Chapter 1

Introduction

1.1 Motivation

In a time of rapid population growth and increasing globalization, the human race is facing widespread issues of malnutrition, which consists of both under and over nourishment, and is quantified by the degree to which a population satisfies minimum daily dietary requirements. Despite the fact that annual food production and agricultural capacity is sufficient to meet the dietary requirements of the entire global population[Kijne], approximately 10%, or between 600-800 million people[19], still suffer from some form of malnutrition.

Historic methods of analysis for consumption patterns are typically based on daily caloric and energy intake, which are not satisfactory for identifying malnutrition, as they lack precision in identifying insufficient levels of specific nutrients. Therefore, there is a need for a more detailed approach to break down dietary consumption into the composition of nutrient parts. By preforming analysis at the micronutrient level, suggestions for dietary additions can be made that are not intuitive on the macronutrient level and increase the scope of potential solutions. As a result of "the weaknesses in current methodology, attention is turning to strategies that automate the dietary assessment process to improve the accuracy and reduce the costs and burden to participants and researchers"[7].

The *Ideal Nutrition Construct* (INC) is the tool we designed in response to these needs, to identify the micronutrient deficiencies between the average dietary intake and the ideal recommended daily targets for an individual or sample population. This allows for dietary suggestions that complement the existing diet to be made by exploiting traditional optimization techniques in combination with the strength of metaheuristic evolutionary algorithms that can successfully incorporate notions of food access. Such an approach has profound implications for nutrition intervention on the local level, as it provides an innovative solution to sustainably meet dietary needs that promote overall health and disease prevention in a culturally relevant way.

The merit of INC lies in its ability to do what dietary software and nutrition intervention strategies do not. The tool deviates from current methods in the depth of its analysis and provides a service that is either not being automated or approached through these methods. Additionally, most popular personal applications are not designed for at-risk demographics or for intervention on the population level.

While we developed this model to exploit the food items that are currently accessible to at-risk populations, INC is ultimately part of a larger system that is motived by the reality that economic and physical access to an adequate food supply is not always guaranteed. Food security, as defined by the Food and Agriculture Organization of the United Nations (FAO), only exists when "all people, at all times, have physical, social, and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy lifestyle"[50]. The absence of food security is a direct result of the consequences of social and economic factors such as poverty, political corruption, agricultural policy, and economic instability, most pervasive in developing regions. Regardless of the efforts being made, by organizations like the United States Department of Agriculture (USDA), World Health Organization (WHO), and the FAO to combat food insecurity on a global level, it is imperative that geographical regions be self-sufficient in their ability to obtain proper nutrition for local populations, in a consistent and sustainable manner. As a result, we are in the process of gathering survey data from additional populations that fit this demographic, in order to expand our efforts.

1.2 Problem Statement

The scope of the INC project can be divided into two main phases. The first is the creation of a dietary evaluation tool to answer the following questions:

> *Given a sample diet composed of food items and quantities for an individual or population, can the diet be broken down into the composition of micronutrient parts and compared against the ideal recommended values for a user specified demographic, in order to accurately identify micronutrient*

deficiencies and overconsumption? Additionally, can the software perform complementary analysis functions to provide a comprehensive view of the overall diet in a way that provides valuable insight to nutritionists?

The second is the comparison of optimization algorithms to determine which is most suitable of the following task:

Given a decomposed sample diet and identified deficiencies, what set of items in the local food composition database and corresponding quantities should be recommended for a given individual or population, such that the micronutrient values of the overall diet are as close as possible to recommended ranges, and culturally relevant and economically available foods are suggested?

In general, the INC optimization problem can be defined in terms of a multi-objective knapsack problem and modified version of Stigler's classic diet problem that includes both cost and food access considerations.

1.3 Challenges

In the process of developing INC, we faced the practical limitations of working with intended populations. Due to the fact that INC is designed for the dietary optimization of at-risk populations, efforts were made to acquire survey data from individuals in Monte Cristo, El Salvador. However, the timeline required to obtain such data proved unrealistic for this initial phase of the INC project. As a result, a secondary representative athletic population was chosen.

This shift exposed the difficulties to effectively design nutrition surveys for specific populations, as the survey designed for El Salvador was uniquely different than that used in practice for the Springfield College American football team. Despite removing the language barrier from the initial survey design, every population presents its own obstacles in securing complete and adequate responses.

In terms of optimization, the two main challenges faced were the nonexistence of a comprehensive ideal micronutrient dataset and the a lack of available sample implementations of the NSGA-II algorithm for a problem remotely similar to a diet problem.

1.4 Contributions

While nutrition applications and dietary analysis techniques are not new concepts, through the development of INC, our contributions include:

1. **Constructing a new data resource.** Due to the desired depth, a dataset of comprehensive ideal micronutrient values for 22 categorical groups based on age, gender, and reproductive status was constructed to meet the needs of INC analysis and optimization. The details of this process are found in chapter 3.

2. **Designing a nutrition survey.** This survey captures four day weighted food records from the Springfield College American football team, which was used as a representative population in order to appropriately test the functionality of INC in a real-world context. Collected data was then processed to be compatible with INC input requirements. Such details are also outlined in chapter 3.

3. **Developing a dietary analysis platform.** One of the primary goals of INC is to automate the nutrition analysis process. The software is designed to allow the user to simply select the appropriate demographic and input all foods and quantities consumed over the four day period, while the program does the rest. Chapter 4 contains a description of the analysis functions and how the user interacts with the INC GUI.

4. **Constructing a local food composition database.** A database was constructed and populated through API REST access to the USDA online database. Not only does it provide a central location for INC to store and manipulate necessary data, but it exclusively contains the values of the desired set of micronutrients per 100 grams of a given food item. All optimization algorithms are designed to select from the foods in this database. Further details of how the database is populated and used by INC can be found in Chapters 4 and 5.

5. **Implementing and modifying classic optimization techniques.** Classic linear programming techniques for menu planning and dietary optimization were implemented first, and then modified to include access related considerations. Chapter 5 provides an explanation of this process and its limitations, while a performance evaluation of each algorithm on survey data is found in Chapter 6.

6. **Performing a literature review and comparison of alternative algorithms.** Due to the discovered limitations of the baseline algorithms, the exploration of

more complex algorithms to satisfy all range constraints and incorporate multiple objective functions was needed. The details and results of this comparison process can be found in Chapter 5.

7. **Implementing a pilot version of a multi-objective evolutionary algorithm for a modified diet problem.** A new implementation of the NSGA-II algorithm was used for the INC modified diet problem, as the select few existing menu planning implementations of multi-objective evolutionary algorithms are not being used for nutrition intervention purposes, The design of the NSGA-II implementation is found in Chapter 5 and performance results in Chapter 6.

Further explanation of all contributions listed above can be found in the section below.

1.5 Overview of the INC Project

In response to shortcomings in nutrition research[1], existing data for ideal micronutrient values was collected and cross-referenced across numerous sources, and all absent values were calculated based on available micronutrient and body metric information. The lack of available data, suggests that current technology and dietary analysis software are not suited for analysis at the micronutrient level. An abbreviated version of this dataset is used for all analysis and optimization we perform. Prior to these functions, INC executes API retrieval of food composition data, quantity scaling and unit conversions, local database storage to decrease retrieval time and encourage the suggestion of common food items within the population, and the generation of a sample diet that includes all consumed values for each micronutrient. The user can then choose from a selection of basic analysis functions that calculate the distance from upper and lower bounds of recommended ranges, overall protein consumption, consumption levels that are in danger of reaching known tolerable limits, and a visual representation of deficiencies.

Typically, menu planning and dietary optimization are most effectively calculated by a nutritionist who can take personal preferences and geographical access related constraints into account. INC is designed to automate this process, that cannot be realistically or precisely achieved, when a large set of food items is being considered.

[1] `http://www.globaldietarydatabase.org/the-global-dietary-database-measuring-diet-worldwide.html`

The initial implementation of the baseline algorithm is inspired by these problems, but did not align with INC goals, as most menu planning problems are based on food preferences and meal rotation over week long periods, and diet problems are just cost minimization functions. Although the influence of food preference in altering consumption patterns cannot be entirely disregarded, the introduction of any preference related ranking system would be considered a secondary task. As a result, the modified baseline algorithm is primarily concerned with meeting nutrition requirements in the most ideal way through physically and economically accessible foods.

The research prompted by the performance limitations of these algorithms showed that multi-objective evolutionary algorithms are more suited for the INC task. As a result, an in-depth comparison of the performance of five MOEAs for a 1/0 knapsack problem was conducted to determine which was most appropriate for INC, and the NSGA-II algorithm was ultimately selected. Despite the challenges we observed in the inability of NSGA-II to comply with upper bound constraints, the algorithm has shown promising implications for the INC model, as it is able to generate a set of pareto solutions, ideal for nutrition intervention by allowing for the selection of the diet most preferable for an individual or population. Additionally, results show that NSGA-II consistently outperforms both baseline algorithms in mean deficiency reduction and access rating.

To summary, although INC has the capability to perform analysis and is often explained on the individual level for the sake of simplicity, it is not a personal application. It is primarily designed to identify the deficiencies of a population, based on the collective analysis of individuals, and to optimize diets based on physical and economic access principles, which has not previously been done. The current prototype for INC is a minimum viable version of a larger system and was developed as a de-risk proposition in order to understand the opportunities this type of data analysis will present as it grows in scope. Additionally, while INC was developed with the intention of dealing with at-risk populations, our results show that it can be applied to a wide demographic of individuals.