacking patterns among high school students: associations with school food environment and policies. International Journal of Behavioral Nutrition and Physical Activity. 2005;2(1):14.

[31] Larson N, Story M. Are 'competitive foods' sold at school making our children fat? Health Affairs. 2010;29(3):430-435.

[32] Burgoine T, Jones AP, Brouwer RJN, Neelon SEB. Associations between BMI and home, school and route environmental exposures estimated using GPS and GIS: do we see evidence of selective daily mobility bias in children? International journal of health geographics. 2015;14(1):8.

[33] Engler-Stringer R, Le H, Gerrard A, Muhajarine N. The community and consumer food environment and children's diet: a systematic review. BMC Public Health. 2014;14(1):522.

[34] Pearce A, Kirk C, Cummins S, et al. Gaining children's perspectives: A multiple method approach to explore environmental influences on healthy eating and physical activity. Health & place. 2009;15(2):614-621.

[35] Barr M, Signal L, Jenkin G, Smith M. Capturing exposures: using automated cameras to document environmental determinants of obesity. Health promotion international. 2014;30(1):56-63.

[36] Sobal J, Bisogni CA. Constructing food choice decisions. Annals of Behavioral Medicine. $2009; 38(suppl_1) : s37 - s46$.

[37] Kearney J, Hulshof K, Gibney M. Eating patterns–temporal distribution, converging and diverging foods, meals eaten inside and outside of the home–implications for developing FBDG. Public Health Nutrition. 2001;4(2b):693-698.

[38] Boushey CJ, Kerr DA, Wright J, Lutes KD, Ebert DS, Delp EJ. Use of technology in children's dietary assessment. European journal of clinical nutrition. 2009;63 Suppl 1(Suppl 1):S50-S57.

[39] Heard AM, Harris JL, Liu S, Schwartz MB, Li X. Piloting an online grocery store simulation to assess children's food choices. Appetite. 2016;96:260-267.

[40] Hillier A, Smith T, Cannuscio CC, Karpyn A, Glanz K. A discrete choice approach to modeling food store access. Environment and Planning B: Planning and Design. 2015;42(2):263-278.

[41] Williams J, Scarborough P, Matthews A, et al. A systematic review of the influence of the retail food environment around schools on obesity-related outcomes. Obesity reviews. 2014;15(5):359-374.

[42] Doherty AR, Hodges SE, King AC, et al. Wearable cameras in health: the state of the art and future possibilities. American journal of preventive medicine. 2013;44(3):320-323.

[43] Gemming L, Doherty A, Utter J, Shields E, Mhurchu CN. The use of a wearable camera to capture and categorise the environmental and social context of self-identified eating episodes. Appetite. 2015;92:118-125.

[44] Briggs L, Lake AA. Exploring school and home food environments: perceptions of 8–10-year-olds and their parents in Newcastle upon Tyne, UK. Public health nutrition. 2011;14(12):2227-2235.

[45] Prosser J, Loxley A. Introducing visual methods. 2008.

[46] Schrempft S, van Jaarsveld CH, Fisher A. Exploring the Potential of a Wearable Camera to Examine the Early Obesogenic Home Environment: Comparison of SenseCam Images to the Home Environment Interview. Journal of medical Internet research. 2017;19(10):e332-e332.

[47] Signal LN, Smith MB, Barr M, et al. Kids' Cam: an objective methodology to study the world in which children live. American journal of preventive medicine. 2017;53(3):e89-e95.

[48] Nebeker C, Lagare T, Takemoto M, et al. Engaging research participants to inform the ethical conduct of mobile imaging, pervasive sensing, and location tracking research. Translational behavioral medicine. 2016;6(4):577-586.

[49] Mok TM, Cornish F, Tarr J. Too Much Information: Visual Research Ethics in the Age of Wearable Cameras. Integrative Psychological and Behavioral Science. 2015;49(2):309-322.

[50] Kelly P, Marshall SJ, Badland H, et al. An ethical framework for automated, wearable cameras in health behavior research. American journal of preventive medicine. 2013;44(3):314-319.

[51] Bradski, G., 2000. The OpenCV Library. Dr. Dobb's Journal of Software Tools.

[52] Xiang, Jia, and Gengming Zhu. "Joint Face Detection and Facial Expression Recognition with MTCNN." 2017 4th International Conference on Information Science and Control Engineering (ICISCE). IEEE, 2017.

[53] Trattner, C., & Elsweiler, D. (2017). Food recommender systems: important contributions, challenges and future research directions. arXiv preprint arXiv:1711.02760.

[54] Rokicki, M., Trattner, C., & Herder, E. (2018, June). The impact of recipe features, social cues and demographics on estimating the healthiness of online recipes. In Twelfth International AAAI Conference on Web and Social Media.

[55] Gao, X., Feng, F., He, X., Huang, H., Guan, X., Feng, C., ... & Chua, T. S. (2018). Visually-aware collaborative food recommendation. arXiv preprint arXiv:1810.05032.

[56] Turmchokkasam, S., & Chamnongthai, K. (2018). The Design and Implementation of an Ingredient-Based Food Calorie Estimation System Using Nutrition Knowledge and Fusion of Brightness and Heat Information. IEEE Access, 6, 46863-46876.

[57] Liang, H., Gao, Y., Sun, Y., & Sun, X. (2018). CEP: calories estimation from food photos. International Journal of Computers and Applications, 1-9.

[58] Ahn, D., Choi, J. Y., Kim, H. C., Cho, J. S., Moon, K. D., & Park, T. (2019). Estimating the Composition of Food Nutrients from Hyperspectral Signals Based on Deep Neural Networks. Sensors, 19(7), 1560.

[59] Aziz, A. S. B. A., & Habeeb, R. A. A. Classification of Food Nutrients Composition using Deep Learning.

[60] Elsweiler, D.; Trattner, C.; and Harvey, M. 2017. Exploiting food choice biases for healthier recipe recommendation. In Proc. of SIGIR '17, 575–584.

[61] Hellock. (2020, June 3). icrawler. Retrieved from https://github.com/hellock/icrawler