AN ONTOLOGY DRIVEN HORMONE IMBALANCE DIAGNOSTIC SYSTEM

BY

MOHAMMED LAWAL

DEPARTMENT OF COMPUTER SCIENCE

AHMADU BELLO UNIVERSITY,
ZARIA, NIGERIA

JUNE, 2018

AN ONTOLOGY DRIVEN HORMONE IMBALANCE DIAGNOSTIC SYSTEM

BY

MOHAMMED LAWAL

Bsc. Computer Science (U.M.Y.U.K., 2012)

P14SCMT8080

A THESIS SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES, AHMADU BELLO UNIVERSITY, ZARIA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER DEGREE IN COMPUTER SCIENCE

DEPARTMENT OF COMPUTER SCIENCE,

FACULTY OF PHYSICAL SCIENCES,

AHMADU BELLO UNIVERSITY,

ZARIA, NIGERIA

JUNE, 2018

DECLARATION

I declare that the work in this Dissertation titled "An Ontology Driven Hormone Imbalance Diagnostic System" has been carried out by me in the Department of Computer Science. The information derived from the literature has been duly acknowledged in the text and a list of references provided. No part of this dissertation was previously presented for another degree or diploma at this or any other Institution.

Mohammed Lawal_____ _____

(Name of Student) Signature Date

CERTIFICATION

This dissertation titled "An Ontology Driven Hormone Imbalance Diagnostic System" by Mohammed LAWAL meets the regulations governing the award of the degree of Masters of Science in Computer Science of the Ahmadu Bello University, and is approved for its contribution to knowledge and literary presentation.

Dr. S. Aliyu _____ _____

Chairman, Supervisory Committee Signature Date

Dr. B.I. Ahmad _____ _____

Member, Supervisory Committee Signature Date

Prof. S. B. Junaidu_____ _____

Head of Department Signature Date

_____ _____ _____

External Examiner Signature Date

Prof. S. Z. Abubakar _____ _____

Dean, School of Postgraduate Studies Signature Date

DEDICATION

This research work is dedicated to my entire family

ACKNOWLEDGEMENT

Many thanks to The Almighty God for bestowing upon me the capacity to carry out this research work.

I sincerely acknowledge the intellectual guidance, encouragement and supervision of my supervisors; Dr. S. Aliyu and Dr. B. I. Ahmed who tirelessly through thick and thin, painstakingly ensured that this work is balanced and completed in good time. Sirs, I am most grateful, may the Almighty God guide and preserve all that relates to you and your families in your race through life.

My gratitude also goes to the Head of Department, Prof. S. B. Junaidu and other lecturers of the Department. May the Lord keep you and grant all that your heart desire in this profession.

My heartfelt appreciation goes to my family – for instilling in me the discipline that brought me this far.

ABSTRACT

Hormone imbalances are known tohaveclosely related symptoms. Thyroid hormone imbalance, for example, shares 75% of its symptoms with Testosterone and cortisol hormone imbalancesmaking it difficult to accurately diagnose the hormone imbalance. A review of anontology based system that diagnoses hormone imbalanceonly diagnoses thyroid hormone imbalance without taking into account of the similarities of symptoms with testosterone and cortisol imbalance. This can lead to an error in diagnosis. This workpresents an ontology based diagnosissystem that diagnoses thyroid hormone imbalance taking into account similar hormone imbalances (cortisol and testosterone) thereby improving the number of correct diagnosis the system makes. The systemmakes use of a knowledge base built using Web Ontology Language (OWL), a carefully designed interface that allows users to interact with the knowledge base implemented with JavaFX. Jena reasoning engine and SPARQL were used to provide both reasoning and querying services respectively.The findings of this study reveal an improved ontology model as a result of adding related hormone imbalances.A simulation program was developed to measure the number of correct diagnosis the developed system makes, and the results revealed the system made7 times more correct diagnosis than the existing system.

Table of Contents

DECLARATION iii
CERTIFICATION iv
DEDICATION v
ACKNOWLEDGEMENT vi
ABSTRACT vii
LIST OF FIGURES x
LIST OF TABLES xi
CHAPTER ONE: INTRODUCTION 1
1.1 Background to the Study 1
1.2 Statement of the Problem 2
1.3 Motivation 3
1.4 Aim and Objectives 3
1.5 Research Methodology 3
1.6 Contribution of Study 4
1.7 Organization of Thesis 4
CHAPTER TWO: LITERATURE REVIEW 5
2.1 Hormones and Possible Ailments due to their Imbalances 5
2.1.1 Cortisol 5
2.1.2 Thyroid 5
2.1.3 Testosterone 6
2.2 Overview of E-Health Systems 7
2.3 Diagnostic Systems 7
2.3.1 Questionnaire Based Approach 8
2.3.2 Machine Learning Based Approach 9
2.3.3 Ontology Based Approach 10
2.4 Technologies Used in Developing Ontology Based Applications 15
2.4.1 Jena API 15

2.4.2 JavaFX 16
2.4.3 Protégé 17
2.5 Evaluation Metrics 18
2.6. Gaps in Literature 19
CHAPTER THREE: RESEARCH METHODOLOGY 20
3.1 Introduction 20
3.2 System Architecture 20
3.3 System Requirements 21
3.4 Designing the Ontology 21
3.5 Designing the Inference Rules 24
3.6 Designing the User Interface 25
3.7 Queries for Statistical Analysis 26
3.8 Evaluation Metrics 27
3.9 Inference Capability Evaluation 27
CHAPTER FOUR: RESULTS AND DISCUSSION 29
4.1 Introduction 29
4.2 System Overview 29
4.3 Evaluation 31
4.4 Result Analysis 33
CHAPTER FIVE: SUMMARY, CONCLUSION AND RECOMMENDATIONS 36
5.1 Summary 36
5.2 Conclusion 36
5.3 Recommendations for Future Work 37
REFERENCES 38
APPENDIX 41

LIST OF FIGURES

Fig 2.1: Androgen Deficiency in Aging Men Questionnaire (Morley *et al.*, 2000) 8

Fig. 2.2: Thyroid disease Ontology Model (Rawte and Roy, 2015) 11

Fig. 2.3: Inference Rules (Rawte and Roy, 2015) 12

Fig. 2.4: System Architecture (Rawte and Roy, 2015) 12

Fig. 2.5: Jena API Architecture (Jena Architecture Overview, 2017) 16

Fig. 2.6: JavaFX Architecture (Castillo, 2013) 17

Fig. 2.7: Protégé Interface (Musen, 2015) 18

Fig 3.1: System Architecture of the Proposed System 20

Fig. 3.2: Relationships of Classes in Ontology 22

Fig. 3.3: Hormone Imbalance and patient ontology model 24

Fig 3.4: Inference rules used by system for diagnosis 25

Fig. 3.5: SPARQL queries 26

Fig. 3.6: Simulation pseudocode 28

Fig. 4.1: Bio-Data tab 29

Fig. 4.2: Diagnosis tab 30

Fig. 4.3: Statistics tab 31

Fig. 4.4 Evaluation of system performance with 2 inference rules against Rawte and Roy (2015) 32

Fig. 4.5 Evaluation of system performance with 3 inference rules against Rawte and Roy (2015) 33

LIST OF TABLES

Table 4.1: Evaluation of knowledge base of the proposed system and Rawte and Roy (2015) ...32

CHAPTER ONE: INTRODUCTION

1.1 Background to the Study

The human body is a complex system and hormones play a vital role in ensuring that it operates optimally. Hormones are molecules released by glands into the blood to regulate human physiology(internal functions)and behaviour. They regulate bodily functions like blood pressure, hunger cravings, mood swings, immune responses, sexual behaviour, and so on (Marieb, 1992). The glands may produce excess or insufficient hormones resulting in an imbalance. This imbalance results to different symptoms depending on the particular hormone.A review of symptoms of hormone imbalances by Smith (2005) revealed that some hormone imbalances have closely related symptoms. Current approaches in diagnosing these imbalances include screening using questionnaires, which is slow as observed by Morley *et al.* (2000). Ontology based systems which only target a single disease as reported in Rawte and Roy (2015), and machine learning based systems that need to be trained over time as noted in Berner and La Lande (2016).

Bucci*et al*. (2011) found that applying ontologies in medicine improves interoperability between systems, and knowledge reuse. They also found that ontologies can define a knowledge base that evolves over time by allowing qualified users to contribute to the knowledge base, using an ontology editor, without the intervention of a system developer.

Chandrasekaran*et al*. (1999) defined ontology as a body of knowledge describing some domain, typically acommon sense knowledge domain, using arepresentation vocabulary.

Ontological analysis clarifies the structureof knowledge. Given a domain, its ontologyforms the heart of any system of knowledgerepresentation for that domain.

Rawte and Roy (2016) used an ontology in the diagnosis of thyroid disease (hyperthyroidism and hypothyroidism). Chang *et al*. (2015) developed an ontology that was used in the diagnosis of depression. Their system was designed to work on mobile devices. Dieng-Kuntz *et al*. (2006) explored using ontologies for medical diagnosis and sharing of information across medical personnel.

Therefore, ontology based systems provide the quickest way to develop diagnostic systems. This research aims at developing an ontology driven diagnostic system that utilizesan ontology in the diagnosis of threesimilar hormone imbalances (thyroid, cortisol, and testosterone hormones).

1.2 Statement of the Problem

E-Healthsystems have opened the way for ontology based systems that aid in diagnosis. Rawte and Roy (2015) developed an ontology based system that diagnoses thyroid disease (both hyperthyroidism and hypothyroidism). However, a review of hormone imbalances by Smith (2005) revealed that 75% of the symptoms of thyroid hormone imbalance are the same as the symptoms of cortisol imbalance and testosterone deficiency making it difficult to accurately diagnose the hormone imbalances.

This research aims at improvingon the work of Rawte and Roy (2015) by diagnosing thyroid hormone imbalance together with testosterone deficiency and cortisol hormone imbalance thereby improving the number of correct diagnosis the system makes.

1.3 Motivation

This research is motivated by the difficulties encountered while diagnosing ailments that have closely related symptoms.The approach of using an ontology as a knowledge base for diagnosis taken in this research was motivated by the easewith which new knowledge can be added to the ontology without the need for a programmer to change the main system.

1.4 Aim and Objectives

The aim of thiswork is to develop an ontology based systemthatdiagnoses hormone imbalances with similar symptoms. The objectives are to:

i. Buildan Ontology for capturinghormone imbalance and patient information
ii. Construct rules that will guide the diagnosis of the ailments
iii. Design an interface for data collection and report presentation
iv. Implement the system and evaluate it against that of Rawte and Roy (2015)

1.5 Research Methodology

i. The ontology used for the system was designed using the Web Ontology Language (OWL) and Protégé v5.1 as the ontology editor
ii. The inference rules were defined using the Apache Jena reasoner that comes bundled with the Jena API
iii. JavaFX was used for the user interface design
iv. SPARQL query language was used in defining the queries for performing the statistical analysis

v. The evaluation was done based on the metrics provided by Manouselis et al. (2010). A simulation program was developed to measure the number of correct diagnosis the developed system makes against that of Rawte and Roy (2015).

1.6 Contribution of Study

The following contributions were made to knowledge:

i. An enhanced ontology model that captured information about three hormone imbalances (testosterone, cortisol, and thyroid hormone) and patient information
ii. Four inference rules used for performing the diagnoses of thyroid disease (hyperthyroidism and hypothyroidism), hypogonadism (testosterone deficiency), and cortisol imbalance
iii. SPARQL queries used in retrieving information about previous users of the system in the form of age and gender distribution.

1.7 Organization of Thesis

This dissertation is organized as follows:chapter one gives the general introduction to the work. Chapter two gives a review of related literature. Chapter three describes the research methodology. Chapter four presents findings and discussions. Chapter five provides summary, conclusion, and recommendations on future works.

CHAPTER TWO: LITERATURE REVIEW

2.1 Hormones and Possible Ailments due to their Imbalances

Hormones may be defined as chemical substances, secreted by cells into the extracellular fluids, that regulate the metabolic functions of other cells in the body (Marieb, 1992). Hormones have the following effects on the human body: stimulation or inhibition of growth, wake-sleep cycle and other circadian rhythms, andmood swings. Hormones also help in theinduction or suppression of apoptosis (programmed cell death), activation or inhibition of the immune system, regulation of metabolism, preparation of the body for mating, andpreparation of the body for fighting or fleeing.The hormones considered in this work are the cortisol, thyroid, and testosterone hormones.

2.1.1 Cortisol

Cortisol is the hormone that is released under stress or when the blood sugar level goes below normal (Raven and Johnson, 2001). Smith(2005) studied the symptoms of cortisol imbalance and found the following: difficulty losing weight, weak bones, weak muscles, and thin skin (easily bruised and dry skin) among others.

2.1.2 Thyroid

The thyroid's main role in the endocrine system is to regulate your metabolism, which is your body's ability to break down food and convert it to energy (Sargis, 2017).Hypothyroidism is the condition when the thyroid gland is not able to produce enough thyroid hormone.

Smith (2005) reviewed the symptoms of hypothyroidism and reported the following: fatigue, constipation, depression, and weight gain. Hyperthyroidism is the condition when the thyroid gland produces too much thyroid hormone.Rawte and Roy (2015) reviewed hyperthyroidism and reported the following: sleep problems, heat intolerance, anxiety, hand tremor, and sexual problems among others.

2.1.3 Testosterone

Testosterone is an important hormone necessary for proper development. Exposure to the hormone begins even before birth. Higher prenatal androgens in males than in female have been shown to masculinize the genitalia, develop the neural systems and behaviour as presented in Alexander and Wilcox (2012). Exposure to testosterone has been shown to affect behaviour, as reported in Hines*et al.* (2015). They showed that females exposed to high levels of testosterone showed more behaviour typically identified as male-type behaviour while exhibiting less of female-type behaviour.

The symptoms of low testosterone are quite numerous, according to Dandana and Rosenberg (2010). The main symptoms include: reduced libido, erectile dysfunction resistant to medication, reduced muscle mass and strength, depressed mood, lack of energy and reduced vitality, osteoporosis. These symptoms are the major signs of low testosterone, but other indicators exist. These suggestive signs include: loss of axillary hair, increase body fat and mass index, diminished work performance, sleep problems, poor concentration and memory problems, and infertility which can be azoospermia or oligospermia. Azoospermia means no sperm in semen while oligospermia means few sperm cells in semen.

2.2 Overview of E-Health Systems

Oh *et al*.(2005) defined e-health as the use of information and communication technologies to provide and or improve healthcare. E-Health is a new field of medical informatics that is finding applications in many aspects of medicine. It has been used in managing patient records (Garets and Davis, 2006)

Catwell and Sheikh (2009) reviewed the benefits and drawbacks of e-health systems and found that E-health systems are convenient and have the potential of cutting costs when it comes to accessing healthcare remotely, but the initial cost of setting up may be prohibitively high. They also found that e-health systems have the potential of reducing the number of people affected by medical errors, and that e-health systems have the potential of providing equality of service. They, however, warn of the possibility of a badly engineered system providing incorrect diagnosis.

2.3Diagnostic Systems

Ledley and Lusted (1959) described the steps followed by physicians in diagnosing a disease. The process includes a step called differential diagnosis. Normally, physicians only consider the disease that matches all the symptoms specified by the patient. The problem arises when more than one disease match the symptoms specified by the patient. When this happens, the physician needs to consider all diseases that match the specified symptoms. This process was necessitated because of the difficulty in accurately diagnosing a single disease that has symptoms in common with other diseases.

There are systems that perform this diagnosis. Berner and La Lande (2016) reviewed the types of diagnosis systems and found most systems are either ontology based or machine learning based. Questionnaire based systems were also reviewed byMorley et al. (2000) and found them to be a useful tool in diagnosis.

2.3.1 Questionnaire Based Approach

The Androgen Deficiency in Aging Males (ADAM) is a questionnaire published by Morley*et al.*(2000) that screens for androgen deficiency without laboratory testing. The questionnaire consists of ten questions as shown in Fig. 2.1.

	Yes	No
1. Do you have a decrease in libido (sex drive)?		
2. Do you have a lack of energy?		
3. Do you have a decrease in strength and/or endurance?		
4. Have you lost height?		
5. Have you noticed a decreased 'enjoyment of life'?		
6. Are you sad and/or grumpy?		
7. Are your erections less strong?		
8. Have you noted a recent deterioration in your ability to play sports?		
9. Are you falling asleep after dinner?		
10. Has there been a recent deterioration in your work performance?		

Fig 2.1: Androgen Deficiency in Aging Men Questionnaire (Morley *et al.*, 2000)

The result of the questionnaire is positive when question one or question seven is a 'Yes'. The result is also positive when any other three questions are 'Yes'.Questionnaires have also found application in measuring the effects of a disease on patients. Slok*et al.*(2016) found that the assessment of burden questionnaire for chronic obstructive pulmonary disease reliably measured the effects of the disease on the patient. The questionnaire is made up of 14

questions and each question accepts a value in the range of 1-7 indicating the severity of that particular symptom.

Jorm (2004) reviewed the informant questionnaire on cognitive decline in the elderly. He found the questionnaire to be highly reliable in screening for dementia and is not affected by education level. Kroenke *et al.* (2010) reviewed questionnaires for measuring and monitoring levels of depression and anxiety. They found the questionnaires to be brief well-validated measures for detecting and monitoring depression and anxiety.

The problem with this approach is the latency (time before system produces output when provided with input data) because of the pen and paper nature of the questionnaires. This research intends to solve this problem by automating data collection, processing, and result presentation.

2.3.2 Machine Learning Based Approach

Artificial intelligence has been used in the development of medical systems that diagnose diseases (Berner and La Lande, 2016). According to Hamet and Tremblay (2017), Artificial Intelligence (AI) is a general term that implies the use of a computer to model intelligent behavior with minimal human intervention. The term is applicable to a broad range of items in medicine such as robotics, medical diagnosis, medical statistics, and human biology. Artificial intelligence systems work by analysing patterns in the patient's data to derive association between symptoms and diseases.The system learns patterns of symptoms and diseases by performing statistical analysis on large amounts of data. Data containing both symptoms and diseases are prepared, the symptoms are fed to the system and a guess from

the system is observed. Parameters are tweaked afterwards, in the case of incorrect diagnosis, to achieve the correct diagnosis. The repeated application of this process of providing data and tweaking parameters is called training the artificial neural network. The system can be trained to diagnose multiple conditions.Atkov*et al.* (2012) developed an artificial neural network based system that diagnoses heart disease. Their system accepts age, cholesterol concentration, and arterial hypertension as input data for the diagnosis. Uguz (2012) developed an artificial neural network based system that diagnoses heart valve disease. His system uses glucose concentration as input data for performing the diagnosis.

The limitation of this approach is that training the artificial neural network takes a lot of time (Berner and La Lande, 2016). Explicitly defining inference rules for the system solves this limitation.

2.3.3 Ontology Based Approach

Chandrasekaran*et al*. (1999) provided two definitions of ontology.First,ontology was defined as a representation vocabulary,oftenspecialized to some domain or subject matter.More precisely,it is not the vocabulary as suchthat qualifies as an ontology, but the conceptualizations that the terms in the vocabularyare intended to capture. They added that it is this conceptualization that allows an ontology to remain the same even after translation from one language to another.Ontology was also described to refer to a body of knowledge describing some domain, typically acommon sense knowledge domain, using arepresentation vocabulary.

Ontological analysis clarifies the structureof knowledge. Given a domain, its ontologyforms the heart of any system of knowledgerepresentation for that domain. Withoutontologies, or the conceptualizations thatunderlie knowledge, there cannot be a vocabulary for representing knowledge. Thus, thefirst step in devising an effective knowledge representation system, and vocabulary, is toperform an effective ontological analysis ofthe field, or domain. Weak analyses lead toincoherent knowledge bases (Chandrasekaran*et al.*, 1999).

Rawte and Roy (2015) used an ontology in the diagnosis of thyroid disease (hyperthyroidism and hypothyroidism). The ontology shown in Fig. 2.2 was designed using the Web Ontology Language (OWL). The ontology has four main classes for staring information about age, gender, disease, and diagnosis.

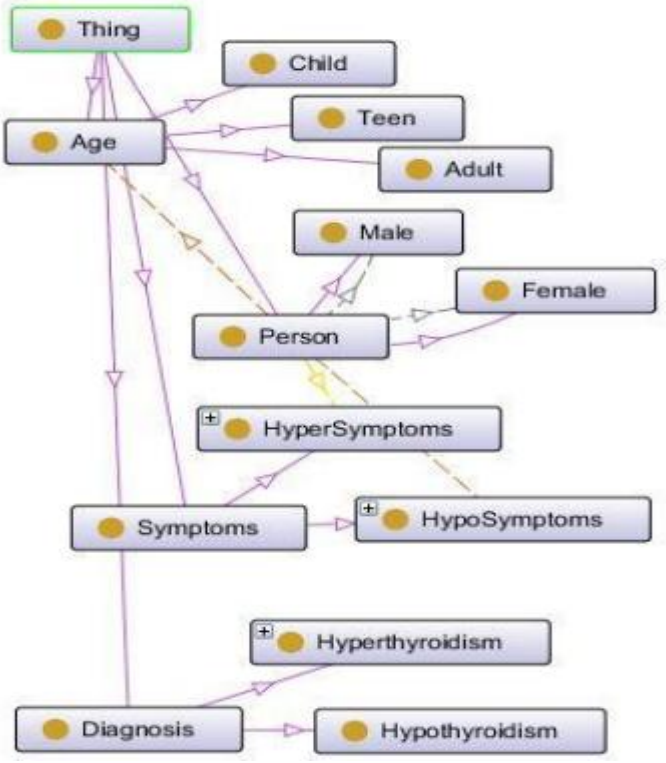


Fig. 2.2: Thyroid disease Ontology Model (Rawte and Roy, 2015)

Rawte and Roy (2015) used the Semantic Web Rule Language for defining the inference rules as shown in Fig. 2.3, and Jena API for interfacing between the system and the ontology.

```
[thd: (?patient odd:hasThis odd:Depression),
      (?patient odd:hasThis odd:IncreasedBodyFat),
      (?patient odd:hasThis odd:Constipation),
      (?patient odd:hasThis odd:Fatigue)
      -> (?patient odd:mayHave odd:Thyroid_Deficiency)]

[hypert: (?patient odd:hasThis odd:SleepProblem),
      (?patient odd:hasThis odd:Anxiety),
      (?patient odd:hasThis odd:HeatIntolerance),
      (?patient odd:hasThis odd:Fatigue)
      -> (?patient odd:mayHave odd:HyperThyroidism)]
```

Fig. 2.3: Inference Rules (Rawte and Roy, 2015)

Rawte and Roy (2015) stored patient information in a MySQL database with the front-end web interface designed with JSP as shown in Fig. 2.4.

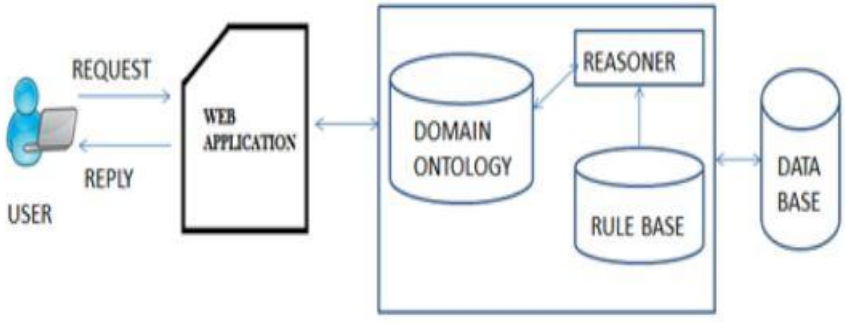


Fig. 2.4: System Architecture (Rawte and Roy, 2015)

Chang *et al*. (2015) developed an ontology that was used in the diagnosis of depression. They used the Description Logics variant of the Web Ontology Language (OWL) in developing the ontology, and GeNIe and SMILE were used as reasoning engines for making inferences on the ontology. The system worked by fetching the patient's personal ontology from the cloud. It then presents a user interface for getting data that will be used to infer whether the patient is suffering from depression.They used a Bayesian Network as a tool for computing the probability of a user having a disease from the user's description.

Bucci*et al.* (2011) used an ontology in the diagnosis of flu. They also made use of a Bayesian Network in computing uncertainty.They made use of RacerPro and pellet as reasoners.Theypostulated ontologies are ideal for diagnosis because of the relative ease of expanding them. Ontologies are very good for sharing knowledge and new symptoms can easily be added to existing ones. Sometimes, additional tests are required to fully diagnose a problem.Bucci*et al.* (2011) made use of decision networks which are extensions to Bayesian Networks for selecting appropriate medical tests to be performed to help in diagnosis.

Dieng-Kuntz *et al.* (2006) explored using ontologies for medical diagnoses and sharing of information across medical actors. They concluded that ontologies are needed for storing and sharing data because some diseases like cystic fibrosis require attention from a plethora of actors like physicians and nurses.

Paganelli*et al.* (2010) investigated using ontologies to improve home care for patients with chronic diseases that require constant care from different caregivers. They arguedthat there is a need for better long term care because most chronic diseases develop slowly overtime. The diseases have a great impact on the patient's personal life and relationships. They built a

system around a care network. A care network is all the people involved with patient care such as nurses, doctors, and relatives. The ‘person’ entity takes care of all the people in the care network, and the ‘Place’ entity models the location like home, hospital, and test centre. The ontology also stores biomedical properties of the patient, like temperature, and social context of the patient like information regarding members of the care network. The system also had the ability of alerting the patient’s care network when some medical parameters get out of range.

Farooq *et al.* (2011) made use of an ontologyto aid in the diagnosis of chest pain. Prior to performing tests to ascertain the kind of pain the patient is presenting, a pre-test is administered. This pre-test is to know whether the chest pain is of a cardiac origin, and is usually in the form of an interactive interview. They argued that the interview is not flexible, and is difficult to maintain and update. Their proposed framework not only stores patient’s data in an ontology, but also the patient’s medical history. The framework also makes use of the patient’s history in generating a questionnaire with questions relevant to that patient. The ontology is also used in chest pain classification, risk assessment, and decision making.

Ontologies are also useful in the field of collaborative diagnoses. Collaborative diagnoses is when experts in a field gather together to solve a problem, or prove that the problem is unsolvable. Iantovics(2009) explored this idea and developed a system where intelligent agents assist human experts in collaboration. His system made use of a shared memory location where experts could post problems and solutions. When a problem is available, experts can send automatic agents to look for probable solutions; these automatic agents interact with ontologies behind the scene and return their findings to the human expert. The

human, then, reports his findings to the shared memory for further improvement. The process is repeated until a solution is found, or until it is proven that the problem cannot be solved.

2.4 Technologies Used in Developing Ontology Based Applications

The technologies used in developing the ontology driven application are the Jena API for interfacing between user and knowledge base, JavaFX for designing the user interface, and Protégé v5.1 for development of the ontology.

2.4.1 Jena API

Jena is an open source Semantic Web framework for Java (Ameen *et al.,* 2014). Jena has an API to extract data from and write toRDF graphs and OWL ontologies. Model represents a graph in Jena. A model can be created by using datafrom URLs, files, databases or by combining different sources. In memory and persistent storage for storinglarge number of RDF triples is provided in Jena. Jena consists of a number of APIs as shown in Fig. 2.5. The Ontology API provides Java classes that allow communicating with ontologies. The SPARQL API provides a model for executing queries on ontologies. These queries retrieve information about entities in the ontology. Jena has support for internal reasoners as well as including external ones. The inference API provides access to these reasoners. These reasoners allow new facts to be inferred from existing ones. The store API provides options for storing triples in memory, SQL database, native tuple store, or a user defined store.

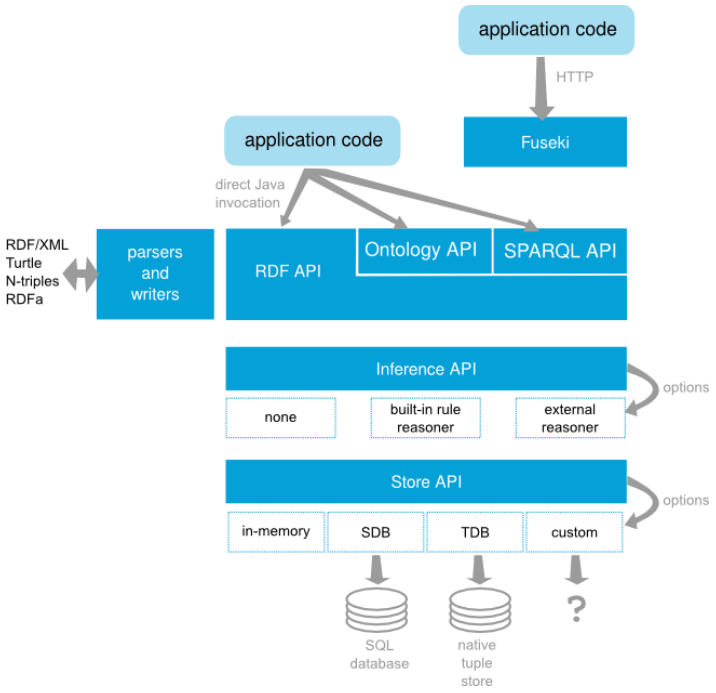


Fig. 2.5: Jena API Architecture (Jena Architecture Overview, 2017)

2.4.2 JavaFX

JavaFX is a set of Java libraries designed to enable developers to create and deploy rich client applications that behave consistently across platforms (Castillo, 2013).JavaFX is an Application Programming Interface (API) that is built on the Java Development Kit (JDK) as shown in Fig. 2.6. Traditionally, the abstract window toolkit was used for developing desktop applications, but Swing came along and replaced it. Now, JavaFX has replaced Swing as the API for desktop application development. JavaFX also provides the added benefit of displaying the results of the statistical analysis in charts and styling controls using syntax similar to Cascading Style Sheets (CSS).

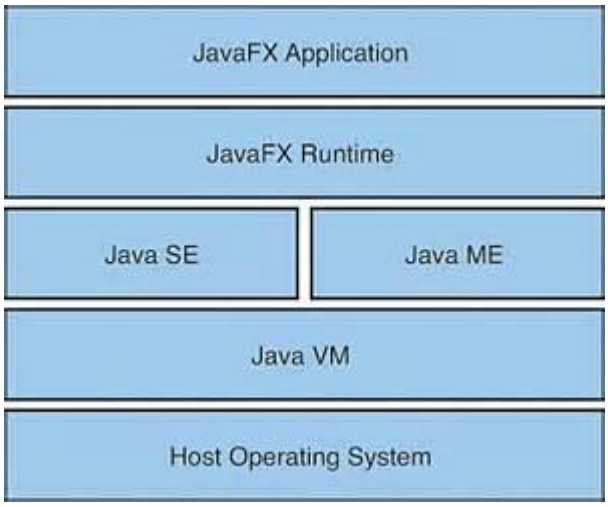


Fig. 2.6: JavaFX Architecture (Castillo, 2013)

2.4.3 Protégé

Protégé is a project that started in the 1980s at Stanford University, and is still going strong (Musen, 2015). It is an open-source platform that aids in building and maintaining ontologies. It can interact with ontologies in different formats. For example: RDF/XML syntax, turtle syntax, OWL/XML syntax, and LaTex syntax. Protégé automatically manages and displays metrics for the active ontology as shown in Fig. 2.7. Itprovides a graphical interface for the definition of classes, class hierarchies, variables, variable-value restrictions, and therelationships between classes and the properties of these relationships. It comes loaded with a reasoner that allows for inferring new facts from existing ones. It also provides an interface for adding plugins that extend the capabilities of the application.

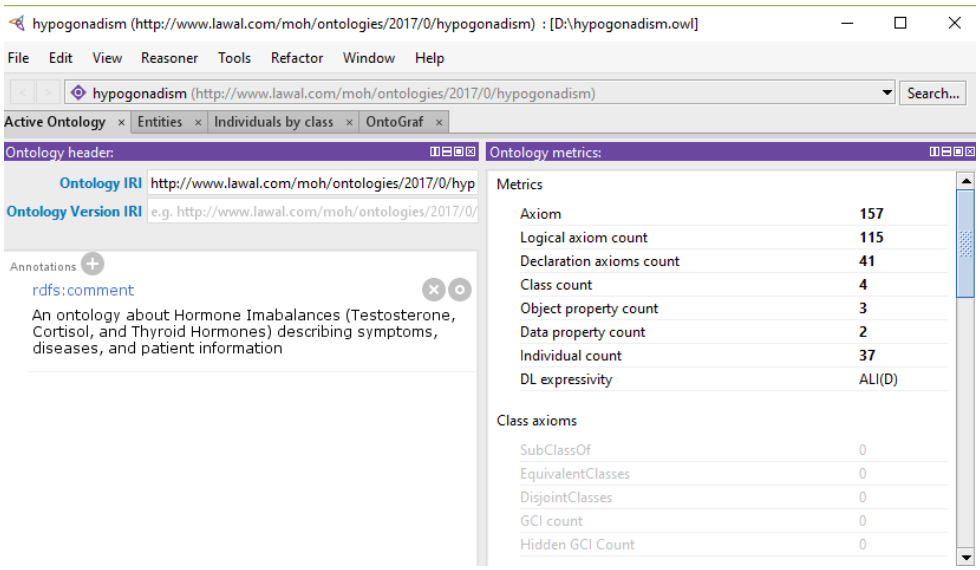


Fig. 2.7: Protégé Interface (Musen, 2015)

2.5 Evaluation Metrics

Manouselis et al. (2010) carried out an empirical analysis of 75 medical ontologies and developed the following metrics for comparing medical ontologies:

i. Class Richness (CR): This is a ratio of the number of classes that have instances and the total number of classes. An ontology that has a high class richness (close to 100%) would indicate that the data in the ontology represents most of the knowledge in the domain. Mathematically, it is given as:

$$CR = C \div C' \tag{2.1}$$

Where C = Number of classes with instances in the ontology
C' = Total Number of classes in ontology

ii. Relationship Richness (RR): This is the ratio of the number of relationships defined in the schema and the sum of all subclasses. Relationship richness values closer to one indicate a richer taxonomy. Mathematically, it is given as:

$$RR = R \div SC \quad (2.2)$$

Where R= Number of relationships defined in the schema

SC = Number of subclasses

iii. Inheritance Richness (IR): This metric describes the distribution of knowledge across levels of the ontology's inheritance tree. This metric is used to differentiate between a horizontal ontology and a vertical ontology. Inheritance richness values close to zero indicate a flat ontology with general knowledge while larger values describe a vertical ontology with detailed knowledge of the domain. Mathematically, it is given as:

$$IR = \frac{\Sigma|H^c(C_1,C_i)|}{|C|} \quad (2.3)$$

Where H^c (C_1 , C_2) is a relation that means C_1 is a sub-concept of C_2. Therefore, the dividend is defined as the total number of subclasses (C_1 – one level) for a class C_i. The divisor (C) is the number of all classes in the ontology.

2.6. Gaps in Literature

The review of questionnaire based approaches revealed that the time the systems take to produce output when provided with input is high. The drawback of artificial intelligence based systems was found to be the long time required to train the system.The limitation of the ontology based system developed by Rawte and Roy (2015) did not account for diseases with similar symptoms thereby reducing the number of correct diagnosis the system makes.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

This chapter discusses the system architecture, how the ontology was designed, and how the inference rules and SPARQL queries were defined. It also talks about the metrics used in evaluation and how the simulation was used in measuring the number of correct diagnosis the system makes.

3.2 System Architecture

The system is made up of several modules: the ontology, the inference engine, query engine, and the user interface. The user interacts with the system through the interface by selecting the appropriate symptoms, and pressing a button for processing. The system then connects to the ontology that is loaded from diskthrough the Apache Jena Application Programming Interface (API) and sends the selected symptoms to the inference engine for reasoning.

The inference engine returns its result indicating whether the user may or may not have a condition. The architecture of the system is shown in Fig. 3.1.

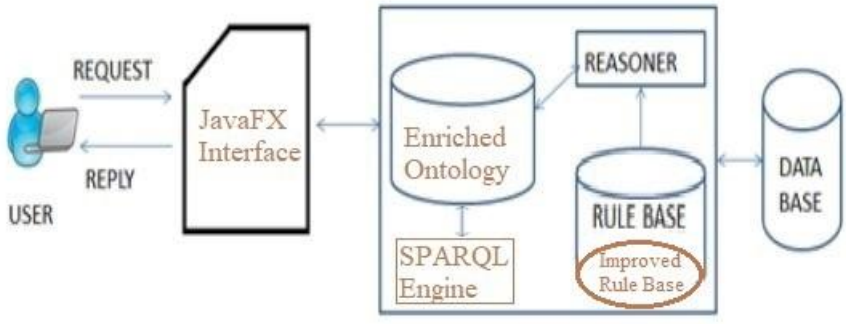


Fig 3.1: System Architecture of the Proposed System

3.3 System Requirements

The system was developed in a computing environment with the following specifications:

i. Windows 10 64-bit edition
ii. Intel Core i3-7100u @ 2.40 GHz
iii. 1 Terabyte hard drive
iv. 4 Gigabyte RAM
v. Java development kid 8 update 131
vi. NetBeans IDE 8.1

3.4Designing the Ontology

Ontologies are used to provide explicit representation of some domain. The Web Ontology Language(OWL) is used in modelling the ontology. OWL-DL was used in this research. The DL stands for description logics, and is the OWL variant that provides maximum expressiveness without losing computational completeness. The ontology was designed using Protégé (v5.1).

An OWL ontology is made up of Individuals, Properties, and Classes. Individuals, here, correspond to objects in the domain the ontology is describing: in our case, hormone imbalance and its symptoms. Properties are relationships between two individuals describing the link between them. Properties can have inverses; for example 'Eats' can have an inverse of 'isEatenBy'. They can also be transitive; *a* is related to *b*, and *b* is related to *c*, it can be concluded that *a* is related to *c*. Properties can also be symmetric; if a and b are related by a symmetric property P, individual *a* is related to *b*, and individual b is related to *a*. A class, in OWL, is interpreted as a set containing Individuals.It explicitly describes the requirements of

membership for the class in question. Classes can be organized into superclass-subclass hierarchies and relationships between them defined as shown in Fig. 3.2.

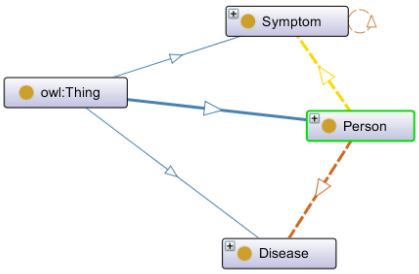


Fig. 3.2: Relationships of Classes in Ontology

Every ontology starts with 'Thing' at the top of the class hierarchy. 'Thing' is a class that is the set of all individuals in the ontology.This means every class to be defined will be a subclass of the Thing class.The subclass relationship between classes implies that all individuals of the subclass are also individuals of the superclass. Fig.3.3 shows the ontology model of the system.

The ontology model can be broken down into the following components:

i. Classes: Defining a Person class. An individual of type Person will be created during system initialization to hold the information of the current user of the system. The information that will be stored will be age, gender, symptoms, and possible hormone imbalances. A Symptom class is defined to be the set of all symptoms in the system. A Disease class is defined to be the set of all hormone imbalances in the system.

ii. Individuals: Symptoms are defined as individuals in the system. The following individuals are defined as symptoms of hormone imbalances; BreastDiscomfort,

LowBoneDesnity, Constipation, BreastEnlargement, Anxiety, ThinSkin, ReducedShaving, Fatigue, FrequentClearUrination, Weakness, IncreaseBodyFat, ReducedMuscleMass, FertilityProblem, ErectileProblem, Thirst, Dehydration, LossOfHair, ReducedLeanBody, Depression, and SleepProblem. The individuals are related to the individual of type of Person to indicate whether the user has the symptoms. Hormone imbalances are defined as individuals of type Disease. These individuals (CortisolImbalance, TestosteroneDeficiency, HypothyroidismDeficiency, and Hyperthyroidism) are also related to the individual of type Person to indicate whether the user has any of the hormone imbalances.

iii. Relationships: Two relationships are defined as object properties. The first one is called 'hasThis'. This property relates an individual of type Person with an individual of type Symptom and is used to indicate whether the user has a particular symptom. The second one is called 'mayHave'. This property relates an individual of type Person with an individual of type Disease and is used to indicate the possibility of the user having a particular imbalance. Two data properties are defined to hold the age and gender of the patients.

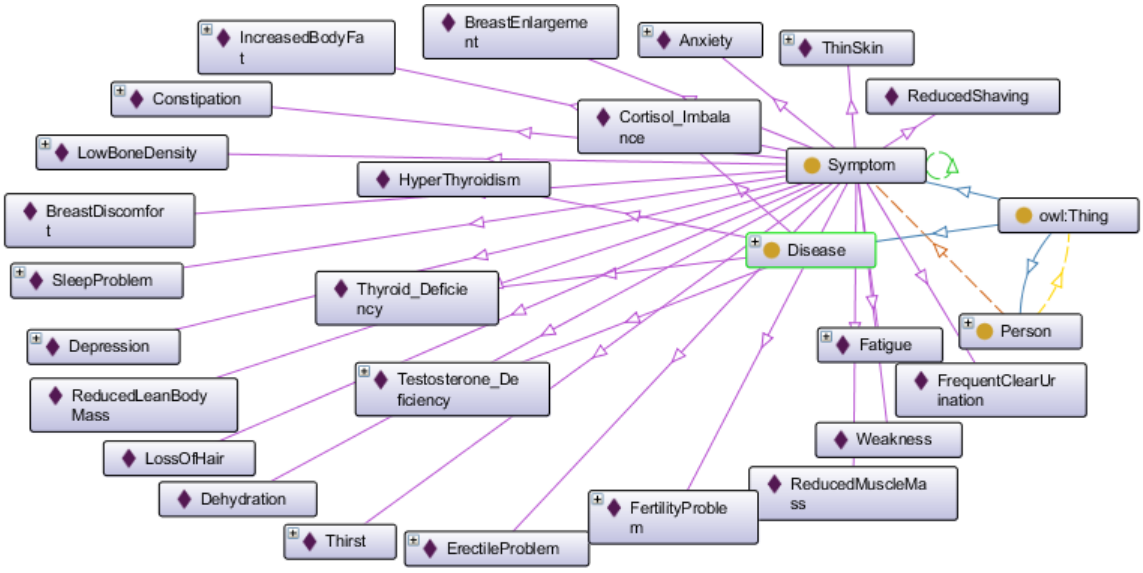


Fig. 3.3: Hormone Imbalance and patient ontology model

3.5Designing the Inference Rules

Data, in the form of symptoms, is collected from the user through the graphical user interface and the Jena API will use inference rulesshown in Fig. 3.4 to infer whether the patient has a disease or not. The shaded rules are the proposed inference rules and the unshaded rules are the inference rules from Rawte and Roy (2015). The rules for determining hypogonadism were based on the Androgen Deficiency in Aging Males (ADAM) questionnaire found in Morley*et al.* (2000). The rules for determining cortisol and thyroid hormone imbalance were based on investigationsfrom Smith (2005).

```
@prefix odd: <http://www.lawal.com/moh/ontologies/2017/0/hypogonadism#>

[hypo: (?patient odd:hasThis odd:Depression),
       (?patient odd:hasThis odd:ErectileProblem),
       (?patient odd:hasThis odd:Fatigue),
       (?patient odd:hasThis odd:Anxiety)
       -> (?patient odd:mayHave odd:Testosterone_Deficiency)]

[hypo2:(?patient odd:hasThis odd:ErectileProblem)
       -> (?patient odd:mayHave odd:Testosterone_Deficiency)]


[cort: (?patient odd:hasThis odd:IncreasedBodyFat),
      (?patient odd:hasThis odd:ThinSkin),
      (?patient odd:hasThis odd:Depression),
      (?patient odd:hasThis odd:Anxiety),
      (?patient odd:hasThis odd:ErectileProblem),
      -> (?patient odd:mayHave odd:Cortisol_Imbalance)]

[thd: (?patient odd:hasThis odd:Depression),
      (?patient odd:hasThis odd:IncreasedBodyFat),
      (?patient odd:hasThis odd:Constipation),
      (?patient odd:hasThis odd:Fatigue)
      -> (?patient odd:mayHave odd:Thyroid_Deficiency)]

[hypert: (?patient odd:hasThis odd:SleepProblem),
      (?patient odd:hasThis odd:Anxiety),
      (?patient odd:hasThis odd:HeatIntolerance),
      (?patient odd:hasThis odd:Fatigue)
      -> (?patient odd:mayHave odd:HyperThyroidism)]
```

Fig 3.4: Inference rules used by system for diagnosis

3.6 Designing the User Interface

The ontology and the inference engine provide the backbone of the application, but an interface to the user is needed for collection and presentation of data. The interface was designed using JavaFX. The system is made up of three tabs. The first tab is the patient bio-data tab. This tab is made up of a text field for username input and widgets for age and gender input. The second tab is the diagnosis tab. This tab is made up of checkboxes representing symptoms, a button for initiating processing, and a text field for displaying results. The third tab is the statistics tab. This tab is made up of bar charts that report the patient, age, and gender distributions.

3.7Queries for Statistical Analysis

The symptoms users select during their interaction with the system are stored in the ontology. The age and gender of the patients are also stored alongside these symptoms. Queries define a way of retrieving information from the ontology for performing the statistical analysis; information like the number of patients that have ever interacted with the system, number of patients that may have any of the four hormone imbalances, and the age and gender of the patients.

The SPARQL Protocol and RDF Query Language (SPARQL) which is a semantic query language was used in defining the queries used by the system. SPARQL queries consist of a single or collection of triples and are of the form 'subject-predicate-object'. A sample query for retrieving the number of patients in the ontology is given in Fig. 3.5.

```
PREFIX odd: <http://www.lawal.com/moh/ontologies/2017/0/hypogonadism#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?y
WHERE {
       ?age odd:mayHave odd:Testosterone_Deficiency
       ?age odd:age ?y
      }

SELECT ?x
WHERE {
       ?gender odd:mayHave odd:Cortisol_Imbalance
       ?gender odd:gender ?x
      }

SELECT (COUNT(*) AS ?z)
WHERE {
       ?patients rdf:type odd:Person
      }
```

Fig. 3.5: SPARQL queries

SELECT is a keyword defining what to retrieve from the triples that match the condition specified in the WHERE clause. Count is a function that returns the number of matching

triples and all identifiers following a question mark are variables. It is the job of the query engine to execute these queries. The Jena API chosen for this research comes bundled with a SPARQL implementation.

3.8 Evaluation Metrics

The ontology of the system was evaluated using the class richness metric as defined in section 2.5. The class richness measures knowledge representation. Higher values are better and they indicate an ontology that captures more information about a particular domain. Inheritance richness was also used. This defines the verticality of the ontology. Horizontal ontologies model general knowledge while vertical ontologies model specific knowledge. Lower values indicate a horizontal ontology while higher values, which are better, indicate a vertical ontology. Relationship richness which measures the classification of information in the ontology was also used, and higher values are better.

3.9 Inference Capability Evaluation

The inference rules were also an evaluation metric because Han *et al*. (2011) found a relationship between the efficiency of an expert system and the number of inference rules. A simulation program was developed to measure the number of correct diagnoses the developed system makes based on data collected from Watt *et al*. (2006). They reviewed patients diagnosed with thyroid disease and noted their symptoms. They devised a list of symptoms from the patients reviewed noting the number of patients presenting with a particular symptom. They ended up with a list of symptoms and a percentage of patients that presented with those symptoms. This list of symptoms served as the input data for the simulation. The

list was stored using an array, as shown in Fig. 3.6, and was used in generating a list of probabilities of patients presenting with those symptoms. The list of probabilities was used to generate a list of cumulative probabilities. This list of cumulative probabilities was then used to generate random number assignment ranges for each symptom. An array of 25 elements was created and all elements were initialized to zero. The array holds information about symptoms selected to be passed to the reasoning engine. A zero indicates a particular symptom is not selected while a value of one indicates a selected symptom. A loop was set up to model the interaction of a patient selecting any of the 25 symptoms. The inference rules of the developed system and that of Rawte and Roy (2015) were used to perform the diagnosis based on the selected symptoms, and the number of correct diagnosis was stored. An outer loop executes 100 times modeling the interactions of 100 patients.

```
set array symptoms_probabilities {values_from_Watt}
set array symptoms_cummulative_probabilities {values_from_symptoms_probabilities}
set array symptoms_random_number_assignment {values_from_symptoms_cummulative_probabilities}
set array size 25 selected_symptoms 0

repeat for 100
for i=1 to 25 {number of symptoms} do
 x=generate_random_number_1-100
 y=symptoms_random_number_assignment[x]
 set selected_symptoms[y] to 1 if 0
end for

a=check_correct_diagnosis_rawte(selected_symptoms)
b=check_correct_diagnosis_proposed(selected_symptoms)
end repeat

return a,b
```

Fig. 3.6: Simulation pseudocode

CHAPTER FOUR: RESULTS AND DISCUSSION

4.1 Introduction

This chapter provides an overview of the system including its interface, the results of the evaluation carried out, and the analysis of the result.

4.2 System Overview

The system was implemented in Java, and is made up of three tabs. The first tab is the Bio-Data tab, shown in Fig. 4.1, used in entering patient information to be added to the ontology. The second tab is used for entering symptoms the patient is experiencing and the third tab displays statistical reports.

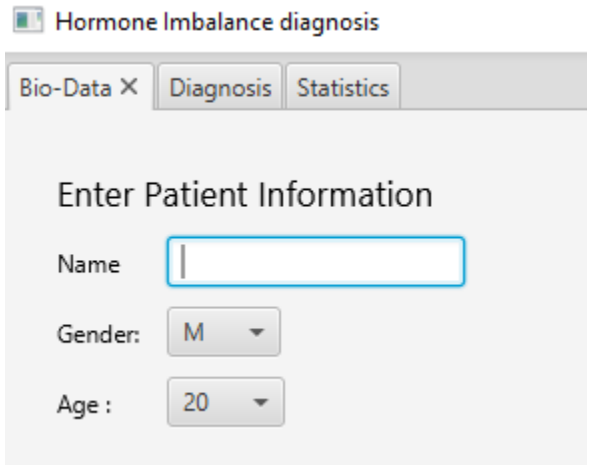


Fig. 4.1: Bio-Data tab

The user interacting with the system selects the symptoms the user is experiencing from the array of checkboxes. The user then specifies gender and age from the bio-data tab as shown in Fig. 4.1, and then clicks on the 'Process' button as shown in Fig. 4.2. The system sends the

selected data to the ontology for storage, and then forwards them to the inference engine to determine if the user has any of the four hormone imbalances. The inference engine sends what it found back to the ontology for storage and to the user interface for presentation.

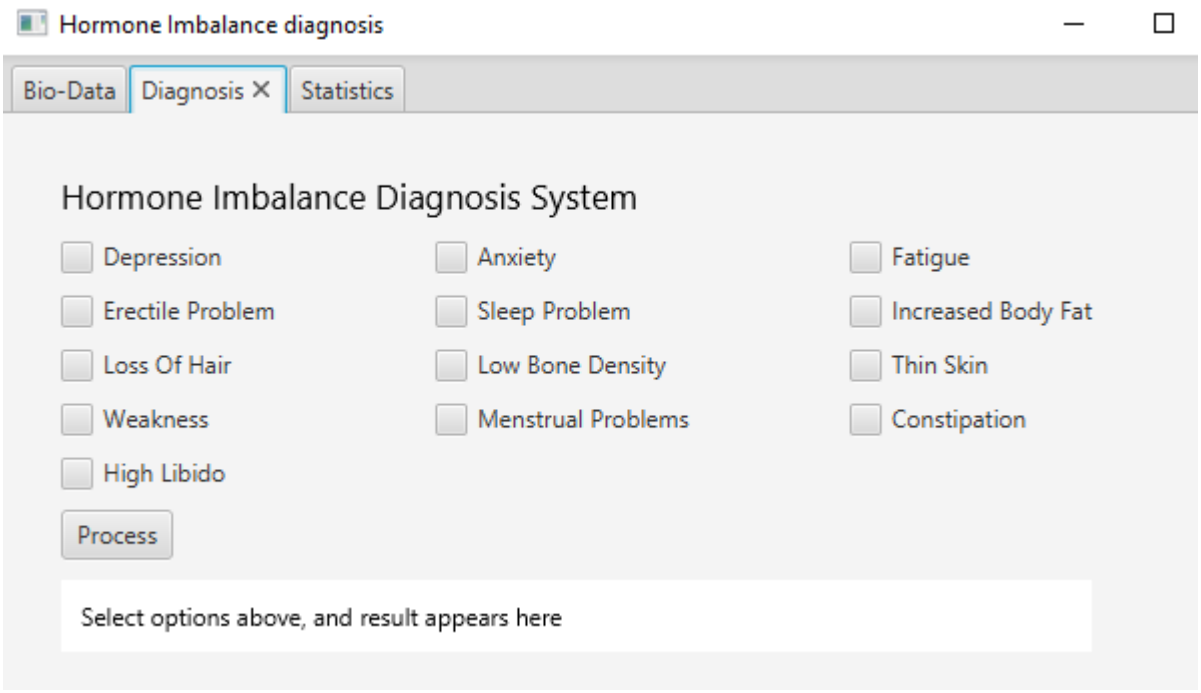


Fig. 4.2: Diagnosis tab

The user can also view statistical information about the users that have previously interacted with the system. The user can view information about users that have any of the four hormone imbalances, their age and gender distribution. This action is achieved by clicking the ‘Statistics’ tab as shown in Fig. 4.3.

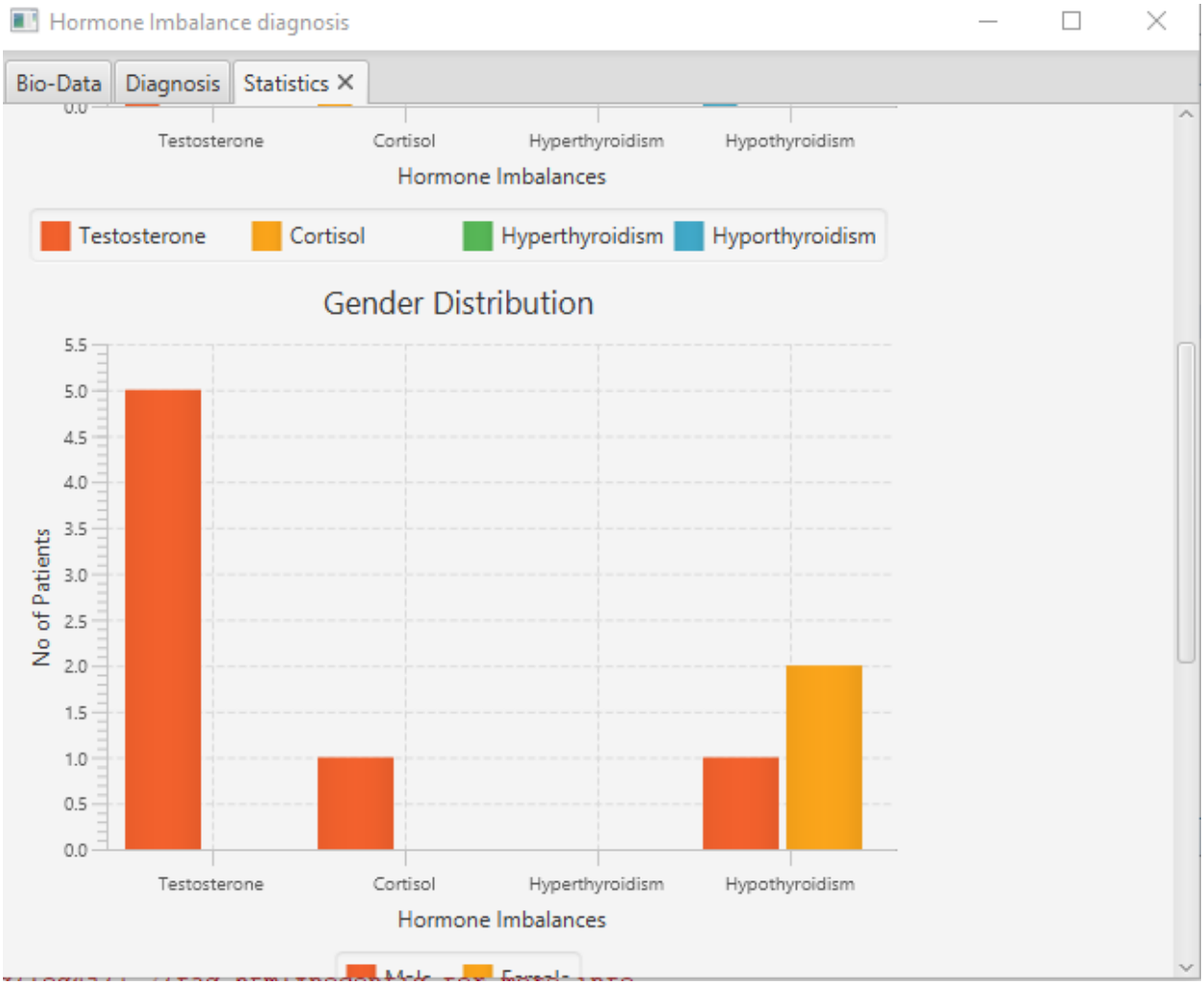


Fig. 4.3: Statistics tab

4.3 Evaluation

The ontology of this system and that of Rawte and Roy (2015) were evaluated using the metrics from Manouselis et al. (2010). Table 4.1 presents the findings of this evaluation.

Table 4.1: Evaluation of knowledgebase of the proposed system and Rawte and Roy (2015)

	No of Classes (noc)	No. of Instances (noi)	No. of Properties (nop)	No. of Classes with Instances (ci)	Class Richness (cr)	Inheritance Richness (ir)	Relationship Richness (rr)
Proposed System	4	20	4	3	0.75	0.08	0.6667
Rawte and Roy (2015)	14	16	6	3	0.2143	0.14	0.3077

The number of correct diagnosis the system makes was measured using the simulation program developed with data retrieved from Watt *et al.* (2006). Fig. 4.4 shows the number of correct diagnosisthe proposed system makes using two inference rules (thyroid disease and testosterone imbalance), and that of Rawte and Roy (2015).

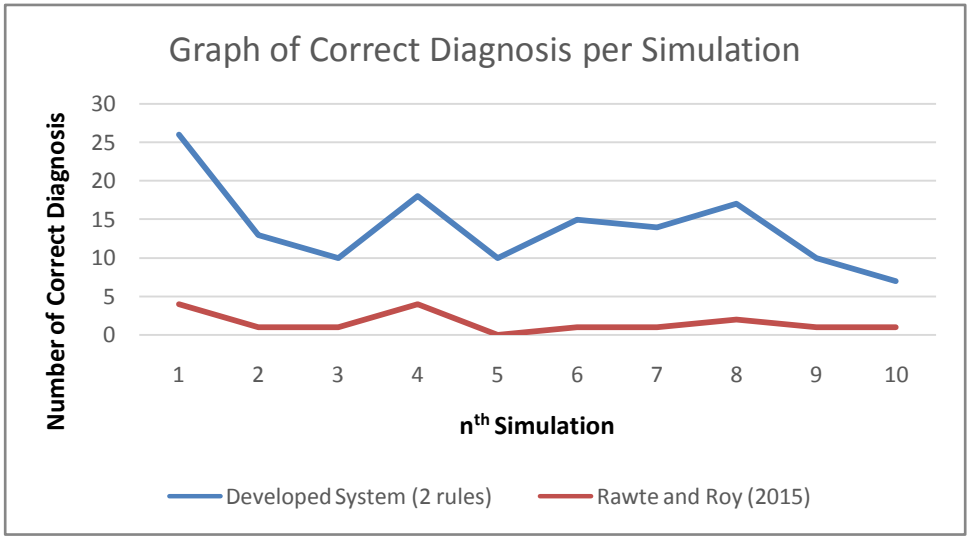


Fig. 4.4 Evaluation of system performance with 2 inference rules against Rawte and Roy (2015)

Han *et al.* (2011) found that increasing the number of inference rules improves the accuracy of the system.

Fig. 4.5 shows the number of correct diagnosis of the proposed system using three inference rules (thyroid disease, testosterone imbalance, and cortisol imbalance).

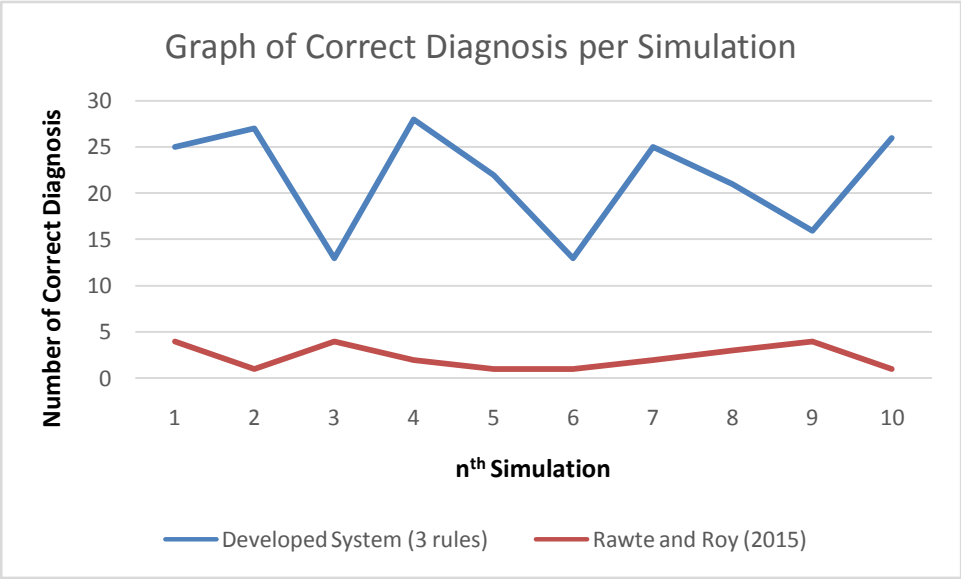


Fig. 4.5 Evaluation of system performance with 3 inference rules against Rawte and Roy (2015)

4.4 Result Analysis

The three metrics defined in section 2.5 on page18 indicate how well the ontology designed were used in the evaluation.These metrics are the class richness (cr), inheritance richness (ir), and the relationship richness (rr). Higher values of the class richness metric indicate an ontology that represents most of the knowledge in the domain. The proposed system has a

class richness value of 0.75 or 75% while that of Rawte and Roy (2015) had a class richness value of 0.2143 or 21.43%. This means the system is 2.5 times better than that of Rawte and Roy (2015). This improvement was anticipated because the ontology used by the proposed system modelled more knowledge about hormone imbalances. The proposed system has more symptoms and more diseases in the ontology. The inheritance richness metric measures the depth of the ontology, and the nature of the knowledge stored in it. An ontology with a low inheritance richness metric indicate a horizontal ontology that models general knowledge. Conversely, an ontology with a high inheritance richness metric indicate an ontology that models specific knowledge. The proposed system presented with an inheritance richness metric of 0.08 while that of Rawte and Roy (2015) had an inheritance richness metric of 0.14. Their system is 42.86% better in this regard, and this is to be expected as the specificity of the ontology is reduced by adding more symptoms and diseases to it. The relationship richness metric is a measure of how well a particular domain is defined in an ontology. Higher values of this metric indicate a richer taxonomy. The proposed system has a relationship richness of 0.6667 or 66.67% while that of Rawte and Roy (2015) has a relationship richness of 0.3077 or 30.77%. This means the proposed system is 1.2x better than that of Rawte and Roy (2015)in this regard. This improvement was also expected as describing closely related symptoms and diseases improves the domain of hormone imbalance.

The results of the simulation show that using a single inference rule results in an average of 2.6 correct diagnoses out of a possible 100 for both the developed system and the system designed by Rawte and Roy (2015). Increasing the number of inference rules two increases the number of correct diagnoses to an average of 14, and increasing the number of inference rules to three increases the number of correct diagnoses to an average of 21.6. This means the

developed system is approximately 7times more accurate in terms of the number of correct diagnosis made.

The system developed by Rawte and Roy (2015) employed the use of a MySQL database for storing patient information. The proposed system used an ontology instead of the MySQL database for storing patient information. One of the benefits of using an ontology instead of a database is the ease of sharing information. Patient information can easily be shared with other expert systems. The system proposed by Rawte and Roy (2015) does not perform any statistical analysis on the data collected from the user. The proposed system is able to perform analysis like getting the percentage of users that have tested positive for any of the hormone imbalances and present the findings on a bar chart. The system can also present the same statistics based on the ‘age’ and ‘gender’ of the users.

CHAPTER FIVE: SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Summary

In this dissertation,an ontologywas used as a backend for a system that diagnoses three hormone imbalances: cortisol imbalance, testosterone deficiency, and thyroid disease. The ontology described the symptoms of the hormone imbalances, and also stored information about the patient. Adding more diseases to the ontology resulted in improved richness and knowledge representation by a factor of 2.5 and 1.2 respectively, but it also reduced the specificity of the ontology by 42.86%. The increased number of inference rules resulted in a system that was 7 times more accurate than that of Rawte and Roy (2015) in terms of correct diagnosis made. The query engine that is part of the expert system was able to perform queries on the ontology which allowed us to fetch the number of patients in the ontology with their characteristics. It was this data that was used in reporting the statistics of the users that have interacted with the system.

5.2 Conclusion

The improved richness and knowledge representation of the ontology aided the system in diagnosing three hormone imbalances: thyroid disease, cortisol imbalance, and testosterone deficiency. Ontologies are specifications about a particular domain. While adding to the ontology improved the representation of the domain of hormone imbalances, it also made it less specific as evidenced by the reduction in the inheritance richness. The ontology was also used in storing patient information, this makes them easy to share with other expert based

systems. The increased number of inference rules improved the number of correct diagnosis the system makes.

5.3 Recommendations for Future Work

The results of this research have led to the following recommendation:

i. The efficiency of the system may be improved by the addition of more relevant reasoning rules.

ii. The ontology of the system may be moved to the cloud to allow access from any location.

iii. A mobile-based approach to ontology based applications can be studied as this will enhance the system accessibility and wide usage.

REFERENCES

Alexander, G. M., & Wilcox, T. (2012).Sex differences in early infancy.*Child Development Perspectives*, *6*(4), 400-406

Ameen, A., Khan, K. U. R., & Rani, B. P. (2014).Reasoning in Semantic Web Using Jena.*Computer Engineering and Intelligent Systems*, *5*(4), 39-47.

Atkov, O. Y., Gorokhova, S. G., Sboev, A. G., Generozov, E. V., Muraseyeva, E. V., Moroshkina, S. Y., &Cherniy, N. N. (2012). Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters.*Journal of cardiology*, *59*(2), 190-194.

Berner, E. S., and La Lande, T. J. (2016).Overview of clinical decision support systems. In E.S. Berner (Ed.); *Clinical decision support systems* (pp. 1-17). New York, USA: Springer International Publishing.

Blaya, J. A., Fraser, H. S., & Holt, B. (2010). E-health technologies show promise in developing countries. *Health Affairs*, *29*(2), 244-251.

Bucci, G., Sandrucci, V. & Vicario, E. (2011). Ontologies and Bayesian Networks in Medical Diagnosis: Proceedings of the 44th Hawaiian Conference on System Sciences, (Edited by Sprague, R.H. (Jr)). *IEEE Computer Society Press*, pp. 1-8

Castillo, C. (2013). *JavaFX Architecture.* RetrievedAugust 5, 2017, from http://docs.oracle.com/javafx/2/architecture/jfxpub-architecture.htm

Catwell, L., & Sheikh, A. (2009).*Evaluating eHealth interventions: the need for continuous systemic evaluation*. Retrieved June, 20, 2017, from https://doi.org/10.1371/journal.pmed.1000126

Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What are ontologies, and why do we need them?.*IEEE Intelligent Systems and their applications*, *14*(1), 20-26

Chang, Y., Fan, C., Lo, W., Hung, W. and Yuan,. S. (2015): Mobile Cloud-Based Depression Diagnosis using an Ontology and a Bayesian Network, *Future Generations Computer Systems, 32*, 87-98

Chueh, K. S., Huang, S. P., Lee, Y. C., Wang, C. J., Yeh, H. C., Li, W. M., ... & Huang, C. H. (2012). The Comparison of the Aging Male Symptoms (AMS) Scale and Androgen Deficiency in the Aging Male (ADAM) Questionnaire to Detect Androgen Deficiency in Middle- Aged Men.*Journal of andrology*, *33*(5), 817-823.

Dandona, P. & Rosenberg, M. T. (2010). A Practical Guide to Male Hypogonadism in the Primary Care Setting, *The International Journal of Clinical Practice, 64*(6), 682-696

Dieng-Kuntz, R., Minier, D., Ruzicka, M., Corby, F., Corby, O. &Alamarguy, L. (2006). Building and using Medical Ontology for Knowledge Management and Cooperative work in a Health Care Network, *Computers in Biology and Medicine*,*36*(7), 871-892

Farooq, K., Hussain, A., Leslie, S., Eckl, C., & Slack, W. (2011, May). Ontology-driven cardiovascular decision support system. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on* (pp. 283-286). IEEE.

Garets, D., & Davis, M. (2006). *Electronic medical records vs. electronic health records: yes, there is a difference*. Retrieved on August, 1, 2017, from https://www.himss.org/electronic-medical-records-vs-electronic-health-records-yes-there-difference-himss-analytics

Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism-Clinical and Experimental*, *69*, S36-S40.

Han, J., Jeong, Y. K., & Lee, I. (2011). Efficient building energy management system based on ontology, inference rules, and simulation. In *Proceedings of the 2011 International Conference on Intelligent Building and Management, Singapore* (pp. 295-299).

Hines, M., Constantinescu, M. & Spencer, D. (2015). Early Androgen Exposure and Human Gender Development, *Biology of Sex Differences*,*6*(1), 3

Iantovics, B. L. (2009). Cooperative Medical Diagnosis Elaboration by Physicians and Artificial Agents. In *From System Complexity to Emergent Properties* (pp. 315-339). Springer, Berlin, Heidelberg.

Jena Architecture Overview. (n.d.). Retrieved August 5, 2017, from http://jena.apache.org/about_jena/architecture.html

Jorm, A. F. (2004). The Informant Questionnaire on cognitive decline in the elderly (IQCODE): a review. *International Psychogeriatrics*, *16*(3), 275-293.

Kroenke, K., Spitzer, R. L., Williams, J. B., &Löwe, B. (2010). The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *General hospital psychiatry*, *32*(4), 345-359.

Ledley, R., & Lusted, L. (1959). Reasoning Foundations of Medical Diagnosis. *Science*,*130*(3366), 9-21.

Manouselis, N., Sicilia, M. A., & Rodríguez, D. (2010). Exploring ontology metrics in the biomedical domain. *Procedia Computer Science*, *1*(1), 2319-2328.

Marieb, E. N. (1992). *Human Anatomy and Physiology*: Benjamin/Cummings Publishing Company, San Francisco, USA, pp. 548-549

Morley, J. E., Charlton, E., Patrick, P., Kaiser, F. E., Cadeau, P., McCready, D. & Perry, H. M. III (2000). Validation of a Screening Questionnaire for Androgen Deficiency in Aging Males, *Metabolism,49*(9), 1239-1242

Musen, M. A. (2015). The Protégé project: A look back and a look forward. *AI matters*, *1*(4), 4-12.

Oh, H., Rizo, C., Enkin, M. &Jadad, A. (2005). What Is eHealth (3): A Systematic Review of Published Definitions, *The OfficialJournal of International Hospital Federation*, *41*(1), 32-40

Paganelli, F., &Giuli, D. (2011). An ontology-based system for context-aware and configurable services to support home-based continuous care. *IEEE Transactions on Information Technology in Biomedicine*, *15*(2), 324-333.

Pike, C. J., Rosario, E. R., & Nguyen, T. V. V. (2006). Androgens, aging, and Alzheimer's disease. *Endocrine*, *29*(2), 233-241.

Raven, P.H. & Johnson, B.H. (2001). *Biology*.McGraw-Hill, Boston, MA. pp 1125-1145

Rawte, V., & Roy, B. (2015). Thyroid Disease Diagnosis using Ontology based Expert System. *International Journal of Engineering Research & Technology*, *4*(6), 365-70

Sargis, R.M. (2017). *Thyroid Gland Overview.* Retrieved August 7, 2017, from https://www.endocrineweb.com/endocrinology/overview-thyroid

Smith, P. (2005). A Comprehensive Look at Hormones and the Effects of Hormone Replacement. In *14th Annual International Congress on Anti-Aging Medicine, Orlando, Fla*(pp. 229-238)

Tuck, S. & Francis, R.(2009). Testosterone, Bone and Osteoporosis. In S. Tuck (Ed.);*Advances in the Management of Testosterone Deficiency*, pp. 123-132. Basel, Switzerland: Kerger Publishers

Uğuz, H. (2012). A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases. *Journal of medical systems*, *36*(1), 61-72.

Watt, T., Groenvold, M., Rasmussen, Å. K., Bonnema, S. J., Hegedüs, L., Bjorner, J. B., &Feldt-Rasmussen, U. (2006). Quality of life in patients with benign thyroid disorders. A review. *European Journal of Endocrinology*, *154*(4), 501-510.

APPENDIX

```
import java.util.Random;
public class HypoSimulation {
   public static void main(String[] args){
int n=100;//Number of patients in simulation
System.out.print("Symptom\t\tProbability\t\tCummulative Probability");
System.out.print("\t\tRandom Digit Assignment\n");
      double [] hypernumbers = {(26+69)/2,(20+62)/2,(31+50)/2, (29+34)/2,(36+46)/2,
      25+41/2,40,(35+41)/2, (39+58)/2, 32, (6+79)/2, (15+52)/2, (27+36)/2,
      52, 39, 16, 33, 15,32, 55,50,52.5,83,23,57,41.5};
      double hypertotal=0;
      for(inti=0;i<hypernumbers.length;i++){
hypertotal+=hypernumbers[i];
      }
      double[] hyperprob=new double[hypernumbers.length];
      for(inti=0;i<hyperprob.length;i++){
hyperprob[i] = hypernumbers[i]/hypertotal;
      }
      double hypercum[] = new double[hyperprob.length];
int j=0,k=0;
      for(inti=0; i<hyperprob.length;i++){
         double sum=0.0;
         while(j<=k){
            sum+=hyperprob[j];       j++;
         }
hypercum[k]=sum;
         k++;j=0;      }
```

```
intstartRandomDigitAssignment[] = new int[hyperprob.length];
startRandomDigitAssignment[0]=0;
        for(inti=1;i<hyperprob.length;i++){
startRandomDigitAssignment[i]=(int)(hypercum[i-1]*100)+1;
        }
        for(inti=0;i<hyperprob.length;i++){
            System.out.format("%d\t\t\t%.3f\t\t\t%.3f\t\t\t\t%d-%d",i,hyperprob[i],hypercum[i],
startRandomDigitAssignment[i],(int)(hypercum[i]*100));
System.out.println();
        }
        System.out.println("*****************************************************");
System.out.println("The simulation will execute for "+n+" (User customizable) patients");
int[] symptoms = new int[hyperprob.length];
        Random gen = new Random();
int[] meAll= new int[10],rawteAll=new int[10];
int me= 0,rawte=0;
        for (int x=0;x<10;x++){
            me= 0;rawte=0;
            for(int z=0;z<n;z++){
                for(int y=0;y<symptoms.length;y++){
                    symptoms[y]=0;
                }
                for(inti=0;i<startRandomDigitAssignment.length;i++){
int l=0, random = gen.nextInt(101);
                    while(l<startRandomDigitAssignment.length&&
                            random>=startRandomDigitAssignment[l]){
                        l++;                    }
```

```
                if(symptoms[l-1]!=1){
                    symptoms[l-1]=1;
                }
            }
            me += checkMe(symptoms);
rawte += checkRawte(symptoms);
          }
meAll[x]=me;   rawteAll[x]=rawte;
      }
System.out.println("Successful Diagnosis:");
System.out.println("This work: ");
      for(int x: meAll){
System.out.print(x +" ");
      }
System.out.println("\nRawte and Roy (2016): ");
      for(int x: rawteAll){
System.out.print(x +" ");
      }
   }
private static intcheckRawte(int[] a){
      //Check for hyperthyroidism
      //Check for hyperthyroidism           Deression=3
      //sleep problem=12                  Sex problem=9
      //anxiety=5                        Weight problem =11
      //Heat intolerance=14               Skin Changes=16
      //Fatigue=8                        Menstrual problem=25
      //Selected symptoms will have a value of 1 and unselected symptoms
```

```
        //will have a value of 0
int correct=0;
        if(a[12]==1 && a[14]==1 && a[5]==1 && a[17]==1 && a[9]==1){
            correct++;       }
        else if(a[12]==1 && a[14]==1 && a[5]==1 && a[17]==1 && a[9]==1 && a[25]==1){
            correct++;       }
        return correct;
    }
    private static intcheckMe(int[] a){
        //Check for hyperthyroidism            Deression=3
        //sleep problem=12                    Sex problem=9
        //anxiety=5                          Weight problem =11
        //Heat intolerance=14                 Skin Changes=16
        //Fatigue=8                          Hand Tremor=17
        //Palpitations = 18                   Diarrhea = 19
        //Weight Loss=10
int correct=0;
        if(a[12]==1 && a[14]==1 && a[5]==1 && a[17]==1 && a[9]==1){
            correct++;       }
        else if(a[12]==1 && a[14]==1 && a[5]==1 && a[17]==1 && a[9]==1 && a[25]==1){
            correct++;       }
        if(a[3]==1 && a[9]==1 && a[8]==1 && a[5]==1){
            correct++;       }
        if(a[3]==1 && a[5]==1 && a[11]==1 && a[16]==1){
            correct++;       }
        return correct;
    }}
```