**A Mobile Application for Diagnosing Celiac Disease using XGBoost Algorithm**

A Thesis   
Presented to the Faculty of the  
College of Computer and Information Sciences  
Polytechnic University of the Philippines

In Partial Fulfilment  
of the Requirements for the Degree

Bachelor of Science in Computer Science

Baleta, Gabriel Mori

Pajanustan, Joven

# CHAPTER 1 The Problem and Its Background

This chapter mainly discusses the background of the study, statement of the problem, the conceptual framework and the scope and limitation about the research. It also includes the introduction, hypothesis, significance of the study and the definition of terms used. It does provide information the conceptualization of this study.

## INTRODUCTION

A Mobile Application for Diagnosing Celiac Disease using XGBoost Algorithm Mobile Application for Diagnosis of Celiac Disease is an android application that evaluates a user’s risk on having Celiac Disease based on their symptoms and set of risk factors. The application outputs the percentage of risk of the user and interprets whether the user is at classical celiac disease, non-classic celiac disease, or silent celiac disease.

* 1. **BACKGROUND OF THE STUDY**

This Celiac disease can be difficult to diagnose because it affects people differently. There are more than 200 known celiac disease symptoms which may occur in the digestive system or other parts of the body. Some people develop celiac disease as a child, others as an adult.

When people with celiac disease eat gluten (a protein found in wheat, rye and barley), their body mounts an immune response that attacks the small intestine. These attacks lead to damage on the villi, small fingerlike projections that line the small intestine, that promote nutrient absorption. When the villi get damaged, nutrients cannot be absorbed properly into the body.

Celiac disease is hereditary, meaning that it runs in families. People with a first-degree relative with celiac disease (parent, child, sibling) have a 1 in 10 risks of developing celiac disease.

Celiac disease is very dangerous so that if there is a way to determine if person have it better do it. In this modern era smartphones are available in the market with its affordable price, the researchers realized that this will be a helpful application for the community, because through mobile application a person can determine if he/she have a high chance of having a celiac disease, because early diagnosis of a patient’s illness will have a high chance of survivability.

**1.3 STATEMENT OF THE PROBLEM**

The study aims to develop and implement a system that will diagnose if a person has a celiac disease. In addition to this, the researchers intend to answer the following questions:

1. What is the accuracy of the system in diagnosing if a person has a celiac disease?
2. Is there a significant difference in the accuracy in diagnosing if a person has a celiac disease using XGBoost Algorithm from the current methods of diagnosis?
3. What are the methods in gathering information for diagnosing Celiac Disease?

1.4 OBJECTIVES OF THE STUDY

* To measure the level of accuracy of the system in diagnosing if a person has a celiac disease.
* To find if there is a significance of accuracy in diagnosing if a person has a celiac disease using XGBoost Algorithm from the current methods of diagnosis.
* To define methods used in gathering information for diagnosing Celiac Disease.

**1.5 CONCEPTUAL FRAMEWORK**

**1.5.1 CONCEPTUAL FRAMEWORK OF THE SYSTEM**

The figure below illustrates the conceptual framework of the system. The needed tools for the system are shown below. First, in the input phase, the user needs to answer the questions for symptoms. In the process phase includes the validation phase wherein only ages 16 and above can use this system then XGBoost phase, where the data given will converted into numerical value first to process that will give the output of the system which will be the possible initial Celiac Disease diagnosis or the type of it.

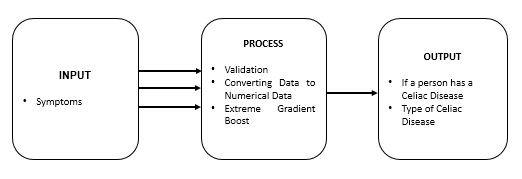
****

Figure 1.1 Conceptual Framework of the system

**1.5.2 CONCEPTUAL FRAMEWORK OF THE STUDY**

Figure 1.2 shows the conceptual framework of the study wherein the required input is the developed system Celiac Disease Diagnosis Mobile Application. Data gathering is through interview method and system testing. Data analysis and presentation is included in this phase. The output of the study will be the accuracy of the system in finding the possible initial disease, conclusions and recommendations of the system.

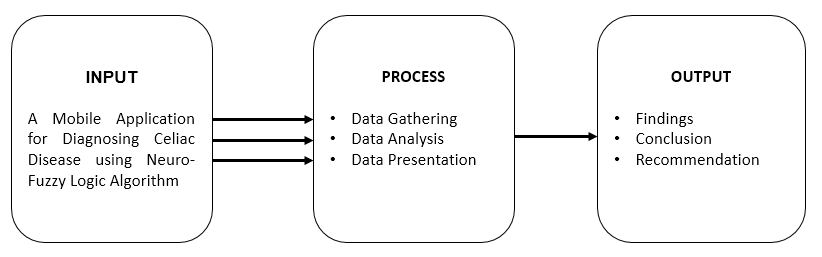
****

Figure 1.2 Conceptual Framework of the system

* 1. **SCOPE AND LIMITATIONS**

This system has the capacity to diagnose if the user has a Celiac Disease and the type of celiac disease only. The age input has a range from 16 years old and above only. The system has no capability of recommending an initial treatment. Although the research has reached its aims, there are still some unavoidable limitations. First, because of the limited time given, the research was conducted only on a small size of population.

**1.6 SIGNIFICANCE OF THE STUDY**

**Celiac Disease Patients**

This can be beneficial to the patients with Celiac disease for the early diagnosis. This can also help them decide to have an immediate consultation with the physician.

**Celiac Disease Experts**

The developed application will be a useful tool to give assistance for the experts in Celiac Disease. The tool can be used as an initial test / initial screening for their first-time patients to assess them if they are already at risk of having Celiac Disease.

**Future Researchers**

This will help the future researchers with the same topic of interest. This will also serve as guide or reference for them especially on the topics related to diagnosing of Celiac Disease.

**1.7 DEFINITION OF TERMS**

*Celiac Disease* - is a serious autoimmune disorder that can occur in genetically predisposed people where the ingestion of gluten leads to damage in the small intestine.

*Gradient Boosting* - is a machine learning technique for [regression](https://en.m.wikipedia.org/wiki/Regression_(machine_learning)) and [classification](https://en.m.wikipedia.org/wiki/Classification_(machine_learning)) problems, which produces a prediction model in the form of an [ensemble](https://en.m.wikipedia.org/wiki/Ensemble_learning) of weak prediction models, typically [decision trees](https://en.m.wikipedia.org/wiki/Decision_tree_learning).

*Extreme Gradient Boosting (XGBoost)* - is similar to [gradient boosting](https://www.analyticsvidhya.com/blog/2015/09/complete-guide-boosting-methods/) framework but more efficient. It has both linear model solver and tree learning algorithms. So, what makes it fast is its capacity to do parallel computation on a single machine. This makes XGBoost at least **10 times faster** than existing gradient boosting implementations. It supports various objective functions, including regression, classification and ranking.