JIMMA UNIVERSITY

JIMMA INSTITUTE OF TECHNOLOGY

SCHOOL OF GRADUATE STUDIES

FACULTY OF COMPUTING AND INFORMATICS

DEVELOPING A SELF-LEARNING EXPERT SYSTEM FOR RESPIRATORY DISEASES DIAGNOSIS AND TREATMENT

A Thesis Submitted to the School of Graduate Studies, Jimma University, Jimma Institute of Technology, Faculty of Computing and Informatics in Partial Fulfillment of the Requirement for the Degree Master of Science in Information Science (Information and Knowledge Management)

BY:

TAMIRAT GELANA

JUNE, 2024

JIMMA, ETHIOPIA

JIMMA UNIVERSITY

JIMMA INSTITUTE OF TECHNOLOGY

FACULTY OF COMPUTING AND INFORMATICS

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF INFORMATION SCIENCE

MSC IN INFORMATION AND KNOWLEDGE MANAGEMENT

(IKM)

DEVELOPING A SELF-LEARNING EXPERT SYSTEM FOR RESPIRATORY DISEASES DIAGNOSIS AND TREATMENT

A Thesis Submitted in Partial Fulfillment of the Requirement for Degree of Science in Information Science (Information and Knowledge Management)

BY:

TAMIRAT GELANA

Principal Advisor: Dr.Amanuel Ayde

CO-Advisor: Mr.Muktar Bedaso (MSc)

JUNE, 2024

Jimma, Ethiopia

JIMMA UNIVERSITY

JIMMA INSTITUTE OF TECHNOLOGY

FACULTY OF COMPUTING AND INFORMATICS

DEPARTMENT OF INFORMATION SCIENCE

MSC IN INFORMATION AND KNOWLEDGE MANAGEMENT

(IKM)

DEVELOPING A SELF-LEARNING EXPERT SYSTEM FOR RESPIRATORY DISEASES DIAGNOSIS AND TREATMENT

BY: TAMIRAT GELANA

NAME AND SIGNATURE OF MEMBER OF THE EXAMINING BOARD

	Name	Sign	Date
Principal Advisor:	Dr.Amanuel Ayde	_____	_____
CO-Advisor:	Mr.Muktar Bedaso (MSc)	_____	_____
Internal Examiner:		_____	_____
External Examiner:		_____	_____

June, 2024

Jimma, Ethiopia

DECLARATION

I affirm that this thesis is my own original work and has not been submitted for a degree at any other university. All the sources and materials used in this work have been properly acknowledged.

_____

Tamirat Gelana

JUNE, 2024

This thesis has been submitted to the department for examination with the approval of the university advisors overseeing this work**.**

	Name	**Date**	**Signature**
Principal Advisor: Dr.Amanuel Ayde		_____	_____
Co-Advisor: Muktar Bedaso (MSc)		_____	_____

JUNE, 2024

DEDICATION

This work is dedicated to my beloved family.

JUNE, 2024

Acknowledgement

First and foremost, I express heartfelt gratitude to the benevolent God who has provided me with everything necessary to complete my courses and initiate work on my thesis. I sincerely appreciate the dedication and patience of my advisor, Dr. Amanuel Ayde, for thoroughly reviewing every section of the thesis. His valuable comments, encouragement, and guidance have been instrumental from the initial stages to the final development of the thesis paper, paving the way for my research work. I would also like to extend my appreciationto my co-advisor, Mr. Muktar Bedaso, for supplying crucial information for my thesis work. Lastly, I extend my thanks to all the staff at the JU Information Science department for their moral support and encouragement throughout my studies. I would also like to thanks Jimma University Medical Center pediatrics department and patient record office staff for their kindlyhelpandsupport.

Abstract

Globally, there is a significant health burden associated with respiratory diseases for instance, Chronic obstructive pulmonary disease (COPD) affects over 200 million individuals, asthma impacts 235 million people, and tuberculosis (TB) afflicts 8.7 million annually. Detecting these diseases at an early stage is crucial for effective treatment. However, the availability of experts, especially in developing nations, is often limited. Therefore, the main objective of this research is to develop a self-learning expert system for respiratory diseases diagnosis and treatment. The research utilized a design science research approach to gain insights into the challenges within the field and create a model as a solution. In this study, a primary focus was on employing classification techniques within data mining tasks to extract representative cases from the collected data. To determine the optimal model and select the most effective data mining classification algorithm, three experiments were carried out using J48, PART and JRIP classification algorithms. The data mining algorithm achieved accuracy of 94.5%, 92.5% and 90% for PART, J48 and JRIP respectively. This system aims to identify and offer insights into the diagnosis and treatment of common respiratory diseases. In order to develop the model the researcher collected both implicit and explicit knowledge from various sources, including domain experts, medical literature, research papers, clinical guidelines and patient records. Then organized the acquired knowledge, selected the most important attributes and discarded unnecessary ones by analysing the patient records. The relevant information was transferred to an Excel sheet and saved as a CSV file. The researcher utilized the WEKA 3.9 data mining software to pre-process and classify the data using algorithms such as J48, PART, and JRIP. The performance of these algorithms was evaluated using a 10-fold cross-validation method and an 80/20% data split. Finally, the PART rules were represented in a structured format suitable for computer-based systems using rule-based representation techniques and integrated with a node to run the main interface. The prototype is evaluated using system testing and user acceptance testing. System testing performed in terms of recall, precision and F-measure registered 90.9%, 90.9% and 90.5% respectively. User acceptance testing also performed by involving domain experts and an average of 84% acceptance was achieved. To enhance the widespread adoption of knowledge-based systems by the general public it is recommended to add more local languages to avoid language barriers.

Keywords: *Rule based reasoning, Expert system, Respiratory diseases, Data mining algorithms*

Table of Contents

Acknowledgement vi

Abstract vii

Table of Contents viii

List of table xii

List of figure xiii

Acronym xiv

CHAPTER ONE 1

INTRODUCTION 1

1.1. Background of the study 1

1.2. Statement of the problem/Rationale 3

1.3. Research Questions 6

1.4. Objectives of the Study 6

1.4.1. General objective 6

1.4.2. Specific objectives 6

1.5. Scope of the study 7

1.6. Significance of the study 8

CHAPTER TWO 9

LITERATURE REVIEW 9

2.1. Introduction 9

2.2. Artificial Intelligence 9

2.3. Overview of Knowledge based systems 10

2.4. History of knowledge based systems in respiratory disease 11

2.5. Early significant of expert system 12

2.6. The architecture of Knowledge-based System 13

2.6.1. Domain Expert and knowledge engineer 14

2.6.2. Knowledge-base 14

2.6.3. The inference engine 15
2.6.4. Learning 15
2.6.5. User Interface 16
2.7. Development of Knowledge-Based System 16
2.7.1. Knowledge Acquisition 17
2.7.2. Knowledge Modeling 18
2.7.3. Knowledge Representation 19
2.8. Rule-based reasoning Technique 19
2.8.1. Forward chaining 20
2.8.2. Backward Chaining 20
2.9. Evaluation and Testing of Knowledge-based System 20
2.10. Tools Used in Knowledge-based System Development 21
2.10.1. SWI Prolog 21
2.10.2. Java Eclipse 21
2.10.3. JCOLIBRI 22
2.10.4. JavaScript 22
2.11. Data mining 22
2.12. Data mining classification algorithm 23
2.12.1. J48 23
2.12.2. PART 23
2.12.3. JRIP 23
2.13. Related works 24
CHAPTER THREE 28
RESEARCH METHODOLOGY 28
3.1. Problem Identification 30
3.2. Define the objectives of the solution 30
3.3. Design and Development 31
3.4. Demonstration 34

3.5. Evaluation 35
CHAPTER FOUR 36
4. KNOWLEDGE ACQUISITION, MODELING, EXPERIMENTATION AND EVALUATION 36
4.1. Types of Respiratory disease frequently occurred in Ethiopia 36
4.1.1. The identified domain knowledge 36
Pneumonia 37
4.2. Diagnosis of respiratory diseases 42
4.3. Respiratory diseases Treatment 43
4.4. Knowledge Acquisition from Respiratory diseases cases (Records) 46
4.5. Knowledge source for Data Mining classification algorithm 46
4.6. Data preprocessing 48
4.7. Data cleaning 49
4.8. Attribute selection 49
4.9. Data Formatting 51
4.10. Experimentation 52
4.11. Experiment Design 53
4.11.1. Experimentation One (J48) 53
4.11.2. Experimentation Two (PART) 57
4.11.3. Experimentation Three (JRIP) 60
4.12. Comparison of classification algorithm 62
4.13. Knowledge Representation 64
4.13.1. Rule based representation 64
4.14. System implementation and evaluation 65
4.15. Prototype demonstration 65
4.16. User acceptance testing 76
4.17. System performance testing 78
4.18. Discussion of findings 82
CHAPTER FIVE 85

CONCLUSION AND RECOMMENDATIONS 85

5.1. Conclusion 85

5.2. Recommendation 86

References 87

Appendix I: Interview Questions 94

Appendix II: Letter for data collection 95

Appendix III: Sample of patient card 96

Appendix IV: Patient Information (Individual folder) 97

Appendix V: Assessment of proposed system 98

Appendix VI: Data acquisition tools 101

Appendix VII: Sample of rules 102

List of tables

Table 2.1: Summary of the related works 26
Table 4.1: Removed attribute 49
Table 4.2: Selected attributes and their description 51
Table 4.3: Sample CSV used for classification 51
Table 4.4: The generated results of 10 fold cross-validation J48 tree 53
Table 4.5: The Generated results of J48 percentage split 56
Table 4.6: The generated results of PART 10 fold cross-validation 57
Table 4.7: The generated results of PART percentage split 59
Table 4.8: The generated results of JRIP 10 fold cross validation 60
Table 4.9: The generated results of JRIP percentage split 61
Table 4.10: Comparison of classification algorithm 63
Table 4.11: Summary of user acceptance testing (Adopted from Daniel, 2021) 77
Table 4.12: Confusion matrix of prototype system 79
Table 4.13: Accuracy of prototype system 80
Table 4.14: Comparison of RDDT with the previous related works 83

List of figures

Figure 2.1 : The knowledge-based system's architecture (Akerkar, 2010)..... 14
Figure 2.2: Developments of KBS (Bethegaw, 2019) 17
Figure 3.1: Design science processing Model (Peffers, 2018) 30
Figure 3.2: Architecture of proposed system (Diriba, 2019) 32
Figure 4.1: ROC curve for 10 fold cross validation..... 55
Figure 4.2: ROC curve for J48 Percentage split (80%, 20%) 56
Figure 4.3: ROC curve for PART classifier experiment result for 10-fold cross validation ... 58
Figure 4.4: ROC curve for PART classifier experiment result..... 59
Figure 4.5: ROC curve for 10 fold cross validation JRIP classifier experiment result 61
Figure 4.6: ROC curve for JRIP classifier algorithm 62
Figure 4.7: The code snapshot for rule based representation..... 65
Figure 4.8: User interface of RDDTS 67
Figure 4.9: User name insertion to the RDDTS..... 68
Figure 4.10: User age insertion to RDDTS..... 69
Figure 4.11: User gender insertion to RDDTS 70
Figure 4.12: Query for insertion of family health history..... 71
Figure 4.13: Insertion of patient symptom to RDDTS for diagnosis..... 72
Figure 4.14: User (patient) symptom query..... 73
Figure 4.15: Diagnosis and treatment result 74
Figure 4.16: Review of user information and symptom 74
Figure 4.17: Learning part of the system 76

Acronym

AC=Average

AI =Artificial Intelligence

ALRI =Acute Lower Respiratory Infection

COPD = Chronic Obstructive Pulmonary Disease

COVID-19 = Coronavirus Disease-2019

CCI= correctly classified instances

CSA = Central Statistical Agency

DALYs = Disability-Adjusted Life-Years

EDS = Expert Diagnostic System

FPR=False Positive Rate

GUI = Graphical User Interface

HTML= Hypertext Markup Language

ILD =Interstitial Lung Disease

ICI=incorrectly classified instances

JESS = Java Expert System Shell

JUMC =Jimma University Medical Center

KBS = Knowledge-Based System

KEE = Knowledge Engineering Environment

LISP =List Processing

NPM = Node Package Manager

OSA = Obstructive Sleep Apnea

PR=precision Rate

RR= Recall Rate

SWI-prolog = Software Implementation of Prolog

TB = Tuberculosis

TPR=True Positive Rate

WHO = World Health Organization

YLD = Years Lived with Disability

YLL = Years of Life Lost

CHAPTER ONE
INTRODUCTION

1.1. Background of the study

Respiratory disease refers to pathological conditions that harm organs and tissues, impeding the process of gas exchange. These conditions encompass ailments related to the trachea, bronchi, bronchioles, alveoli, pleurae, pleural cavity, as well as the respiratory muscles and nerves. As indicated by (Alfonso, 2017), the severity of respiratory illnesses varies, ranging from mild and self-limitingissues like the common cold, influenza, and pharyngitis to severe and potentially life- threatening conditions such as bacterial pneumonia, pulmonary embolism, tuberculosis, acute asthma, lung cancer, and critical respiratory syndromes like COVID-19. The classification of respiratory disorders can be based on various factors, including the affected organ or tissue, the type and pattern of accompanying signs and symptoms, or the etiology of the disease (Alfonso,2017).

Respiratory diseases have a significant global impact on health. Asthma affects around 235 million individuals, while chronic obstructive pulmonary disease (COPD) is experienced by over 200 million people, including 65 million with moderate-to-severe COPD. Sleep disordered breathing affects 1–6% of adults, exceeding 100 million individuals. Each year, 8.7 million people contract tuberculosis (TB), and numerous individuals live with pulmonary hypertension. Additionally, over 50 million people grapple with occupational lung diseases, contributing to a cumulative toll of more than 1 billion individuals affected by chronic respiratory conditions (WHO, 2017).

Respiratory illnesses have a significant impact on public health in Ethiopia. In 2019, lower respiratory infections were the 3rd leading cause of years of life lost (YLLs) in the country, accounting for over 1 million YLLs. Chronic obstructive pulmonary disease (COPD) was the 9th leading cause of YLLs, contributing to over 370,000 YLLs. Asthma was the 18th leading cause of YLLs, responsible for nearly 180,000 YLLs. Furthermore, tuberculosis was the 6th leading cause of YLLs, resulting in over 570,000 YLLs. This data highlights the substantial burden that respiratory illnesses place on population health in Ethiopia, underscoring the need for improved prevention, diagnosis and treatment strategies supported by robust medical research and data infrastructure (Gregory, 2022).

To enhance the efficiency of medical diagnosis and treatment, it is essential to create expert systems capable of automating these processes, as suggested by Rahman (Rahman, 2020). Various knowledge-based expert systems have been designed across various medical domains. The incorporation of reasoning procedures alongside self-learning capabilities is expected to significantly boost the accuracy and reliability of these systems, as highlighted by (Shashank, 2022)

Expert systems represent a significant subset within the AI domain. In a society and industry increasingly oriented towards knowledge, reliance on the decision-making process of various experts has become crucial. Expert systems offer the ability to function as an on-demand expert, available at any time and from any location. The utilization of expert systems can lead to cost savings by tapping into expert knowledge, enabling users to operate at an elevated level and ensuring consistency. Essentially, expert systems can be viewed as productivity tools possessing the collective knowledge of multiple experts over an extended period. These computer-based systems extract and generate knowledge from data, information, and existing knowledge. They possess the capability to comprehend information in process and make decisions based on the knowledge embedded within the system. This stands in contrast to traditional computer systems, which lack the awareness or understanding of the data and information they process, as discussed by (Ajiteru, 2020).

Knowledge can be obtained through various means, including conducting interviews with subject matter experts, analyzing documents, making observations, and more. During interviews, experts may not share all of their knowledge, resulting in hidden knowledge related to the problem. The area of expertise of these experts is often vague or ambiguous, and it encompasses a significant amount of procedural knowledge. Therefore, the knowledge engineer responsible for gathering knowledge must possess expertise in the knowledge collection process (Julius, 2021).

To address the challenges posed by ambiguity and hidden knowledge during knowledge acquisition, it is crucial to employ automatic knowledge discovery methods such as data mining. Data mining is a broad technique for uncovering concealed and previously undiscovered knowledge from datasets. Various data mining methods such as classification, association, and clustering are employed to analyze large volumes of data and reveal hidden patterns and relationships that can aid in decision-making (Tariq, 2020).

Moreover, data mining plays a dual role in problem-solving, encompassing prediction and knowledge discovery. Depending on the specific problem at hand, specific methods are recommended. For prediction tasks, techniques such as classification or regression are utilized, whereas for knowledge discovery, clustering, association rules, database segmentation, sequence analysis, or visualization methods are employed (Deshpande, 2019).

As per the findings of Pandey and Mishra in 2009, Expert systems employ reasoning methods such as fuzzy logic, case-based reasoning, and rule-based reasoning to deliver notable effectiveness in the field of diagnosis. A rule-based system manages issues using a clearly defined knowledge base comprising facts and rules. It is apparent that the rule-based approach in artificial intelligence is a suitable methodology for addressing challenges in medical domains and tasks.

The impetus for this research, as outlined by Diriba in 2019, suggests the endorsement of endeavors to incorporate the SWI Prolog interface with emerging commercial tools such as Java Expert System Shell (JESS), GURU, Vidwan, and KEE for the enhancement of expert systems. As it clear that most of the previously developed expert systems on different areas use SWI-prolog and LISP languages as development tools, the user interface of SWI-prolog is not user attractive and also it requires the user who possess knowledge of prolog syntax to use from user perspective. Hence, addressing the identified gaps in previous research, this study aims to create an expert system for diagnosing and treating respiratory diseases. The development utilized the JavaScript programming language, Node.js, and Handlebars to effectively address the aforementioned gaps (Diriba, 2019).

1.2. Statement of the problem/Rationale

Respiratory illnesses, including acute respiratory infections, pneumonia, tuberculosis, asthma, and chronic obstructive pulmonary disease (COPD), are significant contributors to Ethiopia's disease burden. According to data from the Institute for Health Metrics and Evaluation, these conditions lead to millions of years of life lost annually in the country. Lower respiratory infections are the third leading cause of premature mortality in Ethiopia, while COPD, asthma, and tuberculosis also rank among the top causes of health loss. This underscores the substantial impact of respiratory diseases on the population's well-being and the urgent need to enhance prevention, early detection, and effective management strategies (WHO, 2017).

The diagnosis and treatment of respiratory diseases are complex and often require a high level of expertise. With the increasing prevalence of respiratory conditions, there is a pressing need for accurate, efficient, and accessible diagnostic tools. Traditional methods of diagnosis can be time-consuming and heavily reliant on the availability and skill of medical professionals. This can lead to delays in diagnosis and treatment, especially in regions with limited healthcare resources. The development of a self-learning expert system for the diagnosis and treatment of respiratory diseases presents a promising solution to these challenges. Such a system would leverage existing domain knowledge and advanced classification algorithms to provide accurate and timely diagnoses.

The diagnosis and treatment of respiratory diseases face significant challenges due to the complexity of symptoms, the need for specialized expertise, and variability in healthcare resources. Developing a self-learning expert system offers a promising solution but requires overcoming constraints such as integrating comprehensive medical knowledge, ensuring adaptability to different healthcare settings, maintaining accuracy and reliability, facilitating continuous learning and updates, providing a user-friendly interface, and optimizing for resource limitations. Addressing these challenges will enable the creation of an effective system that enhances diagnostic accuracy and treatment outcomes, particularly in resource-constrained environments.

Jimma University medical centre currently operates with a manual patient registration, soloed approach to patient diagnosis and treatment, where different departments handle these processes separately. The diagnosis and treatment of respiratory diseases require extensive medical knowledge and expertise. With the increasing incidence of respiratory conditions worldwide, there is a critical need for advanced diagnostic tools that can aid healthcare professionals in providing accurate and timely care. However, developing a self-learning expert system for respiratory diseases faces significant challenges due to the complexity and variability of medical data and knowledge. One of the primary challenges is identifying and integrating the available domain knowledge necessary for developing such a system. This includes understanding the pathophysiology of various respiratory diseases, recognizing symptoms and signs, interpreting diagnostic tests, and staying updated with the latest treatment protocols and guidelines

Health professionals require access to current, trustworthy health information to enhance their knowledge and deliver evidence-based care to their patients. However, many health professionals rely primarily on their existing training materials and recollections from their educational background, rather than utilizing the latest credible sources. This creates a concerning knowledge gap, limiting the ability of health professionals to provide the most up-to-date and effective health services to their clients. Closing this knowledge gap and empowering health professionals with timely, reliable health information from credible sources is crucial to develop a self-learning expert system which automatically update its knowledge in the process of respiratory diseases diagnosis and treatment.

Prior research efforts have explored the development of rule-based expert systems for the diagnosis and treatment of respiratory diseases, primarily using SWI-Prolog and LISP programming languages. Many problems and situations encountered in the real world are complex and ever-evolving, making it challenging for traditional expert systems to keep up with the latest advancements and changes.

A self-learning expert system can continuously expand its knowledge base by autonomously acquiring and integrating new information, allowing it to adapt to changing environments and tackle novel problems more effectively. Conventional expert systems often require significant human effort and resources to maintain and update their knowledge bases. This can be time-consuming and costly, especially in rapidly changing domains. A self-learning expert system can automate the knowledge acquisition process, reducing the need for manual intervention and lowering the overall maintenance and development costs.

The successful implementation of a self-learning expert system for respiratory diseases hinges on evaluating both its technical performance and user acceptance. Key focuses include assessing the system's diagnostic accuracy compared to traditional methods and expert judgment, evaluating the effectiveness of treatment recommendations, and gathering feedback on usability from healthcare professionals. Understanding user approval through surveys and interviews is crucial to ensuring the system integrates seamlessly into clinical practice. Continuous monitoring and updates are essential to maintain relevance and effectiveness in diagnosing and treating respiratory conditions, ultimately aiming to enhance healthcare outcomes and streamline medical decision-making processes.

This study investigated the formulation of a self-learning expert system for respiratory diseases diagnosis and treatment, employing JavaScript, Node.js, and Handlebar.js. The proposed system incorporated natural language support for Amharic and Afaan Oromo, in addition to English, to ensure better interaction and understanding for local users. This represents an advancement over previous expert system approaches that primarily utilized SWI-Prolog query interfaces.

1.3. Research Questions

In pursuit of this objective, the study addressed the following research queries:

- What domain knowledge is available for the development of a self-learning expert system in respiratory diseases diagnosis and treatment?
- Which classification algorithm is recommending the most accurate approach for diagnosing and treating respiratory disease?
- How to develop a self-learning expert system that can addresses constraints related to respiratory diseases?
- What is the performance and user approval of the developed self-learning expert system in the diagnosis and treatment of respiratory diseases?

1.4. Objectives of the Study

1.4.1. General objective

The general objective of this research is to develop a self-learning expert system for respiratory diseases diagnosis and treatment.

1.4.2. Specific objectives

The specific objectives of the study are as follows:

- To identify the available domain knowledge for developing a self-learning expert system in respiratory diseases diagnosis and treatment.
- To determine the most suitable classification algorithm that provides the highest accuracy.
- To develop a self-learning expert system for respiratory diseases constraints.
- To evaluate the performance and user acceptance of the developed self-learning expert system in diagnosing and treating respiratory diseases.

1.5. Scope of the study

The research is dedicated to the creation of an expert system designed for the diagnosis and treatment of respiratory diseases. As per the World Health Organization (WHO, 2017) there exist more than 100 different types of respiratory diseases globally. However, this study was specifically concentrated on prevalent respiratory conditions in Ethiopia, namely Acute Bronchitis, Asthma, Bronchiectasis, Chronic Obstructive Pulmonary Disease (COPD), Common Cold, COVID-19, Croup, Influenza, Lung cancer, Pertussis, Pneumonia, Rhino sinusitis and Tuberculosis. The proposed expert system used the above listed disease only to check their symptom and by using forward chaining mechanism if matched with user input symptom but if not matched it proceed to next disease and again check for symptom until it get similarity symptom and give treatment.

This study utilized a purposive sampling approach to select domain experts for the acquisition of knowledge. The criteria for choosing domain experts considered factors such as their professional background, educational qualifications, and years of experience in the diagnosis and treatment of respiratory diseases.

The researcher gathered domain knowledge primarily from three respiratory disease professionals such as doctors, pathologists, and residents at Jimma University Medical Center (JUMC). The primary data source was also include interview with domain experts and insights from previously resolved cases extracted from patient records related to respiratory diseases at JUMC. In addition, secondary data sources encompassed document analysis (Patient case) and from various literature forms, including books, journal articles, conference papers, and online resources.

The expert system implementation considered local languages such as Amharic and Afaan Oromo in addition to English language. In the creation of an expert system dedicated to the diagnosis and treatment of respiratory diseases, the study adheres to a process that involves knowledge acquisition, knowledge modeling, and knowledge representation. It further includes the implementation of rule-based reasoning and the evaluation of the developed expert system to assess user acceptance.

1.6. Significance of the study

Utilizing problem-solving techniques and leveraging expert knowledge relevant to a specific application, a knowledge-based system aims to address problems akin to human problem-solving. The objective of this research is to develop a self-learning expert system for respiratory diseases diagnosis and treatment to alleviate the challenge of limited access to specialists for initial diagnosis and treatment of respiratory conditions.

A significant contribution of this study is to aid individuals with respiratory diseases in obtaining professional guidance regarding their symptoms, either through direct or indirect means. This study significant for pathologists and different research. The research is beneficial for pathologists, offering enhanced accuracy through advanced algorithms and databases that analyses patient data for precise diagnostic information and more effective treatment plans. For researchers, the study provides valuable tools through advanced algorithms, enabling data analysis and pattern recognition to gain insights into the relationships between symptoms, diagnoses, and treatment outcomes. Researchers can use the system for hypothesis testing, refining scenarios to validate hypotheses and contribute to the development of new insights.

1.7. Organization of the thesis

This thesis consists of five chapters, each serving a specific purpose. Chapter 1 provides an introduction to the study, including the background, problem statement, research objectives, scope and significance of the study. Chapter 2 conducts a comprehensive literature review on knowledge-based systems, covering topics such as system architecture, knowledge acquisition, modeling and representation, evaluation and testing techniques and development tools. Relevant research studies related to the current research are also discussed in this chapter. Chapter 3 focuses on understanding the methods and processes of the system, including knowledge acquisition, modeling, and representation techniques employed in the proposed system. Chapter 4 delves into Knowledge Acquisition, Modeling, Experimentation and Evaluation including the knowledge acquired from patient case analysis and data mining algorithm such as J48, PART and JRIP. This chapter also discussed construction of rule-based reasoning (RBR) system. Finally, Chapter 5 concludes the thesis by summarizing the findings.

CHAPTER TWO
LITERATURE REVIEW

2.1. Introduction

In a general sense, a disease refers to an anomalous state or disturbance that impacts the body or mind, hindering typical physiological processes. Multiple factors, including pathogens (like bacteria, viruses, fungi, and parasites), genetic mutations, environmental influences, lifestyle decisions, and others, can contribute to the onset of diseases. These factors may disrupt the typical structure or function of an organism, giving rise to distinctive signs and symptoms. Diseases can manifest in different ways, ranging from acute conditions with sudden onset and short duration to chronic conditions that persist over a long period. Preventing, diagnosing, and treating diseases are major goals of healthcare and treatment approaches may include medications, surgery, lifestyle changes, rehabilitation and other interventions, depending on the nature and severity of the disease (William, 2023).

2.2. Artificial Intelligence

Artificial Intelligence (AI) involves the design of systems with the capability to imitate human intelligence. This discipline focuses on crafting computers that reproduce various elements of human cognitive functions. Key characteristics of an intelligent system encompass the ability to carry out actions, engage in reasoning within a specified domain, make decisions, and address problem-solving with a goal-oriented outlook. The effectiveness of an intelligent system or agent is measured by its performance, which should be indistinguishable from that of a human performing identical task (Ziyad, 2019).

Developing an intelligent system inherently presents challenges in terms of acquiring, representing, and inferring knowledge, mainly due to the diversity of individual perspectives. Although experts can communicate essential knowledge to a knowledge engineer, the process of structuring this knowledge for inclusion in the knowledge base is a substantial challenge. The complexity escalates when dealing with the size of the knowledge base, as converting thisinformation into a computer-readable format proves to be a non-trivial undertaking (Lark, 2023)

2.3. Overview of Knowledge based systems

The Knowledge-Based System (KBS) is a significant component within the artificial intelligence (AI) field. As advanced computing resources and other capabilities become more accessible, attention is increasingly directed towards tackling complex tasks that demand a higher level of intelligence. Expert decision-making is becoming increasingly important in both industry and society as they become more knowledge-driven. Numerous fields, including robotics, natural language processing, and medical diagnosis, use knowledge-based systems. They are capable of carrying out operations like analyzing massive volumes of data and making intricate choices that are challenging or impossible for humans to complete. By harnessing expertise, empowering users to operate at an elevated capacity, promoting consistency, and potentially reducing expenses (Sajja & Akerkar, 2018). A knowledge base constitutes the assembly of pertinent knowledge stored in a computer, organized to facilitate inferences. These inferences correspond to the reasoning process within Artificial Intelligence, analogous to the cognitive function in the brain of an Artificial Intelligence system.

KBS is a computer-based system that makes use of data, information, and knowledge to create and produce knowledge. Unlike typical computer systems, which are unable to comprehend the data or information they process, these systems are able to comprehend the information that is being processed and make decisions based on the information that is stored within the system (Ajiteru, 2020).Various scholars and authors provide diverse explanations when referencing knowledge-based systems. From these definitions, one perspective comes from that a knowledge-based system is employed for the organization and storage of intricate structured and unstructured knowledge. This organization allows a computer system to assist in decision-making processes.

The term was originally associated with expert systems, marking their inception as the earliest knowledge-based systems. According to (Mumine, 2017), a knowledge-based system is recognized as a significant branch of artificial intelligence, possessing the ability to make decisions grounded in the knowledge embedded within them, along with an understanding of the context of processed data.

Another scholar, (Hagemeister, 2019) defines a knowledge-based system as software endeavoring toreplicate the expertise of one or more human experts, typically within a specific problem domain. This concept represents a traditional application and subfield within the realm of artificial intelligence.

2.4. History of knowledge based systems in respiratory disease

Subsequently, a brief exploration of the history of Knowledge-Based Systems (KBS) will be presented. This discussion will delve into its historical significance within the realm of Artificial Intelligence, shedding light on some early and noteworthy milestones in the development of expert systems.

Expansion of Knowledge Bases and Rule-Based Systems (1980s): Throughout this timeframe, knowledge-based systems dedicated to the diagnosis and treatment of respiratory diseases witnessed an expansion of their knowledge bases. Domain experts and clinicians contributed their expertise to create extensive rule sets that covered a wide range of respiratory conditions. These systems leveraged the accumulated knowledge to provide more accurate diagnoses and treatment recommendations (Rahman, 2020).

In the early 1990, as electronic medical records became more prevalent, knowledge-based systems started integrating with these systems. This allowed them to access patient data, such as medical history, laboratory results, and imaging reports. By incorporating patient-specific information, knowledge-based systems could provide more personalized and context-aware recommendations.

Introduction of Probabilistic Reasoning (1990s-2000s): In order to enhance the management of uncertainty and probabilistic information, knowledge-based systems focused on the diagnosis and treatment of respiratory diseases started integrating techniques for probabilistic reasoning. Bayesian networks and other probabilistic models were employed to assess the likelihood of different diagnoses based on available evidence, improving diagnostic accuracy (Rahman, 2020).

Machine Learning and Data-Driven Approaches (2000s-2010s): With the rise of machine learning techniques, knowledge-based systems started incorporating data-driven approaches to enhance their capabilities.

These systems began to leverage large datasets of patient records, medical imaging, and clinical research to learn patterns and improve diagnostic accuracy. Data analysis and decision-making in the diagnosis and treatment processes were aided by the utilization of machine learning algorithms, including decision trees, support vector machines, and neural networks (Sima, 2022).

Integration of Natural Language Processing (NLP) and Text Mining (2010s-2020s): As the volume of medical literature increased, knowledge-based systems began incorporating natural language processing and text mining techniques in the area of respiratory disease diagnosis and treatment. These systems could analyze unstructured clinical text, such as medical notes, research articles, and clinical guidelines, to extract relevant information and Keep abreast of the most recent research findings. NLP also enabled systems to interpret patient symptoms and medical histories documented in free-text format (Sima, 2022).

Emerging Technologies and Future Directions: The advancement of knowledge-based systems for diagnosing and treating respiratory diseases is an ongoing and dynamic process. Recent advancements in technologies like deep learning, computer vision, and the integration of multimodal data, where clinical and imaging data are combined, present novel possibilities to enhance both accuracy and efficiency. Additionally, advancements in personalized medicine and genomics may lead to the integration of genetic and molecular data into knowledge-based systems for more precise and personalized respiratory disease management.

2.5. Early significant of expert system

The origins of respiratory disease expert systems can be traced to the initial phases of artificialintelligence (AI) research during the 1960s and 1970s. The objective was to develop computer programs capable of replicating the decision-making abilities of human experts within particular domains, including the field of medicine. Early expert systems were rule-based systems, which relied on a set of predefined rules and logical statements to make inferences and provide recommendations. These systems were typically developed using knowledge elicitation techniques, where subject matter experts would be interviewed and their expertise captured in the form of rules.

A prominent application in the history of expert or knowledge-based systems, MYCIN, holds significance for two key reasons: firstly, it distinguishes itself from DENDRAL, which is well-known in this domain, which derived its reasoning from a model of a specific molecule, MYCIN was built through interviews with different physicians in the relevant field; consequently, it includes several heuristic rules that clinicians use to identify specific infections. Secondly, MYCIN was instrumental in the later development of EMYCIN (Empty MYCIN).Pomykalski *et al.*, 1999).

Early on in the development of expert systems, additional important applications were also created. Among these systems is pulmonary function analysis (PUFF), which in the context of pulmonary illnesses uses EMYCIN. The knowledge is represented by IF-THEN rules. In order to obtain clinical data, it also communicates with a doctor.

The system formulates questions based on accepted knowledge and current hypotheses, then works backwards to prescribe a course of treatment for a given patient. It can identify roughly 100 bacterial illness causes and stores about 500 IF-THEN rules (Fetene, 2017).

2.6. **The architecture of Knowledge-based System**

Architecture serves as a plan or blueprint that outlines the structure of a system. The interfaces between a system's internal components, or subsystems, and its external environment, particularly the user, are the main concerns of a system architecture (Valderde, Roberto, 2019). Depicting a collection of conventions, guidelines, instruments, and standards that ought to be included in the relevant systems is also helpful.

A typical knowledge-based system (KBS) comprises the following components: an inference engine, a user interface, an explanation or reasoning facility, and self-learning. The various elements work together to simulate the act of addressing and resolving problems and making decisions (Akerkar, 2020).

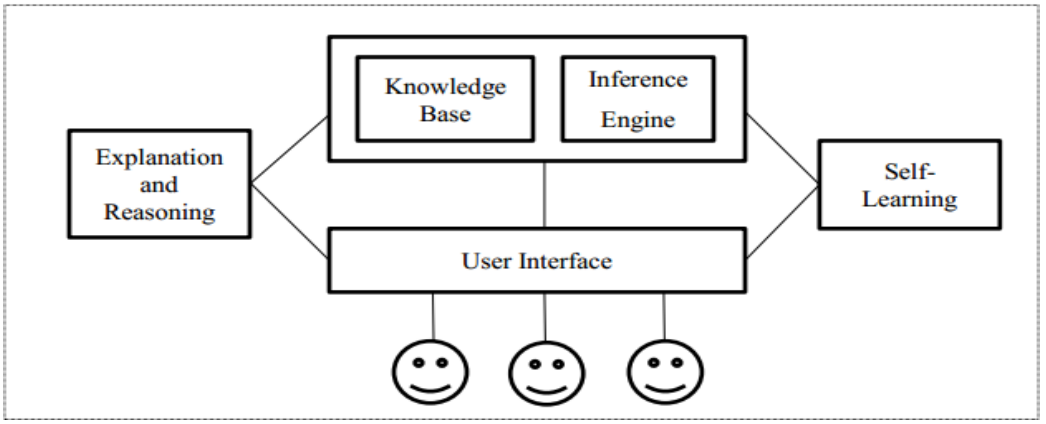


Figure 2.1 : The knowledge-based system's architecture (Akerkar, 2010)

2.6.1. Domain Expert and knowledge engineer

The domain expert, considered the most valuable member of the Knowledge-Based System (KBS) development team, is a proficient individual with extensive knowledge and skills in a specific domain. As their expertise is incorporated into the KBS, it is crucial for the expert to effectively communicate their knowledge, actively engage in the development of the KBS, and allocate a substantial amount of time to the program's advancement (Poole, 2019).

Knowledge engineers are competent of creating, constructing, and testing KBS. To acquire the expertise of the expert, the knowledge engineer conducts an interview and subsequently encodes the gathered information into the knowledge base. The knowledge engineer inquires about the expert's approach to solving specific problems, identifies the reasoning strategies employed for handling facts and regulations, and determines the appropriate representation for incorporation into the Knowledge-Based System (KBS). The knowledge engineer then chooses the development software or KBS shell and assesses programming languages for encoding the acquired knowledge (Bethegaw, 2019).

2.6.2. Knowledge-base

The core repository of information for a specific subject, known as the knowledge base, stands as the essential component in any knowledge-based system or expert system. It includes information obtained from the subject matter expert that qualifies people as precise experts in a given field.

In every knowledge base, there is an abundance of facts, rules, and meta-knowledge. Using various knowledge representation approaches, this knowledge is encoded in a knowledge-based system after being extracted from human specialists. Rule-base and case-base systems are the primary methods for representing knowledge in knowledge-based systems. The expertise of multiple human experts can be consolidated within a system based on rules.

2.6.3. The inference engine

The inference engine serves as the component that offers a methodology for reasoning and drawing conclusions. It guides the utilization of the system's knowledge to address problems, as mentioned by (Margaret, 2017) The primary responsibility of the inference engine is to explore the knowledge base for connections and relevant information, then use those connections to generate recommendations, answers, and predictions just like a human expert would. The right data, interpretations, and rules must be located and correctly assembled by the inference engine. Rule-based systems employ two primary types of inference engines: forward chaining and backward chaining mechanisms. By linking conclusions to their original sources and demonstrating how a particular result was arrived at, the explanation subsystem assists in providing justifications for the conclusions reached by the knowledge-based system.

The explanation facility stands as a crucial component within a knowledge-based system. It can explain how it comes to its findings with this skill. Questions on the what, how, and why of an issue might be posed by the user. The knowledge-based system will furnish the user with a trace of the consultation process, highlighting the primary reasoning paths taken during the interaction. The explanation facility aids the knowledge-based system in elucidating and substantiating the necessity for certain deviations. This feature enables users to comprehend the rationale behind the results produced by the knowledge-based system (Margaret, 2017) .

2.6.4. Learning

Learning is a component of Knowledge-Based Systems (KBS) that strives to emulate human learning. The KBS's knowledge base can be updated manually, semi-automatically, or automatically through the use of machine learning algorithms (Neri, 2022).

As per (Neri, 2022) ,a scientific job known as "self-learning" allows the KBS to automatically learn from its surroundings, instances that are handled, and inference process.

According to (Castillo, 2022) there are three learning methods: structural learning, parametric learning, and learning by memorization. Structural learning involves features related to the structure of knowledge, such as rules and probability distributions. This can include discovering new symptoms related to a particular disease or incorporating additional rules into the knowledge base. Parametric learning involves estimating the parameters needed for constructing the knowledge base, such as predicting probabilities associated with symptoms or diseases. Learning by memorization refers to the knowledge-based system's ability to learn from experience based on existing data. Through this method, the system can perform activities like storing or memorizing knowledge and learning from the facts base.

2.6.5. User Interface

The user interface stands out as a pivotal component within a Knowledge-Based System (KBS), facilitating interactive communication between the user and the system. Considered vital to the efficiency of an expert system, the user interface encompasses all the screens through which the user engages with the KBS. Its objective is to enhance the ease of use for developers, users, and administrators involved with the KBS (Diriba, Developing Self-Learning Knowledge Based System for chickens disease diagnosis , 2023)

The input/output interface delineates how the Knowledge-Based System (KBS) interacts with both users and other systems. Predominantly graphical, these interfaces include elements like screen displays, windowing, and mouse control. They receive user input and display corresponding output. Although some systems incorporate natural language interfaces accepting English-like responses, most opt for a graphical user interface (GUI) with a mouse device. This configuration allows users to choose from selections presented in dialog boxes and menu bars (Shahd, 2023)

2.7. Development of Knowledge-Based System

Transferring knowledge and experience from various specialists and codified resources to a computer program is the primary goal of building a knowledge-based system. Knowledge engineers, or KBS developers, have the responsibility of transforming the system to enable it to function at the intended level.

Identifying the precise requirements, representing knowledge, constructing the knowledge base, modeling knowledge, verifying knowledge, and validating knowledge are the stages within the Knowledge-Based System (KBS) development process (Dirriba, 2019).

Figure 2.2: Developments of KBS (Bethegaw, 2019)

2.7.1. Knowledge Acquisition

Gaining knowledge serves as the primary and pivotal phase in the development of Knowledge-Based Systems (KBS). The expert system development life cycle underscores the importance of knowledge acquisition. Knowledge acquisition involves the procedures of obtaining, organizing, and formalizing knowledge from a source to establish a KBS. This undertaking poses challenges and potential errors for knowledge engineers, necessitating the utilization of techniques such as surveys, record reviews, interviews, observation, and document analysis to gather factual and explicit knowledge (George, 2018).

Interview

The primary approach to acquiring knowledge from subject matter experts involves interviews, which may take the form of unstructured, semi-structured, or structured interviews, as indicated by (Ali, 2022) structured interviews are systematic and goal-oriented, and promoting organized communication between the expert and the knowledge engineer. Semi-structured interviews incorporate a mix of closed-ended and open-ended questions, allowing flexibility to modify and elaborate based on the participant's responses(Ali, 2022).

Questionnaire

A survey instrument, known as a questionnaire, is an economical means of conducting research, involving a series of questions designed to collect information from respondents. Questionnaires offer advantages such as cost efficiency, reduced effort compared to verbal or telephone surveys, and standardized answers for simplified data collection. However, the limitation of necessitating respondents to read and respond constrains the widespread application of questionnaires (Ali, 2022).

Observation Acquiring knowledge encompasses direct observation by knowledge engineers or developers of Knowledge-Based Systems (KBS) while domain experts execute tasks. Methods of observation include on-site observation, where people, objects, and information are identified and recorded; discourse analysis, involving the study of real language use in actual situations; and active participation, supporting an individual's ability to independently engage in daily life activities (Weston, 2021).

Document Analysis Involves collecting information from existing documentation. Collaborating with a human expert during document analysis can aid in verifying or elaborating on pre-existing knowledge or information. Documents such as professional literature, pamphlets, instructions, employee handbooks, reports, glossaries, course texts, and other relevant materials are considered during this process (Hani, 2022).

2.7.2. Knowledge Modeling

Knowledge modeling is the methodical depiction of acquired information utilizing symbols to represent knowledge components and their interconnections, as outlined by (Wei, 2021). Symbolic character-based languages such as logic, diagrammatic representations like ladders and networks, tabular representations like matrices, and structured text like hypertext serve as examples of knowledge models. This model efficiently conveys information for validation and adjustment when necessary, ensuring that all stakeholders grasp the language and terminology in use. Knowledge models prove highly beneficial when cross-validating with other experts (Wei, 2021)

2.7.3. Knowledge Representation

The procedure of translating human information into a format comprehensible by computers is referred to as knowledge representation. This process, emphasized by (Sayantini, 2024), revolves around establishing and employing knowledge storage systems. The objective of knowledge representation is to organize acquired and modeled knowledge into a format readily usable by a Knowledge-Based System (KBS) for decision-making. It serves as a means to encode the expertise of domain experts in a suitable format. Common methodologies for knowledge representation include semantic networks, rule-based reasoning, frame-based reasoning, and case-based reasoning (Sayantini, 2024).

2.8. Rule-based reasoning Technique

Rule-based reasoning stands out as the most prevalent and favored method for knowledge representation in Knowledge-Based Systems (KBS). Rule- based representation techniques have maintained prominence, particularly in the development of medical knowledge-based systems. This approach involves expressing knowledge through aset of facts and regulations. Symbolic rules are among the widely adopted techniques for reasoning and knowledge representation. The fundamental form of a rule follows the structure IF <condition> THEN <conclusion>, where <condition> represents premises and <conclusion> signifies the corresponding action for the given premises. Logical connectives such as AND, OR, NOT, and others are employed to link the conditions of the rules. A standard rule-based system comprises a list of rules, a set of facts, and an interpreter, as outlined by (Wasyihun, 2023).

The general forms of rules can be illustrated as:

IF condition 1,
and condition 2,
and condition 3, ..., condition nTHEN action 1,
action 2, ..., action n

Types of rule-based reasoning

.

2.8.1. Forward chaining

During forward chaining, the inference engines initially establish the criterion and then incrementally incorporate the criteria one by one until the entire chain is processed. Facts within the system are depicted in a working memory that undergoes continuous updates under data-driven control. The rules of the system suggest potential courses of action to be taken when specific conditions persist with objects in the working memory. Typically, the conditions are patterns that must align with the objects in the working memory. Actions in forward chaining often involve adding or removing objects from the working memory. The interpreter of the inference engine oversees the application of rules based on the working memory. The system first checks to identify all the rules whose conditions are satisfied (Nalepa, 2019).

2.8.2. Backward Chaining

This approach focuses on its efforts by considering only those rules pertinent to the specific objective. Its distinction from forward chaining lies in its attempt to identify the premises leading to the conclusion after receiving the problem description as a set of conclusions rather than conditions. Upon receiving a goal state, the system endeavors to verify if the objective aligns with the initial facts. The goal is achieved upon finding a match. However, if no match is found, the inference engine proceeds to examine subsequent rules whose conclusions—formerly termed actions—align with the provided fact. Goal-driven control is commonly known as top-down or backward chaining, as highlighted by (Nalepa, 2019). Therefore, in this study, the forward chaining inference mechanism employed.

2.9. Evaluation and Testing of Knowledge-based System

A comprehensive definition of evaluation involves the systematic process of determining the value of a system through assessment and effort. Similarly, Anumba and Scott characterize the evaluation of knowledge-based systems as "an examination of the knowledge-based system's ability to address real-world issues within a specific domain." The evaluation of a knowledge-based system serves to demonstrate whether or not it successfully accomplishes its stated goals. The prototype system will undergo testing and evaluation to fulfil the specified objectives of the study. This assessment will encompass both the system's performance and the challenges related to user acceptance. The accuracy of the prototype system gauged using performance metrics such as precision, recall, and F-measure. Additionally, user acceptability testing conducted to evaluate the system's guidance level and its alignment with the requirements of subject matter experts.

2.10. Tools Used in Knowledge-based System Development

Knowledge-based systems are typically implemented using specialized programming languages or tools, with shells being the most commonly employed tools for their development. A shell is a software component that furnishes a development framework, encompassing the user interface, a structure for declarative knowledge in the knowledge-base, and the inference engine.

In the development of knowledge-based systems, various tools are utilized, including both general-purpose programming languages like Java and frameworks such as .NET, as well as specific-purpose pre-built utilities for self-learning, explanation, inference, etc. Examples of such tools include Expert System Shell (JESS), CLIPS, Vidwan, Prolog, Recall, and JCOLIBRI (Burnashev, 2019) .However, in this research, the development of the knowledge base and user interface for respiratory disease diagnosis and treatment leveraged the JavaScript programming language, rools (a small rule engine for Node.js), and Handlebars.js (a semantic template language for JavaScript)

Some of these tools are discussed as follows:

2.10.1. SWI Prolog

SWI-Prolog stands as a comprehensive and portable implementation of the Prolog programming language, aspiring to be a robust and scalable solution accommodating a diverse array of applications. Moreover, it features a compatibility framework, developed collaboratively with YAP and specifically instantiated for YAP, SICStus, and IF/Prolog. (Winstanley, 20120).

2.10.2. Java Eclipse

Java Eclipse serves as an integrated development environment (IDE) utilized in computer programming. Comprising a foundational workspace and an adaptable plug-in system for environment customization, Eclipse is predominantly crafted in Java. Its principal application lies in the development of Java applications, yet it possesses the versatility to support the creation of applications in alternative programming languages through the incorporation of plugins.

2.10.3. JCOLIBRI

JCOLIBRI, an acronym for Java Constraint Library for Biologically Inspired Optimization and Research in Informatics, originates from the collaborative efforts of the Research Group on Artificial Intelligence and the Department of Computer Science at the University of Valencia, Spain. Serving as a Java library, JCOLIBRI establishes a framework specifically tailored for Constraint-Based Local Search (CBLS) algorithms. Its primary application lies in tackling combinatorial optimization problems through the utilization of constraint programming techniques. Furthermore, JCOLIBRI represents a technological advancement over COLIBRI and functions as an object-oriented Java framework designed for constructing Case-Based Reasoning (CBR) systems.

2.10.4. JavaScript

JavaScript is a powerful computer programming language primarily used for front-end web development. It works hand in hand with HTML and CSS, forming the trifecta that powers most websites on the internet. Unlike other programming languages that require compilation before execution, JavaScript is an interpreted language. You can execute JavaScript code directly within the web browser without additional compilation steps. This unique feature enables developers to make real-time changes to web pages and instantly see the results without the hassle of a lengthy development cycle. Additionally, Node.js supports full-stack JavaScript development, enabling seamless communication between server-side and client-side components. It offers popular web application frameworks like Express.js, Koa.js, and Hapi.js, streamlining the development of efficient and modular web applications (Bluehost, 2018).

2.11. Data mining

Data mining leverages an amalgamation of explicit knowledge bases, advanced analytical abilities, and domain expertise to unveil concealed patterns and trends. These patterns and trends serve as the foundation for predictive models that empower analysts to generate fresh insights from existing data. The process entails sifting through extensive repositories of data and employing technologies such as pattern recognition, alongside statistical and mathematical techniques, to uncover significant correlations, patterns, and trends (Liu, 2021).

2.12. Data mining classification algorithm

Classification is the process of developing a function or model that can assign objects to different classes based on their attributes. To accomplish this, a training set is used, which includes a collection of objects represented by attribute vectors along with their respective classes. By analyzing the relationship between the attributes and classes in the training set, a classification function or model is constructed. This function or model can then be applied to classify new objects and provide a deeper understanding of the object classes within the database (Safae, 2018).

2.12.1. J48

J48, also referred to as C4.5, is a classification algorithm that utilizes decision trees to perform tasks like classification and prediction. Decision trees offer a straightforward but effective means of representing knowledge. When constructing a decision tree, certain branches may be affected by irregularities in the training data, which can arise from noise or outliers. To address this concern, tree pruning techniques employ statistical measures to eliminate the least dependable branches. By taking attribute values and case descriptions as input, a classifier system built on J48 can generate a classifier capable of accurately predicting the classes of new cases (Aakanksha, 2021).

2.12.2. PART

A rule-based classifier utilizes a set of IF-THEN rules to carry out classification tasks. Each rule comprises an expression in the IF condition THEN conclusion format. The IF-part, also known as the rule antecedent or precondition, represents the condition that must be met for the rule to be applied. The THEN-part, referred to as the rule consequent, defines the resulting classification. This classifier has the ability to generate a decision list and adopts a separate-and-conquer methodology (Aakanksha, 2021).

2.12.3. JRIP

JRIP is another rule-based classifier that employs a set of IF-THEN rules to perform classification tasks. For this study, the default parameters from the WEKA software will be used, and the algorithm will generate a model composed of rules. This model will then be utilized to identify instances that were correctly classified as well as instances that were incorrectly classified (Aakanksha, 2021).

2.13. Related works

Various scholars have endeavored to employ rule-based and case-based reasoning methodologies in the domain of medical diagnosis and treatment, specifically concerning respiratory diseases. The researcher examines several studies relevant to knowledge-based systems and rule-based reasoning.

(Pachpande, 2020), Introduced "Expert System for Diagnosis of Pulmonary Disorders". In the development of the expert system, Prolog served as the primary tool, exclusively utilized for diagnosis.

(Aseffa, 2020), introduced a study titled "Accurate Tuberculosis Diagnosis through a Hybrid Expert System Utilizing Fuzzy Logic and Rule-based Approaches." The design methodology for the expert system involved the use of Python, a high-level, object-oriented programming language with an interpreter.

Razuan and Anderson (2016) introduced a study titled "Evaluation of Clinical Broncho-pneumonia Suspicion Using a Belief Rule-Based Expert System." The expert system created in the research employed Visual Basic 6.0 for the system interface development. The knowledge engineering process involved incorporating data from domain knowledge, and the expert constructed the knowledge base using belief rule-based constructs (IF…THEN). Notably, this research stands out by utilizing Node.js and Handlebars for the system interface development. The outcomes produced by the Belief Rule-Based Expert System (BRBES) demonstrated greater reliability compared to conventional expert opinions (Anderson, 2016).

Almir (2018) introduced a study titled "Automated Identification of Asthma and Chronic Obstructive Pulmonary Disease in Clinical Environments Using an Expert Diagnostic System." The methodology employed in this research involved the utilization of a feed forward neural network for the classification of Chronic Obstructive Pulmonary Disease (COPD) and asthma. The scope of the system was limited to the identification of Asthma and Chronic Obstructive Pulmonary Disease (COPD), excluding other prevalent respiratory diseases from consideration. In contrast, this particular study distinguishes itself from Almir (2018) by focusing on the diagnosis and treatment of a broader spectrum of sixty (16) common respiratory diseases (Almir, 2018).

Riskiawan (2019) undertook a study titled "Development of an Expert System for Lung Diseases Using Forward Chaining." The author applied a rule-based methodology to depict facts and concentrated solely on expanding the coverage to include six specific types of lung diseases, encompassing 19 symptoms. Their system operated exclusively in the local language, but in the current investigation, 16 common respiratory disease types took into account. According to the experiments and validations conducted by experts, the implementation of the expert system for pulmonary diseases exhibited successful performance, featuring an easily comprehensible user interface (Riskiawan, 2019).

(Zarandi, 2010) utilized the Fuzzy Rule-Based Expert System (FRES) to evaluate the probability of asthma during its initial phases. Additionally, the researchers modified the FRES methodology to specifically target the prediction of fatal asthma by assessing the degree of asthma instability. The results demonstrated a sensitivity of 94% and specificity of 100%. However, the researchers acknowledged the need for additional data to validate the effectiveness of their model (Zarandi, 2010).

Evren and Sabriye (2016) introduced a research project entitled "Application of an Expert System in Medicine for Pulmonary Diseases." CLIPS served as the programming environment for the development, utilizing a binary decision tree for respiratory diseases in constructing the system database, encompassing twenty-eight diseases for the inference engine. The accuracy rates for the diagnosis were calculated by dividing the total number of accurate diagnoses by the total number of patients, resulting in rates of 64.02%, 71.43%, 82.54%, and 96.83%, respectively (Sabriye, 2016).

Lamesgin (2016) introduced a study titled "Deployment of an Expert System for Diagnosing Respiratory Diseases." The system proposed was rule-based, utilizing Visual Prolog and SWI-Prolog (Multi-threaded, 64 bits, Version 7.3.14) as the chosen methodology. The author suggested enhancing user interaction by integrating the system with more user-friendly front-end applications, such as Java or C#. In the present research, JavaScript is employed as the front-end language, and Node.js is utilized for back-end communication and system development(Lamesgin,2016).

Table 2.1: Summary of the related works

Author	Methodology used	Findings	Gap
1. Pachpande(2017)	-Rule based methodology -Used SWI-prolog for development	-According to assessments and validations conducted by specialists, the performance of the expert system for pulmonary disease has been successful, demonstrating satisfactory results and featuring a user- friendly interface that is easy to comprehend.	- It worked only for local Languages. - It has low diagnostic accuracy
2. Aseffa(2020)	-Fuzzy and Rule based methodology -The methodology used to design the expert system was Python.	- The assessment of the hybrid expert system, incorporating fuzzy and rule-based approaches, resulted in an accuracy rate of satisfactory.	- The developed expert system only consider limited number of Respiratory disease. -It is not self –learning system
3.Razuan&Anderson (2016)	-Rule based methodology - The developed expert system used Visual Basic	- The outcomes produced by the system surpassed those derived from expert Opinions.	- The author used Visual basic As development tool which has No attractive interface. -Didn't consider uncertainty Issues for symptom.

	6.0 for the system interface.		
4.Almir(2018)	-Rule based methodology - Feed forward neural network was used for the classification of diseases.	-The expert diagnostic system can precisely identify individuals with asthma and COPD, with a sensitivity of 96.45% and specificity of 98.71%.	-The developed expert system has no treatment for disease only focus on diagnosis part
5.Riskiawan(2019)	-Rule based methodology -SWI-prolog as development tools	-This expert system illustrates that constructing an expert system is not a one- time process; rather, it follows anincremental approach.	-. Unfortunately, thereare no available experimental Results for this expert system.
6.Zarandi(2010)	-Fuzzy and Rule based methodology	- The findings indicated a sensitivity of 94% and a specificity of 100%.	- The developed expert system allows user to respond as only yes/no which is not sufficient to diagnose disease Accurately.
7.Evren&Sabriye(2016)	-Rule based methodology - The Methods used for development was CLIPS as a programming Environment.	-The system achieved 64.02% accuracy.	-It has limited accuracy since It has no self-learning capacity
8.Lamesgin(2016)	-Rule based methodology - The development tool was visual prolog, SWI-Prolog (Multi-threaded, 64 Bits, Version 7.3.14).	-The developed expert system proficiently identified diseases associated with the lungs. The author proposes delving into further literature covering diverse disease types and medical expert systems, integrating their symptoms to improve the system's versatility.	-It has no system performance Metrics which makes the developed System accuracy questionable. -It is not integrated with Data mining Algorithms for better performance Of the system.

CHAPTER THREE
RESEARCH METHODOLOGY

The methodology of a study serves as a roadmap, guiding researchers through the process of conducting their investigation and ensuring the reliability and validity of their findings. It entails the systematic use of approaches, methods, and tools to collect, analyze, and interpret data in order to address research questions or hypotheses. The methodology acts as the comprehensive research strategy, outlining the conduct of the research and specifying the methods to be utilized, among other components (Jilcha, 2019). The study employed design science research methods for developing a self-learning expert system for respiratory diseases diagnosis and treatment. Design science research focuses on the development and evaluation of innovative artifacts, such as software systems, models, frameworks, or methodologies, to address practical problems and improve existing practices. It combines elements of both design and scientific research to create and test these artifacts. Learning is an integral aspect of Knowledge-Based Systems (KBS) aimed at mimicking human learning. The knowledge base of a KBS can be modified manually, semi-automatically, or automatically by employing machine learning algorithms (Neri, 2022).

Research Design

A research design refers to the organized plan, structure, and strategy of an inquiry designed to yield solutions to research questions or problems. In this investigation, a design science research approach was followed to create a Knowledge-Based System (KBS) for diagnosing and treating respiratory diseases. As per Peffers and Hevner (2018), Design science research methods promote an iterative and prototyping approach to system development. This aligns well with the nature of expert system development, which often involves multiple iterations and refinements. Furthermore, this study has opted for the Design Science Research (DSR) approach due to the following justifications. DSR is primarily focused on addressing practical problems and involves the generation of prescriptive or decision making. It allows for the integration of scientific insights, such as predicting, understanding, or explaining phenomena, into the process of designing solutions for complex and pertinent issues within a specific field.

In selecting the Peffers et al. Design Science Research (DSR) process model over others, such as Hevner et al. or March et al., a logical rationale can be derived from its comprehensive approach tailored specifically for information systems research. Peffers et al. emphasize a structured methodology that includes problem identification, objectives clarification, design and development, demonstration, evaluation, and communication of the artifact's utility. This model aligns closely with the iterative nature of developing practical solutions in information systems, focusing on creating and evaluating artifacts that address specific problems or challenges within a defined domain. Its systematic approach ensures a rigorous and methodical progression from problem understanding to solution implementation, making it particularly suitable for guiding research aimed at developing innovative technologies or systems like a self-learning expert system for respiratory diseases.

Process iteration

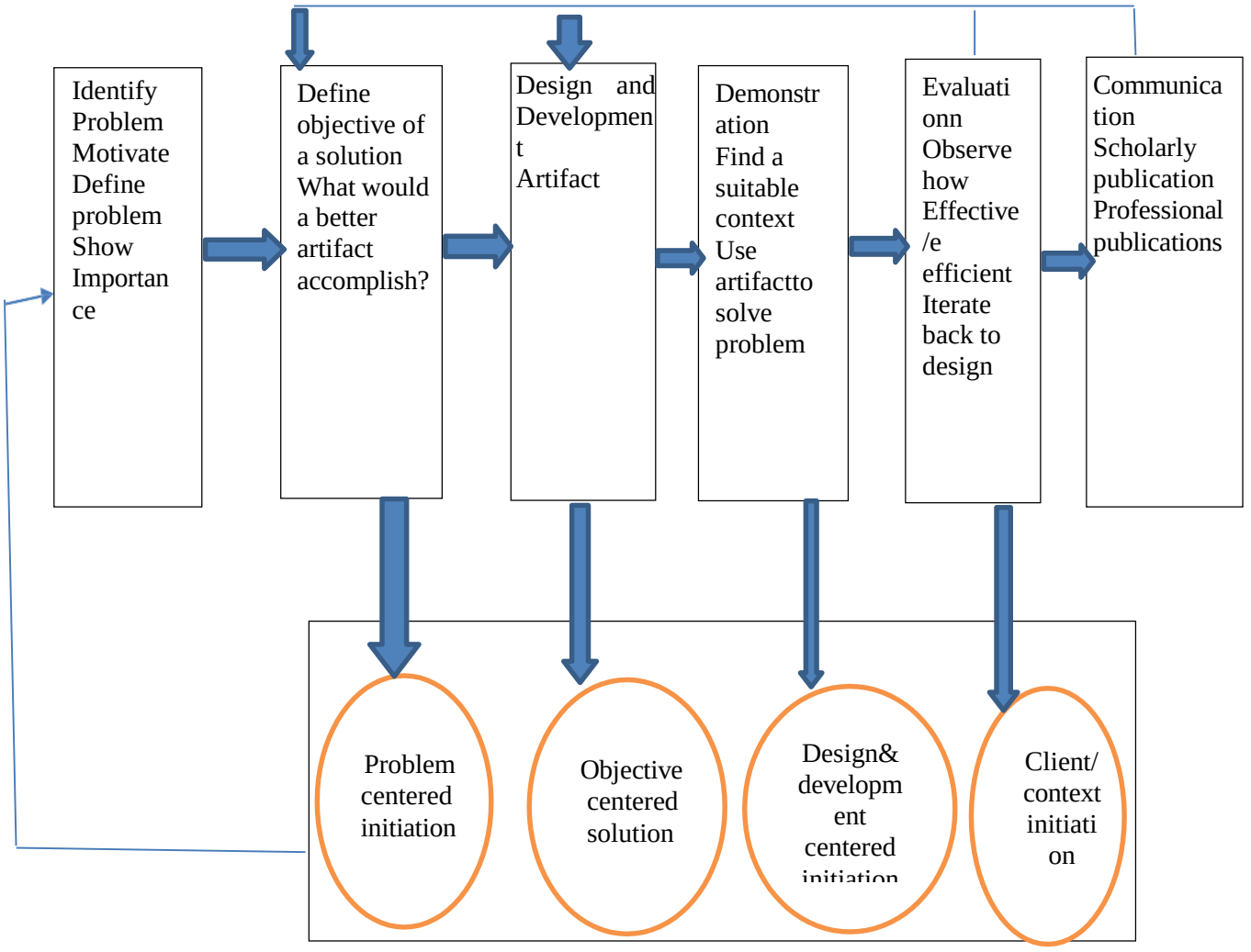


Figure 3.1: Design science processing Model (Peffers, 2018)

3.1. Problem Identification

This research employed various techniques to identify the study's problem and achieve its objectives. The researcher reviewed a range of scholarly sources, including journal articles, books, conference papers, reports, and other academic materials. After reviewing these materials, the researcher found that according to the World Health Organization's 2020 report, respiratory diseases affect over 200 million individuals globally (WHO, 2020).

Study Area

The research was focus on Jimma University Medical Center (JUMC), located in the Oromia regional state of Ethiopia, specifically in the city of Jimma. JUMC is chosen as the study area due to its status as one of the oldest public hospitals in Ethiopia, providing medical services to individuals from the western and southern regions of the country. Consequently, it possesses a substantial amount of documented information compared to other hospitals, making it a valuable source of knowledge.

Moreover, the presence of experienced experts further enhances its suitability for the study. The study's population were consist of respiratory disease specialists, resident doctors, and pathologists directly involved in the pathology field at Jimma University Medical Center. Respiratory infections spread quickly in healthcare settings if proper infection control measures are not in place. Factors like overcrowding, poor ventilation, and high patient turnover contribute to the rapid transmission of viruses and bacteria at the center.

3.2. Define the objectives of the solution

After obtaining the essential knowledge from domain experts, documents and previously addressed cases, the next stage entails defining and outlining the solution objectives. In this study, the researchers used the solution objectives for the identified problems. The first approach involves modelling the accumulated knowledge using a knowledge modelling technique to clarify the knowledge structure. Following that, the knowledge conveyed using appropriate knowledge representation techniques. For this research, the researcher used rule based representation methods scripted in JavaScript.

3.3. Design and Development

The design and development of prototype system involve knowledge acquisition and representation. In this study data mining task and algorithm applied for generating representative rule from the prepared data to design the rule based reasoning prototype for diagnosis and treatment of respiratory diseases.First, In order to develop a self-learning expert system for respiratory diseases diagnosis and treatment, the researcher gathered both tacit and explicit knowledge sources.

These sources included domain experts at Jimma University Medical Centre, medical literature, research papers, clinical guidelines, and patient cards. The acquired knowledge was then organized, and the most important attributes were selected, while unnecessary attributes were disregarded through an analysis of patient cards. Once the necessary attributes and their instances were identified, the researcher transferred the information to an Excel sheet and prepared it for data mining algorithms by saving the file as a CSV. Subsequently, the researcher utilized the data mining software WEKA 3.9 version, where the CSV file was uploaded.

The researcher then performed pre-processing and classification tasks using algorithms such as **J48, PART,** and **JRIP.** These algorithms were tested using a 10-fold cross-validation method and an 80/20% split of the data automatically by using WEKA software. Finally after comparing the result of classification algorithm the generated results of PART rules represented in a structured format suitable for computer-based systems by using rule based representation techniques.

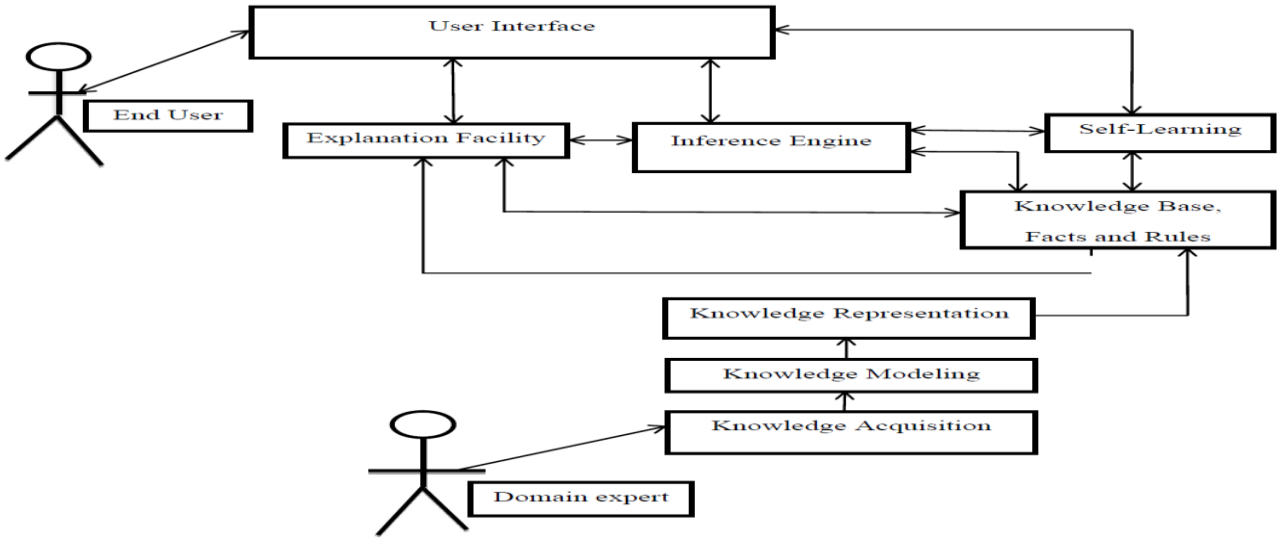


Figure 3.2: Architecture of proposed system (Diriba, 2019)

Knowledge acquisition

Knowledge acquisition involves the procedures of obtaining, organizing, and formalizing knowledge from a source to establish a KBS. In this research, the knowledge necessary for constructing a Knowledge-Based System (KBS) obtained from both tacit and explicit knowledge sources. Tacit knowledge gathered through semi structured interviews with three respiratory disease experts from Jimma University Medical Center.

These experts deliberately selected to engage in in-depth discussions and interviews to gain a comprehensive understanding of the domain knowledge. Their active involvement in the study allow for the verification and validation of the acquired knowledge. Additionally, explicit knowledge were collected from patient case and various sources such as the Internet, manuals, research papers and journal articles.

This research were employed a purposive sampling technique to choose domain experts for knowledge acquisition. The selection of domain experts was based on specific criteria, taking into account their professions, educational qualifications, and years of experience in diagnosing and treating respiratory diseases. Initially, to gain a comprehensive understanding of respiratory diseases, various interviews conducted with experts, including respiratorydisease specialists, resident doctors, and pathologists.

Among these, three respiratory disease experts selected purposively for detail discussions using semi structured interview to discover relevant domain knowledge. For the rule based expert system development 200 previously solved cases of respiratory disease patients from Jimma University Medical Center are collected.

Knowledge modeling

The process of knowledge modeling is an interdisciplinary approach aimed at capturing and organizing information in a reusable format to enhance the preservation, improvement, sharing, aggregation, and processing of knowledge to simulate intelligence (Aronson and Turban, 2007). The iterative nature of generating and adjusting a knowledge model is critical for ensuring understanding among stakeholders, facilitating validation, and allowing swift modifications as needed (Sung-kwan, 2020). In this study, a decision tree employed to conceptually model the

acquired knowledge. A conceptual model represents a system using concepts, ideas, and includes elements such as concepts, control mechanisms, and constraints. Specifically, the concept of a symptoms checking system utilized. Based on identified symptoms, the system will determine the associated disease, which will then be presented within the system. In essence, the essential knowledge for constructing the decision tree model of concepts in the diagnosis and treating respiratory diseases derived from domain experts and relevant patient cases conducted.

Knowledge representation

The process of converting human knowledge into a format that can be understood by computers is known as knowledge representation. For this study rule based representation method used. The reason to choose rule based reasoning is that rule-based systems utilize explicit rules that are easy to understand and interpret.

This method entails representing knowledge by using a collection of facts and rules. Symbolic rules are commonly used in reasoning and knowledge representation. The basic structure of a rule consists of an IF statement with a condition and a THEN statement with a conclusion. The condition represents the premises or conditions, while the conclusion indicates the resulting outcome. In rule-based systems, there are two primary methods of reasoning: forward chaining and backward chaining inference mechanisms

In the forward chaining process, the inference engines start by establishing a condition or criterion and then gradually incorporate individual criteria one by one until the entire chain is processed. The facts within the system are stored in a working memory, which is continuously updated under the control of data-driven processes. Backward chaining approach concentrates its efforts on considering only the rules that are relevant to the specific objective at hand. Its difference from forward chaining lies in its attempt to identify the premises that lead to the conclusion, starting from a set of conclusions provided in the problem description rather than conditions.

The system that developed by using rule-based system for respiratory disease diagnosis and treatment employed a knowledge base that includes a wide range of rules for symptoms and treatment guidelines related to respiratory diseases and this rule integrated with node to run the main interface of respiratory disease diagnosis and treatment. Initially, upon accessing the system the user need to navigate to the directory and open the command line then run the

NPM (node package management) in order to connect the knowledge base and the user interface. Next to access the main respiratory disease diagnosis and treatment interface the user need to select any browser he/she want to use and write https://Localhost:3000. First when user enter to system it ask to fill his or her name then ask for user symptom .When a patient presents with respiratory symptoms, the rule-based system analyzes the provided information using a series of rules. These rules are designed to match the patient's symptoms and test results with known patterns associated with different respiratory diseases.

The system applies forward chaining inference mechanisms to deduce the likely diagnosis based on the available data and recommend treatment. For example, if a patient reports symptoms such as cough, shortness of breath, and wheezing, the rule-based system may match these symptoms with rules associated with asthma and give treatment for asthma and if the provided symptom does not match with disease it iterate to the next disease, check for symptom if it match together and continue to check similarity of symptom until rule n.

The system can also consider additional factors, such as the patient's age, smoking history, and Family history, to refine the diagnosis. Once a diagnosis is established, then inference engine provides treatment recommendations such as bronchodilators or corticosteroids and any necessary lifestyle modifications referrals to users.

3.4. Demonstration

Typically, specialized programming languages or tools are utilized to create knowledge-based systems, with shells being the most frequently used tools in their development. Some tools commonly used for implementing knowledge-based systems include Expert System Shell (JESS), CLIPS, Vidwan, Prolog, Recall, and JCOLIBRI (Burnashev, 2019). However, in this study, the development of the knowledge base and user interface for respiratory disease diagnosis and treatment utilized the JavaScript programming language, rools (a rule engine for Node.js), Handlebars.js (a JavaScript-based semantic template language) and WEKA software version 3.93.The researcher also utilized data mining classification algorithm such as **J48**, which is a classification algorithm that utilizes decision trees for tasks such as classification and prediction.

Decision trees provide a simple yet powerful way to represent knowledge. During the construction of a decision tree, some branches may be influenced by anomalies in the training data, which can be caused by noise or outliers. The other classification algorithm employed is

PART, which is a rule-based classifier employs a collection of IF-THEN rules to perform classification tasks. Each rule consists of an expression in the form of IF condition THEN conclusion. The IF-part, also referred to as the rule antecedent or precondition, represents the condition for the rule to be applied. The third data mining classification algorithm employed for development is **JRIP** which another rule-based classifier that employs a set of IF-THEN rules to perform classification tasks.

3.5. Evaluation

After constructing and showcasing the prototype, it is essential to assess the system's functionality and user acceptance. The evaluation procedures concentrate on both the system's performance and user acceptance of the expert system. Professionals chosen to assess the system and their evaluation of the system's performance conductedt through the use of close-ended questions.

The accuracy of the system determined by calculating its retrieval results through measures such as recall, F-measure, Accuracy and precision. Additionally, the researcher assessed the user acceptance of the system through the use of checklist questions. The reason why use system and user evaluation is System evaluation focuses on objectively assessing the performance and technical aspects of the system. This involves evaluating the accuracy, efficiency, scalability, and reliability of the system. System evaluation helps determine the accuracy of diagnoses and treatment recommendations generated by the self-learning system. It involves comparing the system's outputs with established standards or expert opinions to assess its effectiveness. User evaluation is crucial for gathering feedback and insights from end-users, such as healthcare professionals, patients or other stakeholders who interact with the system.

CHAPTER FOUR

4. KNOWLEDGE ACQUISITION, MODELING, EXPERIMENTATION AND EVALUATION

4.1. Types of Respiratory disease frequently occurred in Ethiopia

Respiratory diseases are medical conditions that impact the functioning of the lungs and airways, resulting in difficulties with breathing and the exchange of gases. The airway systems begin at the nose, extend through the larger and smaller air passages, and ultimately reach the lungs. These systems are categorized into the upper respiratory system and the lower respiratory system. Respiratory ailments encompass both short-term respiratory infections and long-term chronic respiratory diseases (Levine, 2021). In Ethiopia, several common respiratory diseases are frequently encountered. The researcher conducted a study on these respiratory diseases with the aim of developing a rule-based system.

4.1.1. The identified domain knowledge

Tuberculosis

Tuberculosis (TB) is an infectious disease caused by bacteria that can be transmitted from one person to another through airborne particles. While the lungs are the primary target, TB can affect and cause harm to various organs and tissues throughout the body, including the brain, kidneys, and spine. Without proper treatment, TB can be fatal.

Major symptoms of Tuberculosis include general sense of illness or weakness, weight loss, fever, and night sweats. In the case of pulmonary tuberculosis, additional symptoms may include coughing, chest pain, and the presence of blood in coughed-up phlegm. Symptoms of TB disease affecting other parts of the body vary depending on the specific area affected.

Asthma

Asthma is a persistent respiratory condition that can impact individuals of any age. The underlying cause of asthma is inflammation and constriction of the airways, leading to difficulties in breathing. Numerous factors have been identified as contributing to an elevated risk of developing asthma, although pinpointing a single, direct cause can often be challenging. The likelihood of asthma is higher when other family members, particularly close relatives like parents or siblings, also have the condition. People with other allergic conditions such as eczema and rhinitis (hay fever) are more prone to asthma. Increased prevalence of asthma is associated with urbanization, likely due to various lifestyle factors. Early-life events can impact lung

development and raise the risk of asthma. These events include low birth weight, premature birth, exposure to tobacco smoke and other sources of air pollution, as well as viral respiratory infections

Major symptoms of Asthma: The symptoms of asthma can vary from one individual to another, and there can be instances when the symptoms worsen considerably, known as an asthma attack. Symptoms commonly intensify during nighttime or physical exertion. (WHO, 2023)

Typical symptoms of asthma include:

- Persistent cough, particularly at night
- Wheezing when exhaling, and occasionally when inhaling
- Shortness of breath or difficulty breathing, which may occur even at rest
- Tightness in the chest, making it challenging to take deep breaths.

Pneumonia

Pneumonia is an infection that affects the lungs and can be caused by bacteria, viruses, or fungi. It leads to inflammation and swelling of the lung tissue, along with the accumulation of fluid or pus in the lungs. Bacterial pneumonia typically presents more severe symptoms compared to viral pneumonia, with the latter often resolving without specific treatment.

Major symptoms of pneumonia include High fever (up to 105 F or 40.55 C), Cough with yellow, green or Tiredness (fatigue). Rapid breathing. Shortness of breath. Rapid heart rate. Sweating or chills. Chest pain and/or abdominal pain, especially with coughing or deep breathing, Loss of appetite. Bluish skin, lips or nails.

Influenza

Influenza, commonly known as the flu, is an infectious respiratory illness caused by influenza viruses that invade the nose, throat, and lungs. Certain individuals, including those aged 65 and older, young children, and individuals with specific health conditions, face a greater risk of experiencing severe complications from the flu (CDC, 2022).

Major symptoms of Influenza include Fever/feeling feverish or chills,Cough,Sore throat, Runny or stuffy nose, Muscle or body aches,Headaches,Fatigue(tiredness),Some people may have vomiting and diarrhoea.

Lung Cancer

The growth of abnormal cells in the lung that form a tumor is what leads to the development of lung cancer. These tumors can be identified as nodules or masses on chest x-rays or CT ('CAT') scans. While lung cancer is predominantly associated with individuals who have a history of smoking tobacco products, it can also occur in people who have never smoked. Additional risk factors for lung cancer include exposure to radon, asbestos, and second-hand smoke, as well as air pollution and a family history of lung cancer (Respir, 2021).

Major symptoms of Lung Cancer includes Persistent cough, Chest pain, Shortness of breath, coughing up blood, Fatigue, Unexplained weight loss.

Acute Bronchitis

Acute bronchitis is a common respiratory infection that poses a significant health issue, especially among adults. This condition primarily affects the bronchi, which are important air passages responsible for transporting air within the lungs. Acute bronchitis is characterized by inflammation of these bronchi, but it does not involve pneumonia. It typically occurs in individuals without pre-existing chronic obstructive pulmonary disease. Acute bronchitis, which involves inflammation of the bronchi lining, is a common condition seen in emergency departments, urgent care centers, and primary care offices. It is one of the top 10 most prevalent illnesses treated on an outpatient basis affecting around 5% of adults each year (Anumeha, 2023).

Major symptoms of acute bronchitis includes Cough, Chest discomfort, Shortness of breath, Fatigue, Sore throat, Mild fever 100.4°F (38°C),Nasal congestion.

Rhino sinusitis

Rhinosinusitis is the medical term used to describe inflammation of the nose and paranasal sinuses. Acute rhinosinusitis refers to a condition where symptoms completely resolve within 12 weeks of the initial onset. On the other hand, if the symptoms continue for more than 12 weeks, it is classified as chronic rhinosinusitis (CRS) (BSW, 2024).

Major symptoms of Rhinosinusitis includes Nasal congestion or blockage that lasts for more than 12 weeks, Persistent nasal discharge, often thick and discolored, Facial pain or pressure,

especially in the areas mentioned earlier, Reduced sense of smell and taste, Postnasal drip, Cough, particularly upon waking up or lying down, Fatigue and malaise, Headache, Bad breath, Ear pain or fullness, Sore throat.

Occupational lung diseases

Occupational lung diseases (OLDs) are respiratory conditions that are either caused, worsened, or triggered by exposures in the workplace. These diseases encompass a broad spectrum of respiratory disorders that are similar to those occurring outside the workplace environment. The most frequently diagnosed occupational lung disease (OLD) is occupational asthma. Other types of OLDs can include both acute and chronic conditions, such as hypersensitivity pneumonitis, chronic obstructive pulmonary disease (COPD), and pulmonary fibrosis. In the past, research on OLDs has primarily focused on diseases caused by exposures common in high-income countries and occupations with more apparent hazards, such as coal miners developing silicosis (Kevin, 2020).

Major symptoms of Occupational lung diseases includes Cough, Shortness of breath, Wheezing, Chest tightness, Fatigue, Decreased exercise tolerance, Coughing up blood.

Pertussis

Pertussis, also known as whooping cough, is a respiratory infection caused by the bacteria Bordetella pertussis. People of all ages can be affected by this infection. In the time before vaccines were available, nearly all children would contract pertussis. It poses a significant risk, especially for young infants who account for the majority of hospitalizations and deaths due to the disease. However, pertussis can cause a considerable burden of illness at any age. The widespread use of pertussis vaccines has greatly reduced the number of cases. Nonetheless, concerns about potential adverse reactions led to the replacement of the original whole-cell pertussis vaccines with a cellular pertussis vaccines (Michael, 2021).

Major symptoms of Pertussis includes severe, persistent cough that can last for several weeks or even months. The cough is often characterized by a series of rapid coughs followed by a high-pitched "whooping" sound when the person tries to take a breath.

Bronchiectasis

In adults, bronchiectasis is a long-term condition that often leads to a reduced quality of life and

frequent episodes of exacerbation. Until recently, there were no established international guidelines for its management.

However, the European Respiratory Society has developed guidelines specifically for the management of bronchiectasis in adults. These guidelines were formulated based on a systematic review of the available literature, and they provide recommendations on appropriate investigations and treatment strategies (Pieter, 2017).

Major symptoms of bronchiectasis includes chronic cough, Shortness of breath, Chest pain, Fatigue, Wheezing.

Cystic fibrosis

Cystic fibrosis is a genetic disease that is inherited in an autosomal recessive manner. It primarily affects children and young adults of Caucasian descent. The condition is caused by specific abnormal versions of the CFTR (cystic fibrosis transmembrane conductance regulator) gene. These gene variants lead to thickened mucus and impaired clearance of mucus from various organs in the body. The lungs are the most prominently affected organ, but other organs such as the pancreas, sweat glands, intestine, liver, nasal mucosa, salivary glands, and reproductive tract can also be involved (Jaime, 2021).

Major symptoms of Cystic fibrosis includes frequent lung infections, persistent coughing, wheezing, shortness of breath, and decreased exercise tolerance.

Croup

Croup is a short-term inflammatory condition that impacts the upper airways and is frequently observed in young children. It is primarily caused by acute viral infections, although bacterial and atypical agents are less common causes of the condition (Krishna, 2021).

Major symptoms of Croup includes a harsh, barking cough that resembles the sound of a seal or a dog. Other symptoms may include a hoarse voice, difficulty breathing, a high-pitched noise when inhaling (called stridor), and a mild fever. In some cases, the child may also experience a runny nose and nasal congestion.

Common cold

The common cold is a temporary infection of the upper respiratory tract caused by viruses such as rhinovirus, coronavirus, or adenovirus. It typically resolves on its own within a span of 7 to 10

days. Antibiotics are not effective against colds, as there is no clinical evidence to suggest that their use can alter the course of the illness or prevent secondary infections.

Treatment primarily focuses on managing symptoms and can involve the use of over-the-counter medications, antivirals, and herbal remedies (Hanisah, 2017).

Major symptoms of common cold includes runny or stuffy nose, sneezing, sore throat, cough, mild headache, mild body aches, and low-grade fever.

Corona viruses

Corona viruses are a large family of viruses which may cause illness in animals or humans. In humans, several coronaviruses are known to cause respiratory infections ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). The most recently discovered coronavirus causes coronavirus disease COVID-19 (Mantan, 2021).

Major symptoms of Corona viruses includes fever 37.8°C (100°F), Cough, Shortness of breath Fatigue, Feeling extremely tired or experiencing a general lack of energy, Muscle or body aches, Headache.

Bronchiolitis

Bronchiolitis is a medical condition that causes breathing difficulties in young children, typically those under 2 years old. It is identified by initial symptoms similar to those of a common cold, such as a runny nose, which are quickly followed by indications of infection and inflammation in the lower respiratory tract. When the small airways known as bronchioles are affected, it leads to wheezing and/or abnormal lung sounds called rales (Eur, 2023).

Major symptoms of Bronchiolitis includes common symptoms include coughing, wheezing, and a runny or stuffy nose. Children with bronchiolitis may also experience fever, irritability, decreased appetite, and signs of fatigue or lethargy. In severe cases, there may be signs of cyanosis (bluish discoloration of the skin or lips) due to a lack of oxygen.

Chronic obstructive pulmonary disease (COPD)

Chronic obstructive pulmonary disease (COPD) is a significant issue in public health,

characterized by periods of worsening symptoms known as exacerbations, which can lead to severe illness and even death in both developed and developing nations. Typically, individuals experiencing COPD exacerbations seek medical attention at emergency departments, where they receive a diagnosis and appropriate treatment (Nurettin, 2021).

Major symptoms of chronic obstructive pulmonary disease includes Shortness of breath (dyspnea), Chronic cough, Excessive mucus production, Wheezing, Chest tightness, Fatigue.

4.2. Diagnosis of respiratory diseases

Diagnosing respiratory diseases typically involves a combination of patient history, physical examination, diagnostic tests, and sometimes imaging studies. Here's a general outline of the diagnostic steps and procedures commonly used in the evaluation of respiratory diseases:

1. Patient History:

- Detailed questioning about symptoms, their duration, onset, severity, and any factors that worsen or alleviate them.
- Past medical history, including any pre-existing respiratory conditions, allergies, smoking history, occupational exposures, travel history, and family history of respiratory diseases.

2. Physical Examination

Inspection of the patient's general appearance, including signs of respiratory distress or cyanosis.

Examination of the chest for deformities, asymmetry, or abnormalities in shape.

Auscultation of the lungs using a stethoscope to assess breath sounds, including wheezing, crackles, or decreased breath sounds.

3. Diagnostic Tests:

Pulmonary Function Tests (PFTs): These tests measure lung function and capacity. Common PFTs include spirometer, which assesses airflow obstruction, and lung volume tests, which evaluate lung capacity.

Peak Expiratory Flow (PEF) Measurement: Useful for monitoring asthma and assessing the severity of airflow limitation.

Arterial Blood Gas (ABG) Analysis: Provides information about oxygenation, ventilation, and acid-base balance in the blood. Helpful in evaluating respiratory failure and chronic respiratory conditions.

Sputum Analysis: Examination of sputum for the presence of pathogens (bacteria, viruses,

fungi) or inflammatory cells. Useful in diagnosing respiratory infections and identifying specific pathogens. Allergy Testing: Skin prick tests or blood tests (e.g., IgE levels) to identify allergens triggering allergic respiratory conditions like allergic rhinitis and asthma.

4.3. Respiratory diseases Treatment

The treatment of respiratory diseases refers to the medical interventions and therapies aimed at managing and alleviating the symptoms, preventing complications, and promoting the recovery of individuals with respiratory conditions. It involves a range of approaches, including medications, lifestyle modifications, respiratory therapies, and, in some cases, surgical interventions. The specific treatment options vary depending on the type and severity of the respiratory disease and are typically tailored to address the underlying causes and symptoms. The following information about the treatments for respiratory diseases is compiled from various relevant sources and includes insights from domain experts obtained through interviews (Chien, 2022).

Treatment for Tuberculosis

- Administration of several antibiotics such as isoniazid, rifampin, pyrazinamide and ethambutol for the initial phase which lasts for two months, Regular follow-up visits with healthcare providers to monitor progress, usually take only two antibiotics (most commonly isoniazid and rifampin) for the second phase which lasts for four to six months.

Treatment for Asthma

- Quick-relief medications (bronchodilators): Short-acting beta-agonists, such as albuterol, Inhaled corticosteroids, long-acting beta-agonists and leukotriene modifiers, Proper use of inhalers is crucial for effective medication delivery.

Treatment for Pneumonia

- Take Antibiotics if it is bacterial pneumonia, if pneumonia is caused by a viral infection take Antiviral medications, getting plenty of rest to allow the body to heal, Staying hydrated by drinking an adequate amount of fluids, Using cough suppressants or expectorants as recommended by healthcare providers.

Treatment for Influenza

- Getting plenty of rest allows the body to heal and recover, Stay hydrated by drinking an adequate amount of fluids, such as water, herbal tea, or clear broths, Take Acetaminophen or nonsteroidal

anti-inflammatory drugs (NSAIDs) like ibuprofen to reduce fever, Take oseltamivir (Tamiflu), zanamivir (Relenza), or baloxavir marboxil (Xofluza) to reduce the severity within 48 hours of symptom onset.

Treatment for Lung cancer

- Surgical removal of the tumor and nearby lymph nodes considered for early-stage lung cancer, Take High-energy radiation therapy to kill cancer cells, Take (neoadjuvant chemotherapy) to shrink tumors before surgery and take (adjuvant chemotherapy) after surgery to kill remaining cancer cells.

Treatment for Acute Bronchitis

- Get plenty of rest to allow your body to recover, Drink plenty of fluids, such as water, warm tea, or soup, to stay hydrated, Take Over-the-counter medications like acetaminophen or nonsteroidal anti-inflammatory drugs (NSAIDs).

Treatment for Bronchiectasis

- Take Long-term prophylaxis to prevent exacerbations, Take inhaled hypertonic saline or muco active agents like dornase Alfa, Take Inhaled bronchodilators, such as short-acting or long-acting beta-agonists, and Take High-frequency chest wall oscillation (HFCWO).

Treatment for Bronchiolitis

- Encourage frequent breastfeeding or bottle-feeding for infants, or providing plenty of fluids for older children, using a cool-mist humidifier or sitting in a steamy bathroom, raising the head of the child's bed or using a pillow, take bronchodilators or antiviral medications, particularly for children with underlying health conditions or severe symptoms.

Treatment for Chronic obstructive pulmonary disease (COPD)

- Quit smoking and avoid exposure to secondhand smoke, Minimize exposure to air pollution, chemical fumes, dust, and other respiratory irritants that can worsen COPD symptoms, Take Inhaled steroids prescribed for individuals with more severe COPD, Taken Phosphodiesterase-4

inhibitors.

Treatment for common cold

- Getting plenty of rest and stay hydrated by drinking fluids such as water, herbal tea, or clear broths to prevent dehydration, Take Over-the-counter medications like acetaminophen (Tylenol) or nonsteroidal anti-inflammatory drugs (NSAIDs) such as ibuprofen.

The treatment for COVID-19

- Self-isolation, rest, and supportive care, severe cases of COVID-19 may require hospitalization, Supplemental oxygen to individuals with low oxygen levels to support breathing and maintain adequate oxygenation. Using Mechanical ventilation.

The treatment for Croup

- Encourage the child to drink plenty of fluids to prevent dehydration and help thin the mucus, Take Oral or inhaled corticosteroids, such as dexamethasone or prednisolone, a one-time dose of nebulized epinephrine administered in a healthcare setting to rapidly reduce airway swelling.

The treatment for Cystic fibrosis (CF)

- Take Bronchodilators, Take Ivacaftor, lumacaftor, tezacaftor, or elexacaftor/tezacaftor/ivacaftor,

  Take pancreatic enzyme replacement therapy

The treatment for occupational lung diseases

- Avoidance of exposure, Take supplemental oxygen therapy.

The treatment for pertussis

- Antibiotics, such as azithromycin, clarithromycin, or erythromycin, getting plenty of rest to aid in the healing process, Take Prevention and vaccination measures.

The treatment for Rhino sinusitis

- Using a saline solution or nasal sprays to flush out mucus and relieve nasal congestion, drinking plenty of fluids to thin mucus and promote drainage, if symptoms persist for more than 10-14 days, a healthcare provider may prescribe antibiotics.

4.4. Knowledge Acquisition from Respiratory diseases cases (Records)

After reviewing relevant documents and consulting with experts in the field, the researcher gathered a diverse collection of respiratory disease cases from patient records. These cases play a vital role in developing a knowledge representation technique called rule based reasoning. The selection of Jimma University Medical Center (JUMC) was based on its comprehensive and well-maintained repository of respiratory disease cases. To ensure the reliability and quality of the findings, several approaches were utilized. These included extracting cases that had been previously documented with the guidance of experienced pathologists at JUMC, as well as conducting interviews with knowledgeable healthcare professionals such as nurses and health workers.

4.5. Knowledge source for Data Mining classification algorithm

In the context of developing a self-learning expert system for respiratory diseases diagnosis and treatment, acquiring domain knowledge involves gathering information from diverse sources such as patient records, medical documents, manuals, and authoritative medical websites (Irina, 2017). This information is crucial for constructing knowledge-based systems that are comprehensive, accurate, and well-organized. Specifically, in this study, knowledge acquisition utilized data mining techniques, particularly classification algorithms, to analyze and categorize extensive healthcare data (Irina, 2017). These classifications served as the foundational knowledge for developing the self-learning expert system. In the healthcare sector, the proliferation of complex data from patient information, electronic medical records, and medical devices presents an opportunity for extracting valuable insights to enhance decision-making and reduce costs (Irina, 2017). Automated data mining methods are essential for efficiently extracting and utilizing this wealth of information. For instance, data utilized in this study were collected from Jimma University Medical Center, highlighting the application of data mining tools in leveraging institutional data for system development." (Irina, 2017). To collect patient data a

profile of prior diagnosed case collected from a record house by clearly identifying the required diseases. The system design incorporates knowledge obtained from a data mining tool, which effectively extracts valuable information. The data utilized by the data mining tool was collected from Jimma University Medical Center.

Receiving enough and necessary data is the hardest task besides preparing the data for the data mining tool in order to achieve the intended objective of the study. In addition, Knowledge acquisition is a complex and time-consuming stage during case based system development (Amjad, 2022). For case generation and model building, classifier algorithms such as J48, PART, and JRIP are employed and their result is compared to generate best rules and representative model for the rule based system.

Once data has been collected, it is crucial to create a comprehensive and representative dataset. In this particular study, the researcher gathered 1000 instances with their corresponding attributes from Jimma University Medical center. During this process, irrelevant information was removed from the instance list. To address missing values, the researcher employed the WEKA tool, which replaced missing values in the dataset. Nominal and numeric attributes were substituted with modes and means obtained from the training data. Following the preprocessing stage, the dataset consisted of 1000 instances and 14 selected attributes, which included the class attribute. By utilizing more recent data, the study incorporated the evolving diagnostic technologies and modifications in drug dosages, ensuring that the findings were relevant and accurate. In the development of a self-learning expert system for respiratory diseases, acquiring domain knowledge involves gathering extensive information from various sources such as patient records, medical documents The collected data undergoes classification using data mining algorithms to identify patterns and relationships among different variables relevant to respiratory diseases. These classifications not only help organize the data but also serve as the basis for developing algorithms within the expert system that can diagnose and recommend treatments based on learned patterns and historical data. This approach enhances the system's ability to provide accurate diagnostic suggestions and treatment plans, leveraging the wealth of information extracted through automated data mining methods. Thus, the integration of knowledge obtained from data mining into the system design enhances its capability to learn and adapt over time, improving its overall effectiveness in clinical decision-making. After the

relevant attributes from the domain knowledge is selected then it registered on normal Excel file. Since the WEKA software doesn't compatible with normal excel file, the instance file must converted in to csv or arff file extension for data mining classification algorithm. Once the file converted in to csv file extension, the next step is opening WEKA software and upload csv file and start to conduct data mining operation such as preprocessing and classification tasks.

4.6. Data preprocessing

The large size and diverse origins of real-world data make them susceptible to noise, missing values, and inconsistencies. These factors can have a significant impact on the quality of mining results. Therefore, it is essential to preprocess the data before utilizing it in a data mining tool. The data needs to be properly prepared to align with the specific tasks set by the researcher. Data preparation in data mining involves several steps aimed at identifying high-quality data within the provided dataset. These steps include error detection and elimination, redundancy resolution, and the removal of patient identifiers to uphold patient privacy. The ultimate goal of data preprocessing is to ensure the highest level of data quality, enabling accurate and reliable mining outcomes. (Gurram, 2021).

Various data preprocessing techniques exist to enhance the quality and usability of data. Data cleaning is one such technique that focuses on eliminating noise and rectifying inconsistencies in the dataset. Data integration, on the other hand, involves combining data from multiple sources to create a cohesive and unified data store. Data transformations, such as normalization, can be applied to improve the accuracy and efficiency of mining algorithms, especially those involving distance measurements. Additionally, data reduction techniques aim to reduce the size of the dataset by aggregating or eliminating redundant features. It is important to note that these techniques are not mutually exclusive and can be used in conjunction with one another. For instance, during data cleaning, transformations may be applied to rectify incorrect data, such as converting all entries in a date field to a common format.

Applying data processing techniques prior to mining can significantly enhance the quality of the extracted patterns as well as reduce the time needed for the mining process (Gurram, 2021). In this study, the researcher conducted preprocessing tasks to optimize the data for effective utilization of data mining techniques.

4.7. Data cleaning

Raw data commonly contains incomplete records, noisy values, outliers, and inconsistent data. To tackle these issues, data cleaning (also referred to as data cleansing) procedures are employed. These procedures aim to enhance data quality by addressing missing values, smoothing out noisy data, identifying and removing outliers, and resolving inconsistencies (Ihab, 2016).However, it is crucial to recognize that expecting data to be entirely complete and error-free is unrealistic, particularly when dealing with patient data in complex healthcare systems. Data cleaning is a demanding and time-consuming task that involves a combination of automated and manual processes. It requires an iterative approach that leverages the expertise of clinical professionals, statistical experts, and data warehouse engineers (Charlotte, 2021).

The collected data from Jimma University Medical Center, specifically related to the baseline data of respiratory disease, underwent a data cleaning process. This involved consulting domain experts to ensure the data's accuracy and relevance. During this process, attributes that were deemed unnecessary for the decision-making process were removed before proceeding with data preprocessing.

Table 4.1: Removed attribute

Removed attributes		
No	Attribute Name	Reason
1	Patient Name	Not necessary
2	Patient phone number	Not necessary
3	Patient Address	Not necessary
4	Patient card Number	Not necessary

4.8. Attribute selection

In this particular study, the researcher decided to remove specific attributes that were deemed irrelevant to the diagnosis and treatment of respiratory diseases. These attributes were primarily

related to organizational data management and did not directly contribute to the research objectives. The process of attribute selection involves identifying the most suitable subset of features from the original set based on predefined criteria and goals (Getachew et al., 2017). In this study, a total of 14 selected attributes were chosen after eliminating unnecessary attributes.

No	Attributes	Description of attributes	Attributes values
1	Age	The age of the patient	Numeric values
2	Gender	The gender of the patient	F-Female M-Male
3	Family Health History	A record of medical conditions and extended family	Yes No
4	Smoking History	Individual's past and present use of tobacco products	Yes No
5	Chest Pain	Discomfort or distress in the chest area	Yes No
6	Coughing	Reflex action that helps clear the airways of irritants	productive Dry
7	Coughing Blood	Presence of blood in the sputum that is coughed up from the lungs	Yes No
8	Fever	Temporary increase in body temperature	Yes No
9	Rapid Breathing	An increased rate of breathing	Yes No
10	Shortness of Breath	Difficulty or discomfort in breathing	Yes No
11	Rapid Heart beat	Unusually fast heart rate	Yes No

12	Wheezing	High-pitched whistling sound	Yes No
13	Duration of symptoms	Period with symptoms of the disease	<1 Week, b/n 1 and 4 Week, Months and Years.
14	Severity	Level of disease severity	High, Medium and Low

Table 4.2: Selected attributes and their description

4.9. Data Formatting

WEKA (Waikato Environment for Knowledge Analysis) requires data to be prepared in specific formats and file types. The datasets provided for analysis in this software were appropriately prepared to meet the requirements of WEKA. Data transformation refers to the process of mapping and converting data from one format to another, making it suitable for the data mining tool (Mark, 2022). The original data was initially in an Excel format, which is not compatible with WEKA. Therefore, it was necessary to convert the data into CSV and ARFF formats. Once the dataset was successfully converted into CSV format, the next step involved opening the file with the WEKA Data mining software. Additionally, provided below is a sample CSV file format that illustrates how the data mining algorithm is utilized for classification purposes.

Table 4.3: Sample CSV used for classification

	A	B	C	D	E	F	G	H	I	J	K
1	Age	Gender	Family Health History	Smoking History	Chest Pain	Coughing	Coughing Blood	Fever	Rapid Breathing	Shortness of Breath	Rapid Heart beat
2	30	M	N	Y	N	Productiv	Y	N	N	Y	N
3	35	M	Y	N	N	dry	Y	Y	Y	Y	N
4	31	F	N	N	N	dry	Y	N	N	N	N
5	20	M	Y	Y	N	dry	Y	N	Y	N	Y
6	10	F	Y	N	N	Productiv	N	N	N	Y	Y
7	60	F	Y	N	N	dry	Y	N	Y	N	Y
8	38	F	Y	Y	Y	dry	Y	Y	Y	N	N
9	11	F	N	Y	Y	dry	N	Y	Y	Y	Y
10	41	M	Y	N	N	Productiv	N	N	N	Y	N
11	22	M	N	Y	Y	Productiv	Y	N	Y	N	N
12	70	F	Y	Y	Y	dry	Y	Y	Y	N	N
13	55	M	N	Y	Y	Productiv	N	N	Y	N	Y
14	28	M	N	N	Y	dry	N	N	N	Y	Y
15	48	F	Y	Y	N	dry	Y	N	Y	Y	N
16	20	F	Y	N	N	dry	Y	N	Y	N	N
17	30	F	Y	Y	N	dry	Y	N	N	Y	N
18	32	M	N	N	N	Productiv	N	N	N	N	N
19	27	M	Y	Y	N	Productiv	Y	N	Y	N	Y
20	29	M	Y	Y	N	Productiv	Y	Y	N	N	Y
21	21	F	Y	Y	Y	dry	Y	Y	Y	Y	Y
22	37	F	N	N	N	Productiv	Y	Y	N	Y	N
23	59	M	Y	Y	Y	dry	N	Y	N	Y	Y
24	34	M	Y	Y	Y	Productiv	N	N	Y	N	Y
25	36	F	Y	Y	N	Productiv	Y	N	Y	N	N
26	46	F	Y	Y	Y	dry	Y	N	N	N	N
27	39	F	N	Y	N	dry	Y	Y	Y	Y	N
28	43	M	Y	Y	N	Productiv	N	Y	N	Y	N

4.10. Experimentation

After adequately preparing the data for the data mining tool, the subsequent phase entails conducting experiments employing various algorithms such as J48, PART and JRip. The main goal of the data mining aspect in this study was to create a descriptive model for designing a rule based system, utilizing a classification data mining algorithm. The dataset used for sampling comprises 1000 instances, with 14 attributes, including the class attribute. All attributes were employed in all experiments. Default parameter values were utilized for the classifier algorithms in the WEKA Data Mining tool, as they were demonstrated to produce higher accuracy compared to modifying these parameters.

4.11. Experiment Design

Prior to initiating the experimentation phase, it is crucial to establish the methodology for assessing and evaluating the prediction model. In this particular study, two different options were employed for model training and testing: 10-fold cross-validation and percentage splitter. During 10-fold cross-validation, the complete dataset is utilized, with one fold reserved for testing the prediction model while the remaining nine folds are used for training. This process is repeated iteratively, allowing the model to be trained and tested on different folds. Alternatively, the percentage splitter approach involves allocating 80% of the data for model training and the remaining 20% for testing the model.

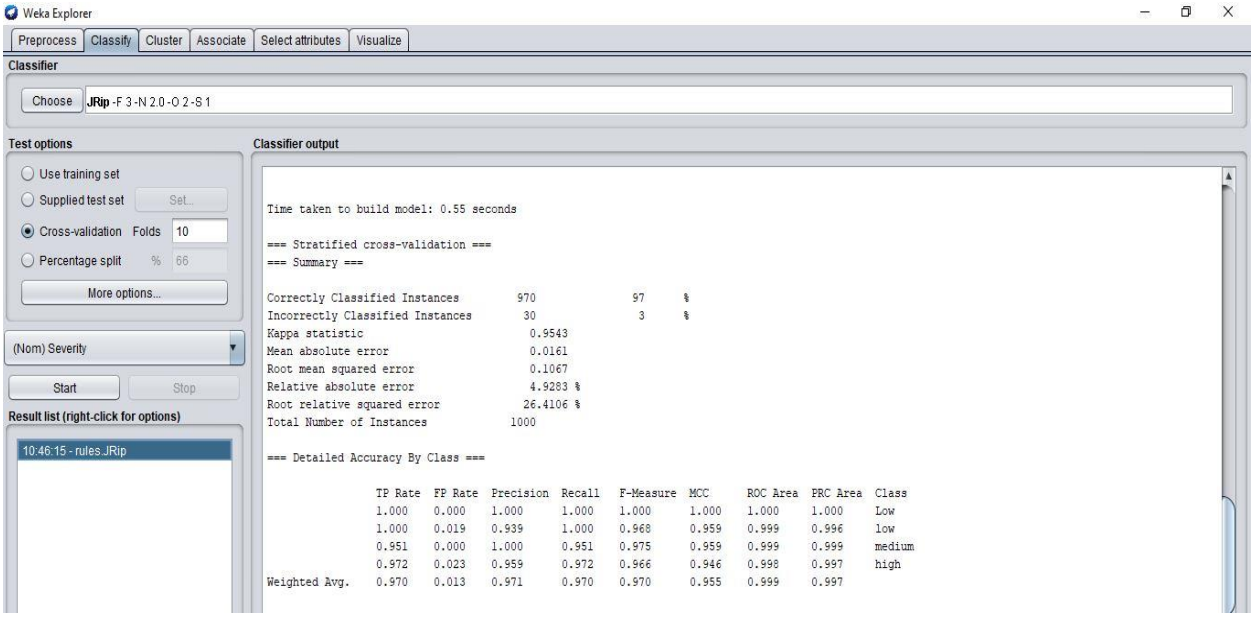
Before finalizing the data split for model training and testing, the researcher conducted several experiments using different splitter settings, such as 70%/30%, 80%/20%, and 90%/10%. After assessing the results, the researcher opted for the 80%/20% splitter. This decision was influenced by the observation that the original data collected from the source did not exhibit sufficient matching characteristics, and the 90%/10% splitter did not adequately encompass classes of respiratory disease in the pre-experimentation phase.

In this study, the researcher selected three classifier algorithms: J48 decision tree, PART, and JRip. These algorithms were chosen due to their capabilities, simplicity, comparable accuracy demonstrated in previous studies, and robustness.

4.11.1. Experimentation One (J48)

In this specific experiment, the J48 decision tree prediction algorithm was utilized with both 10-fold cross-validation and percentage splitter model training and testing options. The initial experiments were conducted using 10-fold cross-validation with the default parameters of the WEKA data mining tool. Out of a total of 1000 instances, the algorithm accurately classified 970 instances (97%), while the remaining 30 instances (3%) were incorrectly classified into other types of respiratory disease classes. Additionally, during this experimentation, the algorithm generated 114 leaves and a tree size of 205 nodes.

Table 4.4: The generated results of 10 fold cross-validation J48 tree



Following the construction of the classifier algorithm, the subsequent step involved evaluating the model in the KDD process. The model's performance was assessed using various criteria, including performance accuracy, time taken to build the model, ROC curves, true positive rate, and false positive rate. The summarized performance evaluation results of the J48 decision tree experimentation are presented in figure 4.1.

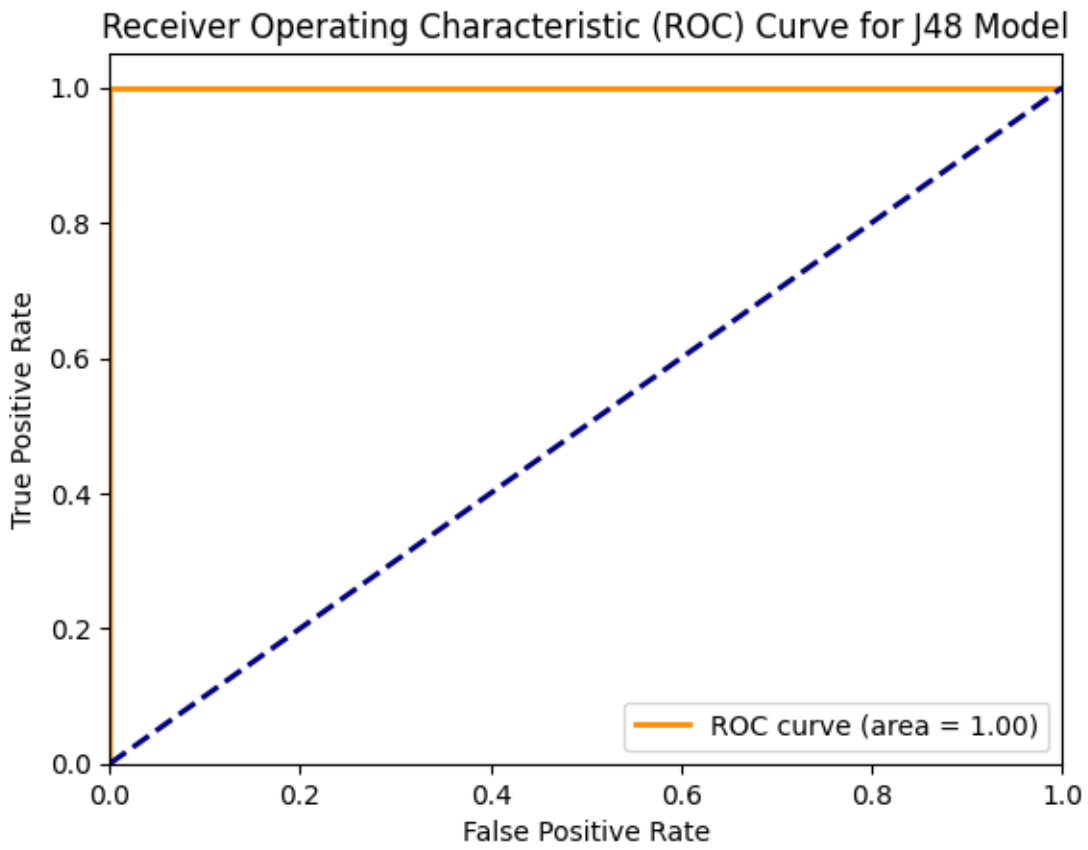
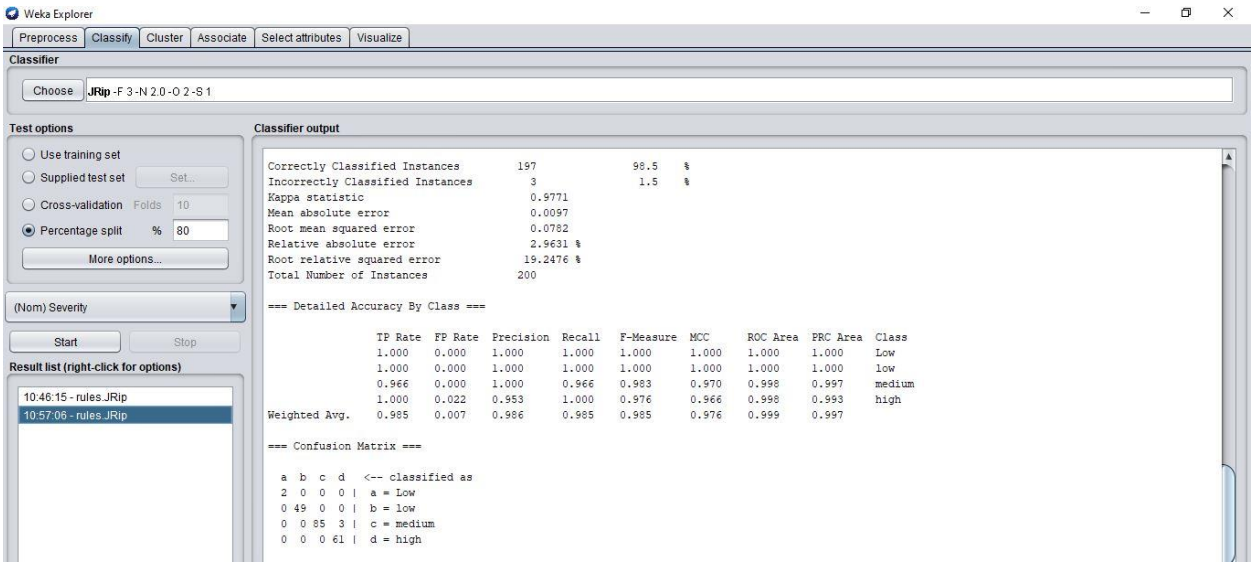


Figure 4.1: ROC curve for 10 fold cross validation

In order to make a meaningful comparison and select the best classifier model, the researcher proposed two approaches for training and testing the models. After completing the initial classifier model with its default parameters (10 fold-cross validation), the next step is to carry out an additional experiment using the percentage splitter method (80%, 20%) for model training and testing, while still using the same algorithm. Consequently, the first experiment entails an extension of the J48 decision tree classifier algorithm, where 80% of the dataset is employed for training the model and the remaining 20% is used to evaluate its performance.

Table 4.5: The Generated results of J48 percentage split



After constructing the classification model, the subsequent step involved evaluating the outcomes of the specific experiments. The provided table displays the results of the J48 classifier algorithm model, which was developed using 80% of the data for training and 20% for testing.

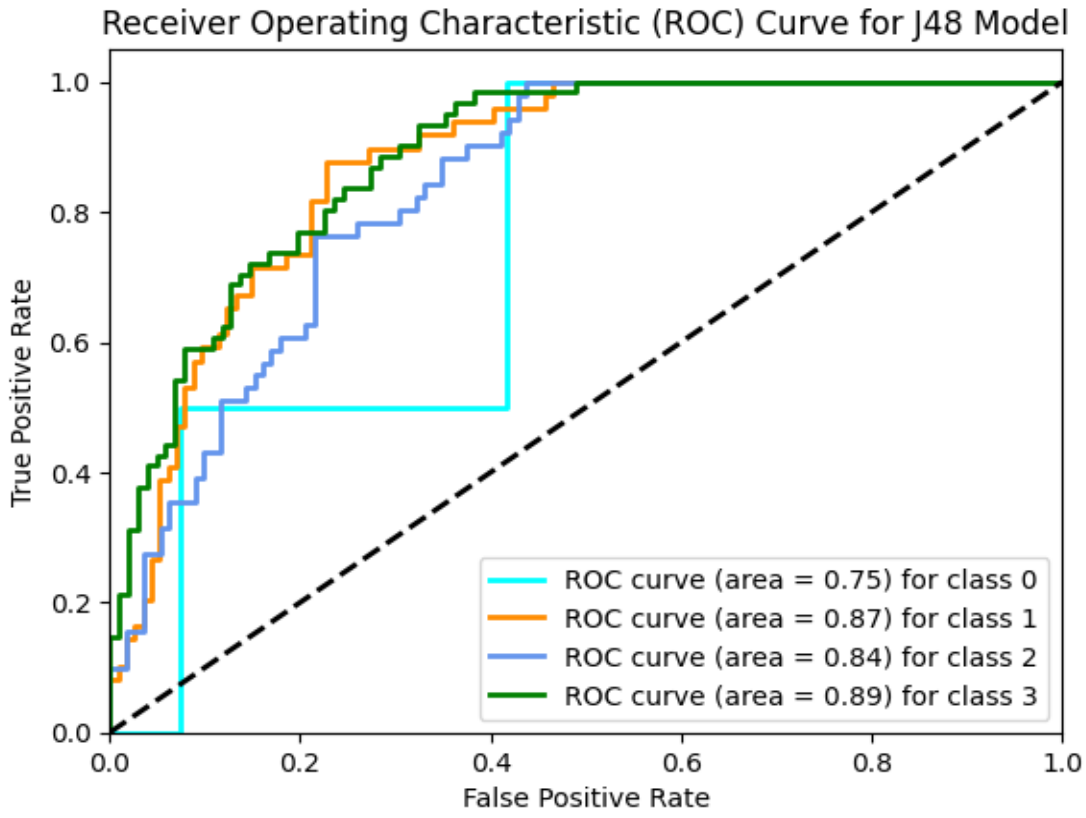


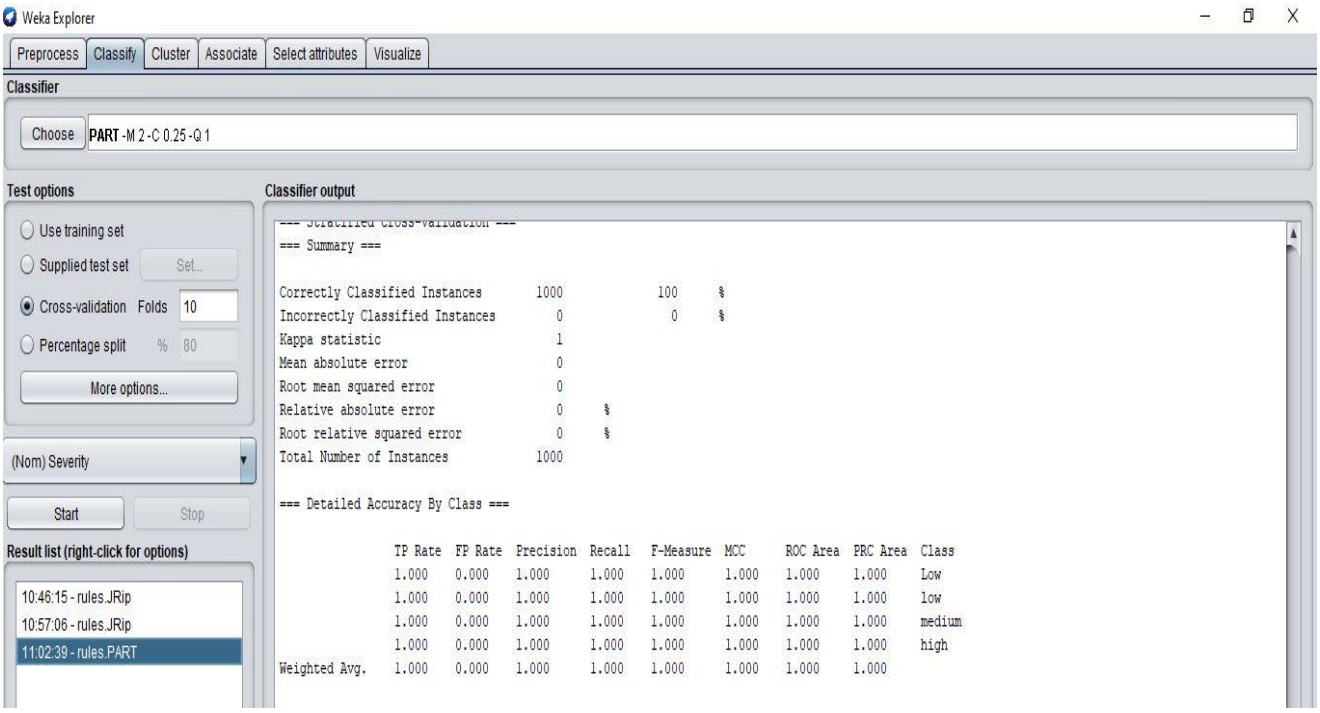
Figure 4.2: ROC curve for J48 Percentage split (80%, 20%)

According to the depicted figure 4.2, out of the 200 test instances, the algorithm accurately classified 197 instances (98.5%) into their respective classes, while only 3 instances (1.5%) were misclassified. The model building process itself took approximately 0.02 seconds.

4.11.2. Experimentation Two (PART)

In this specific experiment, the PART rule classifier algorithm was employed to construct a classifier model. The experiment was conducted using the default or 10-fold cross-validation model training and testing option in the WEKA software. Out of a total of 1000 instances, 1000 cases (100 %) were correctly classified into their respective classes, while the remaining 0 cases (0%) were misclassified.

Table 4.6: The generated results of PART 10 fold cross-validation



In the process of data mining known as KDD, once the classifier model has been built, the next step is to assess the model's performance. To accomplish this, the accuracy of the model is evaluated in a detailed manner. The specific results of the accuracy obtained from the PART classifier in the current experiment are presented in Table 4.6. These results were obtained using a technique called 10-fold cross-validation.

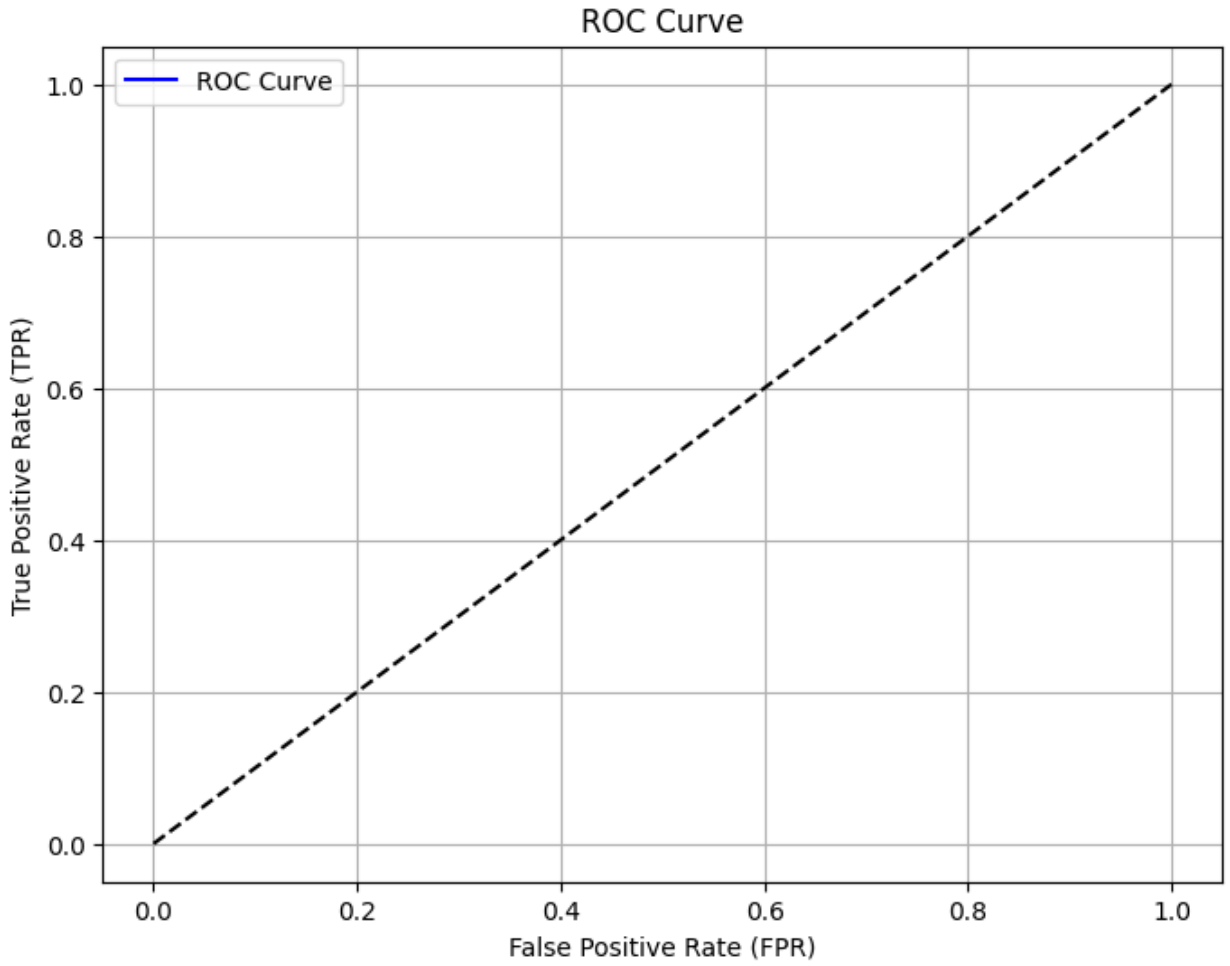
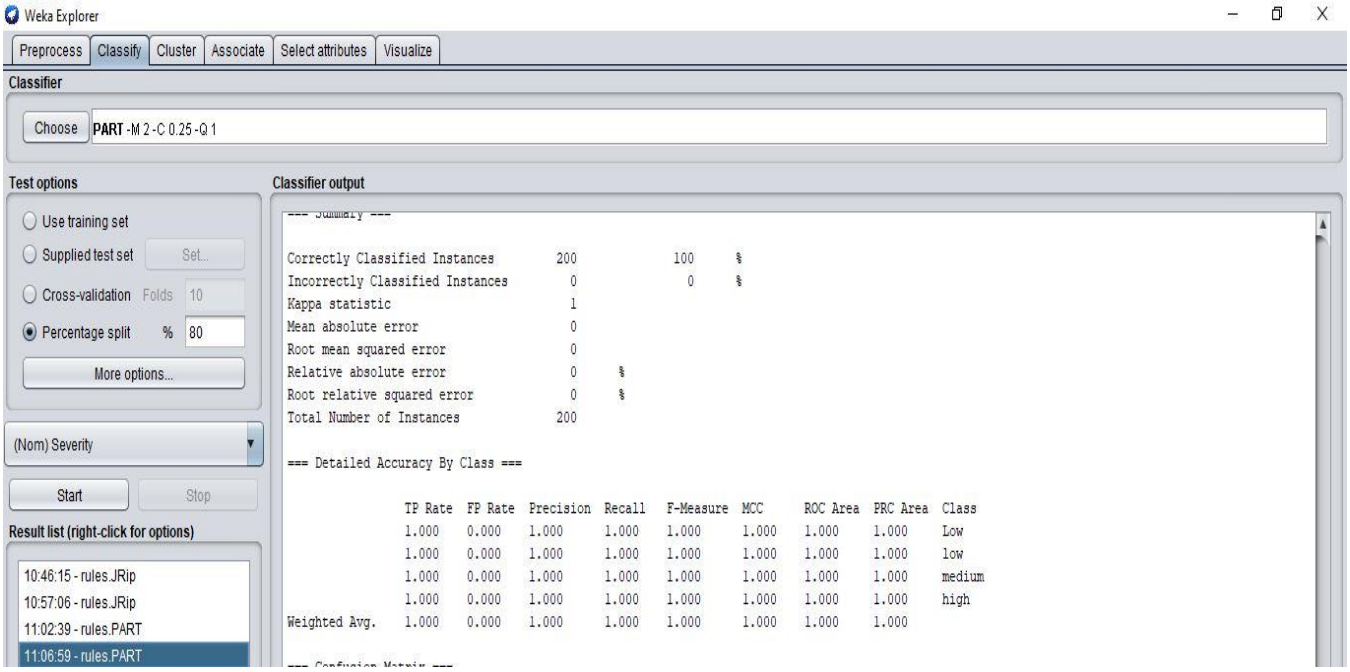


Figure 4.3: ROC curve for PART classifier experiment result for 10-fold cross validation

In this particular experiment, the model constructed using the PART rule classifier algorithm achieved an accuracy of 100%. The model took 0.07 seconds to build. The ROC Area measure indicated that the algorithm performed at 100%.Another experiment was conducted using the PART classifier algorithm, but with a different model training and testing option. Instead of 10-fold cross-validation, the experiment utilized the percentage split model training and testing option. In this case, 80% of the data was used for training the model, while the remaining 20% was used for evaluating the trained model's performance.

Table 4.7: The generated results of PART percentage split



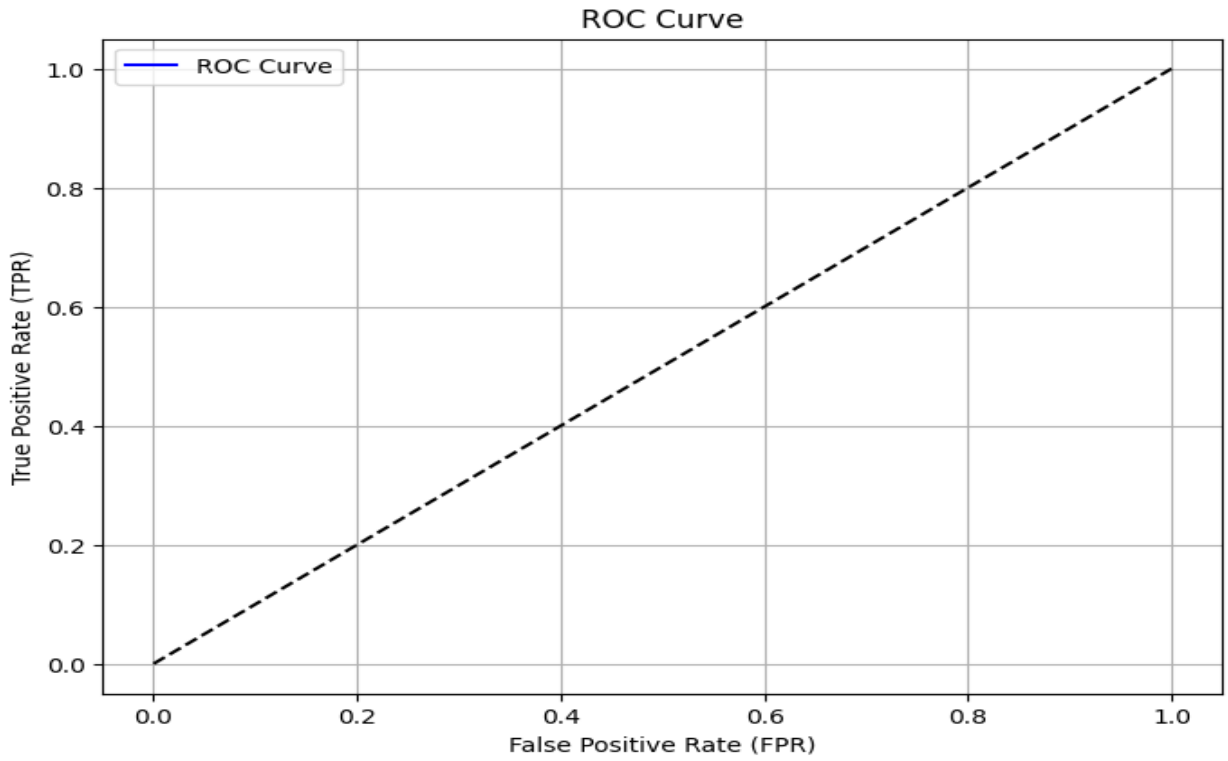


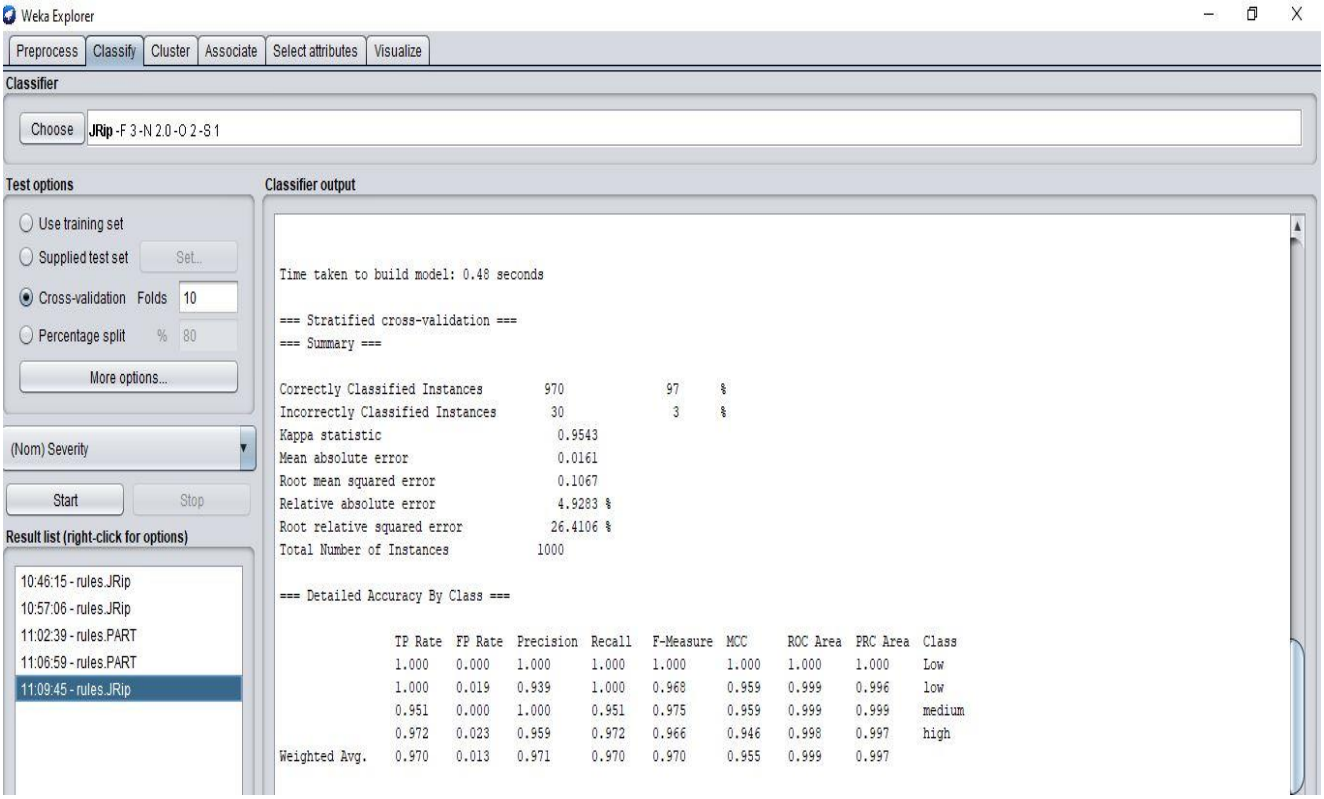
Figure 4.4: ROC curve for PART classifier experiment result

The results of the experiment using the PART rule classifier algorithm demonstrate its effectiveness. The algorithm accurately classified 200 cases (100%) and misclassified only 0 cases (0%) into other classes. The algorithm took 0.07 seconds to build the model.

4.11.3. Experimentation Three (JRIP)

In this specific experiment, the JRIP classifier algorithm was utilized to construct a classifier model. Similar to previous experiments, a 10-fold cross-validation technique and an 80/20 percentage split model training and testing option were employed. Therefore, in this particular experiment, the default choice of using 10-fold cross-validation was applied.

Table 4.8: The generated results of JRIP 10 fold cross validation



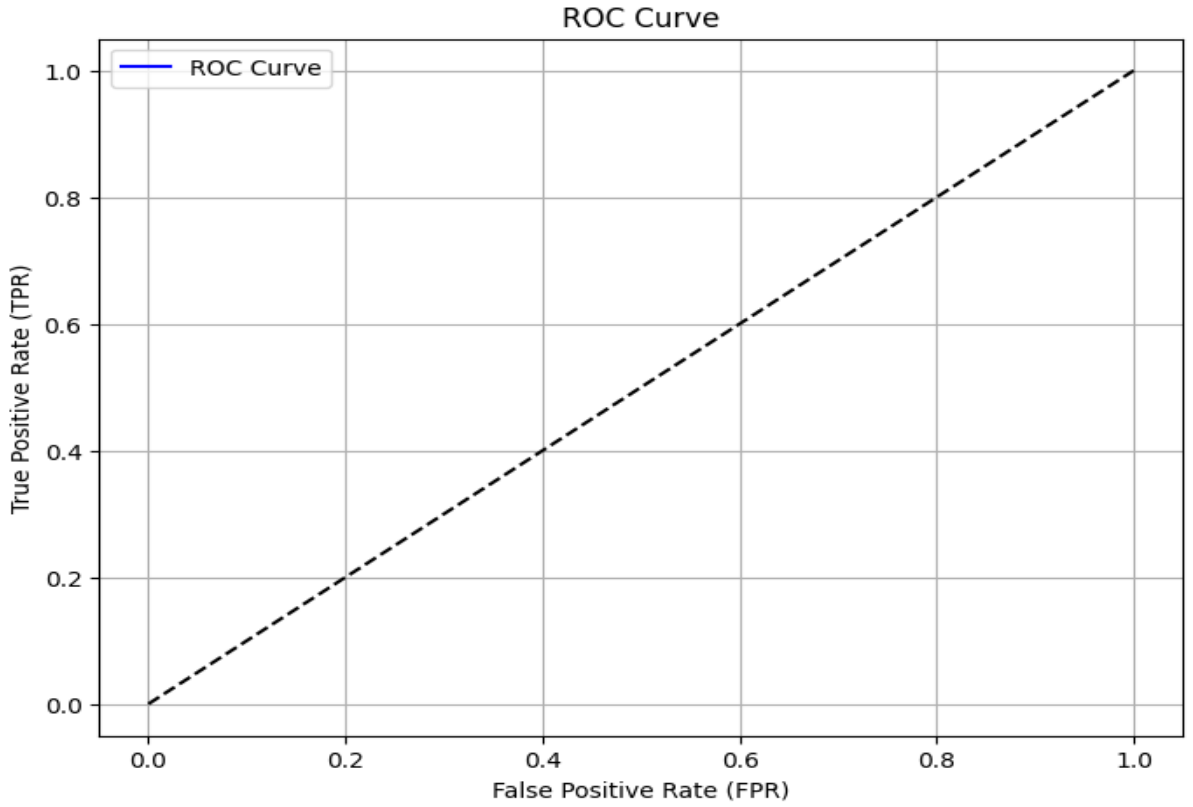
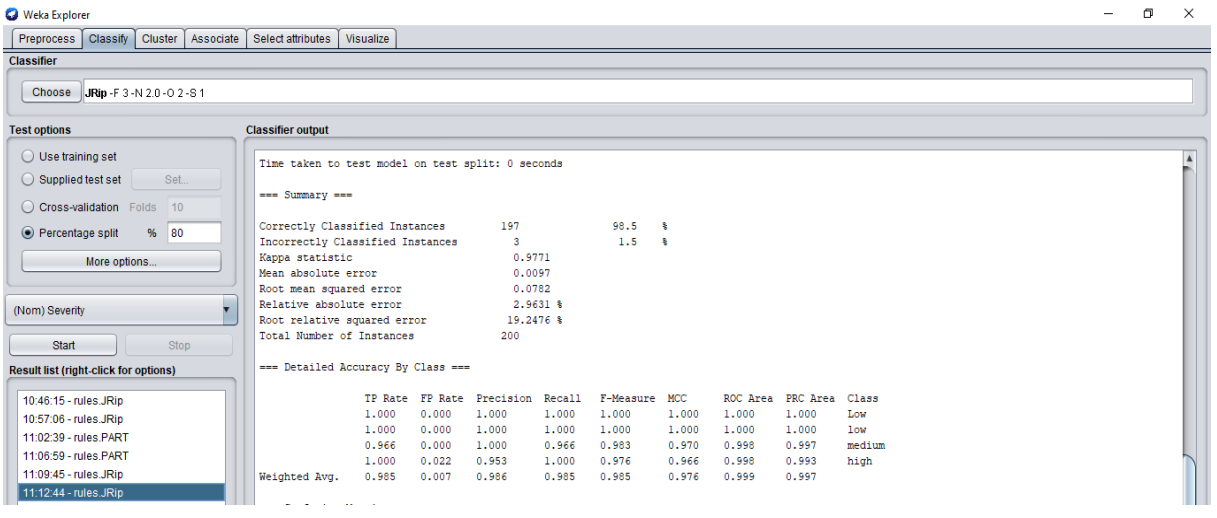


Figure 4.5: ROC curve for 10 fold cross validation JRIP classifier experiment result

The experiment involving the JRIP classifier algorithm was expanded by incorporating a different model training and testing approach known as the percentage splitter. In this approach, 80% of the data was allocated for training the model, while the remaining 20% was utilized for testing purposes.

Table 4.9: The generated results of JRIP percentage split



After build the model the next step or phase on KDD data mining process model was evaluation of the specific model building algorithm performance and accuracy. Therefore the following table 4.9 shows the experiment result and evaluation of JRIP classifier conducted with percentage splitter option.

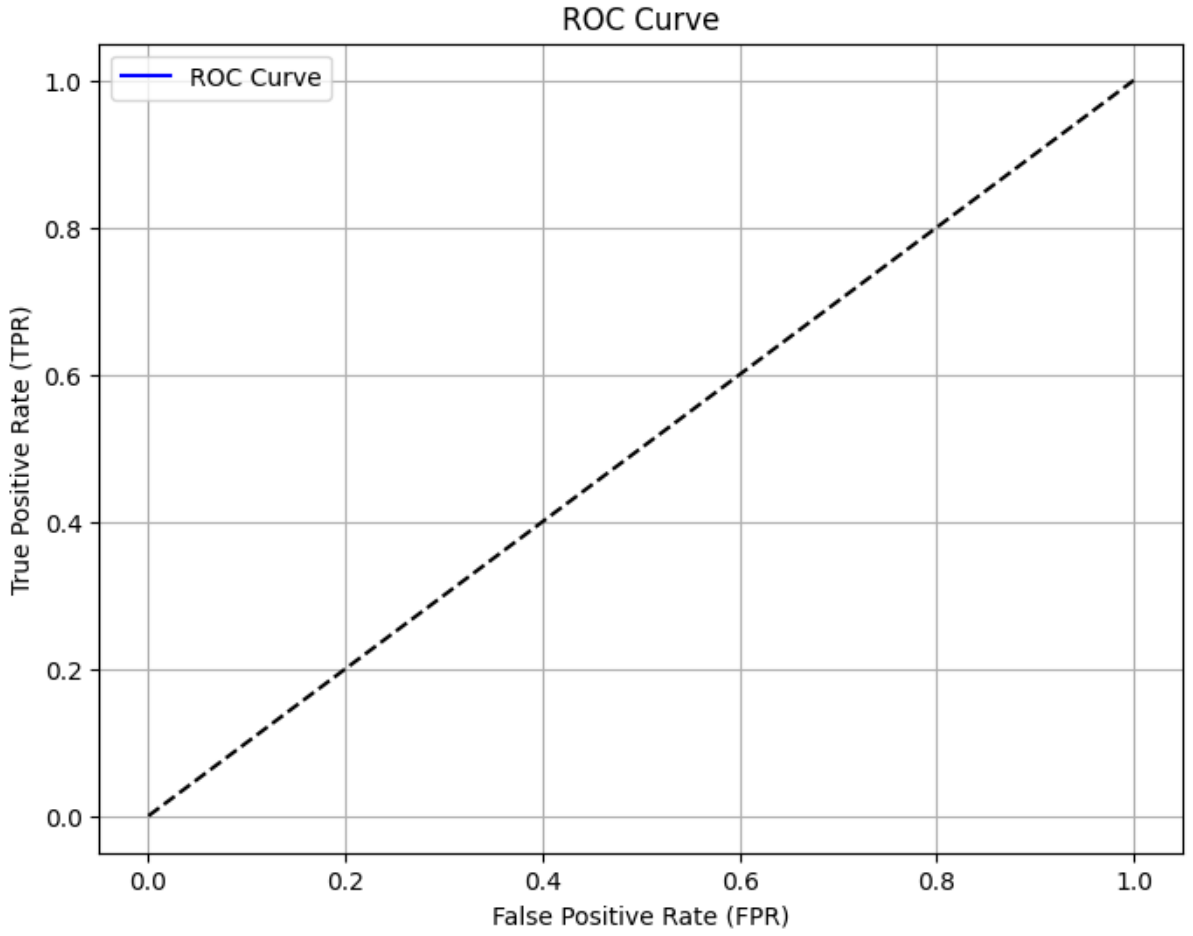


Figure 4.6: ROC curve for JRIP classifier algorithm

In the aforementioned experiment, the JRIP classifier algorithm achieved an accuracy of 98.5% and completed the construction of the classifier model in just 0.47 seconds. Out of a total of 200 test cases, the algorithm accurately classified 197 cases (98.5%) into their respective classes, while 3 cases (1.5%) were misclassified.

4.12. Comparison of classification algorithm

After completing the model construction and testing, it is crucial to choose the most effective classifier model to achieve more precise outcomes in the data mining process. In this particular study, the J48 decision tree, PART rule, and JRIP classifier models were assessed using a range of criteria. The researcher evaluated and compared these three algorithms based on objective measures, aiming to obtain accurate cases for case-based systems.

To identify the superior model, metrics such as overall accuracy, the number of correctly and incorrectly classified cases, model building time, precision, and recall rates were employed in this investigation.

Table 4.10: Comparison of classification algorithm

Classification algorithms				
Evaluation criteria	Experiment setting	**J48**	**PART**	**JRIP**
Overall accuracy	10 -Fold cross validation	97%	100%	97%
	80/20 %	98.5%	100%	98.5%
Correctly classified cases	10 -Fold cross validation	970	1000	970
	80/20 %	197	200	197
Incorrectly classified cases	10 -Fold cross validation	30	0	30
	80/20 %	3	0	3
Time taken	10 -Fold cross validation	0.55	0.05	0.52
	80/20 %	0.06	0.04	0.42
TP rate	10 -Fold cross validation	1	1	1
	80/20 %	1	1	1
FP rate	10 -Fold cross validation	0	0	0
	80/20 %	0	0	0
Precision	10 -Fold cross validation	1	1	1
	80/20 %	1	1	1
Recall	10 -Fold cross validation	1	1	1
	80/20 %	1	1	1

4.13. Knowledge Representation

4.13.1. Rule based representation

Rule-based systems, also referred to as production systems represent the most basic type of artificial intelligence. These systems employ rules as a means of encoding knowledge within the system. The concept of rule-based systems relies heavily on expert systems, which aim to emulate the problem-solving capabilities of human experts in domains rich with knowledge. Rather than expressing knowledge in a fixed and declarative manner as a collection of true statements, rule-based systems represent knowledge through a set of rules that dictate actions to be taken or conclusions to be drawn in various situations (Crina, 2021). The knowledge obtained through patient card analysis, generated results of data mining algorithm (PART rules) and input from domain experts is encoded using the IF-THEN format. In this study, a rule-based system technique is employed for knowledge representation, accompanied by a forward chaining inference mechanism, to develop the system. The knowledge base is constructed by integrating the following set of rules.

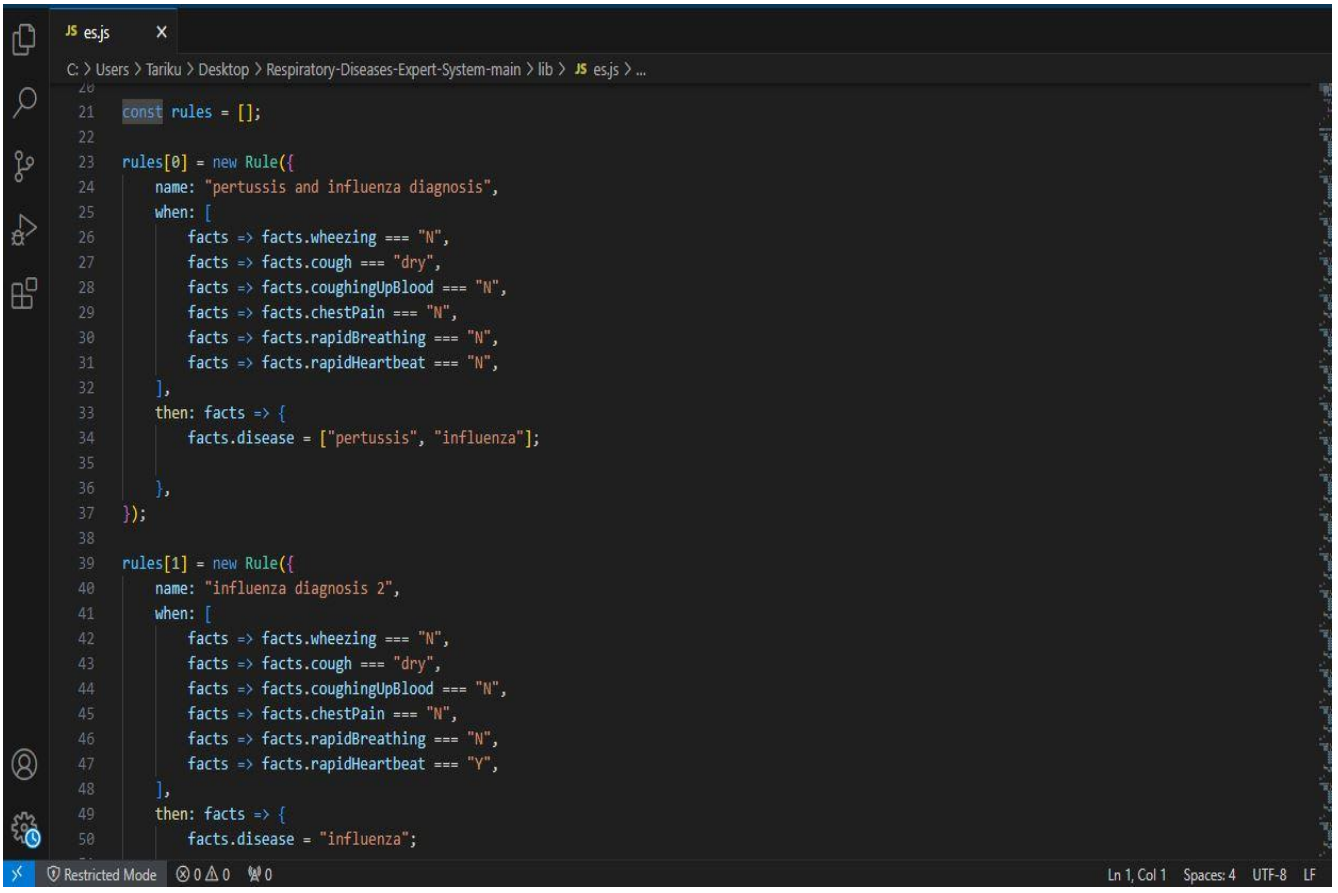


Figure 4.7: The code snapshot for rule based representation

4.14. System implementation and evaluation

The system demonstration serves as an initial representation of the future system, created to showcase its potential applications, feasibility, performance, and approach. The system evaluation is conducted to ascertain whether the proposed system successfully meets the researcher's objectives, while also considering the requirements expressed by the users. In order to achieve the goals of this research, a manual knowledge acquisition method was employed, which involved analyzing documents (such as manuals and regulations for healthcare), conducting interviews with experts from Jimma University Medical Center and gathering information from patient records at the hospital.

In order to develop the system the researcher organized manual knowledge acquired from patient case (cards) and the most important attributes are selected through analysis of patient cases. This manual data also changed in to the format of Excell CSV file for data mining algorithm analysis and the generated results of classification algorithm evaluated. Then, the researcher constructs the rule based reasoning system that contains a set of rules for respiratory diseases diagnosis and treatment. This section provides a presentation of the RDDTS demo along with an examination of its performance.

4.15. Prototype demonstration.

In this study, the development of an expert system for respiratory disease diagnosis and treatment took place by using JavaScript programming language, rools (a small rule engine for Node.js), and Handlebars.js (a semantic template language for JavaScript). Initially, upon accessing the system the user need to navigate to the directory and open the command line then run the NPM (node package management) in order to connect the knowledge base and the user interface. Next to access the main respiratory disease diagnosis and treatment interface the user need to select any browser he/she want to use and write https://Localhost:3000. After the main interface is opened the user is prompted to enter their name and provide information about their symptoms.

When a patient reports respiratory symptoms, the rule-based system examines the given information by applying a set of predefined rules. These rules are specifically designed to identify patterns related to various respiratory illnesses by matching the patient's symptoms

and test results.

The system utilizes forward chaining inference mechanisms to infer the probable diagnosis and suggest appropriate treatment options based on the available data (Knowledge base).

The user interface serves as a means of interaction between the user and the system. Due to the unfriendly nature of the graphical user interface in Prolog, the researcher opted to create the graphical user interface for the proposed system using Node development tools.

This expert system is designed to address language barriers and provide accurate diagnosis and treatment recommendations for patients with respiratory diseases. It utilizes a comprehensive knowledge base of medical information, symptoms, and treatment protocols, and is capable of interfacing with users in three languages: **English, Afaan Oromo and Amharic**.

One of the key features of this system is its multilingual interface. Patients can interact with the system in their preferred language, ensuring that language barriers do not hinder the diagnostic and treatment process. The system guides users through a series of questions to gather information about their symptoms, medical history, and other relevant details, tailoring the questions to the user's selected language.

Using the collected symptom data, the expert system leverages its knowledge base to identify the most likely respiratory disease or condition affecting the patient. It considers factors such as symptom patterns, risk factors, and diagnostic criteria to provide a preliminary diagnosis. Based on the diagnosed condition, the system then offers recommended treatment options, including medication, therapies, and lifestyle modifications, presenting the information in the user's preferred language with clear instructions and guidance. To maintain the system's effectiveness, there is a mechanism in place to regularly update the knowledge base with the latest medical research, treatment protocols, and best practices for respiratory diseases.

This guarantees that the diagnostic and treatment recommendations remain current and relevant. By incorporating these features, the respiratory disease diagnosis and treatment expert system provides a comprehensive and accessible solution for patients, healthcare providers, and the broader community, regardless of their language proficiency. This helps to address the language barriers that often hinder access to quality healthcare, particularly in diverse and multilingual regions.

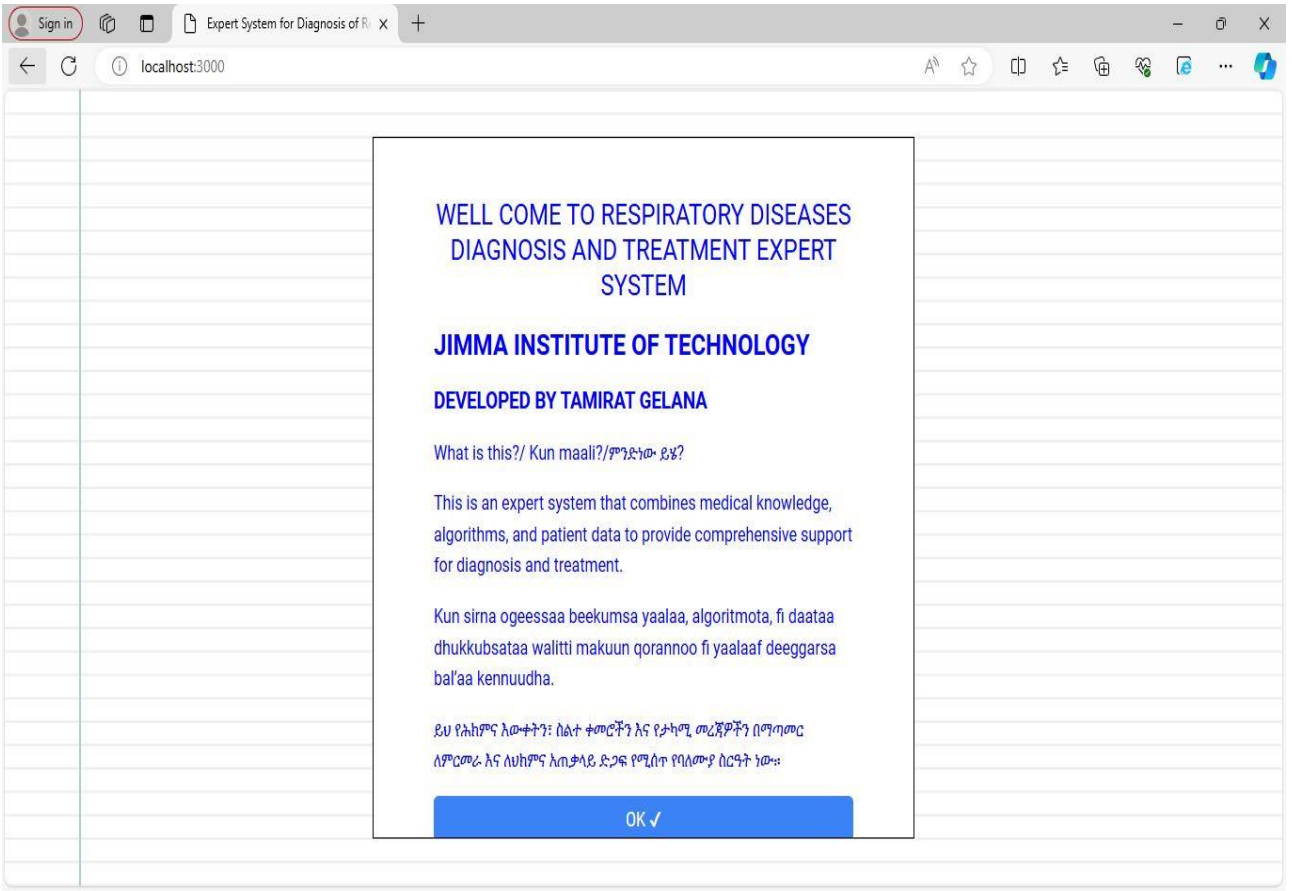


Figure 4.8: User interface of RDDTS

This multilingual approach ensures that users from diverse linguistic backgrounds can fully engage with the system and understand the queries and instructions provided, removing any potential language barriers.

The respiratory disease diagnosis and treatment expert system continues to guide the user (patient) through a comprehensive set of questions to gather essential information for an accurate diagnosis. After the user's name are provided, the system prompts the patient to respond to a series of queries, including their gender, family health history, smoking history, and various respiratory symptoms such as chest pain, coughing, coughing up blood, wheezing, rapid heart rate, rapid breathing, fever, and shortness of breath. The system also inquires about the duration of these symptoms.

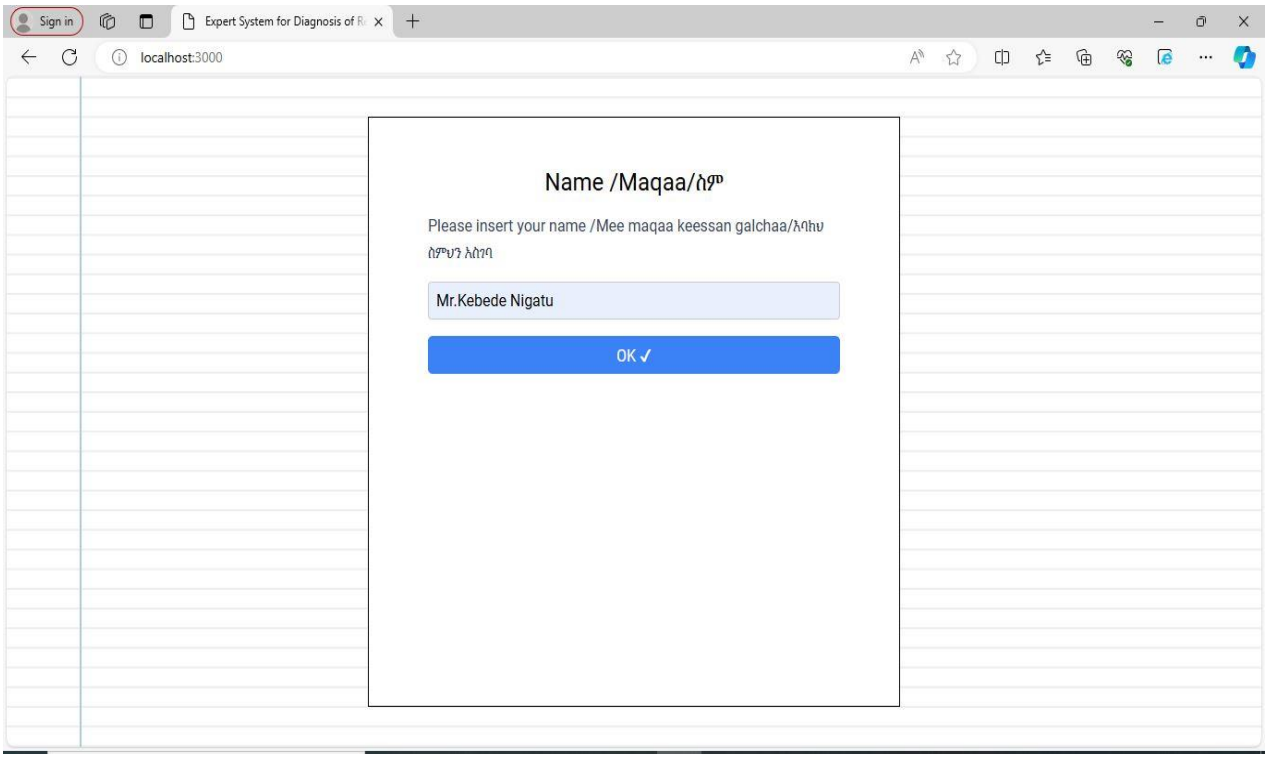


Figure 4.9: User name insertion to the RDDTS

Figure 4.9 shows the respiratory disease diagnosis and treatment system graphical user interface asking the new user to enter his/her name. This intuitive and user-friendly GUI of the respiratory disease diagnosis and treatment expert system begins by prompting the new user to enter their name. This personalized greeting helps to establish a rapport with the user and sets the stage for a tailored interaction. The query for insertion of user name provided with three language options - English, Afaan Oromo and Amharic allowing them to clearly understand the query by their preferred mode of communication.

The next step prompts the user to enter their age, a crucial piece of information that the system will use to tailor its diagnostic process and treatment recommendations based on the patient's stage of life and any age-related factors. The clean and user-friendly design of this graphical interface guides the user through the initial setup, setting the stage for a comprehensive and personalized respiratory health assessment and treatment plan. By prioritizing language accessibility and age-specific considerations, this expert system demonstrates its commitment to providing inclusive and tailored healthcare solutions.

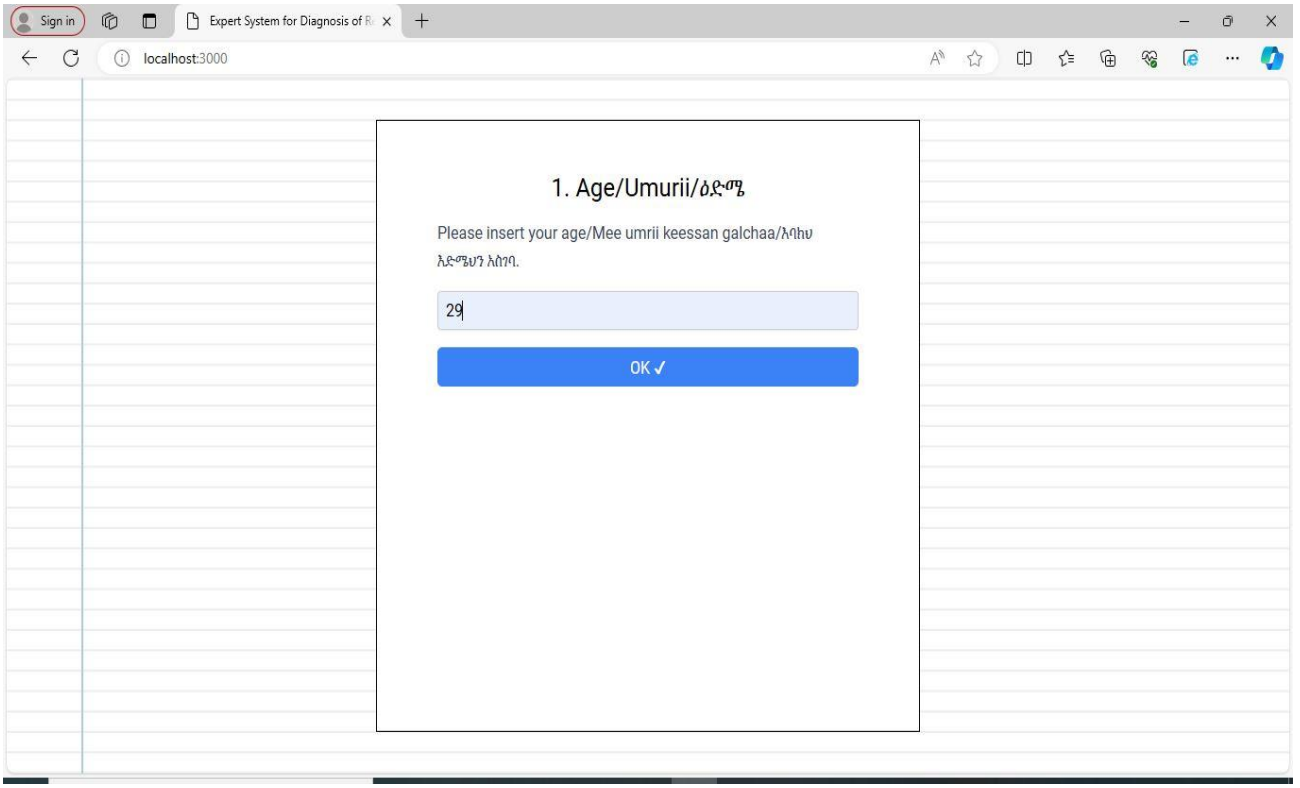


Figure 4.10: User age insertion to RDDTS

The respiratory disease diagnosis and treatment system recognizes the importance of the user's gender in providing accurate and personalized medical assessments and treatment recommendations. As such, the next step in the interface prompts the user to input their gender. This information is crucial for the system to tailor its diagnostic and therapeutic approaches based on the patient's biological characteristics.

By prioritizing language accessibility and gender-specific considerations, the expert system demonstrates its commitment to delivering inclusive and personalized healthcare solutions. This emphasis on inclusivity and personalization underscores the system's dedication to providing high-quality, patient-centric care to users from diverse backgrounds and with varying medical needs.

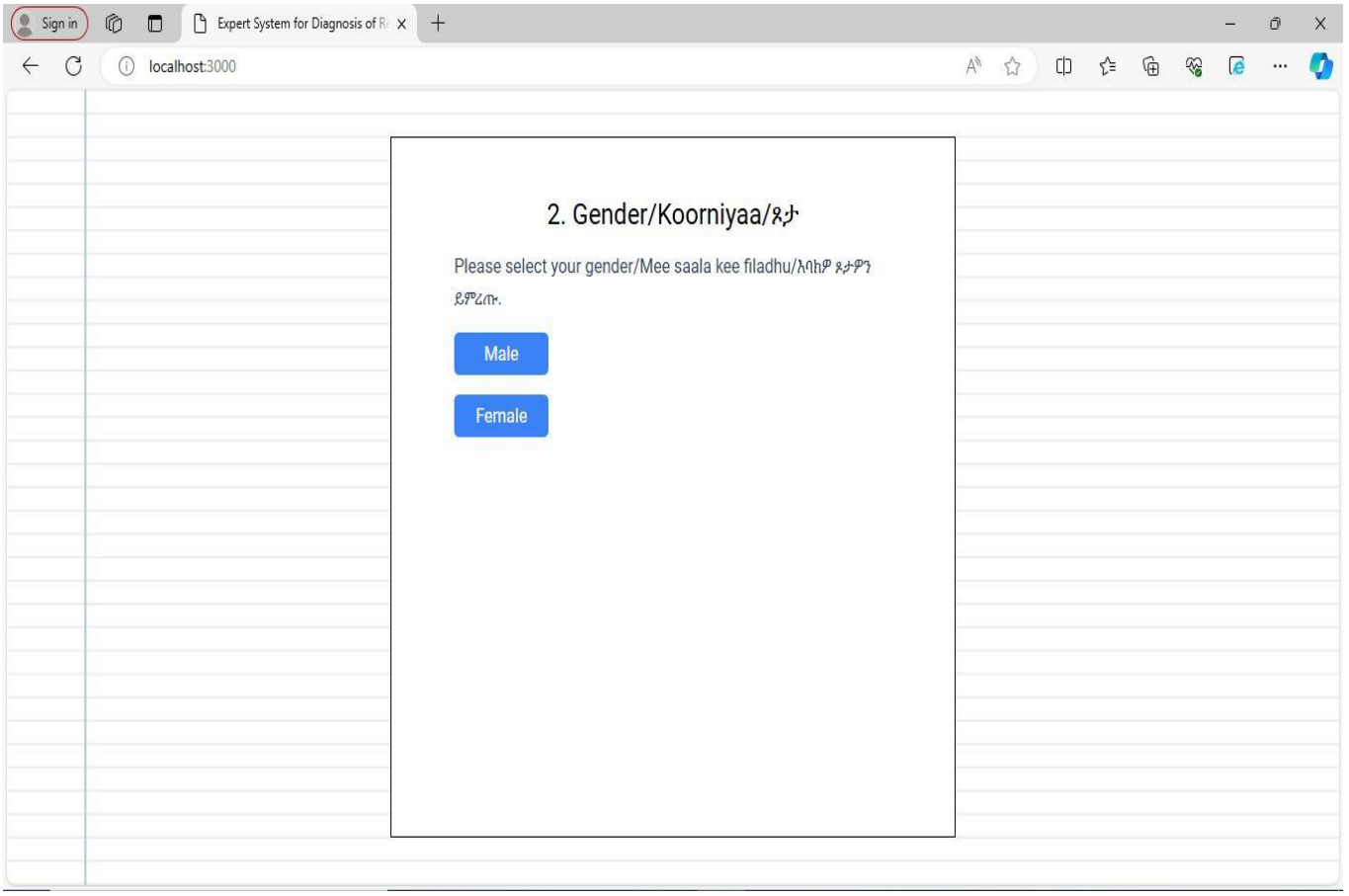


Figure 4.11: User gender insertion to RDDTS

The respiratory disease diagnosis and treatment system presents the user with a visually-appealing and user-friendly graphical interface. The system prompts the user with a crucial question regarding their family health history of respiratory diseases. Displayed in the chosen language, the query asks the user to indicate whether they have any known family members who have experienced respiratory health issues, providing the option to respond either "Yes" or "No."

This family health history information is vital for the system to consider potential genetic or hereditary factors that may influence the user's susceptibility to respiratory ailments. The system can then tailor the diagnostic process and treatment recommendations accordingly, ensuring a personalized and comprehensive healthcare approach.

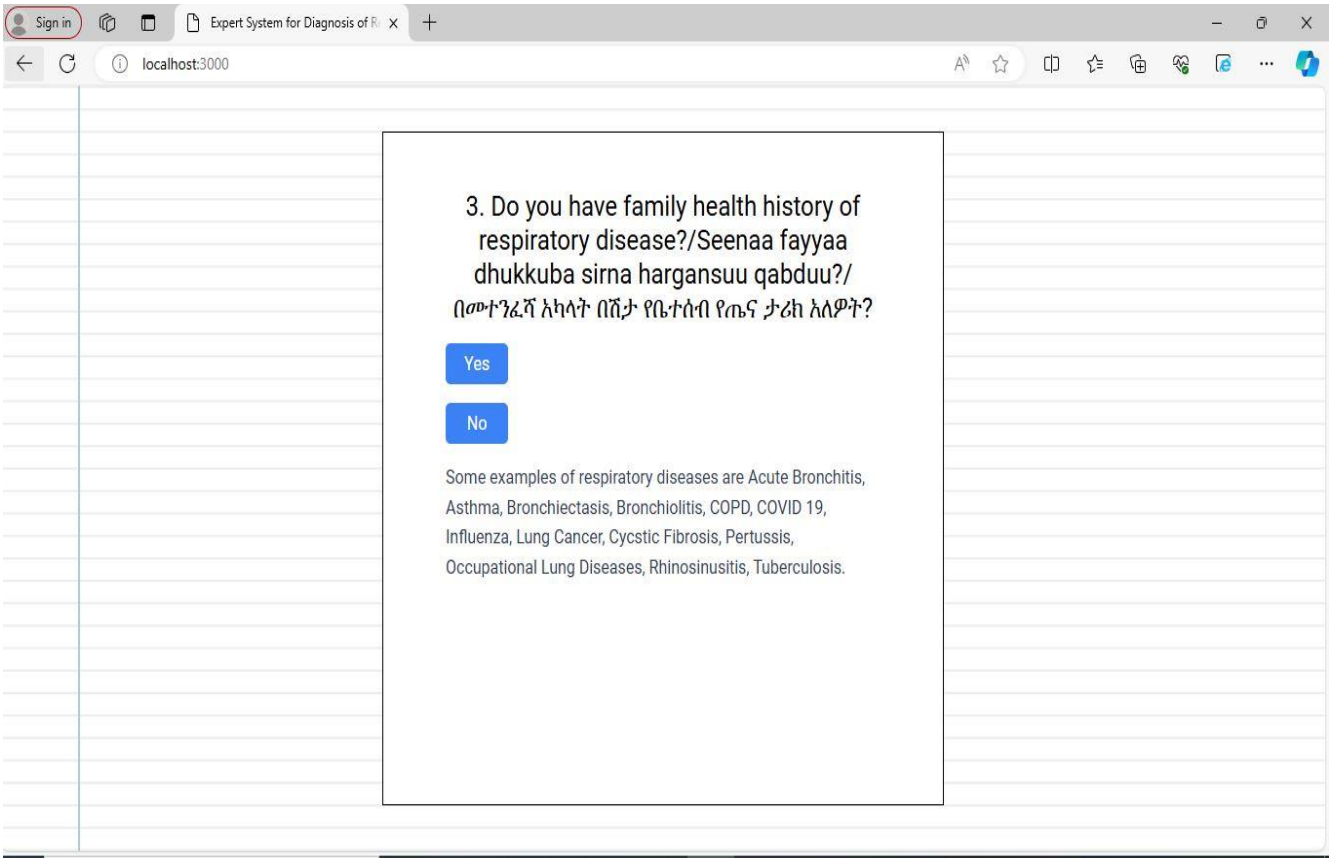


Figure 4.12: Query for insertion of family health history

The respiratory disease diagnosis and treatment system greets new users with a visually attractive and easy-to-use interface. Specifically, the system asks the user to indicate whether they are currently experiencing a cough using their preferred language from the options of English Oromo and Amharic, providing clear and concise response options of "Yes" or "No" displayed in the chosen language. This information is essential for the system to accurately assess the user's respiratory condition and potential respiratory disease indicators.

By offering the query in multiple languages, the system ensures that users from diverse linguistic backgrounds can seamlessly navigate and provide the necessary health information. This allows the system to tailor the diagnostic process and treatment recommendations accordingly, delivering personalized and comprehensive healthcare solutions.

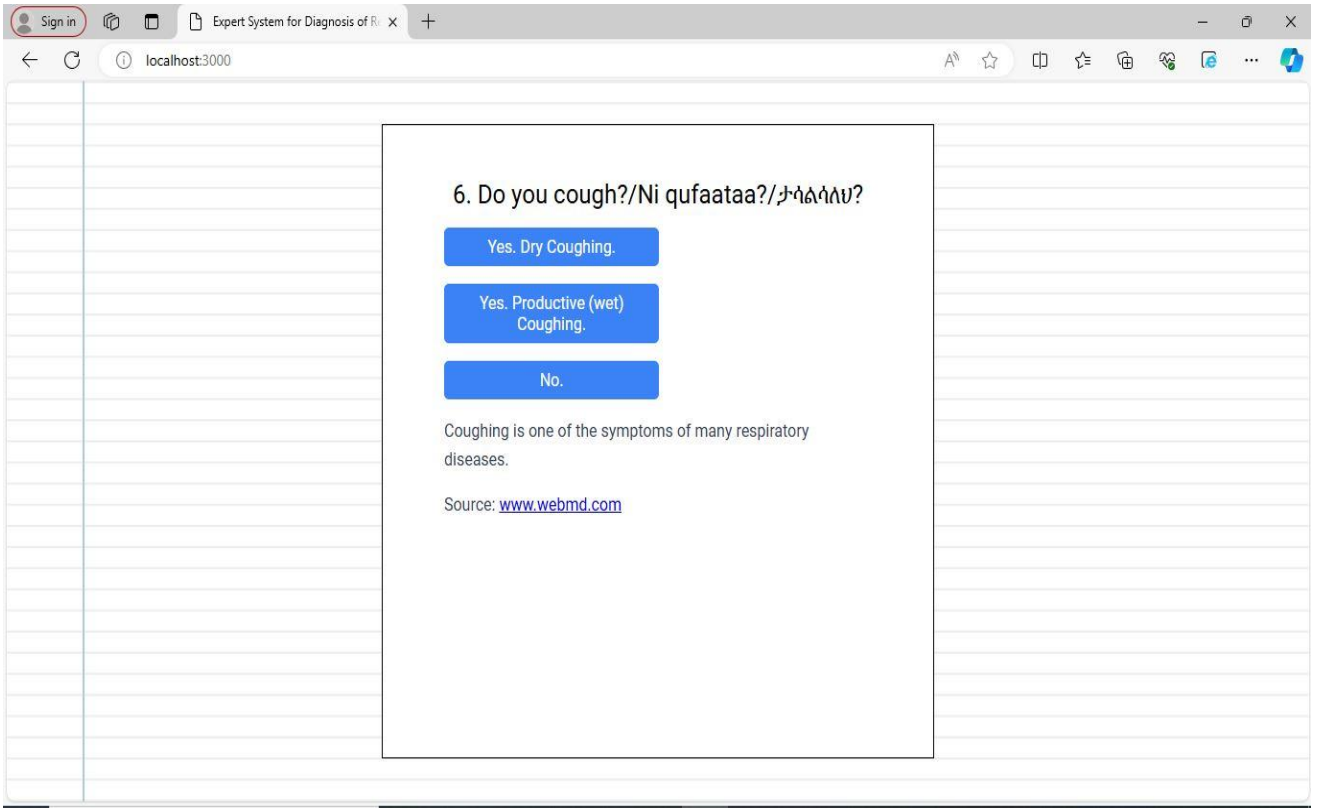


Figure 4.13: Insertion of patient symptom to RDDTS for diagnosis

Finally, the system prompts the user with a key question about the duration of their respiratory symptoms. Specifically, the system asks the user to indicate how long they have been experiencing these symptoms, providing clear response choices of "Less than one week", "Between one week and one month", and "A few months to a few years" in the user's chosen language. This information is crucial for the system to accurately assess the progression and severity of the user's respiratory condition, enabling it to tailor the diagnostic process and recommend appropriate treatment options. By offering this query in multiple languages, the system ensures that users from diverse linguistic backgrounds can seamlessly navigate and provide the necessary health details, reinforcing the system's commitment to delivering personalized and comprehensive healthcare solutions to all individuals.

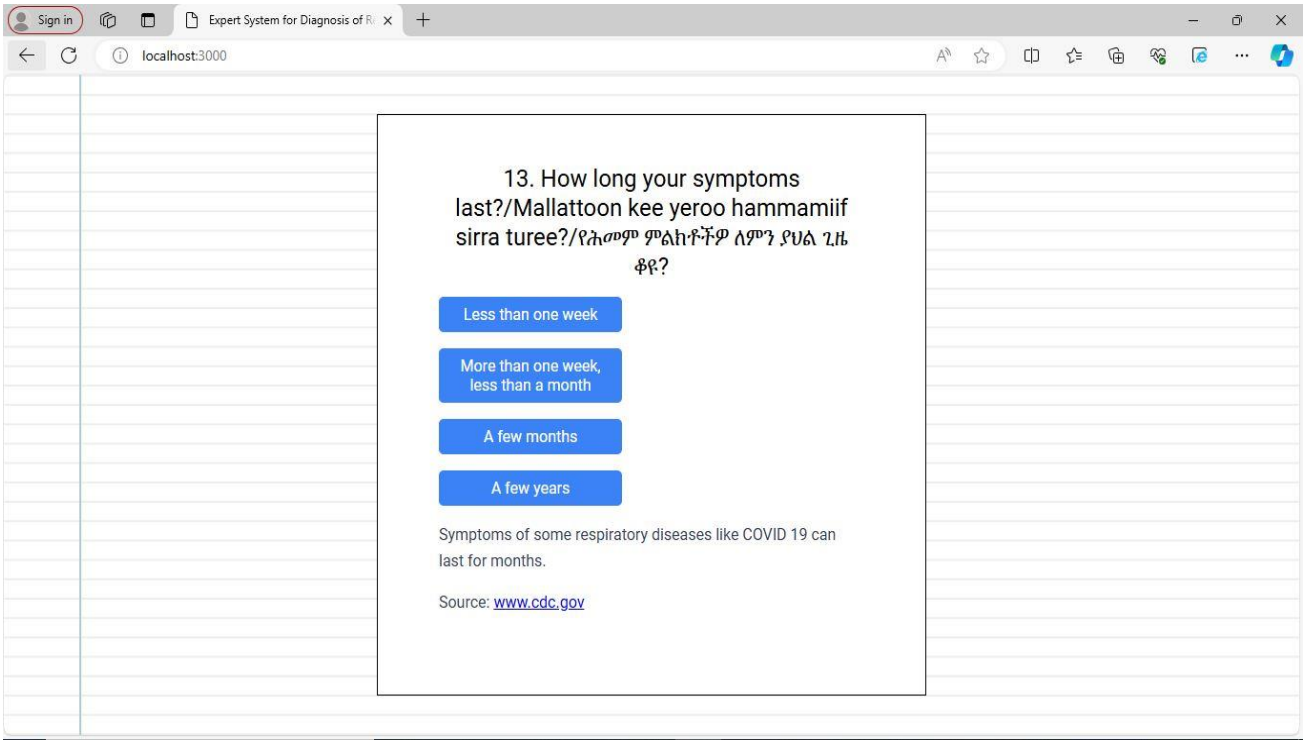


Figure 4.14: User (patient) symptom query

The user is prompted to input the specific symptoms they are experiencing, such as cough, shortness of breath, chest pain, or wheezing. They can also provide details on the duration, severity, and any other relevant factors about these symptoms. Additionally, the system collects information on the user's overall health status, any pre-existing conditions, and recent travel or exposure history.

Using advanced diagnostic algorithms, the system analyzes the user's inputted data and generates a comprehensive assessment of their respiratory condition. Based on the symptoms and other factors, the system provides the user with a detailed diagnosis, explaining the likely underlying respiratory disease or disorder.

Importantly, the system then offers personalized treatment recommendations tailored to the user's specific diagnosis. These may include suggestions for medications, breathing exercises, lifestyle modifications, or referrals to specialized healthcare providers. The treatment plan is presented in clear, easy-to-understand language, empowering the user to take an active role in managing their respiratory health.

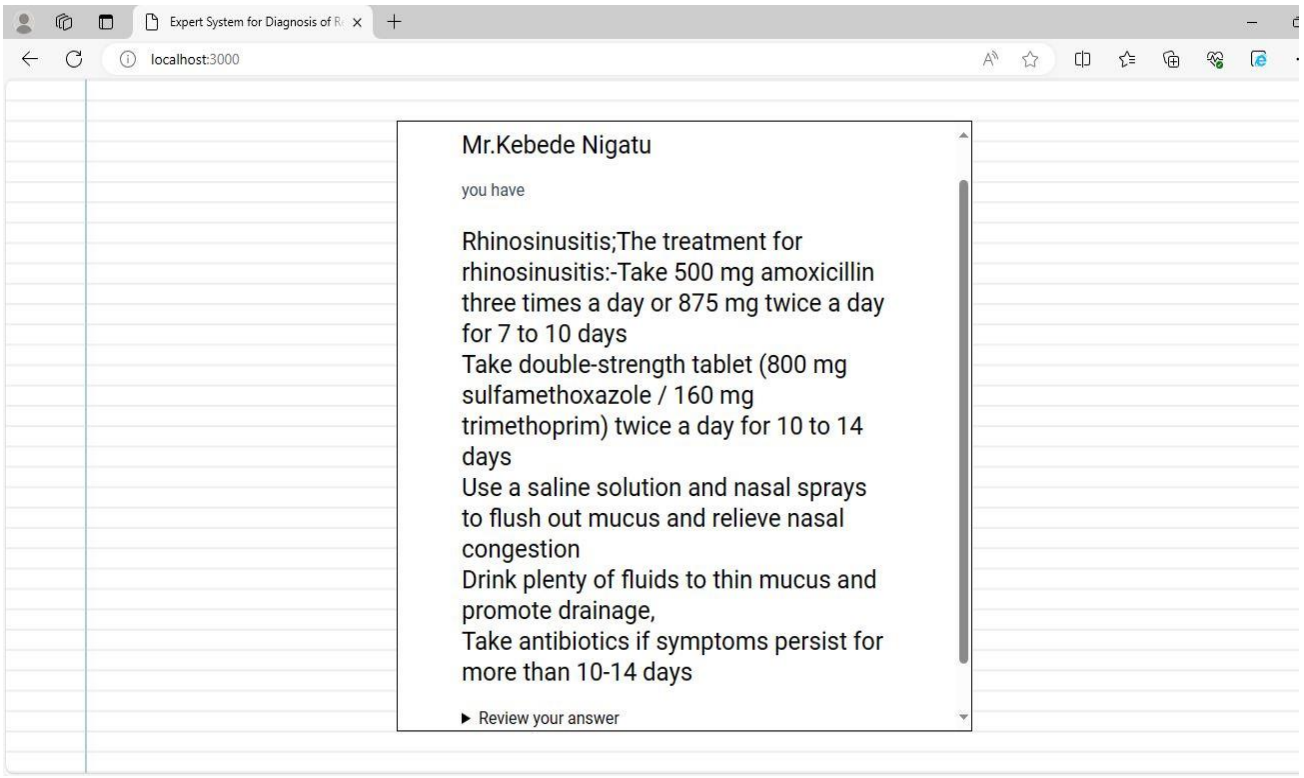


Figure 4.15: Diagnosis and treatment result

Finally if the user is want to back to the homepage he/she can click the try again button and also can review all patient information submitted to RDDT system for diagnosis and treatment.

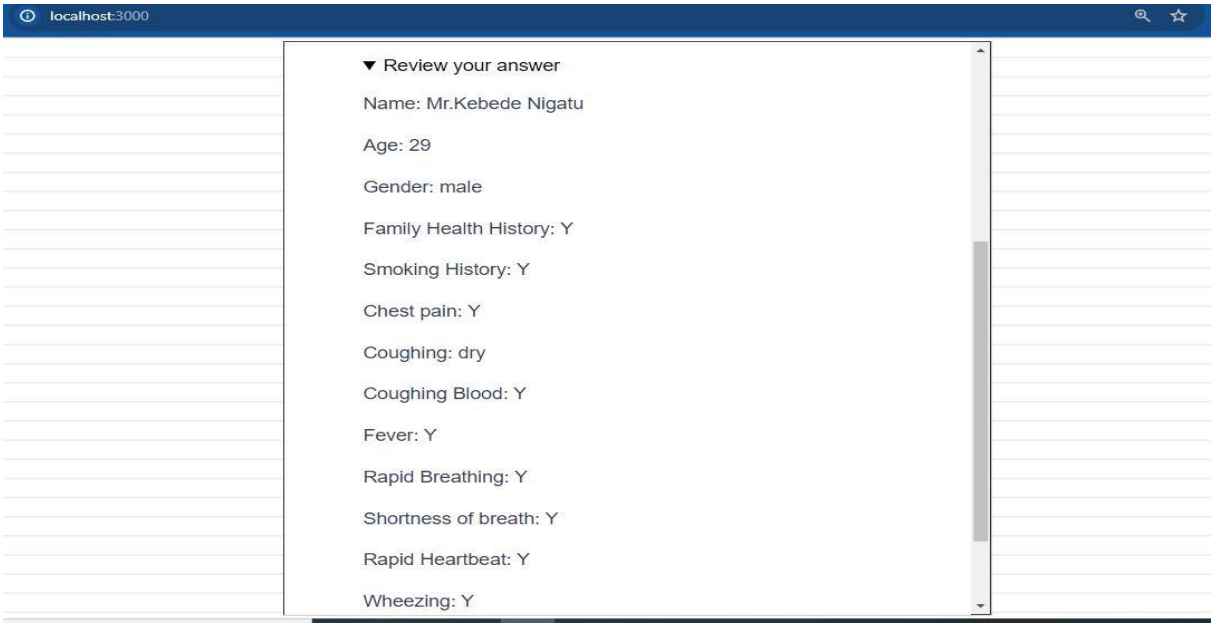


Figure 4.16: Review of user information and symptom

In the realm of expert systems, the incorporation of a self-learning system is a highly valuable feature that allows the system to continuously expand its knowledge. Complementing this knowledge accumulation is the system's pattern recognition capabilities, which employ advanced algorithms to analyze the gathered data identify meaningful relationships and trends, and extract valuable insights. Armed with this expanded knowledge base and enhanced pattern recognition, the self-learning expert system can then improve its reasoning and inference abilities, leading to more informed decisions and more accurate recommendations. Importantly, the system is designed to adapt its decision-making processes, continuously refining and enhancing its performance to better suit the specific needs of its users or the evolving problem at hand. This self-learning, adaptive approach allows the expert system to overcome the limitations of static knowledge bases, becoming a more responsive, versatile, and valuable tool for users seeking reliable and relevant expertise. The next step involves feature extraction, where the system identifies key indicators, such as specific symptoms, vital sign patterns, and imaging markers, and performs dimensionality reduction to simplify the model while retaining essential information. Finally, the system selects appropriate ML algorithms, such as decision trees, neural networks, or support vector machines, and trains the models using the preprocessed and annotated data. This training process can involve supervised learning, unsupervised learning, or reinforcement learning, depending on the nature of the data and the problem at hand.

Self-learning systems for diagnosing and treating respiratory diseases leverage a combination of machine learning (ML), artificial intelligence (AI), and vast amounts of medical data. The next step involves feature extraction, where the system identifies key indicators, such as specific symptoms, vital sign patterns, and imaging markers, and performs dimensionality reduction to simplify the model while retaining essential information. Finally, the system selects appropriate ML algorithms, such as decision trees, neural networks, or support vector machines, and trains the models using the preprocessed and annotated data. This training process can involve supervised learning, unsupervised learning, or reinforcement learning, depending on the nature of the data and the problem at hand.

During a patient visit, a patient presents with respiratory symptoms. The healthcare provider inputs the patient's symptoms, vital signs, and any imaging results into the system. The system then analyzes this data, comparing it with learned patterns to generate a list of possible diagnoses along with associated probabilities.

Based on the most probable diagnosis, the system suggests potential treatment options by referencing past successful cases and predicted outcomes. The patient's response to the treatment is subsequently monitored and fed back into the system to refine and improve future recommendations.

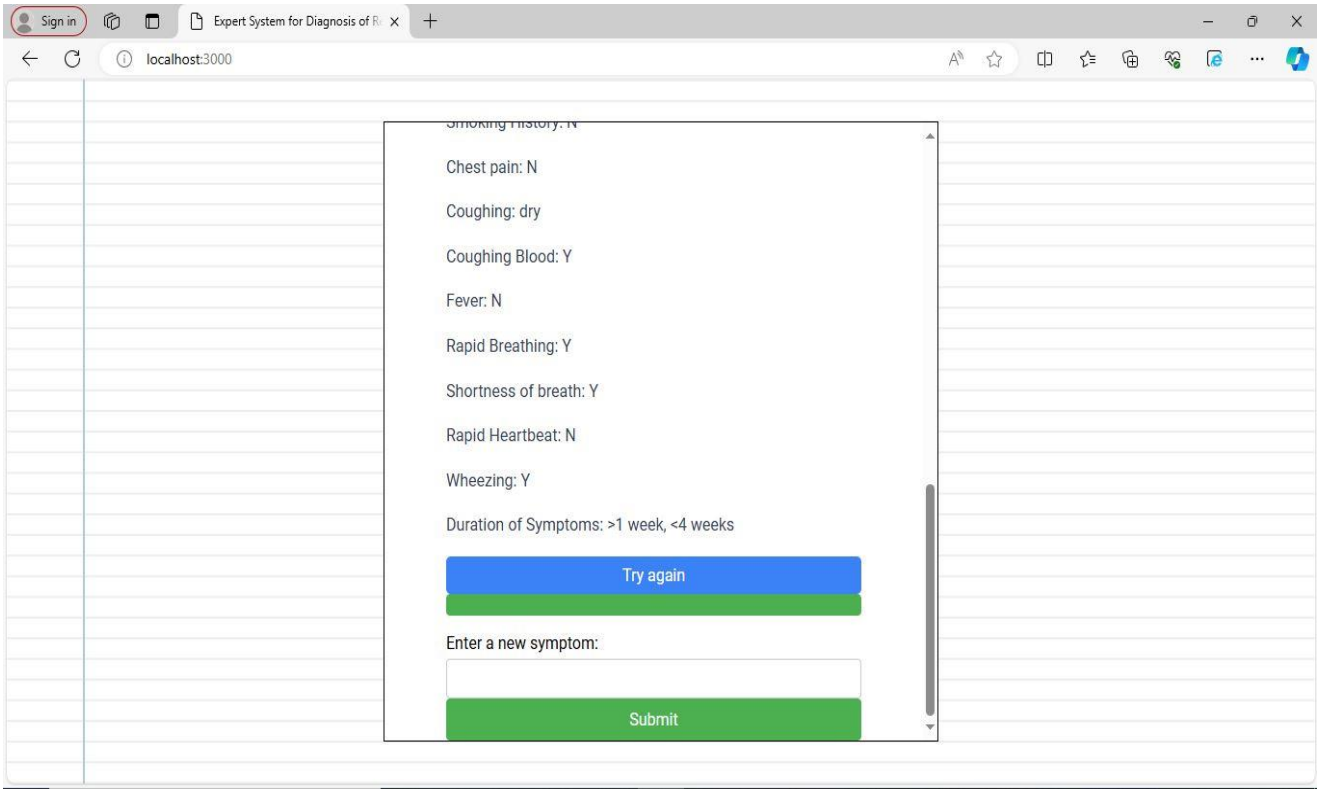


Figure 4.17: Learning part of the system

4.16. User acceptance testing

Domain experts have carried out user acceptance testing. To assess the performance of the knowledge-based system (rule based), a visual interaction evaluation method enables direct interaction between the domain experts and the system. The evaluation of user acceptance in this research employs a checklist mechanism. The checklist was distributed to a group of five domain experts from department of internal medicine and senior nurse workers at Jimma University Medical Center. The evaluators were given the opportunity to rate the options using the following likert scale: Excellent, Very good, Good, Fair and Poor. The reason for using likert scale for user acceptance evaluation is the Likert scale is a widely used method for measuring attitudes, perceptions and opinions.

Table 4.11: Summary of user acceptance testing (Adopted from Daniel, 2021)

Evaluation criteria	Performance Value						
	1	2	3	4	5	Average	Percent
Accuracy of the prototype system in diagnosis and treatment of the respiratory diseases	0	0	1	4	0	3.8	76%
Interactivity of the user interface	0	0	1	3	1	4	80%
Relevance of the system	0	0	0	3	2	4.4	88%
Resource adequacy of the system	0	0	0	3	2	4.4	88%
Attractiveness of the system	0	0	1	2	2	4.2	84%
Does the system have significance contribution for the domain area?	0	0	0	4	1	4.2	84%
Efficiency of the system in time	0	0	0	3	2	4.4	88%
Is the system accurate in analyzing facts and decision making?	0	1	0	3	1	3.8	76%
Is the system's easy to use?	0	0	0	2	3	4.6	92%
Average							84%

Based on the data presented in Table 4:11, the evaluation of the prototype system's performance and user acceptance yielded the following results. In terms of accuracy in diagnosing and treating respiratory diseases, the users rated as the average performance of 3.8 and the overall average performance score for this evaluation measure is 76%.Regarding the interactivity of the user interface, it received an outstanding average rate of 4 indicating that overall average performance score for this evaluation measure is 80%. The relevance of the system was highly regarded, with users assigning it an average rate of 4.4 which indicates overall average performance score for this evaluation measure is 88%. Furthermore, the system's resource adequacy was deemed an average rate of 4.4 and the overall average

performance score is 88% which emphasizing its well-equipped nature. Users found the system to be highly attractive, as indicated by an average rate of 4.2 and the overall average performance score for this evaluation measure is 84%.

The system's significant contribution to the domain area was acknowledged, with users assigning it an average rate of 4.2 which indicate the overall average performance score for this evaluation measure is 84%. When it came to efficiency in time, the system was regarded as an average rate of 4.4 which indicates the overall average performance score for this evaluation measure is 88%. Similarly, the system's accuracy in analyzing facts and making decisions was highly regarded, receiving an average rate of 3.8 which indicates the overall average performance score for this evaluation measure is 76%.Lastly, users praised the system's ease of use and assigning it an excellent average rate of 4.6 which indicates the overall average performance score for this evaluation measure is 92%. These findings demonstrate the positive user acceptance and performance of the system.

In conclusion, the overall evaluation from all respondents (domain experts) indicated that the average user acceptance of the prototype system is 84%.These results suggest that the prototype system exhibits promising potential for the diagnosis and treatment of respiratory diseases, demonstrating its high applicability in this domain.

4.17. System performance testing

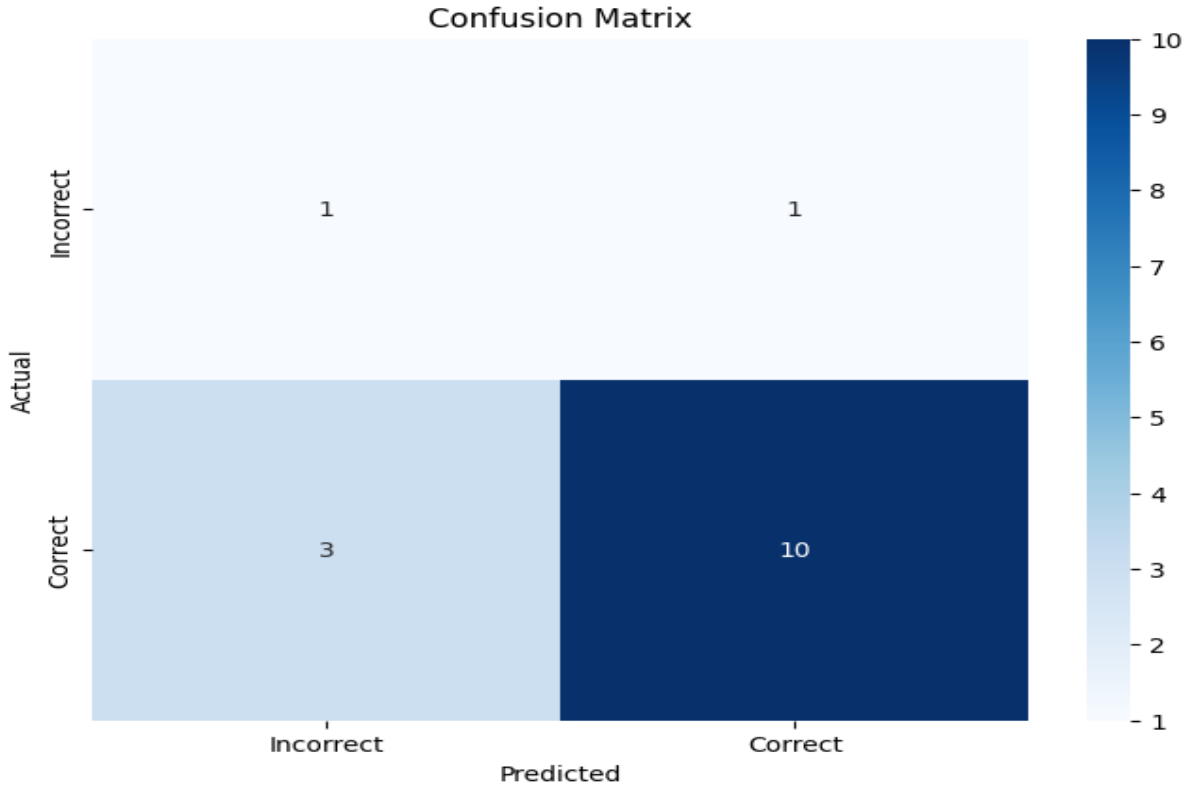
System performance testing is the process of evaluating whether the developed expert system meets the required level of performance and accuracy. It confirms that the right prototype system has been built and is performing as expected.

For the purpose of testing the prototype system, this study selected 3 physicians from the internal medicine department and 2 senior nurses as the evaluators. The criteria for choosing these evaluators were based on their expertise in diagnosing and treating respiratory diseases. The evaluators were selected because they possessed extensive skill and knowledge in the diagnosis and treatment of respiratory diseases, having accumulated this expertise over a long period of time. To thoroughly evaluate the prototype system, the researchers provided 15 respiratory disease test cases and distributed them equally among the 5 evaluators, with each expert receiving 3 test cases to assess.

This approach ensured that the evaluation of the prototype system was conducted by subject matter experts who were well-versed in the domain of respiratory disease diagnosis and management, thereby validating the system's performance against the knowledge and experience of these seasoned healthcare professionals.

In the process of evaluating the performance of the prototype expert system, the domain experts compared the diagnoses and treatment made by the system against their own expert judgments on the same set of respiratory disease test cases. This allowed them to assess the accuracy of the prototype system by identifying the instances where it correctly diagnosed the respiratory diseases, as well as the cases where its diagnoses diverged from the experts' assessments. By conducting this comparative analysis, the domain experts were able to validate the prototype system's performance in terms of its ability to accurately identify and classify different respiratory diseases, relative to the knowledge and clinical expertise of the seasoned healthcare professionals involved in the testing process.

Table 4.12: Confusion matrix of prototype system



The recall, precision, and F-measure were computed using the information from the confusion matrix, which showed the number of true positive, false positive, and false negative predictions made by the prototype system.

Table 4.13: Accuracy of prototype system

	Recall	Precision	F-measure
Results	0.909	0.909	0.905

1. Recall (R) is a performance metric used to evaluate the effectiveness of a search or classification system. It is defined as the proportion of relevant information that has been correctly identified and retrieved, out of the total relevant information available.

The formula for calculating recall is:

Recall (R) = Number of relevant items retrieved / Total number of relevant items

Where:

- "Relevant items retrieved" refers to the number of items that are both relevant and have been correctly identified and retrieved by the system.
- "Total number of relevant items" refers to the total number of items that are actually relevant, regardless of whether they were retrieved or not.

Recall is typically expressed as a value between 0 and 1, or as a percentage between 0% and 100%. A recall of 1 or 100% indicates that the system has retrieved all the relevant information available, while a recall of 0 or 0% indicates that none of the relevant information was retrieved. High recall is desirable in many information retrieval and classification tasks, as it ensures that the system is able to provide users with as much of the relevant information as possible. However, recall must be balanced against precision, which measures the proportion of retrieved items that are actually relevant.

2. Precision (P) is a performance metric used to evaluate the effectiveness of a search or classification system. It is defined as the proportion of retrieved information that is relevant, out of the total information retrieved.

The formula for calculating precision is:

Precision (P) = Number of relevant items retrieved / Total number of items retrieved

Where:

- "Relevant items retrieved" refers to the number of items that are both relevant and have been correctly identified and retrieved by the system.
- "Total number of items retrieved" refers to the total number of items that the system has retrieved, regardless of whether they are relevant or not.

3. F-measure (or F-score) is a performance metric that combines both precision and recall into a single score. It is the harmonic mean of precision and recall, and is calculated using the following formula:

F-measure = 2 * (Precision * Recall) / (Precision + Recall)

Where:

- Precision (P) is the proportion of retrieved items that are relevant.
- Recall (R) is the proportion of relevant items that are retrieved.
- The F-measure ranges from 0 to 1, with 1 being the best possible score.

4. Accuracy is another performance metric used in information retrieval and classification tasks. It is defined as the proportion of correct predictions (both true positives and true negatives) out of the total number of predictions made.

The formula for calculating accuracy is:

Accuracy = (True Positives + True Negatives) / (True Positives + False Positives + True Negatives + False Negatives)

Where:

- True Positives (TP) are the correctly identified positive cases.
- True Negatives (TN) are the correctly identified negative cases.
- False Positives (FP) are the negative cases incorrectly identified as positive.
- False Negatives (FN) are the positive cases incorrectly identified as negative.

The expert system under evaluation demonstrates impressive overall performance based on the provided accuracy metrics. With a recall of 0.909, the system is able to correctly identify 90.9% of the relevant instances, indicating a low false negative rate. Similarly, the precision of 0.909 suggests that 90.9% of the items the system identifies as relevant are indeed relevant, implying a low false positive rate. The F-measure, which combines both precision and recall, stands at a high 0.905, further underscoring the system's excellent accuracy. These metrics

collectively indicate that the expert system is highly effective at the task it was designed for, with a consistently low rate of both false positives and false negatives. The system's performance can be considered quite impressive and suitable for the intended application.

4.18. Discussion of findings

This study aims to address five research questions and has outlined specific objectives based on these questions. The first research question focuses on identifying the domain knowledge available and attributes utilized for the creation of expert system in diagnosing and treating respiratory diseases. To achieve this, the researcher conducted interviews with domain experts, analyzed relevant documents from different sources and gathered cases from patient cards. This study finds out that age, gender, family health history, smoking history, chest pain, coughing, coughing blood, fever, rapid breathing, shortness of breath, rapid heartbeat, wheezing, duration of symptoms and severity are the main attributes to diagnose respiratory diseases.

To answer the second question the researcher acquired knowledge from domain experts at Jimma University Medical Centre, medical literature, research papers, clinical guidelines, and patient cards. Secondly, acquired knowledge organized and the most important attributes are selected through analysis of patient cases.

After identification of attributes and its instances the researcher set out in Excel sheet and prepared for data mining algorithm by saving file as CSV.After this the researcher opened the data mining software WEKA and uploaded the file which is saved as CSV.Next the researcher performed preprocessing and classification by using J48, PART and JRIP algorithm through testing it by 10 fold cross validation and 80/20% and this study finds out that PART classification algorithm achieved high overall accuracy which is 94.5% , 92.5% accuracy for J48 and 90% for JRIP.

Thirdly, acquired knowledge and generated results of PART algorithms represented in a structured format suitable for computer-based systems by using rule based representation techniques. The structured knowledge is then integrated into an expert system framework, encompassing the design and development of software architecture, interfaces, and functionalities required for effective utilization of the domain knowledge in the process of diagnosing and treating respiratory disease.

Finally, thorough validation and testing of the expert system are conducted to ensure its accuracy, reliability, and effectiveness in diagnosing and treating respiratory diseases, involving the use of accuracy, precision measure, and recall and user acceptability test.The researcher conducted an experimentation by using J48, PART and JRIP algorithms in order to select the most accurate algorithm to build the system. After comparing all classification algorithm PART classifier was selected since it achieved high overall accuracy than all others. System testing performed in terms of recall, precision and F-measure registered 90.9%, 90.9% and 90.5% respectively. User acceptance testing also performed by involving domain experts and an average of 84% acceptance was achieved.

In conclusion, an assessment were conducted to evaluate the performance of the suggested system and to gauge user acceptance. As mentioned in the evaluation section, the proposed system yielded promising outcomes, demonstrating a high level of performance. The system's effectiveness in diagnosing and treating respiratory diseases was supported by the research's hypothesis and experimental findings.

According to Aseffa's research in 2020, a Hybrid Expert System that combined Fuzzy Logic and Rule-based approaches achieved an accuracy rate of 80%. In contrast, the study conducted by Evren and Sabriye in 2016 reported that a rule-based expert system achieved an accuracy rate of 64.2%.

Table 4.14*:* Comparison of RDDT with the previous related works

Authors	Technique of reasoning	Accuracy	Recall	Precision
Aseffa, 2020	Fuzzy Logic and Rule-based	80%	Not specified	Not specified
Evren and Sabriye, 2016	Rule-based reasoning	64.2%	Not specified	Not specified
Li, 2022	CNN	82%	Not specified	81%
RDDT	Rule based reasoning	86.7%	90.9%	90.9%

As shown in Table 4.14, when comparing the outcomes of the three aforementioned studies with the rule based respiratory disease diagnosis and treatment it achieved a score of 86.7% in terms of accuracy. This indicates that the proposed system represents a significant enhancement and improvement in the field of diagnosing and treating respiratory diseases. The study's findings highlight the importance of developing an expert system for respiratory disease diagnosis and treatment.

CHAPTER FIVE
CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

In conclusion, developing an expert system for respiratory diseases diagnosis and treatment is of utmost importance in healthcare. It enables accurate diagnoses by utilizing advanced algorithms and medical knowledge, leading to more precise assessments based on symptoms, medical history, and test results. The system facilitates timely interventions, aiding healthcare providers in quickly identifying and assessing respiratory conditions to prevent disease progression.

In this study, developing an expert system for respiratory diseases diagnosis and treatment: A self-learning system has been designed. The two reasoning systems are Rule-based Reasoning (RBR) and Case-based Reasoning (CBR).This study employed rule based reasoning system and manually recorded patient cards and generated results of data mining algorithm(PART) rules are used to construct rules for a rule based reasoning system.

The creation of an expert system for diagnosing and treating respiratory diseases involved the utilization of JavaScript programming language, rools (a Node.js-based rule engine), and Handlebars.js (a JavaScript template language with semantic features).Upon entering the system, the user is requested to provide their name and report their symptoms. The system consistently prompts the user (patient) to provide information regarding their gender, family health history, smoking history, chest pain, coughing, coughing up blood, wheezing, rapid heartbeat, rapid breathing, fever, shortness of breath, and duration of symptoms upon receiving respiratory symptoms from a patient, the rule-based system examines the provided information using a set of predefined rules. These rules are specifically created to identify correlations between the patient's symptoms and test results, and known patterns associated with various respiratory diseases. By utilizing forward chaining inference mechanisms, the system deduces a probable diagnosis based on the available knowledge base and suggests appropriate treatment options.

The performance of the data mining algorithm system shows that 94.5% accuracy with an average Precision of 91.8% and Recall of 92.9% respectively. System testing performed in terms of recall, precision and F-measure registered 90.9%, 90.9% and 90.5% respectively.

User acceptance testing also performed by involving domain experts and an average of 84% acceptance was achieved. The main difficulty in designing the proposed system lies in the limited availability of patient cases to populate the rule base system.

5.2. Recommendation

The current study's results indicate several recommendations to facilitate future advancements and expand the application of Knowledge-Based Systems (KBS) within the healthcare domain:

- In order to address the challenge of obtaining an adequate number of patient case and knowledge for the design of the Knowledge-Based System (KBS), it is recommended to develop a standardized corpus (patient database) that can serve as a foundation for future research in implementing intelligent systems.
- To enhance the widespread adoption of knowledge-based systems by the general public, it is recommended to explore the feasibility of implementing natural language processing (NLP) techniques in developing a user-friendly and localized user interface which is applicable by local languages in addition to Amharic and Afaan Oromo.
- Given the significance of such systems in various health domains, it is advisable for researchers to develop a knowledge-based system that offers diagnosis and treatment solutions for other medical conditions
- In order to increase the accuracy of data mining algorithm, it is recommended to add (collect) more patient case.

References

Aakanksha, J. (2021). Comparative study of JRIP, J48, and Naive Bayes algorithm in flower species prediction. Journal of Physics: Conference Series, 1854(1), 012046. DOI:10.1088/1742-6596/1854/1/012046.

Ajiteru, S. (2020). Design of a problem resolution knowledge-based system for computer diagnosis, repair, and maintenance. International Journal of Engineering Research & Technology, 9(3), 143. DOI: 10.17577/IJERTV9IS030143.

Akerkar, R. (2020). Knowledge-based system architecture (Vol. 2, pp. 67-80). Jones and Bartlett Publishers.

Alfonso, L. (2017). The respiratory system, pleurae, and the mediastinum. In B. J. Sadock, V. A. Sadock, & P. Ruiz (Eds.), Kaplan & Sadock's Comprehensive Textbook of Psychiatry (pp. 243-272). Elsevier. DOI:10.1016/B978-0-323-35775-3.00009-6.

Ali, E. (2022). Conducting an interview in qualitative research: The modus operandi (pp. 46).

Amjad, A. (2022). Big data mining and its challenges: A review. Journal of Computer Science and System Design, 2(2), 302. DOI:10.17303/jcssd.2022.2.302.

Andersson, K. (2016). Evaluation of clinical bronchopneumonia suspicion using a belief rule-based expert system. IEEE Transactions on Fuzzy Systems, 24(3), 0000-0000. DOI:10.1109/FTC.2016.7821675.

Anumeha, S. (2023). Acute bronchitis. Retrieved from https://pubmed.ncbi.nlm.nih.gov/28846312/.

Banda, T. (2019). Community prevalence of chronic respiratory symptoms in rural Malawi: Implications for policy. PLoS ONE, 14(12), e0188437. DOI:10.1371/journal.pone.0188437.

Behzad, H. (2021). Occupational exposures and respiratory symptoms and lung function among hairdressers in Iran: A cross-sectional study. International Archives of Occupational

and Environmental Health. DOI: 10.1007/s00420-020-01645

Bethegaw, D. (2019). Integration of rule-based and case-based reasoning for eye disease diagnosis and treatment.

Bluehost. (2023). what is JavaScript? An introduction to JS for beginners. Retrieved from https://www.bluehost.com/blog/what-is-javascript/.

BSW. (2024). Chronic rhinosinusitis adults and children. Retrieved from https://bsw.icb.nhs.uk/wp-content/uploads/sites/6/2022/06/BSW-CP058-Chronic-Rhinosinusitis-Policy.pdf.

Burnashev, R. (2019). Expert system building tools based on dynamically updated knowledge. Journal of Physics: Conference Series, 1352(1), 012008. DOI:10.1088/1742-6596/1352/1/012008.

Castillo, C. (2022). Machine learning regression and classification methods for fog events prediction. Retrieved from https://doi.org/10.1016/j.atmosres.2022.106157.

Charlotte, P. (2021). An automated data cleaning method for electronic health records by incorporating clinical knowledge. BMC Medical Informatics and Decision Making, 21(1), 21. DOI: 10.1186/s12911-021-01630-7.

Chien, X. (2022). Current status of prevention and treatment of respiratory diseases. DOI: 10.1186/s12890-022-01956-6.

Cordingley, T. (2018). Knowledge elicitation techniques for knowledge-based systems. Journal of Knowledge Management, 16(4), 523-537.

Central Statistical Agency (CSA) [Ethiopia] and ICF. (2016). Ethiopia demographic and health survey 2016.

Diaper, D. (2019). Knowledge elicitation: Principles, techniques and applications. Retrieved from https://dl.acm.org/doi/10.5555/94297.

Diriba, C. (2016). Developing a learning knowledge-based system for diagnosis and treatment

of malaria (Vol. 15).

Diriba, G. (2023). Developing self-learning knowledge-based system for chicken disease diagnosis. Retrieved from https://repository.ju.edu.et//handle/123456789/8921.

Dirriba, A. (2019). Application of knowledge-based system for sustainable agricultural development in Ethiopia (Publication No. 330848755).

Edward, O. (2020). Representing knowledge in artificial intelligence and expert systems. International Journal of Computer Applications, 11(4), 1886-1900.

Eur, J. (2023). Bronchiolitis 2021–2022 epidemic: Multicentric analysis of the characteristics and treatment approach. European Journal of Pediatrics, 182(4), 1921-1927. DOI: 10.1007/s00431-023-04853-0.

Fetene, T. (2017). Developing a knowledge-based system to support debugging computer program source. Retrieved from https://repository.ju.edu.et//handle/123456789/4538.

George, L. (2018). A multi-disciplinary review of knowledge acquisition methods: From human to autonomous eliciting agents, knowledge-based systems, and volume 105. Knowledge-Based Systems, 105, 110-131. DOI:10.1016/j.knosys.2016.02.012.

Gregory, E. (2022). Respiratory health in Africa: Strides and challenges. Journal of the Pan African Thoracic Society. DOI: 10.25259/JPATS_30_2020.

Gurram, B. (2021). Overview of data preprocessing. International Research Journal of Modernization in Engineering Technology and Science, 03(03).

Hagemeister, M. (2019). Knowledge acquisition, training and firm's performance: A theoretical model of the role of knowledge integration and knowledge options. Industrial and Environmental Engineering, 10(3), 127-142.

Hagemeister, M. (2019). Knowledge acquisition, training, and firm performance: A theoretical model of the role of knowledge integration and knowledge options. Journal of Industrial Engineering and Management. DOI:10.1016/j.iedeen.2019.02.003.

Hanisah, B. (2017). Colds and flu - An overview of their management. South African Family Practice, 59(3), 15-22. DOI:10.4102/safp.v59i3.4704.

Hazem, M. (2019). Expert system for pediatric respiratory disease. Retrieved from https://www.researchgate.net/publication/262389728.

Ihab, F. (2016). Effective data cleaning with continuous evaluation. International Journal of Engineering Research and Applications, 6(6), 38-46.

Irina, L. (2017). Applying data mining techniques in healthcare. Journal of Computer Science and Control Systems, 20(3), 385-394. DOI: 10.24846/v25i3y201612.

Jaime, A. (2021). Cystic fibrosis: Current concepts. Boletín Médico Del Hospital Infantil de México, 78(6), 584-596. DOI:10.24875/BMHIM.20000372.

Janice, A. (2018). Investigated the development of an expert system designed for the analysis of pulmonary function data. Concurrency and Computation: Practice and Experience. DOI:10.1002/cpe.6258.

Jilcha, K. (2019). Research design and methodology. DOI:10.5772/intechopen.85731.

Julius, O. (2021). The concept of data mining. DOI:10.5772/intechopen.99417.

Krishna, A. (2019). Design and implementation of a fuzzy expert system for detecting and estimating the level of asthma and chronic obstructive pulmonary disease. Middle-East Journal of Scientific Research, 14(11), 1435-1444. DOI:10.5829/idosi.mejsr.2013.14.11.73192.

Krishna, C. (2021). Croup: A rare manifestation of coronavirus disease. Journal of Child Science, 11(01). DOI: 10.1055/s-0041-1725078.

Lamesgin, A. (2016). Deployment of an expert system for diagnosing respiratory diseases. Retrieved from https://www.researchgate.net/publication/303699292.

Lark, X. (2023). Knowledge engineering in AI. Retrieved from https://www.larksuite.com/en_us/topics/ai-glossary/artificial-intelligence.

Levine, M. (2021). Respiratory diseases. Chest, 120(5), 1153-1154. DOI:10.1016/j.chest.2022.01.014.

Liu, S. (2021). A study on data mining techniques. International Journal of Scientific Research, 12(10), 1287-1292. Retrieved from https://www.ijsr.net/archive/v12i10/SR231014155301.pdf.

Mantan, L. (2021). Pathophysiology of COVID-19. Nature Reviews Immunology, 21(12), 51-764. DOI: 10.1038/s41581-021-00452-0.

Margaret. (2017). what is inference engine?

Mark, H. (2022). The WEKA data mining software: An update. ACM SIGKDD Explorations Newsletter, 11(1), 10-18.

McGraw, L. (2020). Knowledge acquisition, principles and guidelines. Pearson Education.

Md. S. (2019). Designed an expert system utilizing belief rule-based methodology for diagnosing measles in situations involving uncertainty. In Proceedings of the International Conference on Health Informatics and Medical Systems.

Meredith, S. (2018). Occupational asthma: Measures of frequency from four countries. Thorax, 51(4), 435-440. DOI:10.1136/thx.51.4.435.

Mumine, K. (2020). An overview: The impact of data mining applications on various sectors. Retrieved from https://www.researchgate.net/publication/325972292.

Nalepa, J. (2019). Methodologies and technologies for rule-based systems design and implementation: Towards hybrid knowledge engineering. AGH University of Science and Technology.

Neri. (2022). How to implement self-learning system.

Nikam, E. (2017). Role of knowledge base system in higher education. International Journal of Applied Research, 3(11). DOI:10.1002/sres.2782.

Nurettin, O. (2021). Guideline for the management of COPD exacerbations. Breathe, 17(2),

137-176. DOI:10.4103/2452-2473.329630.

Pachpande, S. (2020). Expert system for diagnosis of pulmonary disorders.

Peffers, K. (2018). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45-77.

Pieter, C. (2017). European Respiratory Society guidelines for the management of adult bronchiectasis. European Respiratory Journal, 50(6). DOI:10.1183/13993003.00629-2017.

Poole, D. (2019). Computational intelligence. Retrieved from https://philpapers.org/rec/POOCIA.

Rahman, M. (2020). Knowledge-based systems: A reevaluation. Journal of Knowledge Management, 10(3), 127-142.

Respir, S. (2021). Lung cancer. DOI:https://doi.org/10.1038/s41533-022-00283-x.

Riskiawan, Y. (2019). Development of an expert system for lung diseases using forward chaining. Journal of Physics: Conference Series, 1569(2), 022011. DOI:10.1088/1742-6596/1569/2/022011.

Sabriye, D. (2019). Application of an expert system in medicine for pulmonary diseases. Multidisciplinary Scientific Disciplinary, 3(11), 342. DOI:10.17546/msd.64430.

Sayantini, I. (2024). What is knowledge representation in AI? Retrieved from https://www.edureka.co/blog/knowledge-representation-in-ai/.

Shahd, J. (2023). Development and evaluation of an expert system for diagnosis of kidney disease. International Journal of Academic Engineering Research, 7(6).

Shashank, M. (2022). Designed an expert system for diagnosis of COVID-19 patients. DOI:10.21203/rs.3.rs-2380410/v1.

Sima, C. (2022). Artificial intelligence in technology-enhanced assessment: A survey of machine learning. Journal of Educational Psychology, 116(5), 555-568. DOI: 10.1177/00472395221138791.

Sung-kwan, K. (2020). An integrated view of data: Application of knowledge modeling to data management. Journal of Knowledge Management, 29(2), 72-99. DOI:10.58729/1941-6679.1448.

Tariq, S. (2020). The application of data mining techniques for financial risk. IJCSNS International Journal of Computer Science and Network Security, 20(8).

Wasyihun, S. (2023). Integrating case-based and rule-based reasoning for diagnosis and treatment of mango disease using data mining techniques. DOI: 10.1007/s41870-023-01605-z

Wei, Y. (2021). Knowledge modeling: A survey of processes and techniques. Retrieved from https://doi.org/10.1002/int.22357.

Weston, I. (2021). Using observation to better understand the healthcare context. Quarterly Review of Mobile Healthcare and Medicine, 5(3). DOI:10.4081/qrmh.2021.9821.

World Health Organization. (2020). chronic respiratory disease: Asthma. Retrieved from https://www.who.int/news-room.

William, B. (2023). Disease definition, types and control. Retrieved from https://www.britannica.com/science/disease.

Winstanley, T. (2021). Expert systems in clinical microbiology. Clinical Microbiology, 24(3), 515–556. DOI:10.1128/CMR.00061-10.

Yonas, T. (2019). Creating a model for expert systems to diagnose pulmonary tuberculosis involves designing a pre-screen laboratory process. International Journal of Engineering and Technology, 10(2), 13-18.

Appendix I: Interview Questions

JIMMA UNIVERSITY

JIMMA INSTITUTE OF TECHNOLOGY

FACULTY OF COMPUTING AND INFORMATICS

DEPARTMENT OF INFORMATION SCIENCE (IKM)

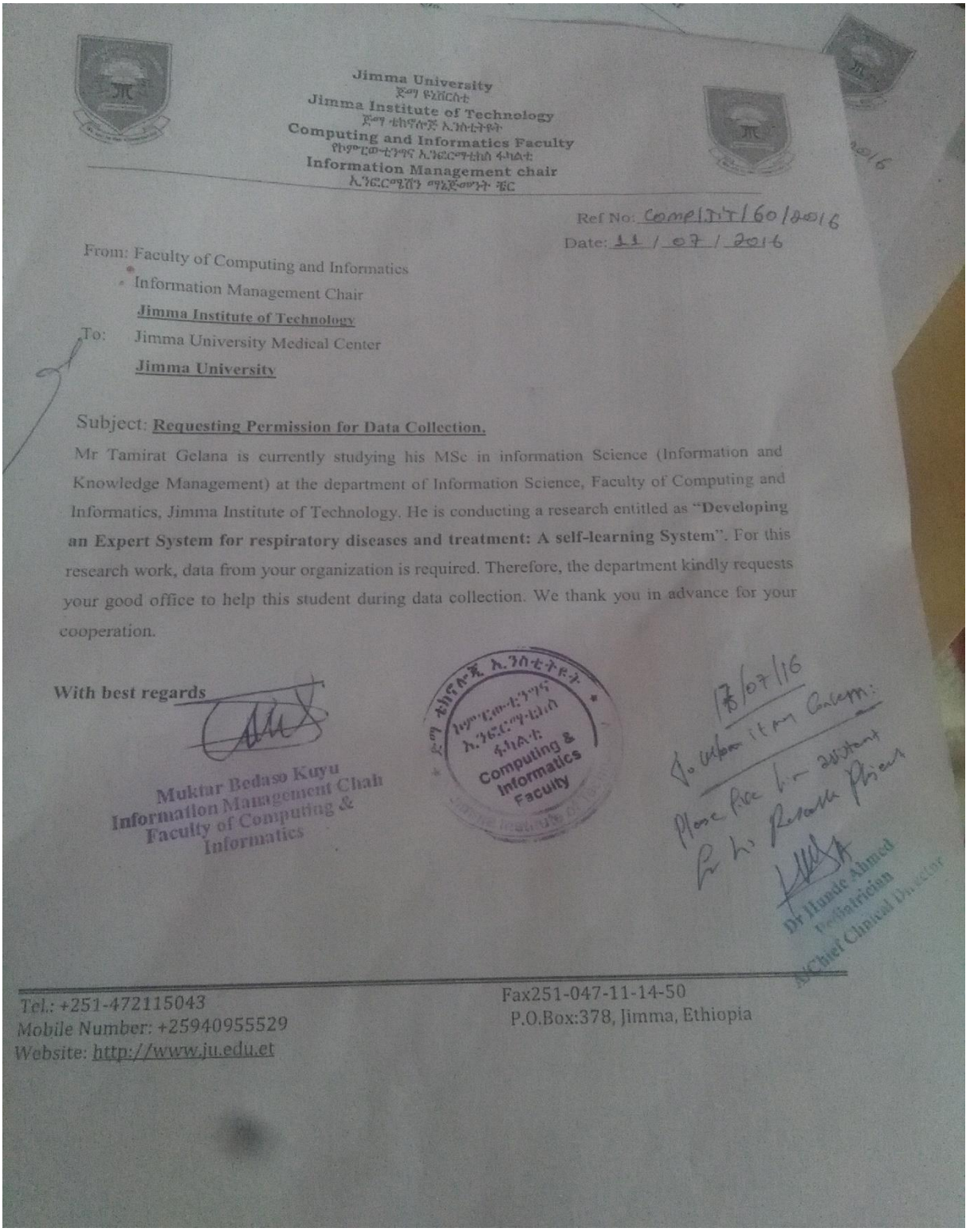
Dear Participants,

To begin, the researcher expresses gratitude for your willingness to participate in the interview. The primary objective of this interview is to gather insights into respiratory diseases. The information gathered will be instrumental in developing a rule based expert system designed to provide diagnosis and treatments for respiratory diseases. This system is specifically tailored for health extension workers, junior nurses, researchers and health officers, who may not possess extensive knowledge of respiratory conditions. Your feedback holds significant importance in ensuring the success of this research, conducted as part of the requirements for the Master of Science in Information Science degree.

Thank you for your valuable contributions to this study.

1. What is respiratory disease?
2. How many respiratory diseases are there in Ethiopia?
3. What are the primary indicators and manifestations of these respiratory diseases, and which symptom is prevalent among most patients?
4. What are the primary methods and protocols for diagnosing respiratory diseases?
5. How do you identify the key symptoms of respiratory diseases?
6. Do you follow any established guidelines for diagnosing patients with respiratory diseases?
7. What treatment options are available for these respiratory diseases?
8. Which attribute are the most important in diagnosing the disease that the clinician should focus respiratory measurement?
9. What are the major difficulties and challenges in respiratory disease diagnosis and treatment?

Appendix II: Letter for data collection



Jimma University
ጅማ ዩኒቨርሲቲ
Jimma Institute of Technology
ጅማ ቴክኖሎጂ ኢንስቲትዩት
Computing and Informatics Faculty
የኮምፒውቲንግና ኢንፎርማቲክስ ፋካልቲ
Information Management chair
ኢንፎርሜሽን ማኔጅመንት ቼር

Ref No: Compl.IIT/60/2016
Date: 11/07/2016

From: Faculty of Computing and Informatics
Information Management Chair
Jimma Institute of Technology

To: Jimma University Medical Center
Jimma University

Subject: **Requesting Permission for Data Collection.**

Mr Tamirat Gelana is currently studying his MSc in information Science (Information and Knowledge Management) at the department of Information Science, Faculty of Computing and Informatics, Jimma Institute of Technology. He is conducting a research entitled as "**Developing an Expert System for respiratory diseases and treatment: A self-learning System**". For this research work, data from your organization is required. Therefore, the department kindly requests your good office to help this student during data collection. We thank you in advance for your cooperation.

With best regards

Muktar Bedaso Kuyu
Information Management Chair
Faculty of Computing & Informatics

ጅማ ቴክኖሎጂ ኢንስቲትዩት
የኮምፒውቲንግና ኢንፎርማቲክስ ፋካልቲ
Computing & Informatics Faculty

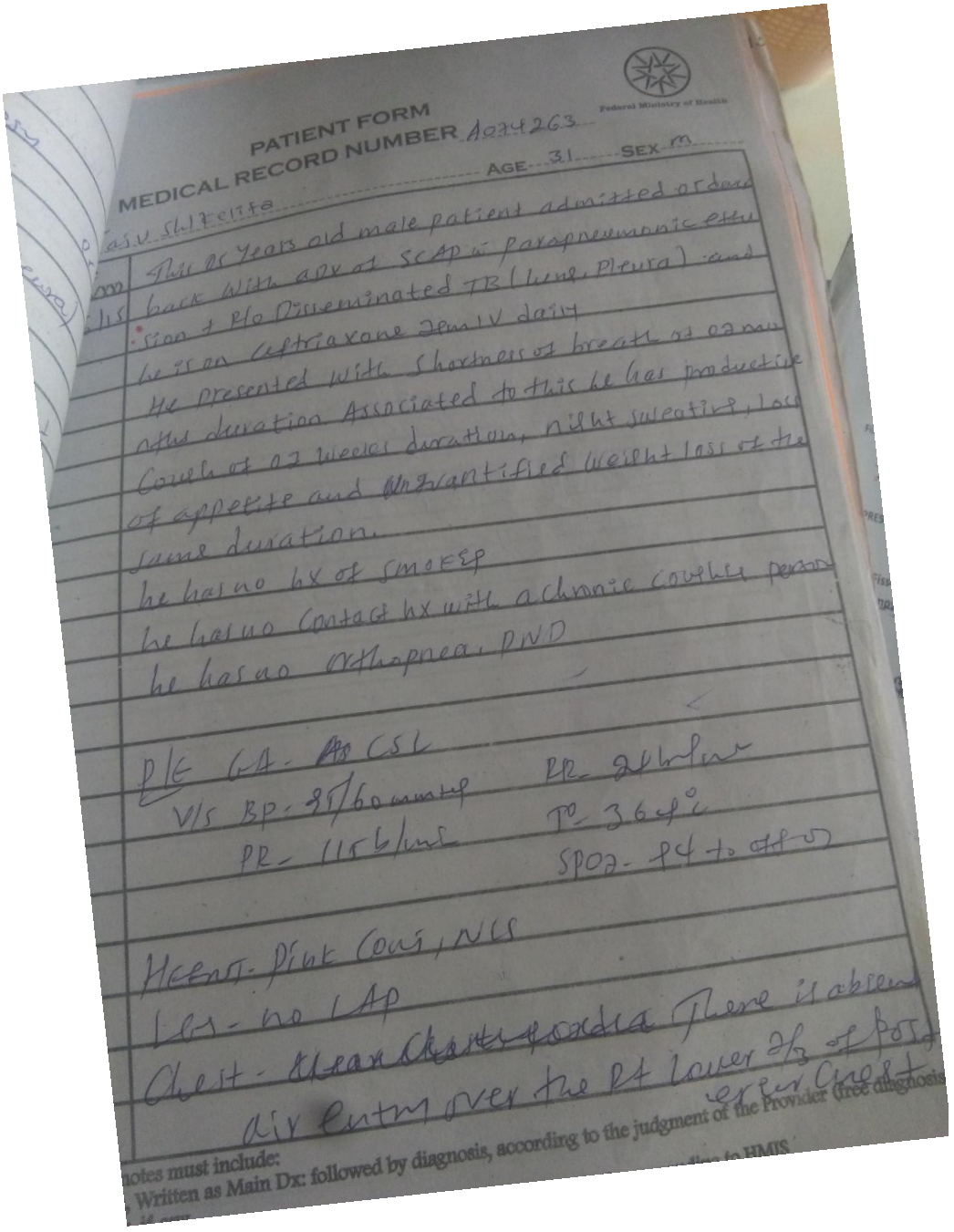
17/07/16
Please Give it my Colleagues
For his Research Please

Dr Hanie Ahmed
Pediatrician
&Chief Clinical Director

Tel.: +251-472115043
Mobile Number: +25940955529
Website: http://www.ju.edu.et

Fax251-047-11-14-50
P.O.Box:378, Jimma, Ethiopia

Appendix III: Sample of patient card



PATIENT FORM

Federal Ministry of Health

MEDICAL RECORD NUMBER A074263 AGE 31 SEX M

CASE V. SLIGHTS

This 31 years old male patient admitted ordered back with dx of SCAP vs Parapneumonic effu sion + R/o Disseminated TB (Lung, Pleural) and he is on Ceftriaxone 2gm IV daily

He presented with shortness of breath of 03 mon ths duration Associated to this he has productive Cough of 02 weeks duration, night sweating, loss of appetite and Unquantified weight loss of the same duration.

he has no hx of smoking

he has no Contact hx with a chronic cough pers on

he has no Orthopnea, PND

P/E GA- Alert CSL

V/S BP- 95/60 mmHg PR- 96b/m

PR- 115b/min T°- 36.4°c

SPO2- 94 to RA

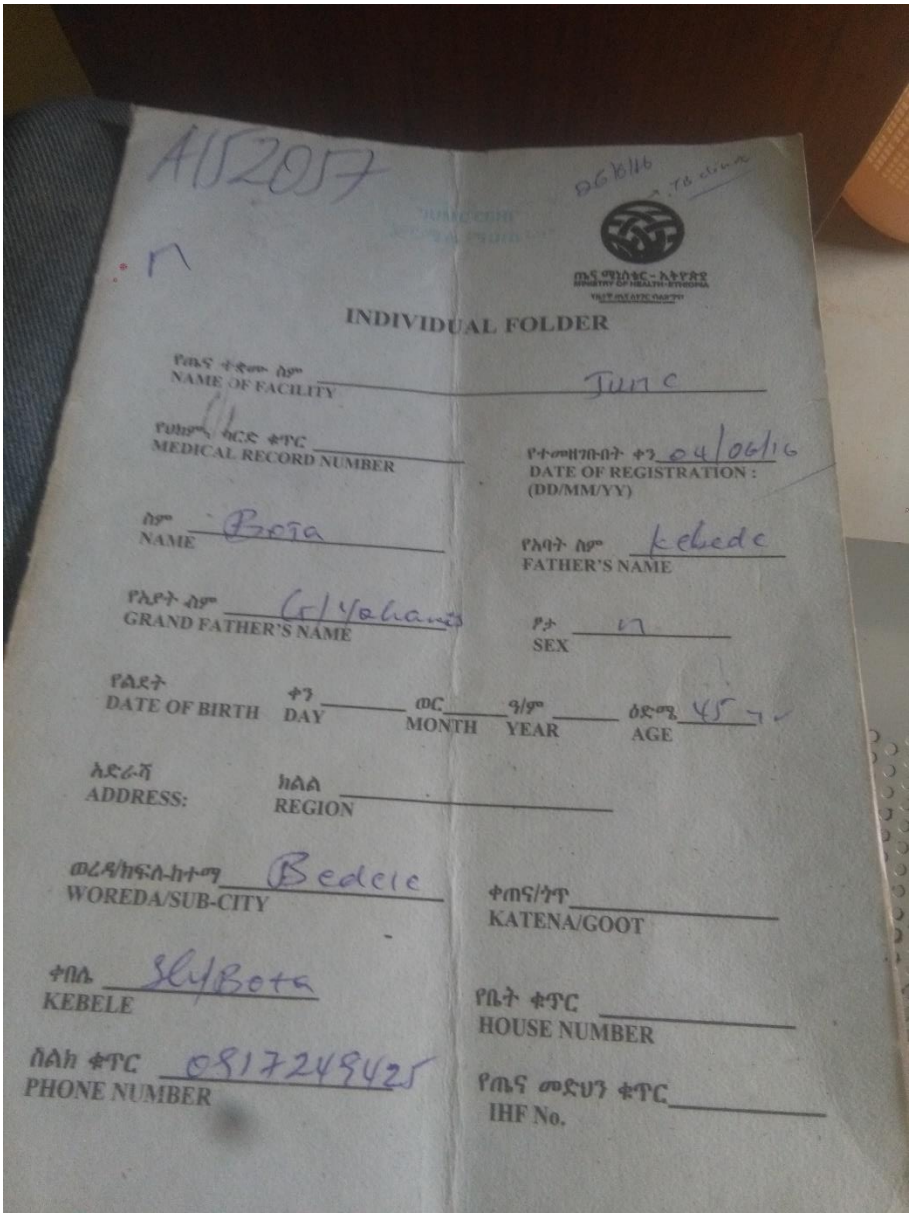
HEENT- Pink Conj, NCS

Lgn- no LAP

Chest- Clear Chest Bilateral There is absent air entry over the Rt lower 2/3 of post. & ant. Chest

Notes must include:
Written as Main Dx: followed by diagnosis, according to the judgment of the Provider (free diagnosis)

Appendix IV: Patient Information (Individual folder)



AIS2057

DG6/16

ጤና ሚኒስቴር - ኢትዮጵያ

INDIVIDUAL FOLDER

የጤና ተቋሙ ስም NAME OF FACILITY Jimc

የህክምና ካርድ ቁጥር MEDICAL RECORD NUMBER

የተመዘገቡበት ቀን DATE OF REGISTRATION : (DD/MM/YY) 04/06/16

ስም NAME Bora

የአባት ስም FATHER'S NAME kebede

የአያት ስም GRAND FATHER'S NAME G/Yohanes

ፆታ SEX M

የልደት DATE OF BIRTH ቀን DAY ወር MONTH ዓ/ም YEAR ዕድሜ AGE 45

አድራሻ ADDRESS: ክልል REGION

ወረዳ/ክፍለ-ከተማ WOREDA/SUB-CITY Bedele

ቀጠና/ጎጥ KATENA/GOOT

ቀበሌ KEBELE Sly/Bota

የቤት ቁጥር HOUSE NUMBER

ስልክ ቁጥር PHONE NUMBER 0917249425

የጤና መድህን ቁጥር IHF No.

Appendix V: Assessment of proposed system

JIMMA UNIVERSITY

JIMMA INSTITUTE OF TECHNOLOGY

FACULTY OF COMPUTING AND INFORMATICS

DEPARTMENT OF INFORMATION SCIENCE (IKM)

Dear participants,

First and foremost, the researcher would like to express gratitude for your willingness to assess the proposed system and provide your valuable feedback. Your feedback is crucial for enhancing the proposed system and ensuring its accessibility to potential users.

The proposed system is a knowledge-based tool designed to offer guidance on the diagnosis and treatment of respiratory diseases. It specifically targets health extension workers, junior nurses, and health officers operating in remote areas of Ethiopia, who may not possess extensive knowledge about respiratory diseases. This survey comprises eight questions aimed at evaluating the proposed system. Please review the questions and select the response that best reflects your opinion from the following options: poor, Fair, Good, Very good and Excellent

Description of the parameter values are as follows.

Performance Value	1	2	3	4	5
Description	Poor	Fair	Good	Very good	Excellent

Instruction: Please, tick on the appropriate value for the corresponding parameter of the knowledge based system in respiratory diseases diagnosis and treatment system user acceptance testing.

Evaluation parameters	Performance Value					
	1	2	3	4	5	Percent
Accuracy of the prototype system in diagnosis and treatment of the respiratory diseases						
Interactivity of the user interface						
Relevance of the system						
Resource adequacy of the system						
Attractiveness of the system						
Does the system have significance contribution for the domain area?						
Efficiency of the system in time						
Is the system accurate in analyzing facts and decision making?						
Is the system's easy to use?						
Average						

Appendix VI: Data acquisition tools

- Patient Surveys and Questionnaires: Structured tools for gathering subjective information from patients regarding their respiratory symptoms, lifestyle, and medical history.
- Mobile Health (health) Apps: Applications designed for smartphones or tablets that allow individuals to track and manage respiratory health.
- Document analysis/Patient record (Previously solved case)
- Interview with domain expert
- Literature review
- Internet sources.

Appendix VII: Sample of rules

```
Rules [5] = new Rule ({
   Name: "croup diagnosis",
   When: [
      Facts => facts.wheezing === "N",
      Facts => facts.cough === "productive",
      Facts => facts.smokingHistory === "N",
      Facts => facts.chestPain === "N",
   ],
   Then: facts => {
      facts.disease = "croup";
Code: 'croup',
 Name: 'The treatment for croup: - Encourage the child to drink plenty of fluids to prevent dehydration and help thin the mucus, Take Oral or inhaled corticosteroids, such as dexamethasone or prednisolone',
   },
});
Rules [6] = new Rule ({
   Name: "tuberculosis diagnosis",
   When: [
      Facts => facts.wheezing === "N",
      Facts => facts.cough === "productive",
      Facts => facts.smokingHistory === "N",
      Facts => facts.chestPain === "Y",
   ],
   Then: facts => {
      facts.disease = "tuberculosis";
Code: 'tuberculosis',
 Name: 'The treatment for Tuberculosis:-Administration of several antibiotics such as isoniazid, revamping, pyrazinamide and ethambutol for the initial phase which lasts for two months'
   },
});
```

```
Rules [7] = new Rule ({
   Name: "rhinosinusitis diagnosis",
   When: [
      Facts => facts.wheezing === "N",
      Facts => facts.cough === "productive",
      Facts => facts.smokingHistory === "Y",
   ],
   Then: facts => {
      facts.disease = "rhinosinusitis";
Code: 'rhinosinusitis',
Name: 'The treatment for rhinosinusitis:-Using a saline solution or nasal sprays to flush out mucus and relieve nasal congestion,  Drinking plenty of fluids to thin mucus and promote drainage, If symptoms persist for more than 10-14 days, a healthcare provider may prescribe antibiotics'
   },
});
Rules [8] = new Rule ({
   Name: "common cold diagnosis",
   When: [
      Facts => facts.wheezing === "Y",
      Facts => facts.smokingHistory === "N",
      Facts => facts.rapidBreathing === "N",
      Facts => facts.chestPain === "N",
   ],
   Then: facts => {
      facts.disease = "commonCold";
Code: 'commonCold',
Name: 'The treatment for commonCold:-Take Over-the-counter medications like acetaminophen (Tylenol, Getting plenty of rest and stay hydrated by drinking fluids'
   },
});

Rules [9] = new Rule ({
```

```
    Name: "acute bronchitis diagnosis",
    When: [
        Facts => facts.wheezing === "Y",
        Facts => facts.smokingHistory === "N",
        Facts => facts.rapidBreathing === "N",
        Facts => facts.chestPain === "Y",
        Facts => facts.coughingUpBlood === "N",
    ],
    Then: facts => {
        facts.disease = "acuteBronchitis";
Code: 'acuteBronchitis',
name: 'The treatment for acuteBronchitis:- Get plenty of rest to allow your body to recover, Drink plenty of fluids, such as water, warm tea, or soup, to stay hydrated, Take Over-the-counter medications like acetaminophen or non-steroidal anti-inflammatory drugs (NSAIDs),
    },
});
Rules [10] = new Rule ({
    Name: "bronchiectasis diagnosis",
    When: [
        Facts => facts.wheezing === "Y",
        Facts => facts.smokingHistory === "N",
        Facts => facts.rapidBreathing === "N",
        Facts => facts.chestPain === "Y",
        Facts => facts.coughingUpBlood === "Y",
    ],
    Then: facts => {
        facts.disease = "bronchiectasis";
Code: 'bronchiectasis',
Name: 'The treatment for bronchiectasis:-Take Long-term prophylaxis to prevent exacerbations, Take inhaled hypertonic saline or muco active agents like dornase alfa, and Take High-frequency chest wall oscillation (HFCWO)',.
    },
});
```

```
Rules [11] = new Rule ({
   Name: "asthma diagnosis 1",
   When: [
      Facts => facts.wheezing === "Y",
      Facts => facts.smokingHistory === "N",
      Facts => facts.rapidBreathing === "Y",
      Facts => facts.cough === "dry",
   ],
   Then: facts => {
      facts.disease = "asthma";
Code: 'asthma',
Name: 'The treatment for asthma:-Quick-relief medications (bronchodilators), Short-acting beta-agonists, such as albuterol, Inhaled corticosteroids, long-acting beta-agonists and leukotriene modifiers.'
   },
});
Rules [12] = new Rule ({
   Name: "copd diagnosis",
   When: [
      Facts => facts.wheezing === "Y",
      Facts => facts.smokingHistory === "Y",
      Facts => facts.chestPain === "N",
      Facts => facts.cough === "productive",
   ],
   Then: facts => {
      facts.disease = "copd";
Code: 'copd',
name: 'The treatment for COPD:-Quit smoking and avoid exposure to second-hand smoke,  Minimize exposure to air pollution, chemical fumes, dust, and other respiratory irritants and  take Inhaled steroids prescribed for individuals with more severe COPD'
   },
});
```

```
Rules [13] = new Rule ({
   Name: "occupational lung diseases diagnosis",
   When: [
      Facts => facts.wheezing === "Y",
      Facts => facts.smokingHistory === "Y",
      Facts => facts.chestPain === "Y",
      Facts => facts.coughingUpBlood === "N",
      Facts => facts.cough === "dry",
   ],
   Then: facts => {
      facts.disease = "occupationalLungDiseases";
Code: 'occupationalLungDiseases',
Name: 'The treatment for occupationalLungDiseases includes Inhaled corticosteroids and
Take supplemental oxygen therapy.
   },
});
```