DEBRE BERHAN UNIVERSITY

COLLEGE OF COMPUTING

DEPARTMENT OF INFORMATION SYSTEMS

Towards Collaborating Data Mining Results with Knowledge-Based System for type Identification and Treatment Recommendation of Glaucoma Ailment: The Case of Borumeda Hospital

BELETE MAMO

November, 2020

Debre Berhan, Ethiopia

DEBRE BERHAN UNIVERSITY
COLLEGE OF COMPUTING
DEPARTMENT OF INFORMATION SYSTEMS

Towards Collaborating Data Mining Results with Knowledge-Based System for type Identification and Treatment Recommendation of Glaucoma Ailment: The Case of Borumeda Hospital

A Thesis Submitted to College of Computing of Debre Berhan University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Systems

BELETE MAMO

November, 2020
Debre Berhan, Ethiopia

DEBRE BERHAN UNIVERSITY

COLLEGE OF COMPUTING

DEPARTMENT OF INFORMATION SYSTEMS

Towards Collaborating Data Mining Results with Knowledge-Based System for type Identification and Treatment Recommendation of Glaucoma Ailment: The Case of Borumeda Hospital

BELETE MAMO

Members of the Examining Board

Name	Role	Signature	Date
Ato Kindie Alebachew	Chair Person	______________	____________
Dr. Solomon Demissie	Advisor	______________	____________
Dr. Kindie Biredagn	Internal Examiner	______________	____________
Dr. Mechael Melese	External Examiner	______________	____________

DECLARATION AND APPROVAL SHEET

I undersigned declare that this thesis is my work and that all sources of materials used in this thesis have been acknowledged following the standard referring practices of the guideline. This thesis is declared and approved by

Declared By:

Name: **Belete Mamo**

Signature: ____________________

Date: ____________________

Approved By:

Name: **Dr. Solomon Demissie (Advisor)**

Signature: ____________________

Date: ____________________

DEDICATION

I would like to dedicate my thesis to my mother W/r Zenebech Beyene, who sacrifices everything for my successful achievement.

Thank you Emeye

ACKNOWLEDGMENT

First and foremost, I would like to thanks **ልዑል እግዚኣብሔር** and his Mother **ንፅኂተ ንፁኃን ቅድስተ ቅዱሳን ድንግል ማሪያም** for giving me the strength, ability, and opportunity to undertake this study. Without His blessings, this achievement would not have been possible and valueless.

My sincere gratitude goes to my advisor **Solomon D**. (**Ph.D.**) for his support, encouragement, and insightful comments.

I would also like to extend my gratitude to Dr. Sissy, who is a manager of Borumeda hospital, Dr. Fozia, Dr. Nefisa, Sister Atnasiya, Sister Leyla, Said Ahmed, and Ahmed Indro for their valuable suggestions and other help.

Name ____________________

Signature ________________

Date ________________

Table of Content

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

***AI**: Artificial Intelligence*

***ANN**: Artificial Neural network*

***ARRF**: Attribute Relation File Format*

***BMH**: Borumeda Hospital*

***CG**: Congenital Glaucoma*

***CRISP:** Cross Industry Standard Process*

***DE**: Domain Expert*

***GDP**: Gross Domestic Product*

***GKBS**: Glaucomatous Knowledge-Based System*

***IOP**: Intra Ocular Pressure*

***JDK**: Java Development Kit*

***JNI**: Java Native Interface*

***KA**: Knowledge Acquisition*

***KBS**: Knowledge-Based System*

***KDD**: Knowledge Discovery in Database*

***MLC**: Machine Learning Classifier*

***OCT**: Optic Coherence Test*

***PACG**: Primary Angle Closure Glaucoma*

***POAG**: Primary Open Angle Glaucoma*

***PROLOG**: Programing in Logic*

***SEMMA**: Sample Explore Modify Model Assess*

***SG**: Secondary Glaucoma*

***SMOTE**: Synthetic Minority Oversampling Technique*

***UI**: User Interface*

***VM**: Support Vector Machine*

***WEKA**: Waikato Environment for Knowledge Analysis*

ABSTRACT

Glaucoma is a type of eye disease characterized by progressive optic neuropathy and visual field defect and categorized by damage to the optic nerve with corresponding visual field loss. It affects through increasing an intraocular pressure (IOP), which is responsible for the glaucomatous optic neuropathy involving the death of retinal ganglion cells and their axons which resulting in blindness. A person, who is responsible to treat glaucoma patients might make an error in glaucoma type identifying and treatment ordering due to subjective decision, knowledge limitation, and visualization through instruments. This results in resource wastage as well as time-consuming. The main aim of this research is to reduce (not removing the problem at all) the biased decision of the ophthalmologist through making an easy, quick and accessible way for glaucoma type identification and order the treatment for each type by developing a better classification model. In this study, data mining techniques are used to discover new knowledge based on the dataset collected from Borumeda hospital. From numerous data mining classification algorithms, this research applied Naïve Baye, Jrip, J48, and PART algorithms that are used with two basic test options based on complete and selected features. Based on the experimental analysis performed, based on the dataset collected from Borumeda hospital with an instance number of 3535 with fourteen attributes under four class labels (namely CG,PACG,POAG and SG), the PART algorithm with a test option of 10 fold cross-validation using the selected feature scored the highest accuracy result which is 74.7%. Finally, Then the output of the data mining result achieved through a classification model based on the PART algorithm incorporated with the knowledge base represented through a rule-based system and a personalized prototype system for type identifying and treatment recommendation of glaucoma ailment is develop. Last of all, the system that was developed through this methodology achieved a remarkable performance by obtaining a user acceptance score of 82.2%. Which, indicate it achieves a good result for classifying and order recommended treatment for each classified types of glaucoma ailment. Based on this finding, in futurity an ensemble method with a hybrid model will recommend to score a good performance outcome. Nonetheless, further researches should be done to upsurge the merits of combining Data mining persuaded knowledge with knowledge base system.

Keywords: Machine Learning, Data Mining, Classification, Eye diseases, Glaucoma, Knowledge Base system

CHAPTER ONE

INTRODUCTION

1.1 Background of Glaucoma

Glaucoma is a category of eye diseases characterized by progressive optic neuropathy and visual field defect [1] and considered by damage to the optic nerve with corresponding visual field loss. In other words, it is characterized by increased intraocular pressure (IOP), which is responsible for the glaucomatous optic neuropathy involving the death of retinal ganglion cells and their axons [2] which in turn causes damage of the optic nerve, resulting in blindness. In glaucoma, the drainage of aqueous humor through trabeculae is blocked, resulting in increased IOP. When the IOP become highly increase, the optic nerve fibers at the optic disk are compressed. Initially it decreases the visual field, which eventually leads to total blindness. Untreated glaucoma leads to permanent damage of the optic nerve and results in blindness [3]. Typically, patients initially loose the mid-periphery of their visual field, while central vision tends to be involved later. The patients become aware of a functional defect when visual field loss impinges upon or involves central vision [4, 5]. Patients with glaucoma may experience hardly in identifying faces, steering, reading, noticing objects in their peripheral vision, and adapting to different levels of lighting. Moreover, they are also at an enlarged risk of falls and accidents [6].

Fifteen percent of the world's blindness is attributed due to glaucoma and around 600,000 people go blind annually [2]. In 2013, the number of people with glaucoma worldwide was projected to be 64.3 million [7]. This is expected to rise to 76 million in 2020 and 111 million in 2040 [7]. Africa accounts for 15% world's blindness burden due to glaucoma [7]. In Ethiopia, glaucoma is the fifth common cause of blindness which results in an irreversible sight loss for an estimated 62,000 Ethiopians [8, 9]. The increasing prevalence of glaucoma is expected to cause a significant economic burden and poor quality of life [10-12] High glaucoma morbidity among some African communities may be attributed to low trained professionals, low awareness, and under-utilization of eye care service as well as limited availability of treatment procedures [13–15]. It has been estimated that half of the glaucoma patients are already blind in at least one eye in presentation in Africa [16]. The situation is worse in sub-Saharan Africa where it is further compounded by poor awareness and knowledge of glaucoma in the region [17].

This research aimed at developing a model through analyzing huge data collected from Borumeda Hospital manual data shelving system for type identification and diagnosis of glaucoma ailments.

1.2 Statement of the Problem

The assessed number of people visually impaired in the world is 285 million, 39 million blind, and 246 million having low vision [2]. Blindness prevalence rates differ widely but the evidence suggests that around 1% of Africans are blind [18]. The prevalence of blindness and low vision in Ethiopia are1.6 and 3.7% respectively [19]. This specifies that the burden of eye disease in Ethiopia posture enormous economic and social influences on individuals, society, and the nation at large. The prevalence is even greater in the rural population with 1.6% as compared with 1.1% in the urban population and this is powerfully emerging in Ethiopia [20].

Blindness and vision impairment has a major impact on local and national economies to decrease capitalization of once country [21, 22]. In developing countries, blindness often eliminates two people from the working staff: the blind patient, and a family member to care for the patient [23]. In the case of Ethiopia glaucoma is the fifth most common root cause of blindness and the disease caused permanent blindness in a projected 62,000 people [24]. Its impact spreads far beyond individuals and their families. For patients living in the country, access to medication is generally limited due to cost and unavailability. Ensuring that patients continue to apply and follow the existing treatment routine postures of an additional task challenge for ophthalmologists [25]. Early detection and treatment of glaucoma are critical to protecting the eyes against serious vision loss [26, 27]. There are around four types of glaucoma that showed similar characteristics [28]; and these are Primary open-angle glaucoma, primary angle-closure glaucoma, secondary glaucoma, and congenital glaucoma. Typically, the trend of treatment procedure in the country is given in such a way that the eye therapist, who is also called Ophthalmologist and considered as the domain expert in this study, utilized three methods to detect, analyze and differentiate those highly characterized glaucoma infection types through visualization. The first one is the assessment of increased pressure inside the eyeball. Second is the assessment of abnormal vision. Finally, the third method is the assessment of the damage to the head of the optic nerve. The first method has not been enough to observe the existence of glaucoma early and is not specific to determine the glaucoma ailment, which rarely occurs without enlarged pressure.

The second statement is the most consistent and dictates an abnormal vision assessment that needs specialized equipment and a trained, specialized person, but it is time-consuming, expensive, and highly subjective. The third assessment of the damaged optic nerve head is more hopeful as compared with the previous two methods and superior to IOP measurement or visual field testing for glaucoma diagnosis, but it is done by highly trained professional which is extremely expensive to achieve due to resource insufficiency in the country [29]. So, the ability to make an accurate diagnosis of glaucoma, to identify its types, determine whether it is an open or closed form, and to assess disease severity and stability, are essential to glaucoma care strategies and blindness prevention [30], and this research is aimed at attaining such points. As Glaucoma is a degenerative disease and the mechanisms to identify its ailments are not fully known, if treatment is carried out at an early stage, it is possible to slow glaucomatous development and improve the quality of life of fatalities.

Despite the great quantity of heterogeneous data that has become accessible for nursing glaucoma, the performance of tests for early diagnosis is still insufficient due to the difficulty of disease identification and the complications in obtaining sufficient measurements. In clinical settings, the techniques most used to assess the onset of glaucoma and identify its ailment are strictly related to clinical knowledge, and they mainly subjectively assess visualization testing according to the Optic Disc examinations. As mentioned above, other factors are also taken into account by the clinician, such as IOP (Intraocular Pressure) and other ocular, systemic, and demographic factors. However, the interpretation of quantitative imaging and functional test measurements is problematic, especially regarding the early detection of glaucoma [31]. So, decisions are based on subjective interpretation of test results by expert clinicians, who are not able to fully exploit all the available data.

Generally, identifying and treatment ordering for each glaucoma ailment is a hot issue and has problems on time consumption when applying several laboratories is done for glaucoma type identification, there may be somehow limited knowledge of ophthalmologist (if they are new) and subjective decisions are conducted when they examine the patient through visualizing and the patient responds. Finally, it resulted in misidentification, mistreatment, and resource wastage.

In this study, the main aim is to investigate how Glaucoma ailment can be identified using the patients' data available in Borumeda hospital by applying of an Artificial Intelligence (AI) techniques and make appropriate treatment of each type and to the best of the present of the study to add new knowledge, there is almost no any work done to solve such problems (the researcher not get similar work until this time) by building a model for Glaucoma type ailment identification and treatment order.

Thus, this research attempts to answer the questions pinpointed as follows.

1) What are the most determinant factors for developing a classification model for glaucoma type identification?
2) Which algorithm achieves the highest performance for developing the classification model for identifying glaucoma types?
3) To what extent the ailment of Glaucoma type can be detected?

1.3 Objectives of the study

1.3.1 General Objective

The general objective of this study is to develop a classification model for type identification and treatment recommendation of glaucoma ailment (four basic types) by applying a data mining technique and incorporate data mining result with a knowledge base system.

1.3.2 Specific Objectives

To achieve the above general objective, the following specific objectives are formulated and these are: To:-

- Review literatures pertinent to the study
- Acquire knowledge from domain experts
- Obtain knowledge from document analysis
- Collect dataset from Borumeda hospital
- Select the appropriate tools for conducting the research.
- Make pre-processing of the dataset collected from BMH via selected tools.
- Identify an appropriate technique and algorithms for developing the classification model
- Develop the model using the selected algorithm based on preprocessed dataset
- Evaluate the model developed by the selected algorithm.

- Collaborating data mining results with a knowledge-based system for classifying the glaucoma ailment and recommending the treatment accordingly
- Develop a prototype based on the selected developed classification model
- Conduct user acceptance testing for evaluating the developed model

1.4 Scope and Limitation of the Study

The main aim of the study is to develop a model through the collaborating of data mining results with the knowledge base system for type identification as well as treatment recommendation of four basic glaucoma ailments namely, primary open-angle glaucoma, primary angle-closure glaucoma, secondary glaucoma, and congenital glaucoma based on the recorded data from Borumeda hospital. This study is only focused on already glaucoma infected patients but does not deal to diagnose glaucoma positivity or not and can't order any glaucoma vaccination mechanism. It's also limited to, identify other types of glaucoma ailment and other eye diseases even if other types of glaucoma and eye diseases are there.

1.5 Significance of the Study

The benefits of this study are to provide medical diagnosis and treatment for four types of glaucoma ailments and reducing the errors of misidentification and mistreatment. Due to mistreatment error ordered via the subjective decision of the ophthalmologist, resources have been wasted. However, this research can handle this mistreatment by adding new knowledge extracted from the data through the selected algorithm and reduce the resource wastage which strays properties based on mistreatment. The resource wastage is handled through wipeout misidentification of glaucoma type and mistreatment handling.

The instrument used by the ophthalmologist for classifying or identifying the glaucoma ailments is very time consuming (take time in laboratory execution). But the model developed by this research can reduce these time-consuming issues because the combination of manual and automatic knowledge is enough to determine or identify types of glaucoma ailment without applying laboratory experiments rather use knowledge combination to identify glaucoma ailment through visualizing or patient response.

This study can also provide an easy and quick way to deliver diagnosis and treatment recommending for glaucoma ailment through developing Knowledge based system by collaborating with data mining technology.

The study also added additional knowledge to prior expert knowledge based on the data analysis performed through different algorithms for easily identifying types of glaucoma and provide a good knowledge for new as well as experienced ophthalmologists. The knowledge-based system nowadays becomes a very hot research area because it can able to provide quick and confident decisions for the users in all activities. So, the developed model in this study also provides quick, confidential, and non-bias decision for identifying and treating the four glaucoma type ailments and this study used for as referring point for the next researchers based on same and related area.

1.6 Methodology of the study

1.6.1 Research Design

Data mining techniques are used through the research process which includes the general approach of the research design, the data collection, analysis, and interpretation of artifacts. So, the general approach for directing this research is to design science research methodology with the reason for discovering new artifacts, clear and reliable artifacts for the user (Ophthalmologist).

1.6.2 Literature review

To conduct this research, the researcher reviews different research works and published papers that are more related to the selected topic. Specifically, some of the papers that were reviewed in this research are related to glaucoma eye infections and its treatments, data mining, Knowledge-based System, knowledge granting area, and collaboration mechanism of data mining with knowledge base systems. The researcher selected those kinds of literature based on the criteria of being more related to the problem that this research is trying to solve.

1.6.3 Knowledge acquisition and representation

Data acquisition is a mechanism of getting knowledge from different sources. In this research, the researcher organized or acquired the knowledge in the form of rule from two basic areas which are manually from expert and/or document and automatically from mining results.

The case for the manual, the knowledge was collected from domain experts through an interview and analysis of the existing document at Borumeda hospital (document analysis).

While the automatic rule or knowledge is extracted (from the dataset recorded from Borumeda hospital) using a data mining technique with the specification of classification.

Finally, acquiring of such types of knowledge from domain experts and document analysis as manually and from mining result automatically become together represented in the form of rule. The definition of knowledge representation is better to be specified before talking about generating rules or adopting a rule-based representation technique [32].

1.6.4 Evaluation methods

To evaluate the developed model, the researcher use performance evaluation techniques like precision, recall, true positive rate, false-positive rate, f-score, and accuracy.

1.6.5 Implementation tool

The researcher used a rapid mining tool to preprocess the dataset like missing value handling, feature selection, and other preprocessing tasks. To extract the hidden knowledge of the pre-processed dataset and compare the performance of classifier algorithms, the researcher used the WEKA tool. To represent rules in the knowledge base and construct the prototype for glaucoma ailment, the researcher used SWI-Prolog for knowledge and rule representation. Java NetBeans IDE 8.2 has been used for developing the user interface for the prototype system. The following table summarizes the tools applied in this research.

Table 1. 1 Tools used in this study

Tools	Function
Rapid miner	For Data preprocessing
SWI-Prolog	For Knowledge-based documenting
WEKA	For Rule generation
Java NetBeans	For Prototype development

1.7 Thesis Organization

This thesis contains seven chapters. The detail description of each chapter is discussed as follows.

Chapter one: - This chapter focuses on the background of the study, the problem of the statement, research questions, objectives of the study, the scope of the study, limitation of the study, and methodologies that this study used to conduct this study.

Chapter two: - This chapter explains about literature review and related works. The main focus area of the chapter is to review previous works that relate to this study.

In this chapter, description and related works about glaucoma disease, terms, and descriptions about different data mining process models, data mining tasks, data mining algorithms, the data mining classification technique, attribute selection measures, basic concepts of the knowledge-based system is explain.

Chapter three: - This chapter clarifies the research methodology and the data organization format for the experiment. The chapter additionally focusses on the research design that this study follows, the techniques and algorithms which adopts to develop the model, how the data is organize, prepare, preprocess, and transform into a usable format to make the data ready for an experiment. Besides, the chapter discusses the feature selection techniques adopted for developing the model for classifying the glaucoma ailments

Chapter four: - This chapter explains the final setup of the experiment using selected tools and the analysis of the result obtain from the experiment.

Chapter five: - In this chapter, the general overview of a knowledge base system, including knowledge acquisition, knowledge representation, and conceptual knowledge representation are deliberate.

Chapter six: - This chapter described the design and development of the user interface for the user (ophthalmologist) and the discussion part (responding to the research questions based on the experimental result and final output).

Chapter seven: - This chapter conclude the thesis by presenting the summary of the research output with some recommended for future work

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Glaucoma is a Neuro-degenerative condition that affects the eye and is associated with increased intraocular pressure (IOP). When left untreated, patients may gradually experience visual field loss, and even lose their sight completely. Glaucoma may be defined as a condition that causes progressive neuropathy in the optic field and is characterized by structural changes to the optic nerve head or optic disk. This may lead to functional changes in the patient's visual field. The pathophysiology of glaucoma is not completely understood but relates to retinal ganglion cell death. A superior comprehension of the pathophysiological systems engaged with the beginning and movement of glaucomatous optic neuropathy is essential in the advancement of better helpful alternatives. The typical physiological harmony between the discharge of fluid humor and the waste thereof is influenced by this condition. Watery humor is emitted by the ciliary body and waste of the humor happens using two free pathways, in particular the trabecular work and the instinctive surge pathway [33].

2.1.1 Glaucoma in Ethiopia

Glaucoma is the fifth most common cause of blindness in Ethiopia that results in an irreversible sight loss for almost an estimated annual number of 62,000 people from the total population. Its prevalence in the country showed an increase in number starting from the year 2012 up to 2019 due to many factors, including living conditions, smoking, simple eye injury, low and high blood pressure, etc. In south wollo zone where this research is based, for example, factors such as eye inflammation, blood pressure, sensitivity to light reflection, charcoal smoke, and eye injuries are some of the main factors for causing blindness due to Glaucoma. There are four basic types of glaucoma ailments that existed in Ethiopia, namely primary Open Angle Glaucoma, Primary Angle Closure Glaucoma, Secondary Glaucoma, and Congenital Glaucoma [34]. Each of these ailments is described as follows.

A. Primary Open Angle Glaucoma:

Primary open-angle glaucoma is a chronic, progressive disease that most often presents with characteristic optic nerve damage, retinal nerve fiber layer defects, and subsequent visual field loss. Intraocular pressure occurs primarily in adults and is generally bilateral, but not always symmetrical, in its presentation.

The majority of persons with intraocular pressure have elevated intraocular pressure. An in intraocular pressure levels are statistically abnormal on damage or loss of vision function ocular hypertension. Primary open-angle glaucoma, in which the intraocular pressure is below a certain level is considered as normal-tension glaucoma. Historically, this has also been expressed through low tension glaucoma. Simply, intraocular pressure is a subset of glaucoma defined by an open, normal-appearing anterior chamber angle and raised intraocular pressure and finally, it results in optic nerve head damage.

B. Primary Angle Closure Glaucoma: Primary angle-closure glaucoma (PACG) was defined as the non-visibility of the filtering trabecular meshwork and affect normal intraocular pressure (NIOP). Primary angle-closure refers to an eye with raised intraocular pressure associated with obstructed filtering trabecular meshwork results in disc damage, or field changes. The term primary angle-closure glaucoma (PACG) is used to represent PAC (primary angle closure) eyes with glaucomatous optic nerve damage or visual field loss. This type of glaucoma is manifested in blocking the drainage of the eye through the trabecular meshwork. As its name indicates, it is primary angle-closure glaucoma (PACG) because its visualized sign is not cleared as compared to other glaucoma classifications. Even though, the problem itself is visible and related to the depth of the drainage angle of the eye. The difference from primary open-angle glaucoma is the closing of the anterior chamber angle of the eye.

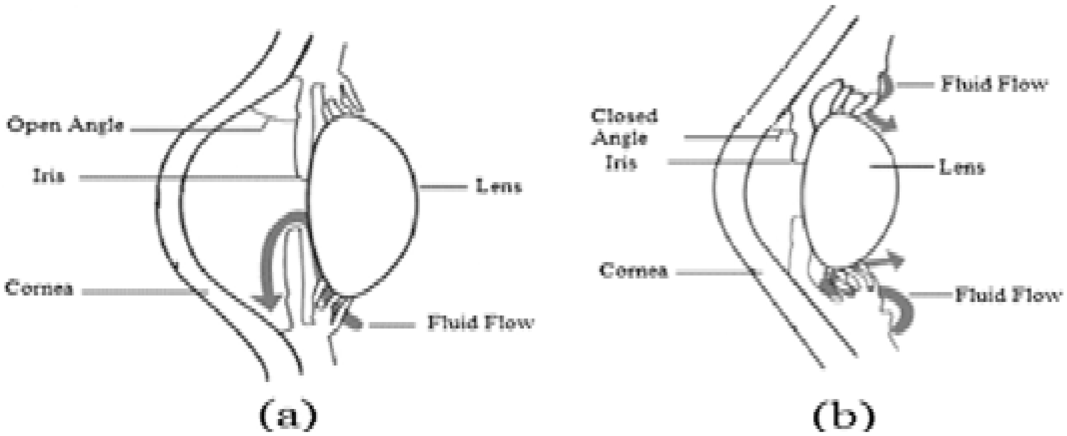

Figure 2. 1 POAG (a) and PACG (b) respectively (adopted from [34])

C. Secondary Glaucoma:

Secondary glaucoma also one major is types of glaucoma caused by fibrovascular proliferation in the chamber angle and Rubeotic fibrovascular proliferation produces adhesions between the iris and the trabecular meshwork, which lead to painful high-pressure glaucoma that is particularly resistant to therapy. The detailed exclamation points of secondary glaucoma and initial case discussed as follows.

- Secondary glaucoma is one of the significant complications of blunt and penetrating ocular trauma and acute an intraocular pressure elevation may occur immediately after blunt injuries, caused by hemorrhage.
- Chronic intraocular pressure elevation is due to permanent damage to the meshwork.
- The main case for secondary glaucoma is penetrating injuries, there is usually a more severe degree of associated inflammation, caused by disruption of intraocular tissues, and possibly also by the retention of intraocular foreign bodies. The outflow channels are thus more likely to be damaged, either because of direct injury or by extensive synechiae that often form as a result of prolonged inflammation and/or anterior chamber flattening.

D. Congenital Glaucoma:

This types of glaucoma are the hard and sometimes it is undetectable one which is characterized on common growing glaucoma types were the color of the eye/s seem to change gradually and become cloudy, the eyeball may become enlarged or gradually change its shape and sometimes an upset of the eye was created when they see bright, lights turning his/her head away from the light or try and bury his/her head into a pillow [35]. Mostly, this type of glaucoma is characterized by Children due to the case of hereditary information from their parents. The reason why this glaucoma type hard for the case hereditary genetics, association, and/ or the most treatment mechanism is obligated to an operational task. Congenital glaucoma is typically associated with severe disease and is mostly diagnosed at birth.

Figure 2. 2 whole site Glaucomatous infected eye (adopted from [35])

2.1.2 Common cases and symptoms of Glaucoma ailment

The four types of glaucoma have their symptoms which are listed as follows.

Cases and symptoms of primary open-angle glaucoma:

The main case as well as the symptoms of primary open-angle glaucoma is as follows.

- This intraocular fluid is pumped into the eye from the bloodstream, where it then cleans and nourishes structures inside the eye. Once the fluid flows through its inner-eye cycle, it is released back into the bloodstream.
- The process that intraocular fluid performs is critical to maintaining good eye health its circulation keeps everything inside your eyeballs balanced and working smoothly.

- In some cases, too much fluid is produced by the ciliary body in the eye, which causes an increase in pressure. However, it's more common in the drainage channels, called the trabecular meshwork, where the aqueous exits the eye to become blocked. Whichever cause, the result is that the pressure inside the eye (intraocular pressure) increases sometimes to dangerous levels. The common symptoms are frequent eye trauma, tear percolate, nausea, vomiting, etc.

Cases and symptoms of primary angle-closure glaucoma

The main causes of primary angle-closure glaucoma are the flow of aqueous humor to the pupil which leads to friction pressure between the anterior and posterior chambers. The main cause of this glaucoma is the existence of the patient in the darkroom.

While the symptoms are:-

- Hazy or blurred vision.
- The appearance of rainbow-colored circles around bright lights.
- Severe eye and head pain.
- Nausea or vomiting (accompanying severe eye pain).
- Sudden sight loss.

Cases and symptoms of secondary glaucoma

The initial cases of secondary glaucoma are mostly the primary glaucoma. Since, this secondary glaucoma has developed through affected an intraocular pressure and develop through a process. The major symptoms of this type of glaucoma are

- Eye hoodness and severe eye pain
- Nausea and vomiting (accompanying the severe eye pain)
- Sudden onset of visual disturbance, often in low light
- Blurred vision
- Halos around lights and reddening of the eye

Cases and symptoms of congenital glaucoma

This glaucoma type mostly occurs at birth time and its basic sign is increasing intraocular pressure and slow development of the eye size. The aqueous humor cannot flow out normally, so the intraocular pressure increases and leads to optic nerve damage. This type of glaucoma majorly occurs as naturally even other little cases are present as another glaucoma type.

The main symptoms of this glaucoma type are listed follows.

Long sight effect, Cornea-cloudy, Eye domineer, sudden eye pain, nausea, headache, blurred vision, itching, Vomiting, and others.

2.1.3 Treatment mechanism of glaucoma ailment

Regarding the mechanisms for making treatment for Glaucoma ailments, instruments such as Tonometry, Perimetry, Gonioscopy, Pachymetry, and ophthalmoscope are applied in recent medical technology. Besides the information obtained through visual observation and patient interviews, especially, the ophthalmoscope instrument is the common tool for Glaucoma treatment in several countries after the ophthalmologist identified its ailments [36].

The Borumeda hospital, where this study is based, is known for providing diagnosis and treatment for several eye diseases such as Glaucoma, Cataracts, Suet Awning, Trachoma, Conjunctiva, Astigmatism, and Color Blindness. The treatment procedure stage ranges from simple medication up to higher operation level. The medication treatment for each glaucoma type is not order all the drug at a time rather ordered according to the Symptoms occurred on the patient. Figure2 .3 below illustrates the procedure for the diagnosis and treatment of each Glaucoma type.

The common drug treatments for each type of glaucoma are as follows.

Primary open-angle glaucoma (POAG)

- 0.25 ml Carbonic anhydrase inhibitors and canaloplasty
- Beta-adrenergic blockers.
- Alpha2 Agonists, Ophthalmic.
- 2 Ml Carbonic anhydrase inhibitors.
- Anti-glaucoma, 0.005-liter Combos.
- 2cc Prostaglandin analogs.

Primary angle-closure glaucoma (PACG)

- Glycerol 1-1.5 g per kg of body weight in a 50% oral solution
- Mannitol 1-2 g per kg of body weight in a 20% solution.
- Intravenous rate of 3-5 mL per minute.
- Acetazolamide 125 to 250 mg every 6-8 hrs.
- Methazolamide 25 to 100 mg every 12 hours
- 10 ml dose prostaglandin inhibitor injection need
- Filtering surgeries with other drug injection

Secondary glaucoma (SG)

- Ciliary body and coronial surgery
- Pilocarpine (2%, 4%) 1 drop every 5 minutes, 2 doses in the acute phase
- Repeat dose every 2-3 hours until iridotomy is performed
- Timolol, Carteolol, Levobunolol, Metipranolol
- Betaxolol 1 drop every 12 hours

Congenital glaucoma (CG)

- Beta-blockers (beta-adrenergic antagonists 3hm).
- 2cc Carbonic anhydrase inhibitors.
- Adrenergic agonists.
- 2.5 cc Prostaglandin analogs.
- Final case, sometimes full eye surgery need.

  To identify the types of glaucoma ailment there are three or four phases performed by an expert. Which are

Phase 1:- an expert can identify types of glaucoma ailment through patient interviews.

Phase 2:- an expert can identify types of glaucoma ailment through visual observation of the glaucomatous patient.

Phase 3:- an expert can identify types of glaucoma ailment through visual observation and applying an instrument by testing either eye bow sweet or others.

Phase 4:- an expert can identify types of glaucoma ailment through applying an instrument for a laboratory experiment.

Finally, the four types of glaucoma infection namely, primary open-angle glaucoma, primary angle-closure glaucoma, secondary glaucoma, and congenital glaucoma are commonly identified, and ordering treatment for each of glaucoma eye disease as shown in fig 2.3 below. But the glaucomatous infected person can't be take all treatment medication at a time rather depend on the symptoms reflected on the patient the treatment medication be ordered.

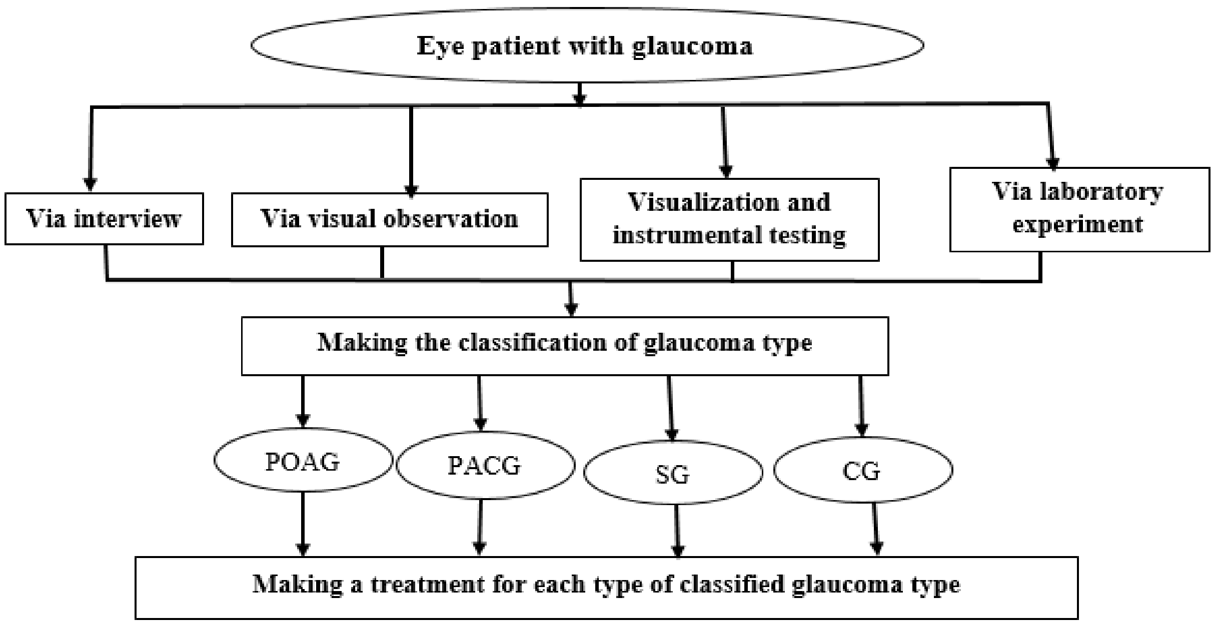

Figure 2. 3 Common Treatment Procedure for Glaucoma Eye disease

2.2 Data Mining Concept

Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large datasets and it deal with what kind of patterns can be mined. These tools can include statistical models, mathematical algorithms, and machine learning methods such as neural networks or decision trees. Consequently, data mining consists collecting data, managing data and also classification of them into appropriate class labels. The major task of data mining is discovering interesting patterns from large amounts of data, where the data can be stored in databases, data warehouses, or other information repositories. Data mining involves the combination of techniques from multiple disciplines such as database and data warehouse technology, statistics, machine learning, high-performance computing, pattern recognition, neural networks, data visualization, information retrieval, image and signal processing, and spatial or temporal data analysis [37].

The final objective of data mining is to identify valid, novel, potentially useful, understandable correlations, and patterns in existing data. Finding useful patterns in data is known in different ways like knowledge extraction, information discovery, information harvesting, data archeology, and data pattern processing [38].

2.3 Data Mining Process Model

Data mining is broadly used for mining the hidden knowledge from the given huge datasets through different analyzing techniques. For conducting and generating the final expected judgment, there are three main data mining processes models which are KDD, CRISP, and SEMMA.

2.3.1 The Knowledge Discovery from Database (KDD) Process Model

The term KDD stands for knowledge discovery in the database which is concerned with extracting a new knowledge of the huge amount of data accumulated with the helping of some specification points or measures and thresholds. The fruit results of the five-stage process model, which are explained in the following list.

Domain and data selection concept: - This phase is the firs and an initial phase, where the essential points are selecting and understanding the problem domain (understanding the business), selecting the targeted data, understand what the data are, and make a full description the features of the data. Variable or attribute determination and dataset accumulation are also the tasks in this phase.

Pre-processing: - This step is the third phase of the KDD process model and the targeted step consists of data cleaning, and all the other data preprocessing activities like missing value handling and feature selection to gain the relevant and usable data. As contrasted from the remaining phases, this phase is the most important and essential one with the reason for making a full and consistent data without any noise and inconsistencies.

Transformation: - The data transformation phase is the third phase, which is mainly concerned on the discretization of the attribute values when the attribute data type is numeric and the attribute values are categorized within some sorts selected by an investigator. In this phase, for the minimum of dimensionality reduction or the transformation method could be implemented on the target data. The respected, needed transformation of data for this research argues in chapter three.

Data mining: - data mining is one part and the fourth process of the KDD process. This phase focuses on selecting both data mining tasks and selecting the appropriate data mining algorithm to discover new knowledge or artifact. The final target of this stage is extracting new knowledge or interest from the mining of the data.

Evaluation/interpretation: - This stage is the final stage, which consists of the interpretation and evaluation of the extracted patterns, though applying different visualization ways, such as graph and evaluation of the generated model and other comparison mechanisms.

The KDD process is interactive and iterative, involving many steps with many decisions being made by the user. Additionally, the KDD process must be obtainable by the development of an understanding of the application area, the pertinent prior knowledge, and the objective of the end-users. It also must be continued by the knowledge consolidation by incorporating this into the model [39].

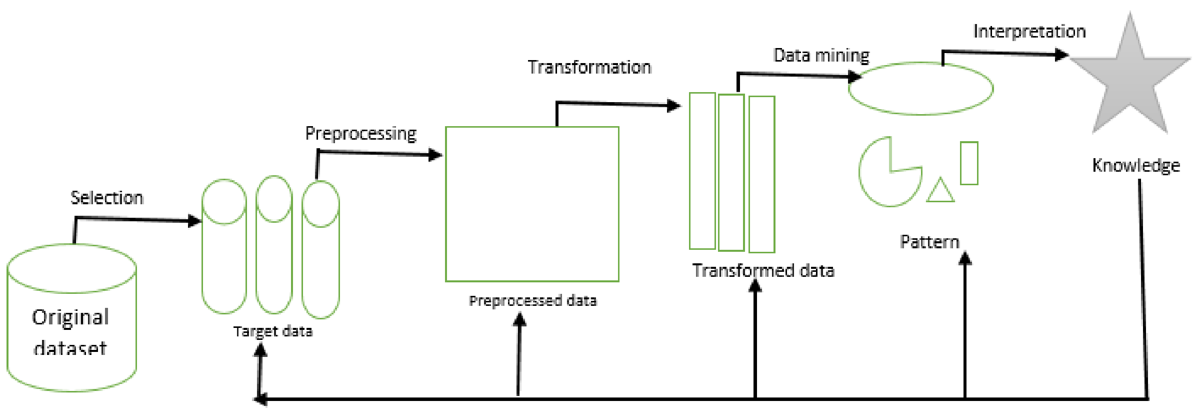

Figure 2. 4 The KDD process model (adopted from [39])

2.3.2 The Cross Industry Standard Process (CRISP) Process Model

As the name indicates, this process model is used for developing a mining model for an industry. It consists of six major steps. Business understanding, data understanding, data preparation, modeling, evaluation, and deployment [40].

Business understanding: This primary stage is concerned with understanding the purpose and requirements from glaucoma ailment, altering this knowledge into a data mining problem concept, and designing a beginner plan to achieve the objectives.

Data understanding: This phase starts from an initial data collection and proceeds with activities to get familiar with the data, to identify data quality and problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information. The total concepts and understanding of the data are manifested in this phase.

Data preparation: The data preparation phase covers all activities to construct the final dataset from the initial raw data to a till executed format.

Modeling: This phase identifies the selection and applying of various modeling techniques and their parameters are standardized to optimal values. The final model is expressed in this stage.

Evaluation: At this stage, the model (or models) obtained are more systematically evaluated and the steps executed to construct the model are reviewed or checked the final developed model is weather achieved their objective with the proper mechanism or not meet its objectives.

Deployment: This deployment phase focuses on the modeling of the final tasks which are needed for modeling.

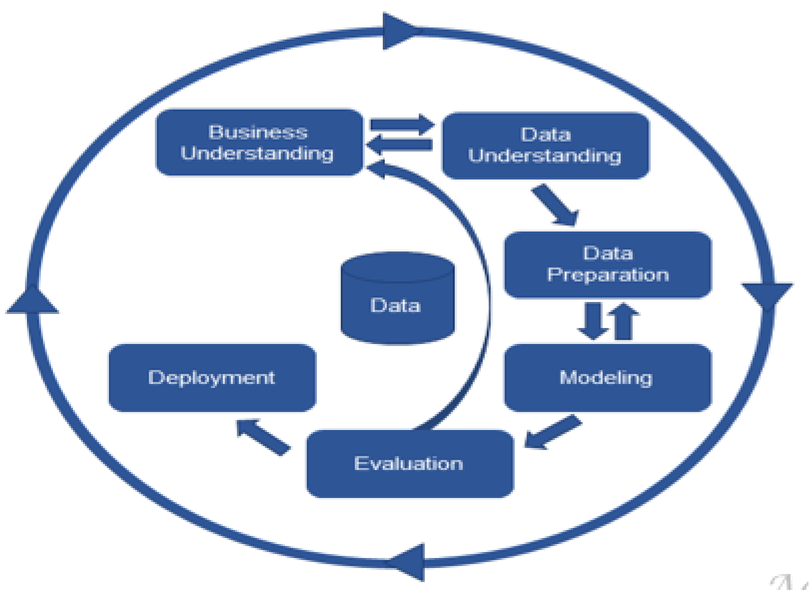

Figure 2. 5 The CRISP process model (adopted from [40])

2.3.3 The Sample Explore Modify Model Asses (SEMMA) Process Model

SEMMA Refers to the core process of conducting data mining. Beginning with a statistically representative sample of your data, SEMMA makes it easy to apply exploratory statistical and visualization techniques select and transform the most significant predictive variables, model the variables to predict outcomes, and confirm a model's accuracy [41].

Sample: -the term sample indicates extracting some significant information about the huge amount of data and mining representative samples instead of the whole data. It reduces the time required and generates a pattern with the representation of a whole data. Generally, sampling indicates talking of some samples form a huge amount of data as representatively.

Explorer: -The term modifies expresses to searching for an anticipated trend and analyzing for the reason of idea getting and understanding of thought.

Modify: -In this stage, the main operation is a modification of the data through creating, transforming, and selecting the variables that focused on model selection procedures.

Model: - The name indicates this stage consists of modeling the data by selecting algorithms for getting the reliability and desired outcomes.

Asses: - This stage consists of assessing the data by evaluating the functionality and reliability of the results from the data mining process and assess how well it performs or analyzing the final model assessment.

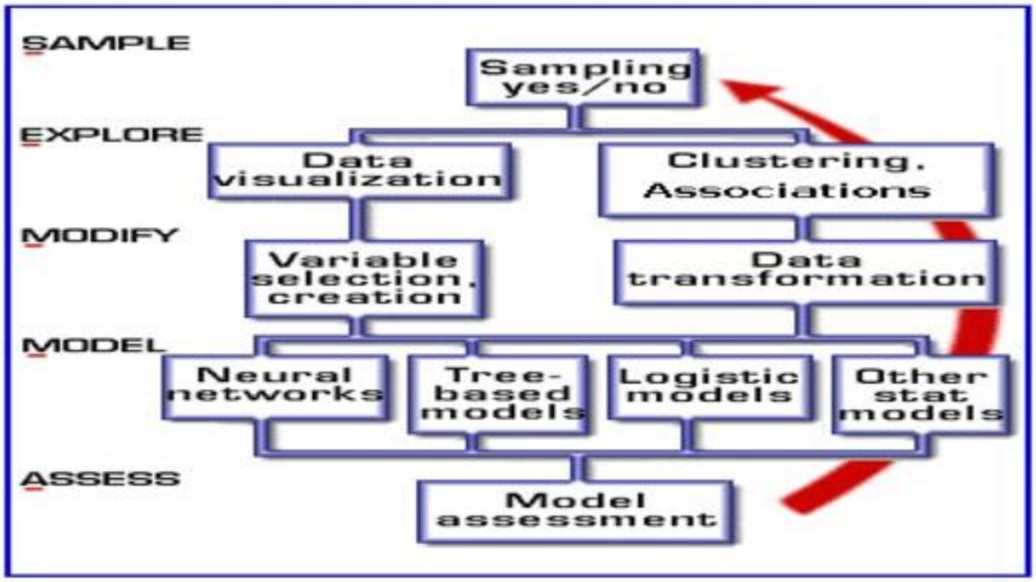

Figure 2. 6 The SEMMA process model (adopted from [41])

The general steps consisting of description and difference between the three major data mining process models are as follows.

Table 2. 1 correspondences between KDD, SEMMA and CRISP-DM [41]

Process model type	KDD	SEMMA	CRISP-DM
Step 1	Pre KDD	------------------------	Business understanding
Step 2	Selection	Sample	Data understanding
Step 3	Preprocessing	Explore	
Step 4	Transformation	Modify	Data preparation
Step 5	Data mining	Model	Modeling
Step 6	Interpretation/ Evaluation	Asses	Evaluation
Step 7	Post KDD	-------------------	Deployment

2.4 Major Data Mining Tasks

Data mining has two basic major tasks which are descriptive and predictive analytics. Descriptive analytics is concerned with the finding of data describing patterns and finish to get new and significant information from the existing dataset and it is described through general properties of data.

While predictive analytics is characterized by generating a model from the available dataset that helps predict unknown and hidden knowledge [42]. Generally, an individual and each of the data mining tasks (predictive analytics) are described as follows.

2.4.1 Association Rule Discovery

Association rule mining is one of the most important fields in data mining and knowledge discovery in databases. Rules explosion is a problem of concern, as conventional mining algorithms often produce too many rules for decision-makers to digest. It is beastly, regarded as a relationship of form A to B, where, A and B are two separate sets of items. Two measures, namely the degree of support and the degree of confidence are used to define a rule [43]. Association discovers the association or connection among a set of items. Association identifies the relationships between objects. Association analysis is used for commodity management, advertising, catalog design, direct marketing, etc. A retailer can identify the products that normally customers purchase together or even find the customers who respond to the promotion of the same kind of products.

2.4.2 Clustering

Clustering is the process of organizing objects into groups whose members are similar in some way or identifying natural groupings or clusters within multidimensional data based on some similarity measures by using some distance calculation like Manhattan, Euclidian, and another measurement [44]. Clustering is used to identify data objects that are similar to one another. The similarity can be decided based on several factors like purchase behavior, response to certain actions, geographical locations, and so on.

2.4.3 Classification

Classification is the task of assigning class labels to the data according to a model learned from the training data where the classes are known. Classification is one of the most common tasks in supervised learning [45].

Data classification is the process of organizing data into categories that make it is easy to retrieve, sort, and store for future use or extraction or Classification derive a model to determine the class of an object based on its attributes. A collection of records will be available, each record with a set of attributes.

One of the attributes will be class attributes and the goal of the classification task is assigning a class attribute to a new set of records as accurately as possible. In another expression, Classification is the process of finding a model that describes and distinguishes data classes or concepts. The models are derived based on the analysis of a set of training data (i.e., data objects for which the class labels are known). The final model is used to classify or predict the class label of objects for which the class label is unknown.

2.4.4 Regression

Regression manages eternally esteemed results and thinks of an incentive for some obscure non-stop factor. The estimation approach has an incredible favorable position that the individual records can be rank arranged by the device. The employments of relapse incorporate forecast, displaying of causal connections, and testing theories about the connections between the factors. Well suited techniques for the task of linear regression models and non-linear regression [46].

2.5 Feature Selection Category and Description

Feature selection is the process of selecting the best feature among all the features because all the features are not useful to develop the final predictive model. The principal goals of feature selection are staying away from over the fitting issue in the data set which thus improves the model execution and to give quicker just as more financially practical models and to decide a negligible element subset from an issue space while holding a reasonably high precision in speaking to the first highlights. Highlight determination considerably affects the exactness of a classifier model. A few points of interest of highlight choice incorporate lessening the dimension of the element space, limit stockpiling prerequisites, and speed up, evacuating the repetitive, immaterial or uproarious information which thus expands information quality, expanding the exactness of the created model [47].

2.5.1 Information Gain

Information gain is the measure of data that is picked up by knowing the estimation of the quality, which is the entropy of the appropriation before the split less the entropy of the conveyance after it. High entropy means the attribute is from a uniform distribution, whereas low entropy means the attribute is from a varied distribution. The attribute with the highest information gain is selected as the splitting attribute.

Let D, the data partition with a training set of class labeled instances. Suppose the class label attribute has k distinct values defining m distinct classes, C_k (for k=1, 2, 3…, n). Entropy is defined as follows. Let Polk be the probability that an arbitrary instance, in D belongs to class cook, estimated by $|C_k$, D|/|D|. Expected information (entropy) needed to classify an instance in D is given shows in the following equation. Entropy (E (D)) - is the average amount of information needed to identify the class label of an instance in D.

$$Entropy\ (E(B)) = \sum_{k=1}^{m} pk\ log(pk) \ldots\ldots\ldots\ldots\ldots\ldots 2.1$$

The information gain value of each attribute depends on the entropy that was calculated. Though, the gain value of each attribute is calculated as follows. Suppose attribute F can be used to split D into different partitions or subsets, {D1, D2… Dk}, where D_K contains those instances in D that have an outcome of A. Information needed (after using A to split D) to classify D:

$$\text{Info}_A\ (\text{D}) = \sum_{k=1}^{n} \left(\frac{DK}{D} \right) x\text{Info}(\text{Dk}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.2$$

The information gain value for information gained by branching on attribute A is calculated as

$$\text{Gain (A)} = \text{E (D)} - \text{Info}_A\ (\text{D}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.3$$

2.5.2 Gain Ratio

The information gain measure is biased toward tests with many outcomes. That is, it prefers to select attributes having a large number of values. A modification of the information gain that reduces its bias towards high-branch attributes. To handle such types of bias, applying the gain ratio is important with normalization to information gain using a "split information" value. The Split Information entropy of the distribution of (unlabeled) instances into branches with Info (D) is

$$Splinfo\ A(D) = -\sum_{j=1}^{v} \left(\frac{|Dj|}{|D|} * \log 2 \frac{|Dj|}{|D|} \right) \ldots\ldots\ldots.2.4$$

This value represents the potential information generated by splitting the training data set, D, into v partitions, corresponding to the v outcomes of a test on attribute A. Note that, for each outcome, it considers the number of tuples having that outcome concerning the total number of tuples in D.

It differs from information gain, which measures the information concerning classification that is acquired based on the same partitioning [48]. The gain ratio is defined as:

$$\text{Gain ratio}(A) = \frac{Gain(A)}{Splitinfo(A)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.5$$

2.5.3 Gini Index

The Gini index is used in the Classification & Regression Trees CART. Using the notation CART written above, the Gini index measures the impurity of B, data partition, or set of training tuple [16].

$$Gini(X) = \sum_{k=0}^{n} Pk^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.6$$

Where, *Pk* is the probability that a tuple in Y belongs to class *Ci* and is estimated by |Ci, Y |/|Y|. The sum is computed over *k* classes. The Gini index considers a binary split for each attribute. When seeing a binary split, the weighted sum of each of the impurity result partition can be calculated. Let, if a binary split on A Partitions Y into Y1 and Y2, the Gini index of Y given that partitioning is

$$\text{GiniA(Y)} = \frac{|Y1|}{|Y|} Gini(Y1) + \frac{|Y2|}{|Y|} Gini(Y2) \ldots\ldots\ldots\ldots\ldots 2.7$$

Confusion matrix

Confusion matrix contains information about actual and predicted classifications done by a classification system and used to evaluate the accuracy of classification. The performance of such systems is commonly evaluated using the data in the matrix. The following table shows the confusion matrix for a two-class classifier.

Table 2. 2 overviews of confusion matrices

		Predicated class	
		Negative	Positive
Actual class	Negative	True negative	False negative
	Positive	False positive	True positive

2.6 Performance Evaluation Techniques

Accuracy

Accuracy of classifier refers to the ability of classifier to correctly classify the instances. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data. Accuracy measures can be explained using the true positive (TP), false positive (FP), true negative (TN) and false negative (FN).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. 2.8$$

Precision

Precision is measuring the proportion of instances that are classified as positive that are really positive.

$$Precision\ = \frac{True\ Posetive}{True\ positive+False\ positive} \ldots\ldots\ldots\ldots\ldots\ldots... 2.9$$

True Positive and False Positive rate

In contrary to Predictive Accuracy TP rate and FP rate values do not depend on the relative size of positive and negative classes.

True Positive Rate (TPR) is the proportion of positive or correctly classified instances as positive or correct instances.

$$True\ Postive\ Rate = \frac{True\ Positive}{True\ Positive+False\ Negative} \ldots\ldots\ldots\ldots\ldots\ldots 2.10$$

False Positive Rate (FPR) rate measures the proportion of negative instances that are erroneously classified as positive.

$$false\ Postive\ Rate = \frac{Flase\ Posetive}{False\ Positive+True\ Negative} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.11$$

ROC

Receiver operating characteristics (ROC) graph is a technique for visualizing, organizing, and selecting classifiers based on their performance. ROC graphs have long been used in signal detection theory to depict the tradeoff between rates of classifiers.

F- Score

F-score is described as a weighted average of precision and recall and it considers both false positives and false negatives into account. It is used to provide a more realistic measure of a test's performance by using both precision and recall.

$$f - score = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.12$$

2.7 Knowledge-Based System

Knowledge-based systems are a part of man-made brainpower, which is a computer program that endeavors to recreate the thinking procedures of a human master, and it can settle on choices, suggestions and perform assignments dependent on client input. Man-made consciousness is about how making the framework think, or act like people. The master's information is accessible when the human master probably won't be thus that the information can be accessible consistently and in numerous spots, as vital. Master frameworks determine their contribution to basic leadership from prompts at the user interface and information records put away. Knowledge-based systems are computer programs intended to take care of issues, produce new data, (for example, a finding), or give exhortation, utilizing an information base and a deduction instrument [49].Most frameworks incorporate a user interface and some clarification ability too. Information based frameworks are described as concentrating on the amassing, portrayal, and utilization of information explicit to a specific assignment,

However, tended to the extended perspectives on such frameworks made conceivable by the capacity to utilize similar information in a few distinctive manners. There are two main parts of any Knowledge-Based System (KBS) the knowledge base and the inference engine. Besides, there are peripheral features designed to facilitate interaction with end-users (user interface), an explanation of a line of reasoning to illustrate the structural design of the general architecture and its advantage of Knowledge-Based System [49] listed as follows.

- *Time-saving*: The time that will be spent doing manually will be minimized.
- *Work Quality*: The work quality of the service will increase due to reducing the errors happen in the center.
- *Complete*: unless there are implementation errors, knowledge-based systems will always produce the desired result as they will not leave out any rule (consideration) in the reasoning processes.
- *Replication*: human experts are scarce resources. They are physically bound to their geographical locations and can only available at one place at a time but the knowledge-based system can be replicated and in effect to be transferred to any other locations to perform other tasks.
- *Knowledge updating*: Knowledge-based systems can be updated easily by editing the rule base, but human experts take to retrain.

A Knowledge-Based System (KBS) is an information stockroom that stores a lot of information that is utilized in critical thinking abilities and the choice taken by the information-based framework ought to be equivalent to those of the master area. Information based frameworks are computer programs that are intended to coordinate crafted by specialists in explicit subject matters. Information-based frameworks or knowledge-based systems execute the heuristic human thinking by utilizing some particular systems, techniques, and components to take care of issues that don't have a customary algorithmic arrangement [50]. The improvement of information based frameworks need the utilization of information and knowledge to conquer the difficulties looked at customary arrangements. The capacity of the smart frameworks to catch and redistribute mastery assumes a crucial job for the advancement of a country, item, or populace from various perspectives. Strong frameworks permit documentation of at least one master's information and use the information for critical thinking in a financially practical way. The architecture of a knowledge-based system demonstrates the general strides to be done beginning from contributing to the framework up to building up the model of the information-based framework.

The knowledge-based system architecture contains basic components [50] such as knowledge engineer, knowledge acquisition, knowledge representation, inference engine, and the knowledge base. The detailed description of those terms as follows.

Knowledge engineer: is a person or an agent who retrieve knowledge from domain experts via interview, document analysis, and by automatic knowledge attainment mechanism. The knowledge engineer should have not a mandatory knowledge about the professional way of systematic development of the knowledge base system rather the knowledge engineer must understand how to develop a knowledge base as well as the knowledge-based system using a development environment.

Knowledge acquisition: it is the gathering of knowledge from different sources through different mechanisms to obtain knowledge. Not only this but also adding new knowledge to a knowledge base or improving the knowledge that was previously acquired in the knowledge base. The knowledge should be acquired in the way of making an interview with a domain expert, document analysis, and technological knowledge (data mining result).

Knowledgebase: The main issue on knowledge base focused on the collection, organization, and retrieval of knowledge and storing knowledge in the knowledge base with a different method of knowledge arrangements as case-based or rule-based format. The rule-based knowledge for the knowledge base is good compatibility for this study.

Knowledge representation: this term identifies the mechanism of representing the knowledge that was acquired from different sources**.** The knowledge data must be represented in an easily understandable format for both humans and computers**.**

In this work, the knowledge has represented the form of rule-based knowledge representation format. The rule-based knowledge representation follows an IF….THEN rule. This rule concludes the conclusion based on the knowledge presented in the premises.

Inference engine: The inference engine could be achieved the functionality of the system by searching through the knowledge base to find rules premises match what the user enters as input.

This process continues until the inference mechanism is unable to match any premise with the inputs in the working memory and it is a basic component of a knowledge-based system that applies the logical rule to the knowledge-based system to infer new information.

This process would iterate as each new fact in the knowledge base could activate additional rules in an inference engine.

Working Memory: The working memory is the working environment of the knowledge-based system which holds user input that is used for classifying the glaucoma type ailment.

User interface: It is a means of communication with the user and to make the dialog user friendly. The user interface is the final prototype developed by using programming languages such as Java-based on the knowledge obtained from domain experts, document analysis, and data mining results (in the form of rule). Developing a simple and easy user interface allows the user of the system to interact easily and make the developed knowledge-based system more enjoyable to use. This user interface could be used to analyzing the final user acceptance evaluation of the model and is the final output of the whole work could be determined on it.

Explanation Facility: is a component that enables the knowledge-based system to provide some explanation to the user about the process why is it asking a question and how it reached on some conclusion. The explanation facility provides help for the user in which the user asks something to be explained and the module responses based on the user's query.

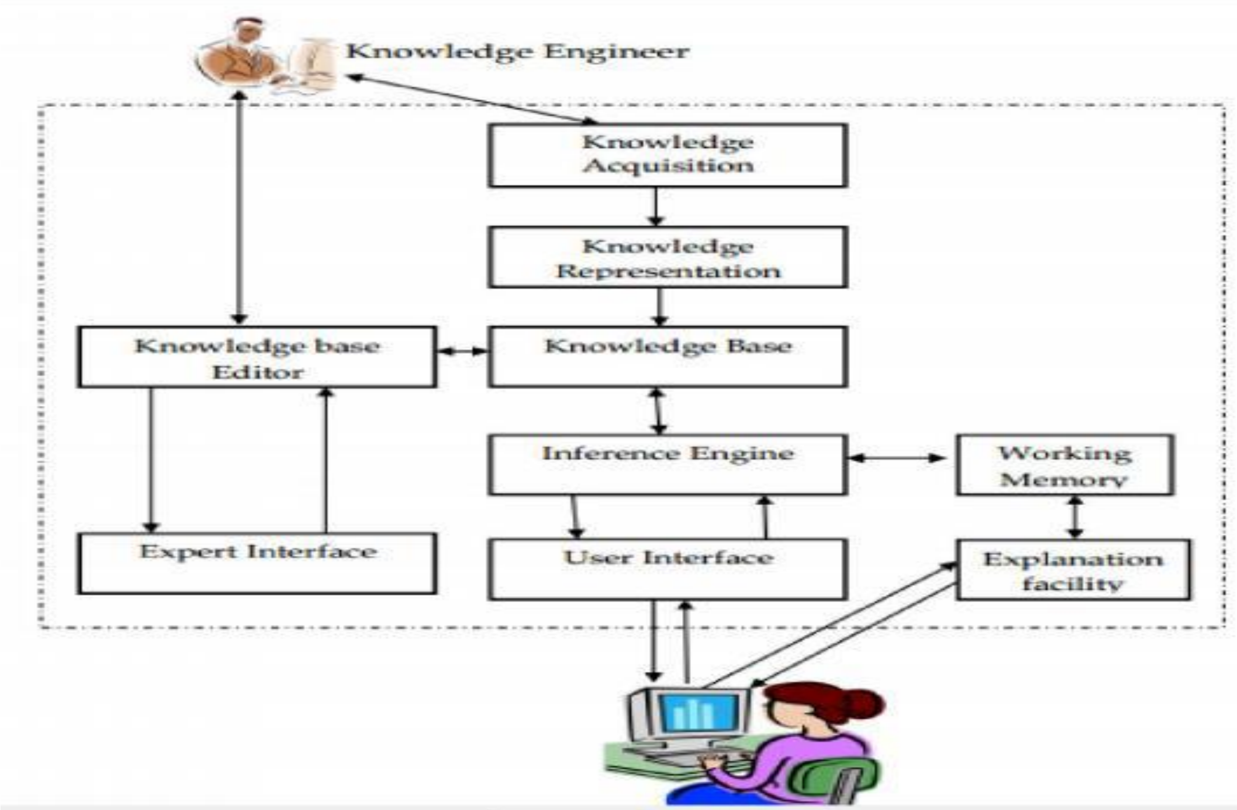

Figure 2. 7 Architecture of a knowledge-based system (Adopted from [50])

2.8 Application of Data Mining Technique in Health Sector

Data mining is good to solve problems based on data and has the greatest role in a health organization. Since health organizations have many recorded data from the patient who get services from there. It also has proven effective in other areas such as the predictive way of disease type, customer relationship management, detection of fraud and abuse, management of health care, and measuring the effectiveness of certain treatments. In the healthcare industry specifically, data mining can be used to reduce costs by increasing efficiencies, improve patient quality of life, and perhaps most importantly, save the lives of more patients and add newly discovered knowledge based on the data [51].Another factor for the need of data mining in the wellbeing division is that the huge measures of information produced by social insurance exchanges are excessively perplexing and voluminous to be prepared and broken down by conventional strategies.

Utilizing information mining to break down such immense measures of information has gotten progressively basic as money related weights have increased the requirement for social insurance associations to settle on choices depending on the examination of clinical and budgetary information. [51]. other factors boost data mining popularity in the health sector one of them is the non-invasive diagnosis and decision support. Some diagnostic and laboratory procedures are invasive, costly, and painful to patients.

2.9 Related Work

Number of researches have been conducted for the diagnosis and detection of glaucoma using data mining technology. Among the data mining based researches conducted on glaucoma, some of them are described as follows.

Maıla C et.al. [52] conducted research entitled automatic glaucoma detection based on optic disc segmentation and texture feature extraction with the major aim of developing an automatic detection method of glaucoma in the retinal image to check whether or not the eye become glaucomatous via classifying the retinal image. To achieve the final objective, the methods used by the researchers were the acquisition of image databases, optic disk segmentation, and texture feature extraction in different color modes based on the dataset collected from public drive dataset.

The classifier algorithms adopted by their research were multiplying perceptron, random committee, random forest, and radial basis function. Finally, the researchers obtained the highest accuracy result of 93.03 % by the Multiline perceptron.

While their study is focused on only detecting glaucoma by applying image processing techniques, this research is focused on classifying types of glaucoma and make treatment for each kind of glaucoma.

Kinjan Chauhan and Ravi Gulati [53] have conducted research adopting a data mining approach with the major aim of detecting and diagnosing glaucoma disease using the Perimetry and Optical Coherence Tomography images and predicting the diagnosis for progression of glaucoma. Additional to the image processing techniques they have used for feature extraction using the OCT data from international OCT data repository, some of the data mining algorithms they adopted were neural network, decision tree, logistic regression, machine learning classifier (MLC), linear discriminant analysis, and support vector machine (SVM).

Finally, the Naïve Bayes algorithms achieved the highest accuracy value, and accordingly, the model is developed for detecting and diagnosing glaucoma based on a decision of whether the glaucoma ailment is absent or present in the eye. On the other hand, the aim of this research is classifying the glaucoma ailments and providing diagnosis and treatment.

The research was done by Sadaf Malik et al. [54] aimed at developing a classification model for classifying the different types of eye diseases based on a dataset obtained from direct analyzation of 344 patients by following a data-driven approach and by adopting machine learning and data mining techniques such as decision tree, naive Bayes, random forest and neural network from which the random forest algorithm achieved the highest accuracy result (86.36%). Finally, the score point of 86.36% has been registered via the algorithm of random forest. The paper was focused on classifying diseases by considering glaucoma as one class label. On the other hand, the proposed work is focused on classifying types of glaucoma ailment to each of the classes and make the treatment for each of the glaucoma ailment types.

In [55] research entitled "classifying chief complaint in eye diseases using data mining techniques", has been conducted to discover new hidden knowledge using an analyzed data from 192 patients and classifying the different types of eye disease using the intelligent capability of data mining technologies like MLP, ANN and naïve baye from those classifier algorithms the Naïve baye algorithms achieves the highest accuracy result. Like all the literature reviewed, this paper also uses glaucoma as one of the class labels, and the last target is only classified the eye disease as one of the types.

But in this research, the primary task is classifying the type of glaucoma into its type and making treatment for each of the classified glaucoma types using the developed knowledge-based system. In [56] research entitled "Data Mining Techniques for Diagnostic Support of Glaucoma using Stratus OCT and Perimetric data from international OCT data repository for the identification of glaucoma from other diseases in the case of retinal image processing. The Parameters obtained from the Perimetry and Stratus Optic Coherence Test (OCT) have been fed to each technique to find out their performance in terms of accuracy, sensitivity, and specificity by using the data mining technique. The data mining algorithms used for diagnostic classification were Decision Tree, Linear Regression, and Support Vector Machine (SVM).

Among those algorithms, the Decision Tree and Linear Regression Model performs much better than other algorithms for the diagnosis of Glaucoma by achieving the highest accuracy of 92.56% and 70.25% respectively. The specificity of Linear Regression and SVM is 97.56% and 96.34% respectively. The difference of the paper from the proposed model developed in this research is that the previous paper covers the only diagnosis the glaucoma disease by applying the data mining techniques from another eye disease rather proposed model developed in this research dictates identifying types of glaucoma by applying data mining technology and make treatment for each type of classified glaucoma.

A research done in [57] applied a machine learning techniques such as fuzzy logic, decision tree based on ID3 algorithms, support vector machine, and k-nearest neighbor. Among those techniques, k-nearest neighbor achieved the highest accuracy score of 86% based on an image data collected though digital image capturing technique. The main difference between this research and the proposed model developed in this research is that while the research conducted in [56] was focused on the detection and prediction of glaucoma ailments the proposed model developed in this research is concentrated on glaucoma type identification and treatment.

In [58] research has been done for glaucoma detection and prediction using data mining classification algorithms namely, KNN and Naïve Baye based on the dataset obtained from health care organization. The final analysis was decided on 65% accuracy result scored through the Naïve Baye algorithm. The basic difference with this work is that the related paper is focused on detect glaucoma disease. While the proposed model developed in this research is identifying glaucoma and order treatment for each classifying item.

Table 2. 3 Summary of related work

S .no	Authors name/year	Title	Objective	Algorithms/techniques/	Accuracy result
1	Maıla C et.al. (2016)	automatic glaucoma detection based on optic disc segmentation and texture feature extraction	developing an automatic detection method of glaucoma in retinal image whether or not the eye become glaucomatous	Acquisition of image database, optic disk segmentation, and texture and feature extraction. classifier algorithms multiline perceptron, random committee, random forest, and radial basic function	93.03 % by Multiline perceptron
2	Kinjan Chauhan at al (2012)	A Proposed Framework for diagnosis of Glaucoma - A Data mining Approach	To detecting and diagnoses of glaucoma disease using the Perimetry and Optical Coherence	Image processing techniques for feature extraction classification algorithms like a neural network, decision tree, logistic regression, machine learning classifier (MLC) linear discriminant analysis, and SVM	Bayes algorithm scores the highest accuracy than other algorithms.
3	Sadaf Malik at al. (May 2019)	Data-Driven Approach for Eye Disease Classification	To develop a classification model for classifying types of eye disease	data mining algorithms like decision tree, naïve baye, random forest, and neural network	86.36% by random forest.
4	Archana L. Rane et.al (march 12)	Classifying chief complaint in eye diseases using data mining techniques	Discovering new hidden knowledge for classifying of different types of eye disease using the intelligent capability of data mining	Classification algorithms ANN, MLP, and naïve baye	Highest accuracy result scored via Naïve baye classifier

5	Kinjan Chauhan et.al (*2016*)	Data Mining Techniques for Diagnostic Support of Glaucoma using Stratus OCT and Perimetric data.	Identification of glaucoma from other diseases in the case of retinal image processing.	Decision Tree, Linear Regression and Support Vector Machine	97.56% Linear Regression
6	Tehmina Khalil et. al (2017)	Applying machine learning techniques for glaucoma detection and prediction	The main objective of this study was to detect glaucoma by itself either of it occurred or not.	Data mining algorithms ID3 algorithms, Support Vector Machine, k-nearest neighbor	86% by k-nearest neighbor
7	Ritu Sindhu (June 2018)	Data mining technique for glaucoma detection	detecting the glaucoma positivity of the eye by using data mining techniques	Naïve Baye and KNN	65% accuracy result via Naïve Baye.

N.B The objective of those related works (almost all) are either of detecting glaucoma or classifying types of eye disease. While the proposed model developed in this research is focused on identifying the type of glaucoma ailment. So, the accuracy result of those related works with this proposed model never be comparable. Since, their objective, feature used and dataset were different from the new proposed model.

CHAPTER THREE

RESEARCH METHODOLOGY AND DATA ORGANIZATION

3.1 Design and Approach of the Research

A research approach is a technique that explores and applies different procedures to gather and assess information. The research procedure includes research design and implementation tools. It is characterized by utilizing those apparatus to accumulate significant data in a particular research study. To conduct this research, the researcher used interviews, document as well as data analysis mechanisms.

Mainly, the research adopted a design science approach based on the following reasons.

1) Design science generally focused around the collected data which is used to deal with the issue looked in the area.
2) Design science is used for creating new artifacts from the given dataset.
3) Design science research accomplishes knowledge and understanding of a problem through the way of establishing a designed artifact and so on.

The research also reviewed different literature and Journal articles related to the research problem that this study is based on, which is papers related to glaucoma type eye diseases and its treatment. The research also applied different tools for data pre-processing activities which are missing value handling, class balancing, and feature selection. These selected tools were also used for extracting the hidden knowledge from the preprocessed data. Among these tools:-

Rapid miner: It is used as a visual environment for building analytics processes: Graphical design environment makes it simple and fast to design better models [59]. The basic function of this tool in this research is for data preprocessing activity of missing value handling, class balancing, feature selection, and others.

WEKA: This tool is used to extract the hidden knowledge from the preprocessed dataset and compare the performance of classification algorithms.

SWI-Prolog: A term stands for Programming in Logic which is the most predominantly used tool in developing a knowledge-based system [60].

It is a declarative language rather than procedural, which means that rather than telling how to reach on the decision, the language focuses on a database that consists of facts, and rules that describe the relationships which hold for the given application. The basic functionality of this tool is to represent knowledge.

Java NetBeans IDE 8.2: The java NetBeans is an open-source tool that was used to incorporate data mining results extracted using the WEKA machine learning tool with the knowledge-based system and develop the prototype of the proposed model.

3.2 Description and Architecture of Proposed Model

The proposed model was drafted from the concept of the general framework of common data mining and knowledge base combination ways. As shown in figure 3.1 below, the general skeletal framework consists of the business area and data understanding, data pre-processing, classification algorithms, model evaluation, knowledge acquisition, knowledge representation, knowledge base, inference engine, and prototype development.

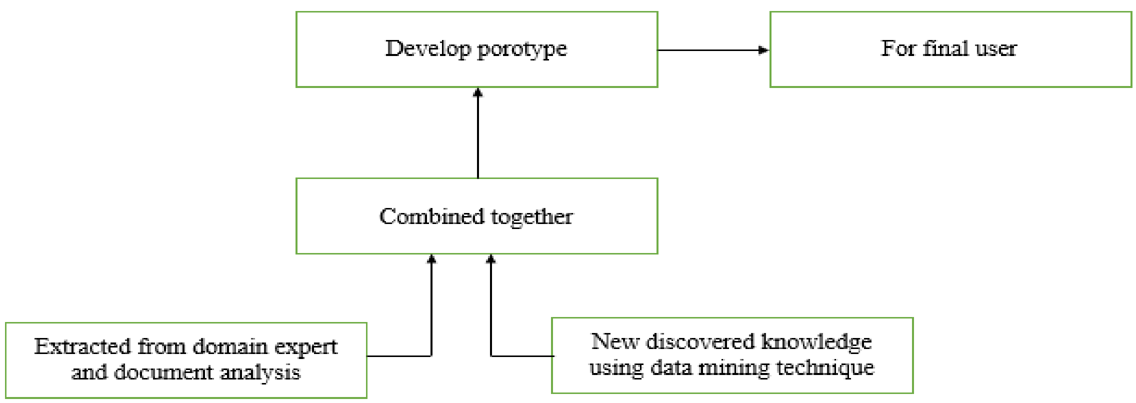

Figure 3. 1 General outline of knowledge combination system

The new proposed model aimed at identifying the glaucoma type from the patient and suggest treatment recommendations accordingly. Hence, misclassification costs can be very high for a sensitive field like disease diagnosis, addressing the class imbalance issue becomes of utmost importance.

Such a misclassification problem is occurred due to the subjective decision of an expert and limited knowledge of the expert (especially an expert who is new for work) to identify types of glaucoma of the patient based on visual examination, the patient responds analysis, and a combination of both visualization and patient response analysis. Another solution reflected through this general proposed model is the problem of time consumption. This time consumption problem is produced due to applying instruments and laboratory examinations for identifying which types of glaucoma is occurred on the glaucomatous patient. At the time of this, instruments for laboratory examination can operate through taking some tear or other jelly flow from glaucoma infected person and requires time to test and display the classified result. But, the model of this research can handle such a problem by adding more knowledge to identify types of glaucoma, namely primary open-angle glaucoma, primary angle-closure glaucoma, secondary glaucoma, or congenital glaucoma by eradicating of laboratory examination and instrument applies even if those are reliable. Hence, instruments used for visualization and laboratory experiment are different. Figure 3.2 shows the general flow chart of the proposed model to fix such difficulty.

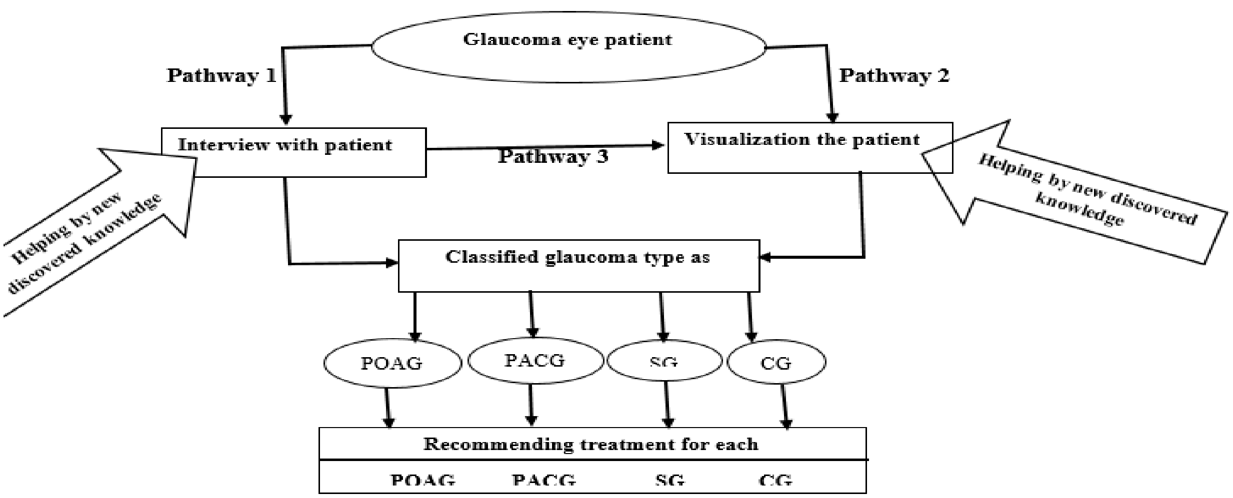

Figure 3. 2 General proposed model flowchart of glaucoma type identification and treatment

The detailed procedure of the proposed model is started from identifying the problem domain which is based on the knowledge collected from two sources and these are: from domain expert and document analysis.

From domain expert: - the researcher collected the knowledge through an interview from domain experts which was quite enough to make analysis and develop the model and finally the research represent this knowledge in the form of rule.

From document analysis: - Hence, the knowledge extracted from the document existed in manual form and the researcher extracts the knowledge from those existed documents. The document by itself contains the knowledge as rule form based on the symptom of each glaucoma type for identifying and ordering the treatment for each glaucoma type and the researcher easily represents the knowledge in the form of rule. I.e. If---Then. Those two knowledge extraction types were manual. Because, the researcher collects that knowledge manually from domain experts and existed document through interview and document analysis, respectively, and represent in the form of rule.

From raw data (hidden knowledge discovery): - the other method of knowledge yielding mechanism is from raw data. The raw data are collected from the selected organization. I.e. BMH. But the data existed in the manual case and the researcher organized these manual cards to Microsoft word excel understanding format. After gathering the data, it needs pre-processing activities of missing value handling, feature selection, data discretization, and other tasks. So, after completing the pre-processing task of the dataset, the model is developed after empirical analysis of four different classification algorithms, namely Naïve Baye, Jrip, J48, and the PART algorithm was used with two basic test options namely, 10 fold cross-validation with its default value and percentage split of the dataset with 80/20. The reason for using only four classification algorithms is that after trying other classification algorithms those four algorithms have better accuracy than others, are easy to understand and interpret the model outcome and classify hidden patterns. The attribute used for model development was either in the whole attribute or selected attribute. From those four algorithms, one algorithm is selected (highest accuracy result) by comparing others based on two test options using either in the whole or selected attribute. Finally, one model is selected (highest accuracy result algorithms based on either 10 fold cross-validation or 80/20 percentage split, using either in the whole or selected attribute).

Generally, the main product point of this study is reflected in this stage by adding new knowledge from data analysis through model development.

Finally, the knowledge extracted from domain experts and/or document analysis and the new hidden discovered knowledge from a preprocessed dataset using data mining techniques became combined and applied for glaucoma type identification and treatment order. The general architecture of the proposed model is displayed in figure 3.3 as follows.

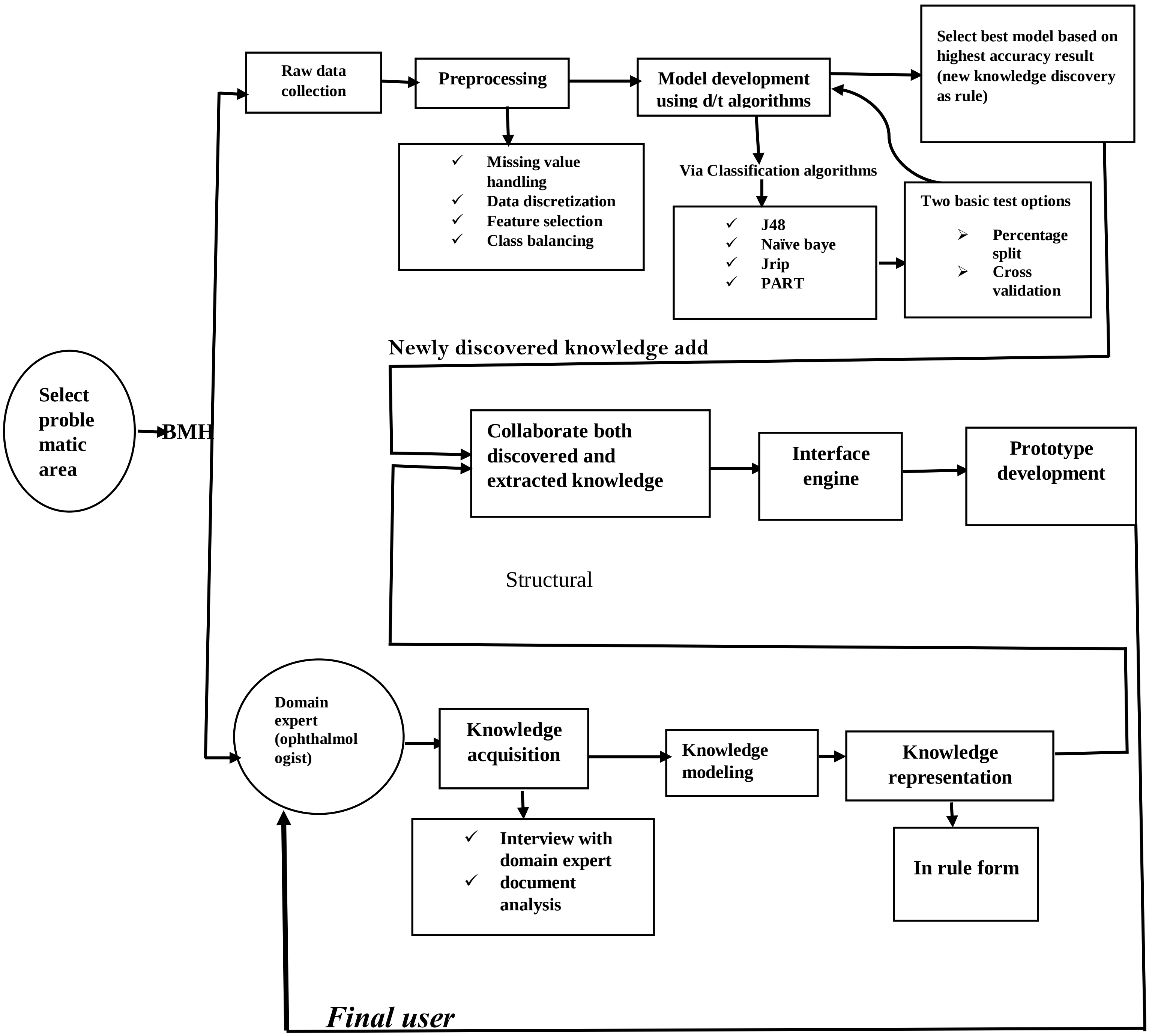

Figure 3. 3 The detailed workflow process

3.3 Research design concerning with KDD Process Model

Conducting a data mining research demand the following process model structure, either KDD, CRISP, SEMMA, or a hybrid model, depending on the nature of the investigation. In this research, KDD is the best congruent process model because of the reason that it is characterized by a nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data which is the main objective of this research. Since the main objective of this research is extracting hidden knowledge from a huge dataset, KDD is suitable in this sense that it is characterized by a non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. In a more specific concept, the KDD process is interactive and iterative, involving numerous steps with many decisions being made by the user and it is preceded by the development of an understanding of the application domain, the relevant prior knowledge, and the goals of the end-user. It must be continued by knowledge consolidation, incorporating this knowledge into the system. The major steps of the KDD process concerning this study are as follows and the diagrammatic representation of this study respect to the KDD process model is also shown at the end of the step description.

3.3.1 Domain understanding concept and targeted data selection

A domain understanding is a primary phase, which is concerned with understanding the whole problem domain or problem area. In this research, the domain understanding phase is focused on understanding the objectives and requirements of the glaucoma ailment perspective, converting the knowledge into a data mining problem definition, and designing an introductory plan to achieve the objective/s. Hence, the researcher understands the problem area by reading different related works that were done on Glaucoma eye disease. The researcher gets information about types of glaucoma, their level of effectiveness, and the way of treatment for each type of glaucoma from the selected research area. Data understanding is the way of systematic knowledge about the data which is target fully taken from an organization (BMH). In this research, the dataset is collected from Borumeda hospital, which is one of a few eye hospitals in the country. The target dataset contains the data from 2012 up to 2019 and most of the data exist in manual form. The amount of manual data not the card from 2012 was over 5419 based on the hospital's numeric report. Within the above-specified range of years, the researcher collected 3142 instances and 19 attributes with four basic class labels. Table 4 below illustrates the summary of the dataset.

Table 3. 1 attributes with their value and data type

Serial Number	Name of the attributes	Data type	Values
1	Age	Nominal	*Child, yeard, young* (nominal values after **discretization**)
2	Sex	Nominal	*Male and female*
3	Long sight effect	Nominal	*Yes, no*
4	Cornea-cloudy	Nominal	*No, yes*
5	Eye domineer	Nominal	*Yes, no*
6	Sudden eye pain	Nominal	*No, yes*
7	Nausea	Nominal	*Yes, no*
8	Headache	Nominal	*No, yes*
9	Blurred vision	Nominal	*Yes, no*
10	Itchy	Nominal	*No, yes*
11	Vomiting	Nominal	*No, yes*
12	Sudden sight loss	Nominal	*Yes, no*
13	Eye casing distent	Nominal	*No, yes*
14	Light sensitivity	Nominal	*Yes, no*
15	Eye hoodness	Nominal	*Yes, no*
16	Frequent Eye trauma	Nominal	*No, yes*
17	Tear percolate	Nominal	*Yes, no*
18	Living Area	Nominal	*Lowland, highland and temperate*
19	Class labels	Nominal	*1) Primary open-angle glaucoma* *2) Secondary glaucoma* *3) Primary angle-closure glaucoma* *4) Congenital glaucoma*

Note: the logic why the attribute age has nominal value since any numerical value of the age could be included in the three categories since it was discretized.

As shown in the above table, the dataset was collected based on these attributes from each of the glaucoma ailment type, namely, Primary open-angle glaucoma, Secondary glaucoma, Primary angle-closure glaucoma, and congenital glaucoma. The following table indicates the full description of the attributes.

Table 3. 2 features description

Attributes name	Descriptions
Age	This feature indicates the age level of each glaucoma patient with three major discretized values. Namely:- Child: ranges from a year up to fifteen years. Young: ranges from fifteen up to fifty-one Yeard: age of patient greater fifty-one [60].
Sex	Indicates the nature sex of glaucoma's patient either of Male or Female.
Long-sight-effect	This feature indicates the reduction of the long sight or effect is occurred on the convex lens for loosing long sight.
Cornea-cloudy	As the name indicates this feature is concerned with loss of transparency of the cornea.
Eye-domineer	This one shows the status of the eye become domineer or the eye is looking like changed the original situation due to the some removing of the eye bow hair and mucus-like liquidation occurred.
Sudden-eye-pain	This feature Points out the immediate shock or pain of the eye that occurred.
Nausea	Nausea indicates the biliousness occurred, but not vomit rather sicken.
Headache	Headache is the action of immediate penetrating of the brain.
Blurred-vision	Seeing the object in unclear status, but the object was clear.
Itchy	High sensitivity for friction creation of an eye bow or eye casing portion.

Vomiting	An immediate reflex action to turn back of food from the stomach the case for relating brain (controller) and eye humor.
Sudden-sight-loss	An immediate limitation of sight or speedy looseness of eyesight characteristics.
Eye –casing-distent	This feature is described as swelling of the whole part of the eye bow. Differ from eye masacaration as, it is easy-full distent of the eye bow.
Light-sensitivity	The occurrence of the fancy or flashy of light.
Eye-hoodness	The color of the eye becomes red.
Frequent-Eye-trauma	The frequent trauma or eye shaken or rolling of cob like beachy.
Tear-percolate	The liquidation occurrence of tear on the glaucoma patient.
Living Area	Indicate the come up the living area of glaucoma's ailment or patients. Which has three basic categories. Namely Highland: hilltop area and cold air pressure. Lowland: desert, some-desert area, and light air pressure. Temperate: moderate air-conditioned area.

As described in the above table, the dataset was collected based on those attributes with their corresponding values. Those attributes or features are formulated based on the data (manual card) which are shelved in an organization after 2012. The following table shows the sample collected dataset from Borumeda hospital.

Table 3. 3 sample of collected dataset from BMH

age	sex	long sig	cornea	Eye do	Sudden	Nausea	Headac	Blurred	Itchy	Vomiti	Sudde	Eye c	Light se	Eye ho	frewue	Tear pe	Living A	Class la
child	female	no	yes	no	yes	no	yes	no	no	yes	no	yes	no	no	no	yes	temprate	secondary glaucoma
Yeard	male	yes	yes	yes	yes	no	no	yes	yes	no	yes	no	no	no	no	no	highland	primary angle closure glaucoma
child	male	no	no	no	yes	yes	yes	yes	yes	no	yes	no	no	yes	yes	no	temprate	secondary glaucoma
child	female	yes	no	yes	no	yes	yes	no	no	yes	no	no	no	no	no	yes	temprate	primary open angle glaucoma
Yeard	male	no	no	no	no	no	no	yes	yes	no	yes	yes	no	yes	no	no	temprate	primary angle closure glaucoma
child	male	yes	yes	no	yes	yes	yes	yes	yes	yes	no	no	no	no	yes	no	temprate	congenital glaucoma
Yeard	female	no	no	yes	no	yes	yes	no	no	yes	yes	no	yes	no	no	yes	temprate	secondary glaucoma
Young	male	no	no	no	yes	yes	yes	yes	no	yes	yes	no	yes	no	yes	no	highland	primary angle closure glaucoma
child	male	no	no	no	no	no	yes	yes	yes	no	yes	yes	no	no	no	no	temprate	congenital glaucoma
child	female	yes	yes	yes	no	yes	no	yes	no	yes	no	no	yes	no	no	no	temprate	secondary glaucoma
Yeard	male	yes	no	no	no	yes	no	no	yes	yes	yes	yes	yes	yes	yes	yes	temprate	secondary glaucoma
child	male	no	no	no	yes	no	no	yes	yes	yes	yes	yes	no	no	no	no	lowland	primary angle closure glaucoma
child	male	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	no	no	no	no	no	lowland	primary open angle glaucoma
Yeard	female	no	yes	no	no	yes	yes	yes	yes	no	no	no	yes	no	yes	no	temprate	primary open angle glaucoma
child	male	no	no	no	no	no	no	no	no	yes	no	no	no	no	no	yes	temprate	secondary glaucoma
Yeard	male	yes	no	yes	yes	yes	yes	yes	yes	yes	no	yes	no	no	no	no	lowland	congenital glaucoma
Yeard	female	no	no	no	yes	yes	yes	yes	yes	yes	yes	no	no	yes	yes	no	temprate	primary open angle glaucoma
Young	male	no	yes	yes	yes	no	yes	yes	no	no	no	no	no	no	no	no	lowland	secondary glaucoma
child	male	yes	no	no	no	yes	no	no	yes	yes	yes	no	no	yes	no	yes	highland	congenital glaucoma
child	female	no	no	yes	yes	yes	no	yes	yes	yes	yes	yes	yes	yes	no	no	highland	primary angle closure glaucoma
Yeard	male	yes	yes	no	yes	no	yes	yes	yes	no	no	no	yes	no	yes	no	temprate	secondary glaucoma

3.3.2 Data Preprocessing

Data preprocessing is a major task of data mining technique that is used to form data analysis and transform the raw data into a useful and efficient format. This data preprocessing is mainly concerned with the task of missing value handling with a method of ignoring or fill the values, though different methods, data cleansing, data transformation with normalization and discretization. In this research, the collected dataset has some missing values, imbalance classes, and needs feature selection. The missing value was handled by replacing the maximum frequent value in the recorded dataset.

In detail discerption, the missing value of each dataset is a full field or handled through replacing by maximum frequent value in the dataset. Whereas feature selection is operated through an information gain value. The information gain value of each attribute was calculated through calculating the entropy of classes (four classes namely, primary open-angle glaucoma, primary angle-closure glaucoma, secondary glaucoma, and congenital glaucoma) and calculating the gain value of each attribute and subtracting it from the entropy and put them orderly from highest to lowest. The other remaining data cleaning procedure is data discretization where data are discrete, making it easier for understanding since it is so tedious to use the data as it is recorded. In this study, the need for discretization is an attribute age to categorize into some ranges. Glaucoma ailment has four basic class labels with classified coverage of the whole dataset. The number of instances in each class label is shown in the table below

Table 3. 4 number of instances for each glaucoma ailment

Glaucoma ailment type	**Number of instances**
Primary open-angle glaucoma	816 instances
Primary angle-closure glaucoma	664 instances
Secondary glaucoma	741 instances
Congenital glaucoma	921 instances

3.3.2.1 Missing Value Handling

Missing values are common challenges in the data mining area and their values might be either of numeric or nominal value. The following table shows the number of missing values and their percent coverage of missing values for each feature.

Table 3. 5 missing value percentage coverage in the dataset

Attributes name	Number of missing value instances	Percent coverage
Age	4	0.12%
Sex	4	0.12%
Long-sight-effect	7	0.22 %
Cornea-cloudy	4	0.12%
Eye-domineer	6	0.19 %
Sudden-eye-pain	4	0.12%
Nausea	8	0.25 %
Headache	9	0.28 %
Blurred-vision	12	0.38 %
Itchy	19	0.6 %
Vomiting	13	0.41 %
Sudden-sight-loss	8	0.25 %
Eye-casing-distent	10	0.31 %
Light-sensitivity	9	0.28%
Eye-hoodness	8	0.25 %
Frequent-Eye-trauma	11	0.35%
Tear- percolate	9	0.28 %
Living Area	11	0.28 %
Total missing value	156	4.96 % missing values

Almost every attribute has a missing value with different percent coverage. The over-all number of missed value for all feature was 156 from a total of 3142 instances. Missing values were handled through replaced by highly frequent values of the existed data.

3.3.2.2 Attribute Selection

Attribute or feature selection is a mechanism of selecting attributes with different methods like through their information gain value, gain ratio value and Gini index result. For this research, a researcher used an information gain value of each attribute. The information gain value of each attribute was calculated and attributes are listed in descending order based on their information gain value. The final experiment was conducted using both the whole attributes and particular attributes. The mechanism for selecting the attribute from the whole feature should be through assigning the threshold value. This threshold value is assigned or selected based on the highest information gain scored value gap of attributes. The information gain value statistical difference of each feature is almost nestled. But, the mechanism to assign the threshold value is finding the highest gain value difference and select the attribute which has the gain value above the high gap of gain value difference and some attribute was included based on domain expert advising point.

The expert advising point is that some common and necessary features (symptoms) are mandatory to identify types of glaucoma ailment. The revolution or the total process for feature selection is shown in the figure below.

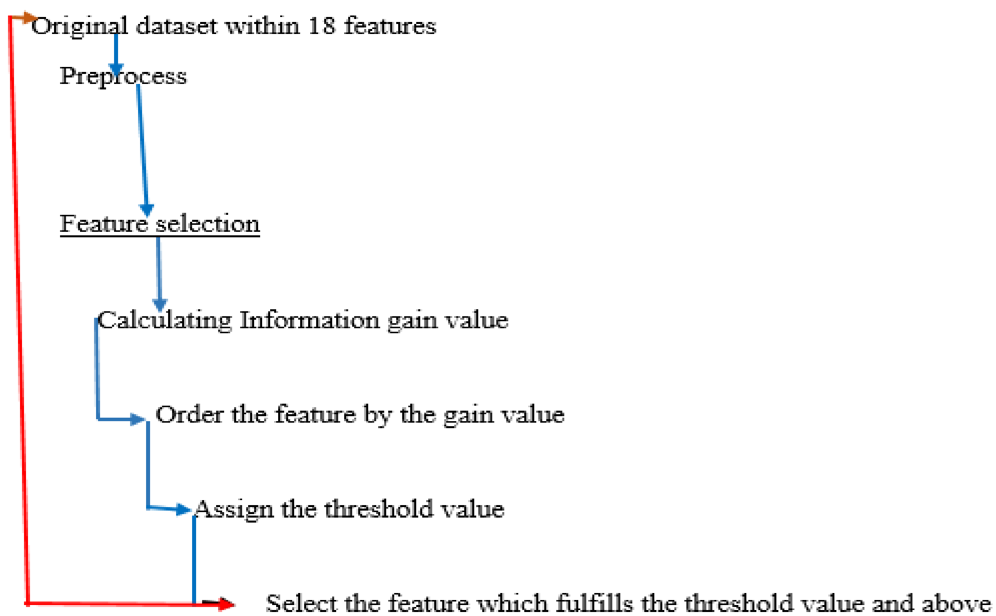

Figure 3. 4 feature selection process

The calculated information gain value of each attribute through the tool of the rapid miner is shown in the table below.

No	Attribute name	Information Gain Value
1.	'Headache'	0.127
2.	'age '	0.091
3.	'Nausea'	0.074
4.	'sudden eye pain'	0.072
5.	'frequent eye trauma'	0.066
6.	'Sudden sight loss'	0.058
7.	'long sight effect'	0.053
8.	'Itchy'	0.047
9.	'Vomiting'	0.040
10.	'Tear percolate'	0.036
11.	'Living Area'	0.035
12.	'Eye domineer '	0.025
13.	'Eye hoodness'	0.023
14.	'cloudy cornea'	0.011
15.	'sex '	0.006
16.	'Eye casing distent '	0.006
17.	'Blurred vision'	0.005
18.	'Light sensitivity'	0.003

Figure 3. 5 information gain value of each attribute through the rapid miner

3.3.3 Data Categorization

The data categorization phase which focused on the categorization of attribute values. Data categorization is characterized as a procedure of changing over consistent information property estimations into a limited arrangement or category of data with negligible loss of it. Discretization is the mechanism of categorizing data into different categories. In this research, the discretization is required for age. According to [60] age can be categorize into three main basic categories. Namely, Child, young, and yeard. Each year category coverage is from a year up to fifteen years, from fifteen up to fifty-one and age of patient greater than fifty-one respectively. The reason for categorizing each age range is that the boundary between the ages has similar index characteristics.

3.3.4 Data Mining

Data mining is the process of searching through large datasets to identify patterns and establish relationships to solve problems through data analysis. Data mining (knowledge discovery in databases) is the extraction of interesting (non-trivial, implicit, previously unknown, and potentially useful) information or patterns from data in large databases. The ultimate goal of the knowledge data mining process is to find the patterns that are hidden among the huge sets of data and interpret them to useful knowledge and information. Data mining is a central part of the knowledge discovery process. The term data mining and KDD uses reversibly but in this case, data mining is one step or part of the KDD process. So, in the KDD process model, data mining is the fifth step which mainly targeted extracting on new artifacts from different mining techniques like classification, clustering, or association techniques with different algorithms. Concerning this research, the researcher uses classification techniques from data mining tasks and classification algorithms like J48, PART, Naïve Baye, Jrip, and such algorithms for mining new artifacts.

3.3.5 Interpretation and Evaluation

After conducting the above all KDD phases, this phase is the last and the final step which is focused on interpreting and evaluating the extracted result or model generated. The identified knowledge is also recorded for further use and overall feedback and discovery results were evaluated.

In this work, an interpretation is on the visualizing mechanism of the diagram and selecting a model that scores the highest performance value.

The KDD process model concerning this research is started with understanding the business concept of the organization and conceptual analysis of the glaucoma patient in the organization. After this stage, the targeted data is selected and collected from the health organization called BMH. After collecting the dataset the next step is to perform the preprocessing activity on the collected dataset like missing vale handling, feature selection, transformation and SMOTE analysis. After those all preprocessing activities are done, the next work is discovering new knowledge through different classification algorithms and selected the one which scored the highest performance result by evaluating the performance through accuracy, recall, f-score, confusion matrix, TP, and FP. Then use the discovered knowledge as a new artifact generated through a selected algorithm.

3.4 Handling Class Imbalance Problem

The class imbalance problem is a major problem for the classification issue. This type of problem is handled by a method of balancing the class called SMOTE. The term SMOTE stands for Synthetic Minority Oversampling Technique, which is used to deal with the problem of having few cases in the minority class in a dataset and used to remove the class imbalance problem. The module creates new examples by taking minority class samples and introduces synthetic examples along the line segments joining any/all of the k minority class nearest neighbors [61]. The class-imbalance problem was anticipated both at the data and algorithmic stages. In the case of the data point, the data imbalance issue is reflected in over-sampling and under-sampling. SMOTE is a mechanism for handling an over-sampling (over-sample the minority Class to address the imbalanced datasets issue).

As discussed on the concept of class balancing problem, the SMOTE technique is used to balance the class. The class label has four basic values with so far distance scores in between them. The dataset before SMOTE and after SMOTE is shown as follows in the figure 3.7 below.

Table 3. 6 values of the class label before SMOTE

Class label name	**Value before SMOTE**
Congenital glaucoma	921
Primary angle-closure glaucoma	664
Primary open-angle glaucoma	816
Secondary glaucoma	741

Figure 3. 6 imbalanced dataset before SMOTE

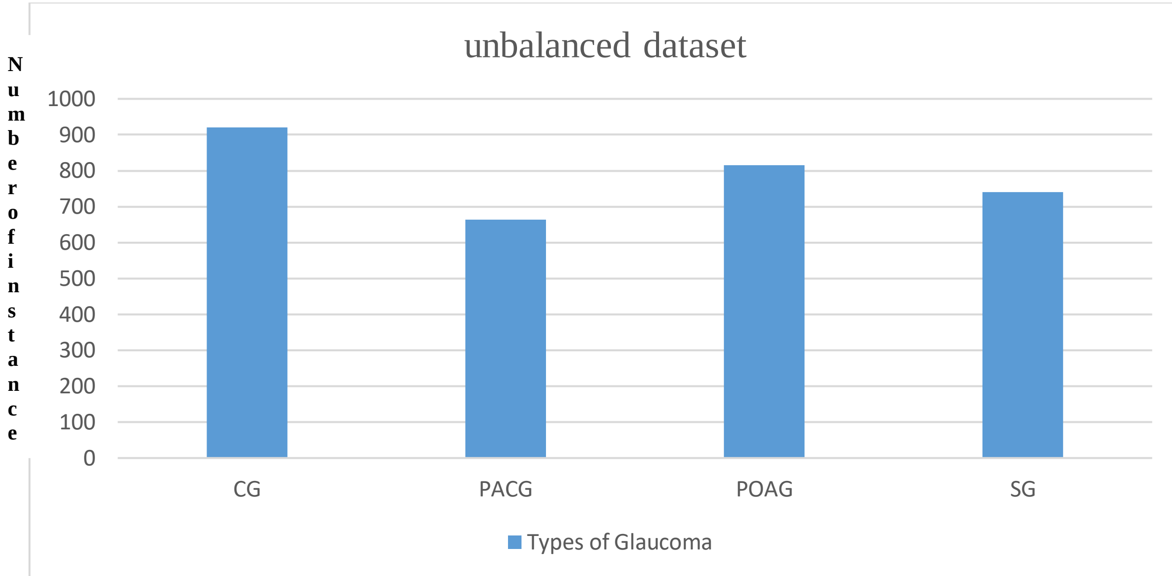

Table 3. 7 value of class label after SMOTE

Class label name	Value after SMOTE
Congenital glaucoma	921
Primary angle-closure glaucoma	803
Primary open-angle e glaucoma	909
Secondary glaucoma	902

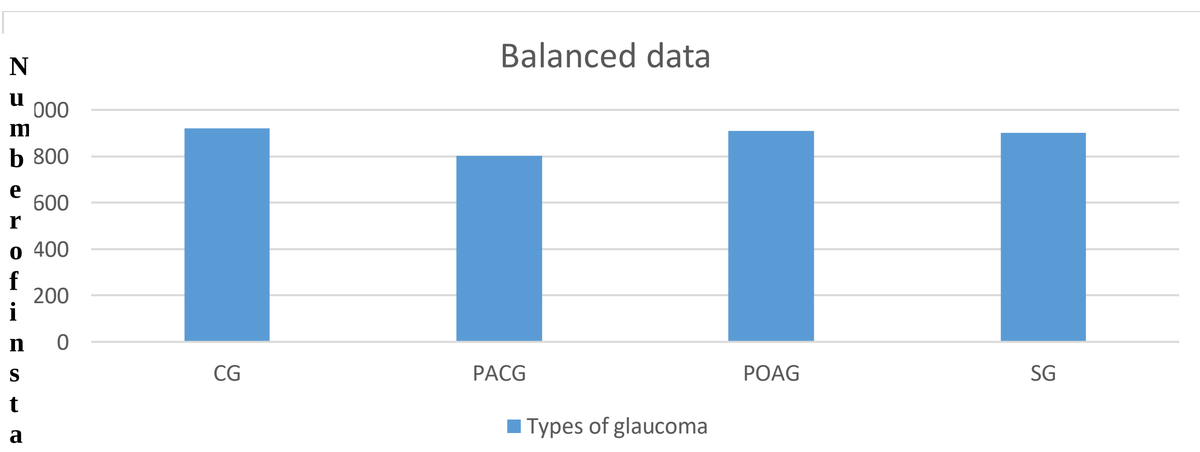

Figure 3. 7 Balanced datasets after SMOTE

3.5 Classification Algorithms

To conduct an experiment of this research, the researcher uses classification algorithms among data mining techniques namely, naïve baye algorithm, PART algorithm, SVM, J48 algorithm, random forest, KNN, Jrip, Naïve baye..etc. Among those classification algorithms the researcher uses Naïve Baye, Jrip, PART algorithm, and J48 under two basic test options with both selected and unselected or whole attributes.

The reason why only four classification algorithms selected is that because from different classification algorithms those selected algorithms have better accuracy results than others within the repetition experiment under this study. Since the researcher did repeat the experiment with different classification algorithms and select the four algorithms than others.

The description of these four adopted classifier algorithms is as follows.

Naïve Baye algorithm: - Naive Bayes is a classification algorithm that is based on Bayes theorem with strong and naïve dependence free assumptions and is the simplest probabilistic classifiers. The Naïve Bayes classifiers are used for any classification in machine learning based on the conditional possibility or probabilities of the features attributed to a class, in which the features are selected by feature selection methods [62].

The procedural steps of the Naïve Bayes algorithm for classification concerning this research are shown as follows.

Step 1: Calculate the prior probability of four class labels namely, primary open-angle glaucoma, primary angle-closure glaucoma, secondary glaucoma, and congenital glaucoma

Step 2: Find Likelihood probability with each attribute for each class label (primary open-angle glaucoma, primary angle-closure glaucoma, secondary glaucoma, and congenital glaucoma)

Step 3: Put these values in Bayes Formula and calculate the posterior probability of the new case to belong to which class label. Use the Baye's formula here

$$P(ci/x) = \frac{p\,(x/ci)P(ci)}{p(x)} \qquad \text{2.14}$$

Step 4: See which class has a higher probability, given the input belongs to the higher probability class either of primary open-angle glaucoma, primary angle-closure glaucoma, secondary glaucoma, and congenital glaucoma.

PART algorithm: - PART (Projective Adaptive Resonance Theory) is simply a separate and conquer rule learner and producing sets of decision lists rules. A new data are compared to each rule in the list in turn, and the item is assigned the class of the first matching rule. PART builds a decision tree in each iteration and makes the best leaf into a rule [63].

The general working procedures of the PART algorithm concerning with this research is shown as follows.

Step 1. Construct a decision tree on the current set of instances of glaucoma patients with seven symptoms included.

Step 2. Create a rule from the decision tree to identify types of glaucoma or classify it into given class labels.

Step 3. Select the leaf with the largest coverage is made into a rule for glaucoma type identification system based on an expert advising ideas.

Step 4. Discarded the decision tree after the rule is selected (rules which have include seven basic symptoms) for glaucoma type identifying.

Step 5. Eliminate the instances covered by the rule if the covered instances do not eliminate the next rule is identical to the previous rule.

Step 6. Repeatedly do the steps until the rules satisfies enough to type identification.

J48: - J48 is an extension of ID3. An added features of J48 are secretarial for missing values, decision trees clipping, continuous attribute value ranges, derivation of rules, etc.

The WEKA tool provides several options associated with tree trimming. In the case of potential overfitting, trimming can be used as a tool for précising. In other algorithms the classification is performed recursively till every single leaf is pure, that is the classification of the data should be as perfect as possible.

This algorithm generates the rules from which particular identity of that data is generated. The objective is gradually generalization of a decision tree until it gains the balance of elasticity and accuracy [64].

The working procedure of the J48 algorithm for this research is shown as below

Step 1. Check and hold the whole dataset.
Step 2. Find the information gain value for each attribute
Step 3. Select the best attribute based on its information gain value.
Step 4. Create a decision node by ordering attributes based on their information gain value through decreasing order.
Step 5. Construct rules based on ordering nodes starting from the root node and write it in the form of IF-THEN rule.

Jrip: - Jrip is Repeated Incremental Pruning to Produce Error Reduction (RIPPER) which is the most popular algorithm and uses sequential covering algorithms for creating ordered rule lists. Classes are examined in growing size and an initial rule for the class is produced using incrementally reduced error Jrip (RIPPER) proceeds by treating all the examples of a particular decision on the training data as a class and finding a set of rules that cover all the members of that class. Thereafter, it proceeds to the next class and does the same, repeating this until all classes have been covered [65].

The major steps of Jrip (RIPPER) algorithm concerning this research is shown as follows

Step 1.Building stage

1.1 Growth phase: - In this phase, a rule is generated (automatically) by greedily adding attributes to the rule until the rule meets stopping criteria.

1.2 Prune phase:- In this phase each rule is incrementally pruned, allowing the pruning of any final sequence of the attributes, until a pruning metric is fulfilled

Step 2. Optimization stage: - In this stage, each generated rule is further optimized by

- Greedily adding attributes to the original rule and
- Independently growing a new rule undergoing a growth and pruning phase.

Step 3. Selection and deletion stage: - In this stage the best rules are kept and the other rules are deleted from the model to determine glaucoma types.

3.6 Performance Evaluation Metrics

Model evaluation centers around assessing all the experimented classifier models acquired by utilizing the chosen classification algorithms within two basic test options based on whole and selected attributes. The evaluation metrics adopted by this research are confusion matrix, false-positive rate, true positive rate, Accuracy and Precision, Recall, F-score, and ROC curve analysis to compare one classifier model to others. After evaluating all generated classifier models, the model with the best performance metrics was selected and the knowledge obtained from the selected model is used to type identification and treatment of glaucoma.

3.7 Knowledge Acquisition Methods

The research acquires the knowledge from extracted data in the form of an automatic rule generation from the dataset, interview of a domain expert, and document analysis for each type of glaucoma ailments. The following points are acquiring the mechanism of knowledge manually.

3.7.1 Interview with Domain Expert

For acquiring of knowledge from domain expert through an interview, the researcher uses both structured and unstructured interviews about the concept of glaucoma, the occurrence of glaucoma, types of glaucoma ailment in the world and those occurred in Ethiopia, treatment mechanism, medication scheme for each type of glaucoma ailment.

To extract the required knowledge from domain experts, a purposive sampling technique was applied. Purposive sampling is the best way to provoke the views of persons who have expertise and knowledge about a specific domain. Thus domain experts were selected purposively based on their working experience, level of education, and current roles in Borumeda hospital.

Table 3. 8 interviewee background

Sex	**Education level**	**Position**	**Work Experience**
male	MGR	MED	Eight years
female	MD	Eye specialist	Eight years
Female	BSC candidate	coordinator	Apparent
Female	BSC candidate	coordinator	Apparent
Female	MD	Normal service provider	Six years

Keys: - MGR: - *Manager,* ***MD****; medical doctor,* ***HO;*** *health officer* ***BM****; Borumeda Hospital,* ***MED****; Manager of Eye Dep't,* ***BSC****; Bachelor of Science*

3.7.2 Document Analysis

The other method of knowledge acquiring mechanism is document analysis. In this health organization, a document is written in the form of a magazine that is reviewed and updated within some period of time interval when the new medication is produced. This reviewed document contains the new characteristics of the glaucoma and treatment mechanism of the updated information. In simple content, it contains the occurrence of glaucoma at what time and at what condition with its manifestation and its treatment mechanism.

3.8 Knowledge Representation

The researcher uses rule-based knowledge representation because it is the most predominant knowledge representation technique used to develop a knowledge-based system, which the researcher used for type identifying as well as treatment ordering, can be easily converted to rules by organizing the knowledge into IF…. THEN format. Similarly, the automatic knowledge was extracted from the raw data (after preprocessing) by analyzing the classifier algorithm and represent in the form of rule. Likely, the knowledge acquired from experts or the documents was also represented in the form of rule. So, the rule-based representation method was implemented to represent the expert and knowledge document. This rule-based knowledge representation has many benefits compared with other knowledge representation practices.

The reason for the choice of rule-based reasoning is that this method is a common one, it can be agreeably powerful from the viewpoint of building useful claims and the most predominant knowledge representation technique. Moreover, the experience or knowledge of domain experts and documents were captured and represented in the form of IF-THEN rules [66]. This makes it easy to represent the existing knowledge and the newly discovered knowledge in the form of rule.

General Procedure of Knowledge acquisition up to representation from two sides is as shown below

1. *Manual knowledge acquisition*

- *Domain expert (ophthalmologist)*
  - *Through Interview*
- *Existing document*
  - *Through document analysis*
- *Then combine both knowledge and represent in the form of rule (rule based knowledge).*

2. *Automatic knowledge generation through Data mining techniques*

- *Using raw data*
  - *Preprocessing activity has done*
    - *Generate rule via different classification algorithms*
      - *Select the one algorithm result based on performance metrics evaluation*
        - *Take rule generated by the selected algorithm*
          - *Select the rules (If-Then) which fulfill experts' advice*
- *Collaborate knowledge represented from two sides (knowledge from 1 and 2).*

3.9 Collaboration of data mining results with knowledge-based system

The researcher needs to combine data mining results with a knowledge-based system where the knowledge was collected from domain experts and document analysis. Whereas the knowledge extracted from the data mining results was used for identifying or classifying types of glaucoma ailments, the knowledge-based system is used to order treatment for each category of glaucoma ailments. The research functionality or final targeted output is done by collaborating the result of data mining with the knowledge-based system.

The knowledge-based system is developed based on the knowledge obtained from domain experts and document analysis coordinately represented in the form of IF… THEN rule for ordering the treatment of each type of glaucoma ailment. The rule extracted from data mining and the knowledge collected from experts was rule-based and follow IF… THEN rule. So, those rules are easily collaborating. The researcher uses the Waikato Environment for Knowledge Analysis (WEKA) machine learning tools to identify or classify types of glaucoma ailment based on the dataset collected from the Borumeda Hospital after preprocessing activity were done. Again the researcher uses the tool called SWI-Prolog to develop the knowledge-based system. To join the results of data mining with the knowledge-based system for developing the system, SWI-Prolog and JAVA tools were applied and the final prototype system was developed.

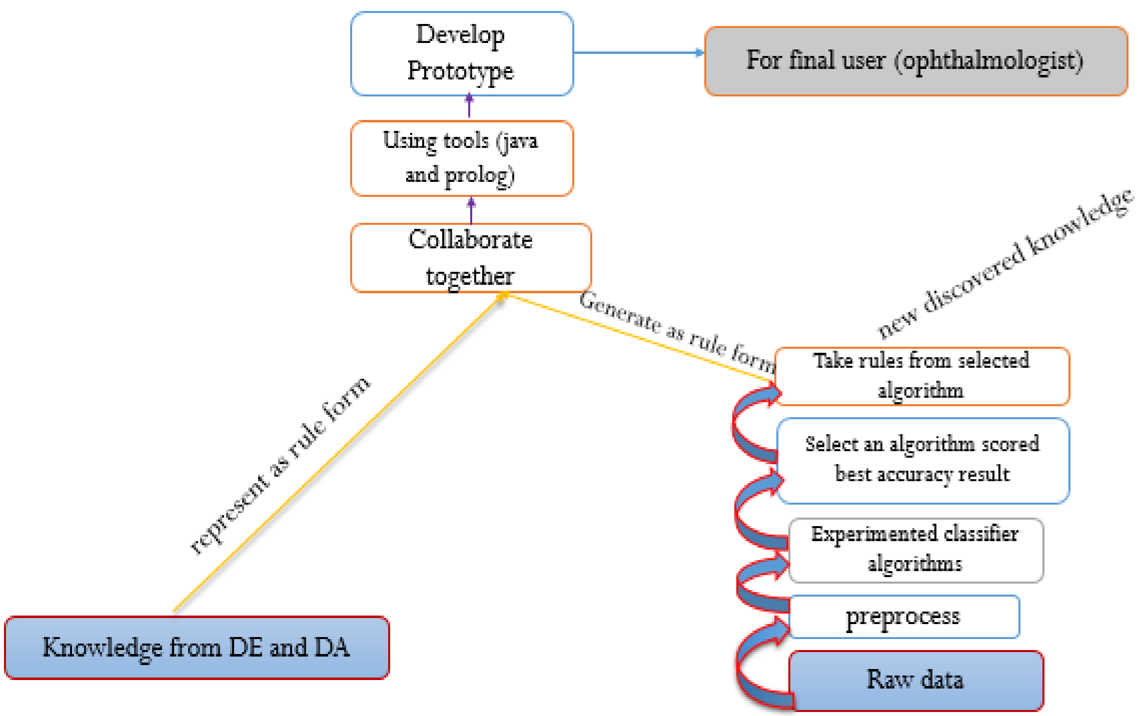

Figure 3. 8 Knowledge combination mechanism from two sides

3.10 User Interface Development

User interfaces are the access points where users interact with designs and the design of an interface is an easy way to communicate the user and developed a prototype. So, this phase is used for developing a prototype system for evaluating the functionality of the final developed classification model. NetBeans Java Development software with a prolog tool was applied to develop the prototype system.

The prolog tool is used for storing the knowledge in the knowledge base and connecting to Java by helping JPL package for automatically rule fetching and the Java tool is used for final user interface development for final users (ophthalmologist).

CHAPTER FOUR

EXPERIMENTAL ANALYSIS AND RESULT

4.1 Introduction

Experiments are done to develop a classifier model and to extract relevant knowledge in the form of rules. In this study, many experiments were conducted using various data mining methods to derive knowledge based on preprocessed data for identifying glaucoma. An experiment is conducted using a total of 3535 instances of 18 features, though both selected and whole attribute by using different classification algorithms that classify the data automatically. Finally, the result of those all accuracy is compared to each other and the best can be selected.

4.2 Experimentation setup

Primarily, to conduct the experiment there must be understanding features with their information gain value as discussed in chapter three. So, the ordering impact of attributes for classifying of glaucoma ailment is conducted by the WEKA tool after the data were preprocessed through rapid miner. The dataset of glaucoma recorded from problem selected area (Borumeda hospital) was converted into an ARFF (Attribute Relation File Format) for the WEKA tool understandable format

4.3 Test options and Threshold boundary

For this study, the hidden knowledge is extracted in the form of rule. To conduct experiments, the researcher uses test options coordinating with different algorithms. An experiment is conducted either in the whole or selected attribute. The attributes are selected based on their information gain value and by assigning threshold value and adding some attributes based on the knowledge of domain experts on selected attributes. So, based on the value of their information gain, fourteen attributes are selected and one attribute was added to selected attributes for an experiment based on the domain expert idea.

Test options

For experimenting, the researcher selects two basic test options with coordinating different (four) types of algorithms. The dataset should be converted to ARRF (Attribute Relation File Format) file to understand the format for analyzing the software.

K- Fold cross-validation: -Cross-validation is a resampling process that has the functionality of evaluating machine learning models on a data section. The procedure has a single parameter called k that refers to the number of groups that a given data to be split. It is making a fold of a dataset into some folds or cataloging the data into k folds. In this research, the researcher uses 10fold cross-validation with its default value.

Percentage split: -This type of test option is applying a partitioning method of a dataset into training and test set in percentage. In the case of this research, the researcher uses 80/20 splitting test option means that 80% of the dataset is used for training purposes and the remaining 20 % is used for testing determination.

Threshold boundary: -Threshold value is a minimum value that is used to select the attributes for experiment case based on the value of information gain value. This threshold value is assigned a border in between the gain value of each attribute, but the border is assigned between the highest gap score valuable points. In this research, the minimum threshold border is the information gain value of 0.023 and above are selected. Some attributes were added to those selected attributes, but their value is less than the minimum threshold value. The reason why add these attributes which did not fulfill the threshold value is according to the domain expert perception or advice.

The based boundary of a threshold value, the selected attributes are Headache, 'age, Nausea, 'sudden eye pain', 'Sudden sight loss', frequent eye trauma, long sight effect', itching, vomiting, tears percolate, living area, eye domineer, eye hoods with decreasing values of their information gain value respectively. Additionally, one parameter, namely “blurred vision” was added even if it can’t fulfill the required threshold value rather based on expert advice since this attribute is very critical and a major symptom of every type of glaucoma ailment.

4.4 Experimentation

To conduct an experiment, types of classification algorithms and selected test options have a great role. So, four basic classification algorithms with two main basic test options were selected for an experiment. There are four basic scenarios. But each scenario has four experiments (the whole and selected attribute within each of two test options). Totally sixteen experiments were conducted.

4.4.1 Scenario I

4.4.1.1 Experiment I

A) Jrip with whole attribute via 10 fold cross validation

This experiment is conducting by the rule-based classifier called Jrip)(RIPPER) with the test option of 10 fold cross-validation (default value) by applying whole attributes and it consumes an execution time of 1.42 seconds to score an accuracy result 63.5%. Another evaluation mechanism is the confusion matrix. The confusion matrix evaluation of an experiment conducted by the Jrip algorithm is shown below.

Table 4. 1 Confusion matrix with Jrip algorithm with the whole attribute via 10 fold cross-validation

	Predicated class					
		A	B	C	D	Classified as class of
Actual class	A	732	143	145	45	A=SG
	B	253	634	77	61	B= PACG
	C	313	78	565	82	C= POAG
	D	164	69	88	717	D=CG

Table 4. 2 detailed accuracies by the class

	TP rate	FP rate	precision	f-measure	ROC area	Class label
	0.687	0.235	0.501	0.279	0.783	Congenital Glaucoma
	0.619	0.092	0.686	0.651	0.820	Primary angle closure Glaucoma
	0.544	0.099	0.646	0.591	0.799	Primary open angle glaucoma
	0.691	0.060	0.792	0.738	0.851	Secondary Glaucoma
Weighted Average	0.636	0.123	0.655	0.639	0.813	

4.4.1.2 Experiment II

B) Jrip with selected attribute via 10 fold cross-validation

This experiment is conducting via the rule-based classifier called Jrip (RIPPER) with the test option of 10 fold cross-validation (default value) with selected attributes with the time consumption of 1 seconds to score an accuracy result 67.5%. The general statistical analysis of this experiment is detailed with the true positive and recall an equal score of 0.67.

Table 4. 3 Confusion matrix with Jrip algorithm with selected attribute via 10 fold

	Predicated class					
		A	B	C	D	Classified as a class of
Actual class	A	639	77	120	128	A=SG
	B	40	541	34	137	B= PACG
	C	107	39	464	207	C= POAG
	D	48	88	92	755	D=CG

Table 4. 4 detailed accuracies via each class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.663	0.075	0.766	0.711	0.834	Congenital Glaucoma
	0.729	0.077	0.755	0.742	0.873	Primary angle closure Glaucoma
	0.544	0.097	0.650	0.593	0.811	Primary open angle glaucoma
	0.684	0.066	0.776	0.727	0.856	Secondary Glaucoma
Weighted Average	0.675	0.112	0.685	0.6675	0.841	

4.4.1.3 Experiment III

C) Jrip with whole attribute via percentage split 80/20

In this experiment, the whole attribute is used for model building but the test option is a percentage split with the splitting of the dataset into eighteen percent (80%) of the data used for training purposes and the remaining twenty (20%) are used for testing.

Table 4. 5 Confusion matrix of Jrip algorithm with whole attribute via percentage split 80/20

	Predicated class					
		A	B	C	D	Classified as a class of
Actual class	A	128	31	28	8	A=SG
	B	66	127	10	14	B= PACG
	C	58	23	114	27	C= POAG
	D	23	10	13	153	D=CG

Table 4. 6 detailed accuracy by the class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.656	0.230	0.465	0.545	0.752	Congenital Glaucoma
	0.585	0.104	0.665	0.623	0.823	Primary angle closure Glaucoma
	0.514	0.053	0.691	0.589	0.781	Primary open angle glaucoma
	0.769	0.077	0.757	0.769	0.853	Secondary Glaucoma
Weighted Average	0.6627	0.122	0.647	0.763	0.802	

4.4.1.4 Experiment IV

D) Jrip with selected attribute via percentage split 80/20

This experiment is based on selected attributes which are used for model building but the test option is a percentage split with the splitting of the dataset into eighteen percent (80%) of the dataset used for training purpose and the remaining twenty percent (20%) were used for testing purpose.

Table 4. 7 Confusion matrix of Jrip algorithm with selected attribute via percentage split 80/20

	Predicated class					
		A	B	C	D	Classified as a class of
Actual class	A	112	25	39	23	A=SG
	B	4	116	9	30	B=PACG
	C	14	14	78	41	C=POAG
	D	5	20	18	162	D=CG

Table 4. 8 detailed accuracies by the class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.563	0.045	0.830	0.671	0.775	Congenital Glaucoma
	0.683	0.068	0.765	0.722	0.849	Primary angle closure Glaucoma
	0.605	0.127	0.593	0.599	0.813	Primary open angle glaucoma
	0.656	0.077	0.755	0.702	0.830	Secondary Glaucoma
Weighted Average	0.659	0.115	0.676	0.659	0.808	

The confusion matrix is not the only evaluation mechanism for one’s experiment rather true positive rate, false-positive rate, recall, precision, F-measure, ROC area and the main one is an accuracy result of classification with time consumption of the model building. The detailed statistical experimental evaluation score is as shown below.

Table 4. 9 experimental summary via Jrip algorithm

Algorithm	Attribute	Test option	Time taken	TP rate	FP rate	Precision	F-measure	Roc area	Accuracy
Jrip	whole	10 fold cross validation	1.42	0.12	0.65	0.85	0.66	0.79	63.5%
	selected		1sec	0.67	0.11	0.68	0.67	0.84	67.5%
	whole	Percentage Split 80/20	1.06	0.52	0.12	0.64	0.62	0.8	62.6%
	Selected		0.81	0.65	0.11	0.07	0.65	0.83	65.9%

Among four experiments of Jrip algorithm, Jrip with selected attribute via percentage split 80/20, Jrip with the whole attribute via percentage split 80/20, Jrip with selected attribute via 10 fold cross-validation, and Jrip with the whole attribute via 10 fold cross-validation, the highest score of accuracy was 67.4% with the time consumption of 0.98 seconds via Jrip with selected attribute under 10 fold cross-validation as a test option.

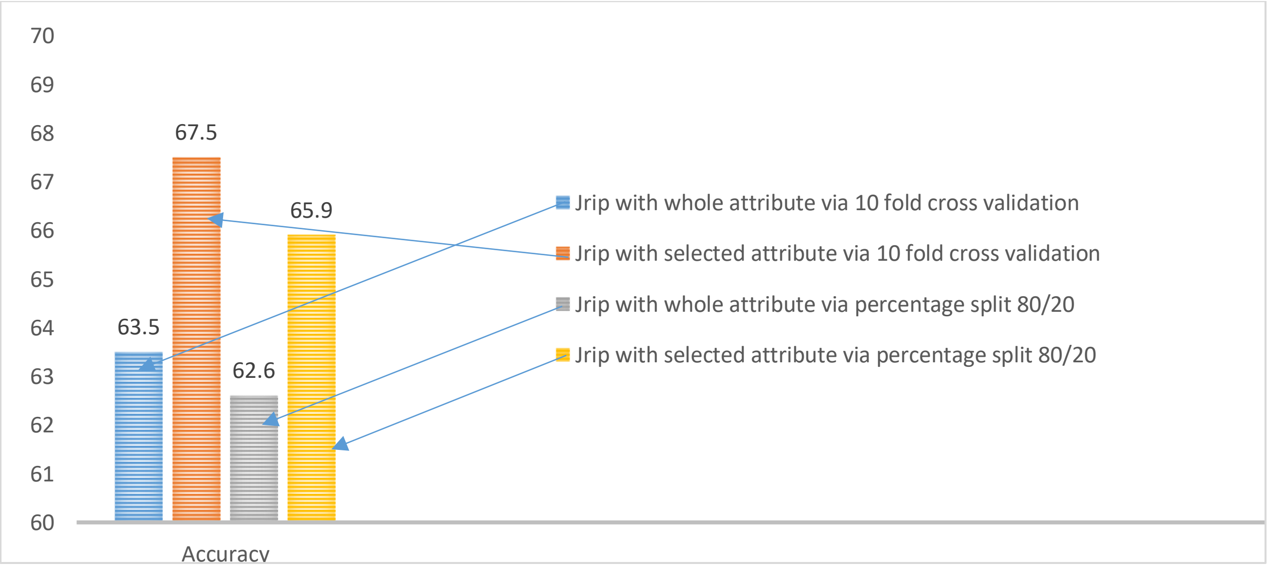

Figure 4. 1 graphical representation of Jrip algorithm accuracy result

4.4.2 Scenario II

4.4.2.1 Experiment I

A) Naïve Baye with whole attribute via 10 fold cross-validation

This experiment is conducting by the classifier called Naïve Baye with the test option of 10 fold cross-validation (default value) with whole attributes with an execution time of 0.03 seconds to score an accuracy result 50.3 %. The general statistical analysis of this experiment the detailed confusion matrix score as shown below.

Table 4. 10 Confusion matrix of naïve baye algorithm with whole attribute via 10 fold

	Predicated class					
Actual class		A	B	C	D	Classified as class of
	A	448	273	150	194	A=SG
	B	123	391	72	275	B= PACG
	C	154	201	313	350	C= POAG
	D	110	60	68	800	D=CG

Table 4. 11 detailed accuracy by the class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.421	0.125	0.537	0.472	0.774	Congenital Glaucoma
	0.514	0.173	0.492	0.503	0.761	Primary angle closure Glaucoma
	0.311	0.093	0.527	0.391	0.731	Primary open angle glaucoma
	0.771	0.271	0.486	0.596	831	Secondary Glaucoma
Weighted Average	0.504	0.16	0.511	0.490	0.774	

4.4.2.2 Experiment II

B) Naïve Baye with selected attribute via 10 fold cross-validation

This experiment was conducted through Naïve Baye classifier with the test option of 10 fold cross-validation with selected attributes and time consumption of 0.1second. It scored an accuracy result of 55.6%. The statistical analysis of this experiment the detailed confusion matrix score as shown below.

Table 4. 12 Confusion matrix of naïve baye algorithm with selected attribute via 10 fold

	Predicated class					
Actual class		A	B	C	D	Classified as class of
	A	587	97	155	125	A=SG
	B	79	406	18	285	B= PACG
	C	212	83	238	284	C= POAG
	D	102	62	73	746	D=CG

Table 4. 13 detailed accuracy by the class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.609	0.152	0.599	0.604	0.802	Congenital Glaucoma
	0.346	0.134	0.470	0.399	0.701	Primary angle closure Glaucoma
	0.526	0.204	0.450	0.485	0.742	Primary open angle glaucoma
	0.398	0.087	0.634	0.489	0.780	Secondary Glaucoma
Weighted Average	0.557	0.156	0.558	0.544	0.806	

4.4.2.3 Experiment III

C) Naïve Baye with whole attribute via percentage split 80/20

In this experiment, the whole attribute is used for model building but the test option is a percentage split with the splitting of the dataset into eighteen percent (80%) of the data used for training purposes and the remaining twenty percent (20%) was used for testing. This experiment results in the accuracy of 47.6% with the execution time taken of 0.01 second.

Table 4. 14 Confusion matrix of naïve baye algorithm with whole attribute via percentage split 80/20

	Predicated class					
Actual class		A	B	C	D	Classified as class of
	A	81	53	21	40	A=SG
	B	34	102	14	67	B= PACG
	C	24	56	61	81	C= POAG
	D	23	10	13	153	D=CG

Table 4. 15 detailed accuracy by the class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.415	0.127	0.500	0.454	0.782	Congenital Glaucoma
	0.470	0.193	0.462	0.466	0.750	Primary angle closure Glaucoma
	0.275	0.079	0.560	0.369	0.745	Primary open angle glaucoma
	0.769	0.297	0.449	0.567	0.831	Secondary Glaucoma
Weighted Average	0.477	0.172	0.494	0.461	0.775	

4.4.2.3 Experiment IV

D) Naïve Baye with selected attribute via percentage split 80/20

For this experimentation selected attribute was used for model building but the test option is a percentage split with the splitting of the dataset as eighteen percent (80%) of the data used for training purpose and the remaining twenty percent (20%) were used for testing.

Table 4. 16 Confusion matrix of naïve baye algorithm with whole attribute via percentage split 80/20

	Predicated class					
Actual class		A	B	C	D	Classified as class of
	A	109	21	37	32	A=SG
	B	10	84	4	61	B= PACG
	C	38	15	48	46	C= POAG
	D	21	10	14	160	D=CG

Table 4. 17 detailed accuracy by the class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.548	0.135	0.612	0.578	0.834	Congenital Glaucoma
	0.605	0.104	0.656	0.629	0.860	Primary angle closure Glaucoma
	0.344	0.082	0.563	0.427	0.792	Primary open angle glaucoma
	0.751	0.288	0.485	0.590	0.812	Secondary Glaucoma
Weighted Average	0.565	0.156	0.567	0.555	0.93	

In Jrip algorithm, there are four typical experiments were conducted based on two basic test options namely cross-validation with a default value and percentage split by splitting the data into 80% for training and 20% for testing data through two basic attribute arrangements of whole attribute and selected attribute. The total summary of all the performance evaluation metrics for those experiments are shown in the table below.

Table 4. 18 experimental summary via naïve baye algorithm.

Algorithm	Attribute	Test option	Time taken	TP rate	FP rate	Precision	F-measure	Roc area	Accuracy
Naïve Bayes	whole	10 fold	0.03	0.5	0.16	0.5	0.49	0.74	50.3%
	selected		0.1	0.557	0.156	0.558	0.544	0.81	55.6%
	whole	Split 80/20	0.01	0.47	0.17	0.49	0.46	0.77	47.6%
	Selected		0.01	0.565	0.156	0.56	0.55	0.83	564%

Note :-Among those all experiment conducted using Naïve Baye algorithm the highest accuracy result has been scored using selected attribute with 10 fold cross validation for an accuracy result of 57.6% correct classification.

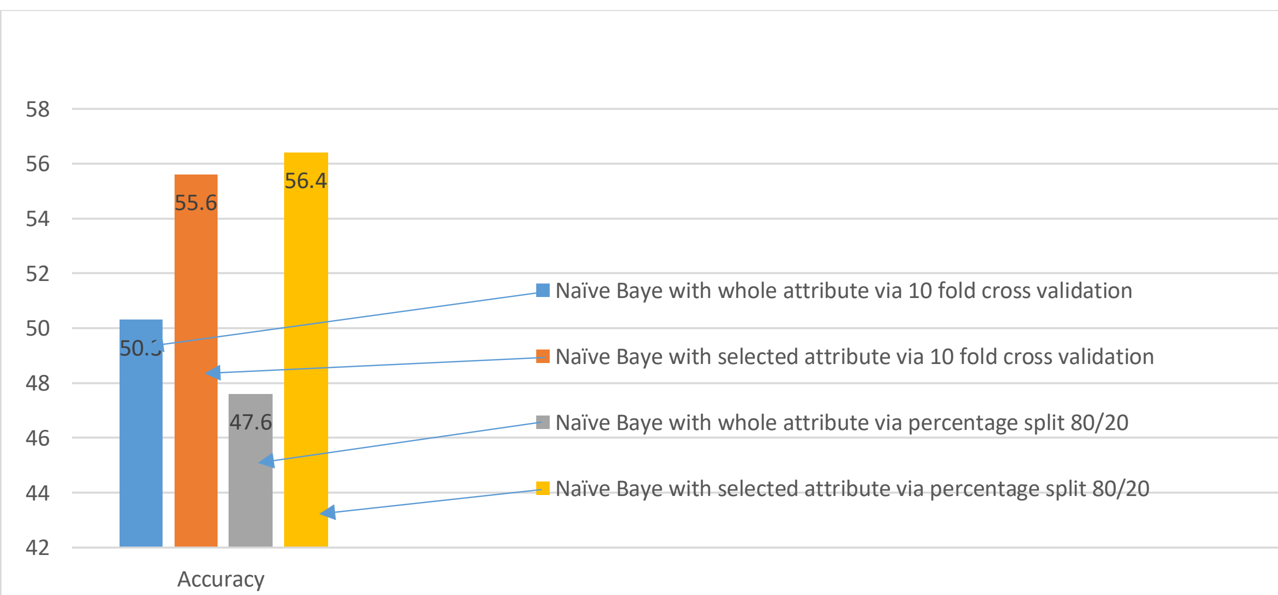

Figure 4. 2 graphical representation of naive baye algorithm accuracy result

4.4.3 Scenario III

4.4.3.1 Experiment I

A) J48 with whole attribute via 10 fold cross-validation

J48 is one of the decision tree algorithms and used for experimenting with this work. But the experiment is conducted on a test option of 10 fold cross-validation using the whole attribute for model development. This experiment result of the score is 68.5 % accuracy with the time taken of 0.06 second and 31.5% of the data were incorrectly classified. The confusion metrics are shown as follows.

Table 4. 19 Confusion matrix of J48 algorithm with whole attribute via 10 fold cross validation

	Predicated class					
Actual class		A	B	C	D	Classified as class of
	A	678	152	176	59	A=SG
	B	111	762	94	58	B= PACG
	C	161	111	654	112	C= POAG
	D	89	78	110	761	D=CG

Table 4. 20 detailed accuracy by the class

	TP rate	FP rate	precision	f-measure	ROC area	Class label
	0.637	0.116	0.653	0.644	0.811	Congenital Glaucoma
	0.743	0.109	0.691	0.716	0.858	Primary angle closure Glaucoma
	0.630	0.121	0.632	0.631	0.806	Primary open angle glaucoma
	0.733	0.073	0.769	0.750	0.874	Secondary Glaucoma
Weighted Average	0.685	0.105	0.686	0.837	0.837	

4.4.3.2 Experiment II

B) J48 with selected attribute via 10 fold cross-validation

In this experiment, the selected attributes were used for model development with the test option of 10 fold cross-validation. To conduct this experiment, 0.04 seconds was required to develop a model. From a total of 3535 instances, 69% were correctly classified while the remaining 31% of the total instance were incorrectly classified. The confusion matrix for this experiment is shown as follows.

Table 4. 21 confusion matrix of J48 with selected attribute via 10 fold cross validation

	Predicated class					
		A	B	C	D	Classified as class of
Actual class	A	707	71	136	50	A=SG
	B	53	588	55	92	B= PACG
	C	150	72	520	75	C= POAG
	D	57	97	89	740	D=CG

Table 4. 22 detailed accuracy by the class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.733	0.1	0.731	0.733	0.863	Congenital Glaucoma
	0.769	0.088	0.741	0.754	0.895	Primary angle closure Glaucoma
	0.679	0.133	0.618	0.633	0.828	Primary open angle glaucoma
	0.715	0.077	0.754	0.734	0.879	Secondary Glaucoma
Weighted Average	0.71	0.09	0.71	0.71	0.866	

4.4.3.3 Experiment III

C) J48 with whole attribute via percentage split 80/20

Another experiment was conducted using the whole dataset on the test option of percentage splitting of the dataset for training and test option means that 80% of the dataset were for training as well as the remaining 20% of the dataset were used for testing purpose. This experiment an accuracy result of 68% correctly classify while the remaining percent of the dataset 32 % incorrectly classify with taken of 0.14 seconds.

Table 4. 23 confusion matrix J48 with whole attribute via percentage split 80/20

	Predicated class					
Actual class		A	B	C	D	Classified as class of
	A	121	30	29	15	A=SG
	B	33	161	13	10	B= PACG
	C	38	28	135	21	C= POAG
	D	17	14	18	150	D=CG

Table 4. 24 detailed accuracy by the class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.621	0.138	0.579	0.599	0.799	Congenital Glaucoma
	0.742	0.117	0.691	0.716	0.891	Primary angle closure Glaucoma
	0.608	0.098	0.692	0.647	0.503	Primary open angle glaucoma
	0.754	0.073	0.765	0.759	0.882	Secondary Glaucoma
Weighted Average	0.681	0.106	0.683	0.681	0.844	

4.4.3.4 Experiment IV

D) J48 with selected attribute via percentage split 80/20

The last experiment on J48 was conducted based on the whole dataset on the test option of percentage splitting of the dataset for training and test option. I.e. 80% of the dataset was for training as well as the remaining 20% of the dataset used for testing purpose. This experiment results in 70% correctly classifying while the remaining 30 % were incorrectly classified with the time taken of 0.04 seconds.

Table 4. 25 confusion matrix J48 with selected attribute via percentage split 80/20

	Predicated class					
Actual class		A	B	C	D	Classified as class of
	A	135	17	35	12	A=SG
	B	10	115	14	20	B= PACG
	C	28	16	93	10	C= POAG
	D	10	17	18	160	D=CG

Table 4. 26 detailed accuracy by the class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.642	0.108	0.670	0.655	0.831	Congenital Glaucoma
	0.707	0.088	0.725	0.716	0.876	Primary angle closure Glaucoma
	0.677	0.169	0.550	0.607	0.818	Primary open angle glaucoma
	0.679	0.065	0.789	0.730	0.838	Secondary Glaucoma
Weighted Average	0.708	0.095	0.712	0.71	0.841	

Table 4. 27 summary of detailed evaluation of J48 Algorithm

Algorithm	Attribute for experiment	Test option	Time taken	TP rate	FP rate	Precision	F-measure	Roc area	Accuracy
J48	whole	10 fold	0.03	0.65	0.1	0.68	0.68	0.83	68.5%
	selected		0.04	0.71	0.09	0.71	0.71	0.86	69%
	whole	Percentage Split 80/20	0.14	0.68	0.1	0.68	0.68	0.84	68%
	Selected		0.04	0.78	0.095	0.712	0.71	0.84	70%

Note:-Among those all experiments conducted using the J48 algorithm the best accuracy result has been scored using selected attribute with 10 fold cross-validation for the accuracy result of 69% correct classification.

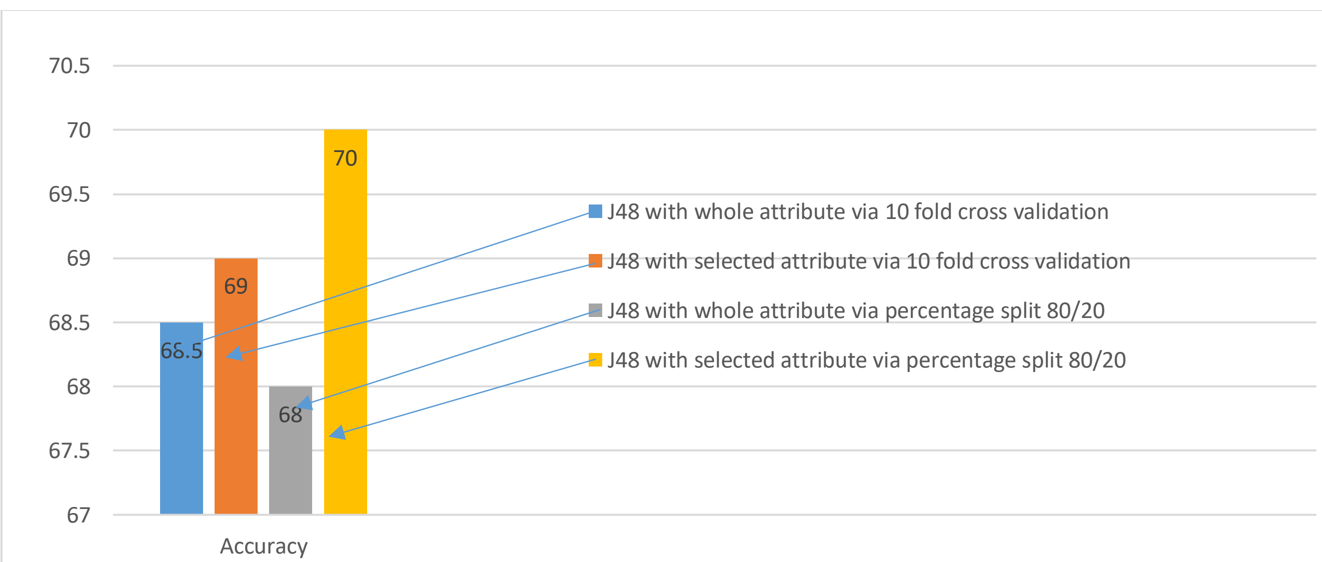

Figure 4. 3 graphical representation of J48 algorithm accuracy result

4.4.4 Scenario IV

4.4.4.1 Experiment I

A) PART with whole attribute via 10 fold cross validation

PART is one of the rule-based algorithms and used to experiment with this study. But the experiment is conducted on a test option of 10 fold cross-validation using the whole attribute for model development. Finally, this experiment scored an accuracy 68.3 % with the time taken of 0.27 seconds, and incorrectly classifying instances was 31.7%. The confusion metrics are shown as below.

Table 4. 28 confusion matrices of PART with whole attribute via 10 fold cross-validation

	Predicated class					
Actual class		A	B	C	D	Classified as class of
	A	643	161	163	98	A=SG
	B	122	736	94	73	B= PACG
	C	137	113	696	92	C= POAG
	D	96	71	100	771	D=CG

Additionally of the confusion matrices the detailed accuracy of each class in terms of true positive rate, false positive rate and other metrics is resulted as shown in the table below.

Table 4. 29 detailed accuracy by each class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.604	0.114	0.644	0.623	0.815	Congenital Glaucoma
	0.718	0.110	0.681	0.699	0.858	Primary angle closure Glaucoma
	0.671	0.114	0.661	0.666	0.3838	Primary open angle glaucoma
	0.743	0.084	0.746	0.744	0.870	Secondary Glaucoma
Weighted Average	0.683	0.106	0.683	0.683	0.845	

4.4.4.2 Experiment II

B) PART with selected attribute via 10 fold cross validation

This experiment selected attributes were used to develop a model with the test option of 10 fold cross-validation. To conduct this experiment, 0.04 seconds is required. From a total of 3535 instances, **74.7**% were correctly classified while the remaining 25.3% of the total instance were incorrectly classified. The confusion matrix for this experiment is shown as follows.

Table 4. 30 confusion matric of PART with selected attribute via 10 fold cross validation

	Predicated class					
Actual class		A	B	C	D	Classified as class of
	A	695	80	135	54	A=SG
	B	66	577	61	84	B= PACG
	C	143	68	493	113	C= POAG
	D	59	86	101	737	D=CG

Table 4. 31 detailed accuracy by each class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.712	0.104	0.722	0.721	0.865	Congenital Glaucoma
	0.762	0.083	0.749	0.755	0.895	Primary angle closure Glaucoma
	0.655	0.117	0.765	0.653	0.845	Primary open angle glaucoma
	0.749	0.081	0.754	0.751	0.886	Secondary Glaucoma
Weighted Average	0.704	0.099	0.704	0.704	0.873	

4.4.4.3 Experiment III

C) PART with whole attribute via percentage split 80/20

Another experiment was directed based on the whole attribute on the test option of a percentage split of the data into 80% and 20% for training and test option respectively. Finally, this experiment results in 66% correctly classifying while the remaining 34 % were incorrectly classified with the time taken of 0.25 second.

Table 4. 32 confusion matrices of PART with whole attribute via percentage split 80/20

	Predicated class					
Actual class		A	B	C	D	Classified as class of
	A	119	34	30	12	A=SG
	B	28	158	21	10	B= PACG
	C	52	23	125	22	C= POAG
	D	23	15	13	148	D=CG

The detailed accuracy of each class in terms of including true positive rate, false-positive rate, and other metrics results as shown in the table below.

Table 4. 33 detailed accuracy by each class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.610	0.161	0.536	0.571	0.785	Congenital Glaucoma
	0.728	0.117	0.687	0.707	0.842	Primary angle closure Glaucoma
	0.563	0.105	0.661	0.608	0.788	Primary open angle glaucoma
	0.744	0.069	0.771	0.757	0.872	Secondary Glaucoma
Weighted Average	0.660	0.113	0.665	0.661	0.821	

4.4.4.4 Experiment IV

D) PART with selected attribute via percentage split 80/20

This experiment was conducted using a whole attribute on the test option of percentage split through splitting of the dataset for training and testing 80% and 20% respectively. This experiment result in 69.4 % correctly classify while 30.6 % of the dataset were incorrectly classified with the time of taking of 0.17 seconds.

Table 4. 34 confusion matrices PART with selected attribute via percentage split 80/20

	Predicated class					
Actual class		A	B	C	D	Classified as class of
	A	136	14	37	12	A=SG
	B	16	1147	16	10	B= PACG
	C	32	15	84	16	C= POAG
	D	11	22	16	156	D=CG

Table 4. 35 detailed accuracy by each class

	TP rate	**FP rate**	**precision**	**f-measure**	**ROC area**	**Class label**
	0.683	0.115	0.697	0.69	0.840	Congenital Glaucoma
	0.717	0.083	0.739	0.728	0.879	Primary angle closure Glaucoma
	0.656	0.157	0.561	0.605	0.838	Primary open angle glaucoma
	0.692	0.092	0.732	0.712	0.683	Secondary Glaucoma
Weighted Average	0.964	0.1	0.697	0.695	0.856	

Table 4. 36 experimental summary via PART algorithm

Algorithm	Attribute	Test option	Time taken	TP rate	FP rate	Precision	F-measure	Roc area	Accuracy
PART	whole	10 fold	0.27	0.68	0.1	0.68	0.68	0.8	68.3%
	selected		0.04	0.704	0.099	0.704	0.704	0.87	74.7%
	whole	Split 80/20	0.25	0.66	0.11	0.66	0.66	0.82	66%
	Selected		0.17	0.694	0.1	0.697	0.695	0.85	69.4%

Note:-Among those all experiments conducted using the PART algorithm, the highest accuracy result has been scored using the selected attribute with 10 fold cross-validation for the accuracy result of 74.7% correct classification with the time taken of 0.04 seconds.

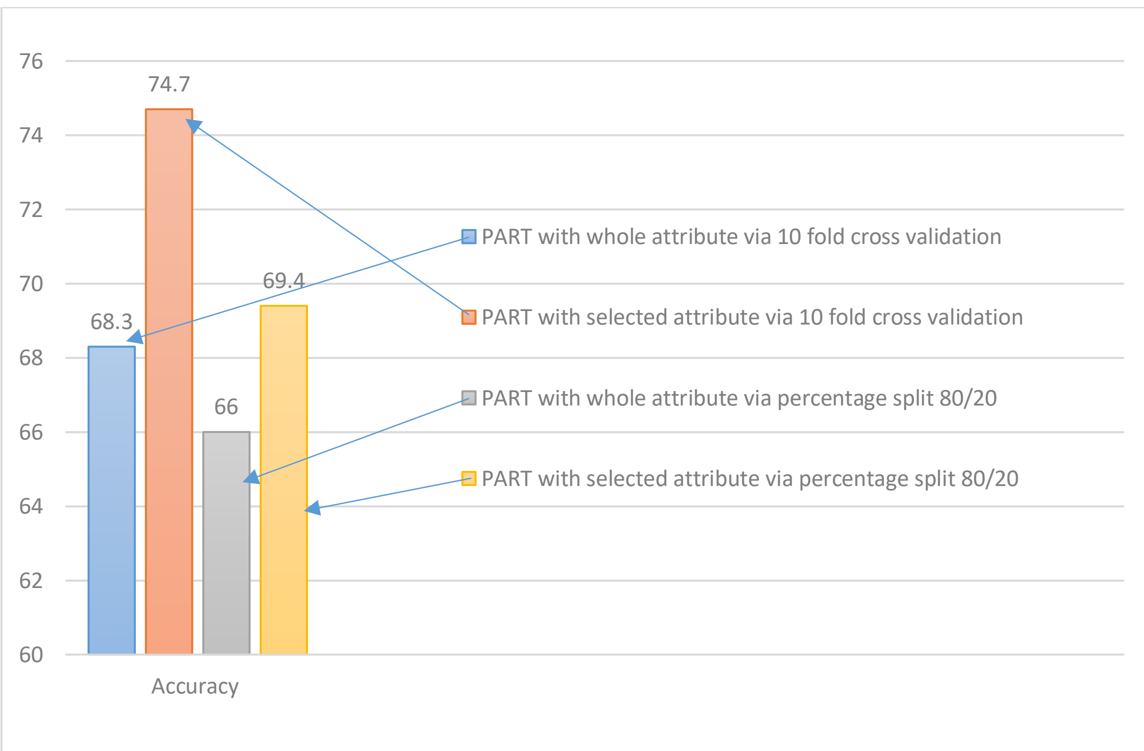

Figure 4. 4 graphical representation of PART algorithm accuracy result

4.5 Comparison between experimental results

Totally ***sixteen experiments*** were conducted. Each experiment has resulted in its accurate classification result (accuracy result). The comparison between sixteen experimental based on selected evaluation matrices is shown in the following table.

Table 4. 37 comparison between models (summary of sixteen experiments)

Test mode with Attribute	**Classifier algorithms**				
	Performance metrics	**Jrip**	**Naive aye**	**J48**	**PART**
10 fold cross-validation/ whole attribute	*Accuracy*	63.5%	50.3%	68.5%	68.3%
	Precision	65%	50%	68%	68%
	TPR	65%	50%	65%	68%
	FPR	12%	16%	10%	10%
	F-score	63%	49%	68%	68%
	ROC area	79%	74%	83%	80%
Percentage Split 80/20/ whole attribute	*Accuracy*	62.6%	47.6%	68%	66%
	Precision	64%	49%	68%	66%
	TPR	52%	47%	68%	66%
	FPR	12%	17%	10%	11%
	F-score	62%	46%	68%	66%
	ROC area	83%	77%	84%	82%
10 fold cross validation/ selected attribute	*Accuracy*	67.5%	55.6%	69%	**74.7%**
	Precision	68.5%	55.8%	71%	70.4%
	TPR	67.5%	55.7%	71%	70.4%
	FPR	11.2%	15.6%	9%	9.9%
	F-score	66.75%	54.4%	71%	72%
	ROC area	84%	71%	76%	79%
Percentage Split 80/20/ selected attribute	*Accuracy*	65.9%	56.4%	70%	69.4%
	Precision	76%	56.7%	71.2%	69.7%
	TPR	65.9%	56.5%	70.8%	69.4%
	FPR	11.5%	15.6%	9.5%	10%
	F-score	65.7%	55.5%	71.2%	69.7%
	ROC area	83%	72%	75.7%	81%

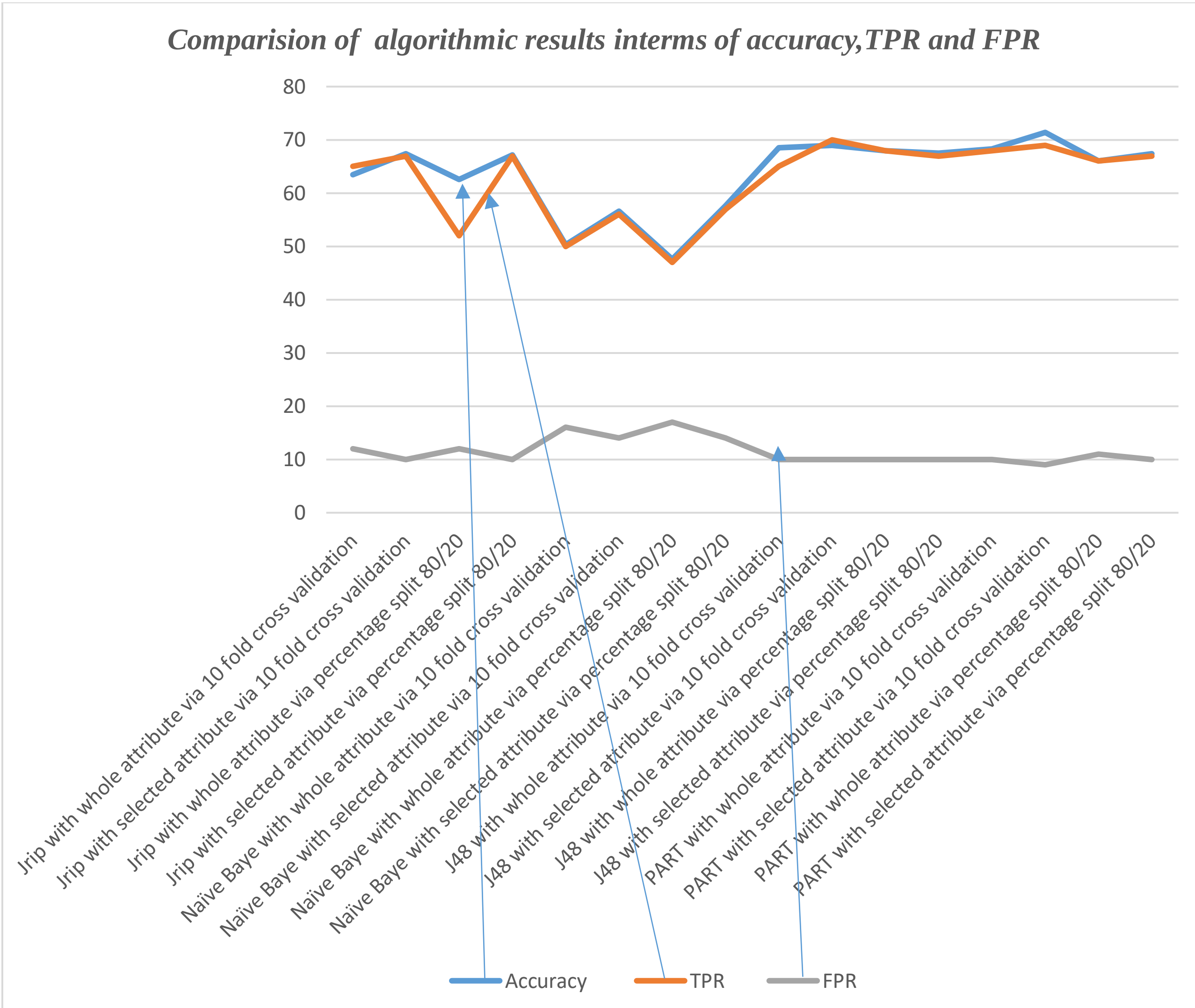

Figure 4. 5 comparison between accuracy, TPR and FPR in all experiments

The above diagram tells us the PART algorithm has resulted in the highest accuracy for classifying the glaucoma disease into its typical glaucoma ailment identifying with the test option of 10 fold cross-validation and using selected attributes.

As discussed early, four basic classifier algorithms were used to develop a model as comparatively. Among those algorithms, the PART classifier results the highest accuracy as compared between them. Figure 4.6 shows the comparison accuracy of each classifier algorithm

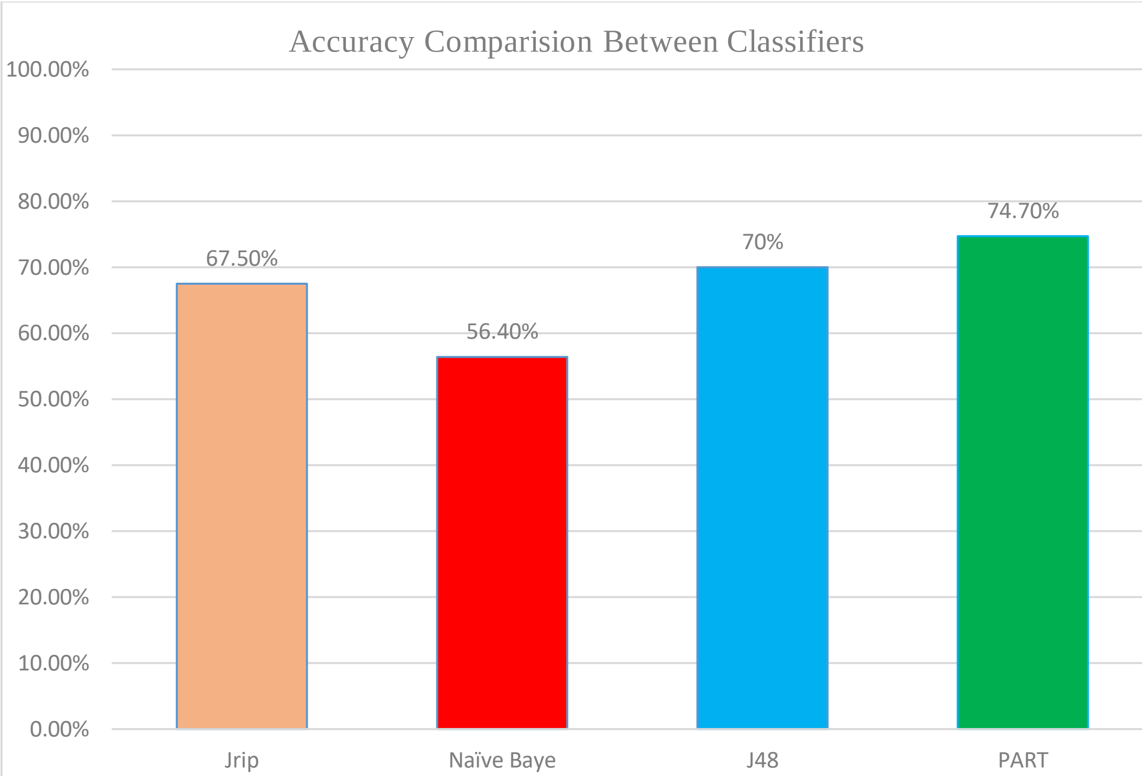

Figure 4. 6 Accuracy comparison between classifier algorithms

The comparison of this study from the previous studies (related works) is shown below even if the objectives and target points were different.

Table 4. 38 Comparison of the experiments with previous studies

Author	Title	Best algorithm	Accuracy result	Objective
Maıla C and scvet.al. (2016)	Automatic glaucoma detection based on optic disc segmentation and texture feature extraction	Multiline perceptron	93.03 %	developing an automatic detection method of glaucoma in retinal image whether or not the eye become glaucomatous
Sadaf Malik at al. (May 2019)	Data-Driven Approach for Eye Disease Classification	Random forest.	86.36%	To develop a classification model for classifying types of eye disease
Kinjan Chauhan and et.al (2016)	Data Mining Techniques for Diagnostic Support of Glaucoma	Linear Regression	97.56%	Identification of glaucoma from other disease in the case of retinal image processing.
Tehmina Khalil and et. al (2017)	Applying machine learning techniques for glaucoma detection and prediction	k-nearest neighbor (KNN)	86%	The main objective of this study was detecting glaucoma by itself either of it occurred or not
Proposed work	Towards Collaborating Data Mining Results with Knowledge-Based System for type Identification and Treatment ordering of Glaucoma Ailment:	PART	74.7%	The main objective of this study this proposed work is identifying types of glaucoma ailment and order appropriate treatment for each identifying glaucoma ailment.

Note:-finally**,** the accuracy result of this model is **74.7%** and is less accurate than the previous related studies. But its **objective**, **features,** and **dataset** are different from the related works. The objective of this study is to identify the type of glaucoma ailment and order treatment accordingly while almost all the whole related studies were focused on detecting glaucoma and classifying types of eye disease.

4.6 Model and Rule Selection

To select a model from the total experiment, the main criteria is an accuracy that an algorithm scored to develop a model and additionally consideration of time taken that an algorithm required to develop a model. Totally sixteen experiments were conducted and each experiment has been scored their accuracy with the different time taken to develop a model. Hence, from those models that all experiments are done, one model should be selected for final prototype development which has the highest accuracy score than the other experiments. To select rules, there must be a prior task to select the model. Consequently, the rules were drafted from these selected models. In this research, the selected model from sixteen experiments was on the rule-based procedure of the PART algorithm with a selected attribute on the test option of 10 fold cross-validation with the highest accuracy of 74.7%. Rules were taken from it because this algorithm scores the highest accuracy result and preeminent for the development of the final prototype to the final users (ophthalmologists) through collaborate with external sources of knowledge.

Classifier algorithm=PART

Test option =10 fold cross validation.

Attribute usage = selected attribute based on threshold boundary assignation.

Accuracy result = 74.7% (since it scored the highest accuracy result than the other classifier algorithms).

CHAPTER FIVE

KNOWLEDGE MODELING AND REPRESENTATION

5.1 Knowledge Acquisition

Knowledge acquisition (KA) is the main body for conducting the research based on extracting knowledge and typically knowledge-based and it is the process of eliciting, structuring, and organizing knowledge from human experts, books, documents, and sensors. To conduct this research, the researcher acquires the knowledge from data mining (in an automatic way) in the form of rule generation, interview of domain experts, and document analysis for each glaucoma ailment. The researchers acquired that knowledge as considering the bottlenecks of this investigation and transferring this knowledge to pure rule form by using logics and way of knowledge representations from the analyzing area even if the knowledge was rule form since recording computer files and transferring to the knowledge base using knowledge representation techniques used in Knowledge-based system [67].

Knowledge elicitation is the interaction between the expert and the knowledge engineer to elicit expert knowledge in some systematic way. The knowledge thus obtained is usually stored in some form of the human-friendly intermediate representation. The intermediate representation of the knowledge is then compiled into an executable form that the inference engine can process.

5.2 Knowledge Representation

As discussed in chapter three, the knowledge was acquired from the domain expert and document analysis. Means that, use an interview with a domain expert about the nature of glaucoma, types of glaucoma, solution as a treatment mechanism of each type of glaucoma, and analyzing the document for the newly revised treatment of each glaucoma. The knowledge acquired from domain experts and documents with their treatment mechanism was represented. There are several commonly used techniques for knowledge representation in the development of knowledge-based systems. These are logic, production rules, semantic nets, frames, and rule-based. In this study, rule-based knowledge representation and reasoning could be used.

5.2.1 Knowledge Represented From DE and DA

As discussed early, the domain expert and the document analysis has a key role for developing a knowledge-based system about the glaucoma ailment and treatment mechanism for each type of glaucoma based on which clinical sign was manifested. The researcher collects the knowledge from domain expert through interview and from document analysis then try to model the acquired knowledge and finally represented the knowledge in the form of rule. Similarly, the researcher read repeatedly and analyzed the knowledge from the document and again the researcher represented the knowledge in the form of rule. Finally, the knowledge collected from a domain expert and the knowledge collected from document analysis has similar characteristics and the knowledge was represented together physically. The common collaborative representation of that knowledge in the form of rule as follows.

IF (condition)…. THEN (result)

IF (condition) THEN (result), where (condition) represents premises and (result) represents an identified class of glaucoma ailment type.

Sample Knowledge representation from **domain experts** (rules numbers 2, 4 and 7) and **document analysis** (rule numbers 1, 3, 5 and 6) respectively in the form of rule with treatment ordering as follows.

1. If the patient has a clinical symptom of vomiting, eye domineers, and eye hoods in a child stage, THEN the glaucoma ailment identified as SG, corneal surgery is needed.
2. If the patient has a symptom of blurred vision, light sensitivity, and eye hoodness THEN glaucoma ailment identified as SG, ciliary body surgery need.
3. If the patient comes from the highland area within the age of young stage on the clinical sign of eye hoodness, tear percolate and cornea-cloudy THEN the glaucoma ailment identified as CG, full eye surgery need.
4. If the patient has the sign of headache and vomiting THEN the glaucoma ailment identified as POAG, Carbonic anhydrase inhibitors and canaloplasty need.
5. If the patient has a nausea sign, sudden eye pain and tear percolate THEN the glaucoma ailment identified as POAG, need injecting Carbonic anhydrase inhibitors.
6. If the patient has cornea-cloudy and sudden eye pain THEN the glaucoma ailment identified as PACG, the patient should need Filtering surgeries.

7. If the patient eye with eye hoodness and cornea-cloudy THEN the glaucoma ailment identified as PACG must be 10 ml dose prostaglandin inhibitor injection need.

5.2.2 Knowledge Represented From Data Mining Result (New Discovered Knowledge)

As discussed earlier, knowledge is represented in two basic areas. Namely, from domain experts and ordering documents manually. From the domain expert and document, the knowledge was represented in the form of rule. The other new knowledge and the contribution of this new study are discovering new knowledge from the collected dataset. This knowledge is represented in the form of rule. To talk out the knowledge of the mining tool, there must be considering the interference advisable thing of the domain expert. That is, from the dataset recorded symptoms, to decide types of glaucoma of the patient, the patient must reflect at least seven medical symptoms. According to this advisable thing, the generated selected rule must contain at least seven parameters. Many rules were generated based on the collected dataset experiment. But, based on expert advisability, seven basic medical symptoms must be faced to identify the glaucoma type. Due to this, more than sixteen rules were generated. Those rules are represented as follows. Sample Rules from the **experiment** (hidden knowledge discovery)

1) IF Tear percolate = yes AND Eye hoodness = yes AND age = Young AND Living Area = lowland AND Vomiting = no AND Eye domineer = no AND Blurred vision = no THEN primary open angle glaucoma
2) IF Tear percolate = yes AND Eye hoodness = yes AND age = Young AND Itchy = no AND Eye domineer = no AND Headache = yes AND Living Area = highland THEN primary open angle glaucoma
3) IF Tear percolate = yes AND Eye hoodness = yes AND age = Young AND Eye domineer = no AND Itchy = no AND Headache = no AND Nausea = no AND Light sensitivity = no AND Living Area = temperate THEN congenital glaucoma
4) IF Headache = yes AND Nausea = no AND Sudden sight loss = yes AND Itchy = no AND Vomiting = no AND Blurred vision = no AND sudden eye pain = yes AND Light sensitivity = yes THEN secondary glaucoma
5) IF Tear percolate = yes AND Eye hoodness = yes AND Headache = no AND frequent eye trauma = yes AND Sudden sight loss = yes AND Nausea = yes AND Vomiting = yes THEN congenital glaucoma

6) IF Tear percolate = yes AND Eye hoodness = yes AND age = Young AND Itchy = no AND Eye domineer = no AND Headache = no AND Nausea = no AND Blurred vision = yes AND Vomiting = no AND Light sensitivity = yes THEN primary open angle glaucoma
7) IF Tear percolate = yes AND Eye hoodness = yes AND age = child AND Eye domineer = no AND Living Area = temperate AND Blurred vision = yes AND Headache = no THEN congenital glaucoma
8) IF Tear percolate = yes AND age = Yeard AND sudden eye pain = yes AND Eye hoodness = no AND Blurred vision = yes AND Headache = yes AND Light sensitivity = no AND Sudden sight loss = yes AND Vomiting = no THEN secondary glaucoma
9) IF Tear percolate = yes AND Eye hoodness = yes AND age = Young AND Eye domineer = yes AND Nausea = no AND sudden eye pain = no AND Light sensitivity = yes AND Blurred vision = no THEN primary open angle glaucoma
10) IF Tear percolate = yes AND Eye hoodness = yes AND age = Yeard AND sudden eye pain = yes AND Headache = yes AND Light sensitivity = no AND Sudden sight loss = yes AND Itchy = no THEN primary open angle glaucoma
11) IF Tear percolate = yes AND Eye hoodness = yes AND Eye domineer = yes AND sudden eye pain = no AND Living Area = lowland AND Nausea = no AND Light sensitivity = no AND age = child THEN congenital glaucoma
12) IF Tear percolate = yes AND Eye hoodness = yes AND Eye domineer = yes AND Blurred vision = yes AND Nausea = no AND Frequent eye trauma = no AND Light sensitivity = yes THEN congenital glaucoma

5.3 Knowledge Collaboration

As discussed earlier, knowledge has been acquired from two basic sides which are manually from experts and documents as well as automatically from data mining as hidden and new discovering knowledge. The new knowledge has one main target which is identifying glaucoma type and order an appropriate treatment for each identified type of glaucoma eye disease. Namely, primary open-angle glaucoma, primary angle-closure glaucoma, secondary glaucoma, and congenital glaucoma. When a new case comes to the model, it responds to a new case based on the knowledge from the manual (expert and document analysis) and automatically from (mining) respectively. To do this, the knowledge could be combined and respond to what the new case needs to identify and treat appropriately. In a simple explanation, the knowledge which represents both in manual from expert and/or document analysis and automatically from the result of machine learning were combined to respond when a new glaucomatous patient case come here.

5.4 Conceptual Knowledge Modeling

As the knowledge is extracted from domain experts and document analysis the knowledge was modeled. In this study, the conceptual knowledge model helps the ophthalmologists for ordering the appropriate treatment based on the glaucoma identification task to differentiate the glaucoma types and criteria's that must be considered during the treatment ordering process. For developing the knowledge model, some features have the highest value according to the knowledge acquired from the domain experts and manual representations. The knowledge engineer then has to develop the conceptual model based on the knowledge obtained during the knowledge acquisition stage. The developed conceptual model interconnects the domain expert and the knowledge engineer who will transform the model into effective computer programs. Due to the prior knowledge acquisition from experts, attributes like age, living area, and others have been highly affected for identifying and ordering the treatment. In this modeling of knowledge from the experts, the parameter living area has a high value next to age. In this study, the conceptual model presentation helps an ophthalmologist to order the appropriate treatment based on the diagnosis was made to differentiate the types of ailment and criteria's that must be considered during the treatment ordering procedure.

The basic criteria to order treatment for a glaucomatous patient is an age and living area for the reason that the treatment ordering could be different within an age gap and their living area condition rather the treatment results in another side effect to the glaucomatous patient. So, the general benefit of this conceptual modeling is ordering an appropriate treatment for each type of classified glaucoma type ailment.

The general and compacted diagrammatic flow of conceptual modeling are shown in the figure below.

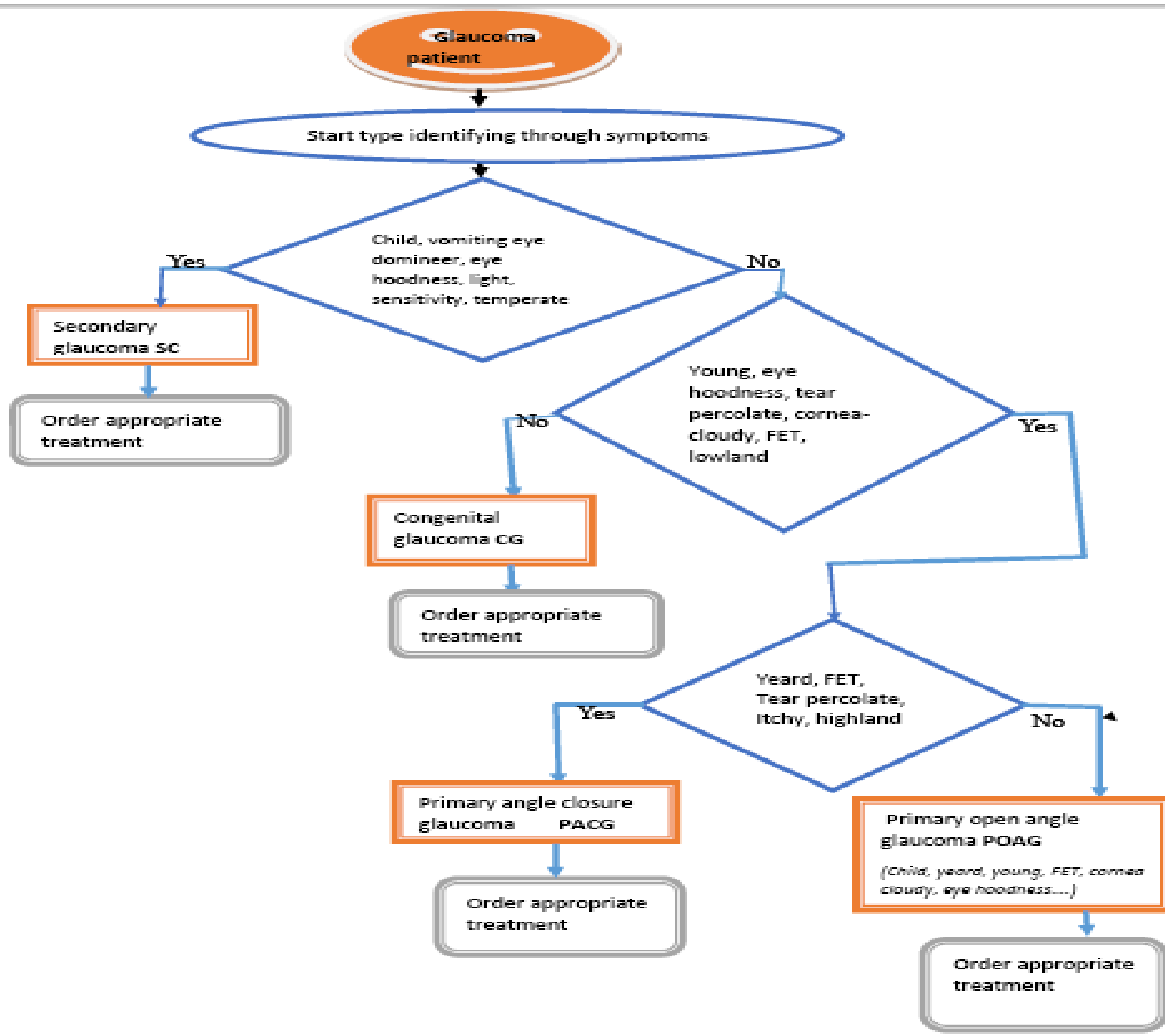

Figure 5. 1 General conceptual modeling of Glaucoma type ailment identification and treatment

Note: - The reason for saying "ordering an appropriate treatment" in a generic way is that the treatment ordering (ordering drug) is depend on each of the condition or each symptoms occur.

CHAPTER SIX

PROTOTYPING AND DISCUSSION

The prototype is the mechanism of interacting the system to the system used for the sake of providing service. Different platform software's are used like computerized platform and others. Nowadays computer software is commonly used in almost every place of human activity as well as in many specialized areas. The usability of the software very much depends on the aesthetics and quality of the user interface which allows the user of the model to provide input usually by using input devices such as a keyboard, mouse, or by touch screen and others. The developed prototype must consider the ability of the users because users' experience, knowledge, and skills to learn how to use the system are different [68]. UI prototype is a concrete foundation to make the discussion between the designers and users of the developed model that gives opportunities to learn more about product features and design and used which enables them to thoroughly validate user formal requirements and informal expectations regarding the development system.

6.1 Tools for Prototyping

Different tools were used to develop an easy and friendly user interface for interacting with the user to the system. Many tools are purposely used for user interface development.

6.1.1 The JAVA and PROLOG tools

Java tool is a widely used tool for user interface development and in any organization since it is an easy and practicable type. It could be integrated into many other systems to facilitate and make result in a good activity. While the term prolog stands for Pro means programming and log means logic and finally thermally determine as programming in Logic and is the most predominantly used tool in developing a knowledge-based system. It is a declarative language rather than procedural, which means that rather than explaining how to reach on the decision and language concentrations on a file that consists of facts and rules that describes the relationships which hold for the given application. The model responds to queries given by the user by analyzing the rules and facts stored in the database. Prolog language is used widely in artificial intelligence applications like natural language interfaces, automated reasoning systems, and expert systems. A knowledge-based system usually consists of a database or data warehouses that is composed of facts, rules, and an inference engine.

Like any other tool, the prolog has its syntax based on those syntaxes the rule could be represented within the likely of forward chinning with some advanced techniques for developing the knowledge base on it. The major prolog syntax contains some special symbols which have their meanings. These are:-

∧ is a symbol that represents, (comma)

←: - is a symbol that represents “if”.

/\: Is a symbol for conjunction

<=>: represent equivalence

∨: a symbol that represents; (semi-colon)

¬ is a symbol that represents not…. Etc.. Amongst many of the syntaxes to represent the rules in the prolog some of them are the above and additionally use other extra functions to develop a knowledge base on the prolog and connectivity mechanism.

6.1.2 Mechanism of Tools Integration

Since the investigation is a collaborative based in between the rule-based and the knowledge-based. So, the integration should be between the development tools. Those tools are the Java tool, WEKA tool, and prolog tools which are used for knowledge representation and interface development structure. There must be an integration between the Java tool and the prolog tool as well as the Java tool and the prolog tool. To concatenate Java with prolog there is a need for JPL Java classes providing an interface between Java and Prolog. JPL uses the Java Native Interface (JNI) to connect to a Prolog engine through the Prolog Foreign Language Interface (FLI) that is more or less in the process of being standardized in various implementations of Prolog. JPL is not a pure Java implementation of Prolog, but it makes extensive use of native implementations of Prolog on supported platforms. The current version of JPL only works with SWI-Prolog. The Java tools are also connected to the WEKA tool to get the selected accuracy (high result accuracy) from the experiment. The model Generator class is used to generate a model by classifying the dataset collected from BMH with the default parameters of the classification algorithm (with the selected algorithm of Parts) from classification algorithms. The Model Generator class contains some basic operations like load Dataset, build Classifier, Evaluate Model, and save the Model.

The knowledge is best for through combining the hidden discovery and expert knowledge as implicitly and explicitly than by using the knowledge individually.

6.2 User Interface

The user interface is the prototype structure used for communication between the developed model and the final user (ophthalmologist). The user interface has been developed by using Java programing language tools integrating with prolog (knowledge representation). The front side of the user interface displays the glaucoma type identification button. When the user clicks on it, the system goes to the next interface which contains some other options. Among the options the classify/identify option has classified and order the final treatment for each glaucoma type identification.

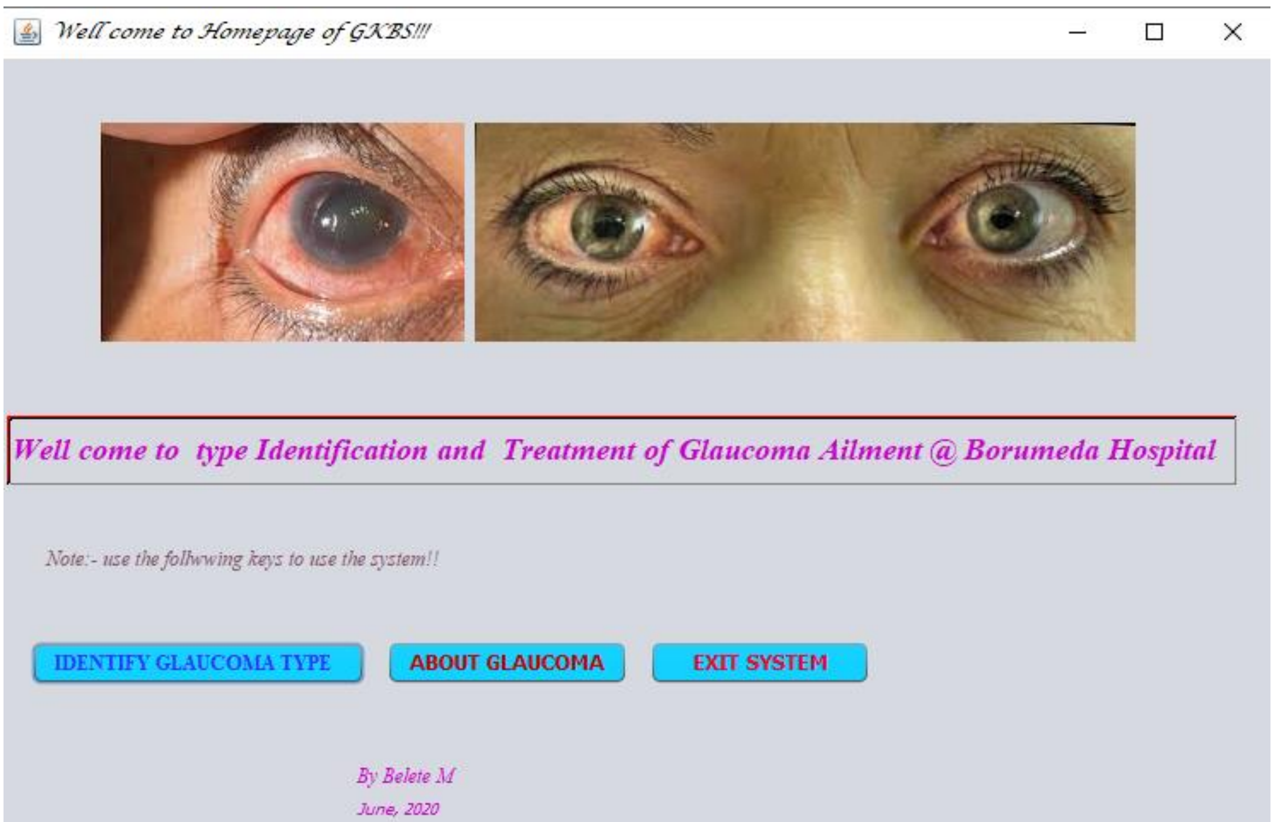

Figure 6. 1 GUI for the home page

After the well come page is displayed, the next interface shows the features required for identifying or classifying the glaucoma aliment into basic settled class labels.

Figure 6. 2 validate correct input needed

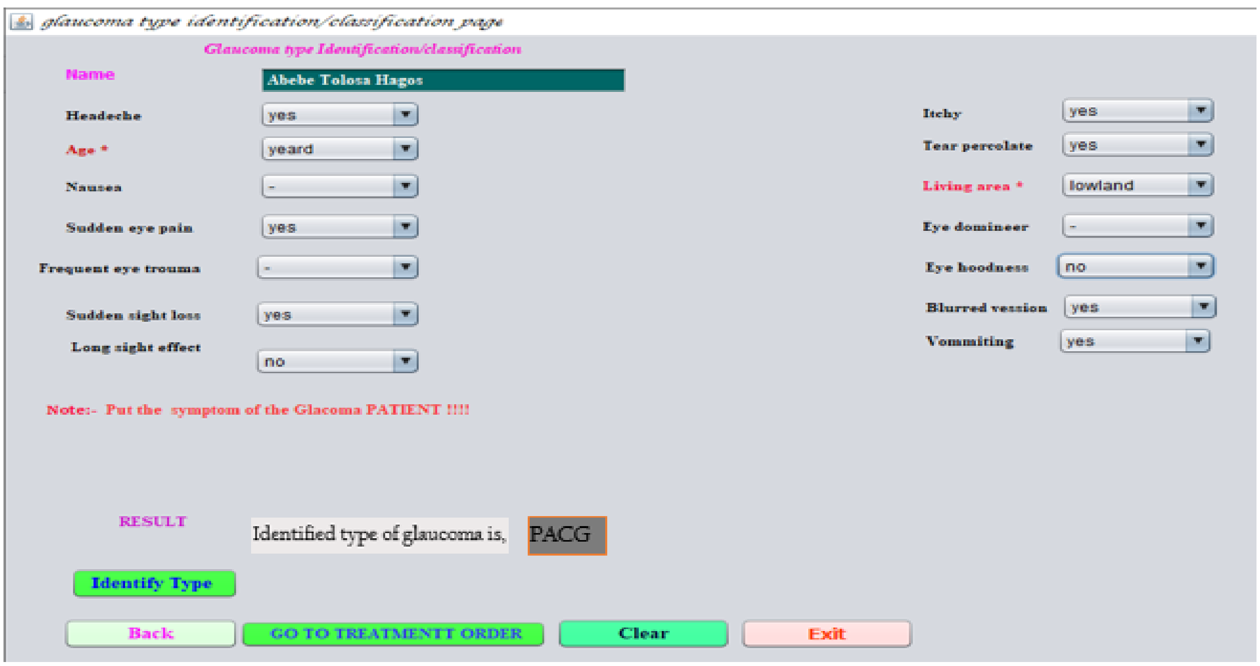

Figure 6. 3 final Glaucoma type identification display

Figure 6. 4 Final Treatment ordering of identified glaucoma type

6.3 user acceptance testing

The main objective of conducting user acceptance testing is how much the model accepted by the users and how much of the accuracy result is applicable and acceptable on the side of service prodders. To make sure how well the developed model can satisfy the organization for glaucoma type identification and treatment for each type and the model should be performed to make sure that the developed model was accepted and usable by ophthalmologist or eye cares.

User acceptance testing ensures how the users or domain experts view the developed system on the bases of the criteria. Different researchers used different types of user acceptance testing, evaluation criteria [69]. The evaluators were allowed to rate the options as excellent, very good, good, fair, and poor with the assigning result of 5,4,3,2 and 1 respectively for the final assigned questions. The final evaluated model is assessed based on the following points.

- Easiness to utilize and interface with the framework.
- The attractiveness of the framework.
- Efficiency in time.
- The exactness of the framework in arriving at a choice to distinguish the kinds of glaucoma infirmity.
- The capacity of the framework to make the right end and proposal and the Importance of the KBS in the space territory are included.

Different researchers used different types of user acceptance testing evaluation criteria. To conduct user acceptance testing, the researcher selects five domain experts to check and test the model. The reason why the researcher selects only five experts were due to the presence CORONA virus (2012 E.c), some of the experts were busy, some stayed in their home and other reasons were BMH become the center of CORONA Virus careers treatment center and the service of eye treatment were transferred to another private hospital **(ሰሊሆም የዓይን ህክምና)** which is somehow narrow and difficult to support many service providers. Due to the case of this, more service providers were not present.

Figure 6. 5 BMH become treatment center for CORONA virus at 2012 E.c

.

Table 6. 1 User acceptance testing evaluation value

User acceptance testing Evaluation(using Likert scale)							
S No	**Criteria's for evaluating the proposed model**	**poor**	**fair**	**Good**	**Very good**	**excellent**	**Average**
1	Simplicity of the model	0	0	1	2	2	4.2
2	Does the model contains the necessary symptoms to diagnosis the glaucoma type?	0	0	2	1	2	4.0
3	Efficiency of the model relative to its response time	0	0	1	3	1	4.0
4	Does the model recommend the appropriate treatment ordering for the diagnosed type of glaucoma type?	0	0	1	2	2	4.2
5	Is the user interface suitable for the user?	0	0	1	2	2	4.2
6	Would you like to use the predictive model frequently?	0	0	1	2	2	4.2
7	Is the model effective in diagnosis and treatment of glaucoma ailment type?	0	0	1	3	1	4.0
		Total Average					4.11

All of the selected respondents, no one says neither poor nor fair for each of the acceptance criteria rather they fill good, very good, or excellent. As shown in the above, the acceptance of this model via domain experts in all criteria as good is 22.8%, as very good is 42.8%, and as excellent 34.2%.

But the developed model has not an acceptable rate value on both poor and fair value. Within the average value of the developed model acceptance out of five respondents (model evaluators) 4.11 acceptance as averagely. This means the developed model has got an acceptance or it achieves 82.2%. Due to this, the researcher understand that, if the model prototype is applied, it will very good for glaucoma type identification and treatment ordering.

6.4 Discussion

As discussed above, the developed model can handle the existing environment way of identifying and treatment of glaucoma and has its objective with a link to solve the problem that was faced before. From that way, the developed model achieves an acceptance of 82.2%. Additionally, the user acceptance testing indicates using a combination of manual and automated knowledge is better than using a separate way either manual or automated knowledge techniques. When the researcher proposed and conduct the research, different research questions were settled. This section is focused on responding to those research questions that the researcher has established. Hence, the researcher outline four basic research questions, and those research questions should be addressed with their response as follows.

Research question 1, what are the most determinant factors for developing glaucoma types of disease classification models?

To develop a model, fourteen attributes were used. After conducting the final experiment some determinants Namely Headache, age, Nausea, sudden eye pain, and frequent eye trauma were highly affected the model development respectively.

Research question 2, which type of algorithm achieves the highest performance result for developing glaucoma type's identification/classification model?

To answer this research question the researcher used four types of classification algorithms namely PART, J48, naïve Baye, and Jrip algorithms under two basic test options (percentage splitting and cross-validation). Among those algorithms, the PART algorithm results in the highest accuracy and it has been the best and selected algorithm for this essence.

Research question 3, to what extent the ailment of Glaucoma type can be detected?

To get or respond to the address for this research question, the analysis of user acceptance evaluation must be necessary. From the point of this, this study has been achieved by 82.2% by the ophthalmologists or the service providers. This indicates the final developed model was good and if it implements, it will be good.

CHAPTER SEVEN

CONCLUSION AND RECOMMENDATION

This chapter dictates the final suggestion points of the research after all the necessary works done or after the investigation is conducted. The other major point in this chapter is the recommendation part. The recommendation points indicate the future work that directs what looks like and what points will include the study for the future.

7.1 Conclusion

In Ethiopia, Borumeda hospital is a huge epicenter of eye treatment and medication centers. As the aim of this study is to develop collaborative way of data mining results with a knowledge-based system to make the model too easy for the user (ophthalmologist) and order the treatment for typed glaucoma ailment. Since the research is data mining based, it requires collections of datasets from the selected area (Borumeda hospital). The researcher collects the dataset and analyzation of this dataset were done through a different tool. Finally, using the data mining technique generating a hidden rule and setting the knowledge base which is from domain expert and manual document. The researcher uses SWI-Prolog for knowledge base setting, a rapid miner for all data preprocessing, and Weka technology for hidden knowledge discoveries through applying different classification algorithms. Among classification algorithms, the researcher uses Jrip, Naive Baye, J48, and PART algorithms with two basic test options (10 fold cross-validation and percentage split of 80%/20%). Among those algorithms, the PART algorithm with a selected attribute on the test option of 10fold cross-validation results in the highest accuracy result which is 74.7%. The researcher uses the rule generated from this highest accuracy resulted algorithm (PART). The rule has been combined with a knowledge-based system for the treatment ordering of Glaucoma ailment. The final model was developed by Java NetBeans integrated with SWI-prolog. Finally, the user acceptance testing was conducted and the implementation of the developed model has been gotten an acceptance of 82.2% via user which implies the developed model is so as good if it implements in the health organization (BMH).

7.2 Final Contribution of the study

The final contribution of this study is providing a good decision support system for the ophthalmologist in order to correctly identifying of glaucoma ailment and order an appropriate treatment for each glaucoma type and reducing the resource wastage through removing mistreatment and handling the time consumption issue when ophthalmologists apply the laboratorial experiment time expenditure problem.

7.3 Recommendation

This research is conducted based on the researcher's belief and direction. furtherly, other researchers will modify and update this research as much as possible, for the sake of this, the following future recommendation points will be settled.

The researcher conducted this research based on the dataset essentially from one area (Borumeda hospital) and it follows this hospital's data shelving system. For the future, the dataset will be collected from different areas and the research will be country-based.

This research was conducted based on the technology process model of KDD with the reason of producing academic learnability for the users. If it will be discussed through using the hybrid form (KDD with CRISP) it will good and more usable as much as possible.

This research is only focused on four basic glaucoma type identification. In the future, the classes of glaucoma will be enlarged and become good studies.

Again the research was done using four classifier algorithms and select only one algorithm which scored the highest accuracy. For the future, if the combination of more classifier algorithms (ensemble) is used, it may be a high accuracy result produce.

REFERENCES

[1] Simmons ST, Cioffi G, Gross R.,"Basic and clinical science course, section 10 Glaucoma", American Academy of Ophthalmology, San Franscisco, 2017.

[2] WHO. Global data on visual impairment, Geneva: WHO; 2012. Available from: "http://www. who.int/about/licensing/copyright_form/en/index.htm, [Accessed 05/ 02/ 2020].

[3] Sembulingam K. and Sembulingam P. "Essentials of medical physiology", 6th edition, Jaypee Brothers Medical Publishers, 2012.

[4] Jampel HD, "Glaucoma patients' assessment of visual function and quality of life Transactions of the American ophthalmological society", Trans Am Ophthalmol Soc. 2001.

[5] Gutierrez P, Wilson MR, Johnson C, Gordon M, Cioffi GA, Ritch R," Influence of glaucomatous visual loss on health-related quality of life", Arch Ophthalmol. 1997.

[6] Heijl A," Delivering a diagnosis of glaucoma: considering the patient eyes", Acta Ophthalmol Scand Suppl, 2001.

[7] Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. "Global prevalence of glaucoma and projections of glaucoma burden through 2040, a systematic review and meta-analysis", Ophthalmology, 2014.

[8] Berhane Y, Worku A, Bejiga A, Adamu L, Alemayehu W, Bedri A, "Prevalence and causes of blindness and low vision in Ethiopia. Ethiop J Health Dev, 2007.

[9] Giorgis AT.," Raising public awareness of glaucoma in Ethiopia" Community Eye Health j. 2012.

[10] Cypel MC, Kasahara N, Atique D, Alcântara MP, Seixas FS," Quality of life in patients with glaucoma who live in a developing country", Int Ophthalmol. 2004.

[11] Do S, Hans L, "Report of the rapid assessment for avoidable blindness in Cambodia", National Program for eye health; 2007.

[12] Cristina F, Kafi S, Otavio C, Dave P, Jonathan C, Marcelo H, "Burden of disease in patients with glaucoma in Brazil: results from 2011 –2012 national health and wellness survey". Milan, Italy; 2015.

[13] Nwosu S. Patients, “knowledge of glaucoma and treatment options”, Niger J Clin Pract. 2010.

[14] Adegbehingbe BO, Bisiriyu L, “Knowledge, attitudes, and self-care practices associated with glaucoma among hospital workers in Ile-Ife”, Osun State, Nigeria. Tanzan J Health Res. 2008.

[15] Tenkir A, Solomon B, Deribew A. “Glaucoma awareness among people attending ophthalmic outreach services in southwestern Ethiopi”,. BMC Ophthalmol. 2010.

[16] Altangerel U, Nallamshetty HS, Uhler T, Fontanarosa J, Steinmann WC, Almodin JM, “Knowledge about glaucoma and barriers to follow-up care in a community glaucoma screening program” ,Can J Ophthalmol. 2009.

[17] Kyari F, Abdull MM, Bastawrous A, Gilbert CE, “Epidemiology of glaucoma in sub-saharan Africa: prevalence, incidence and risk factors, Middle East Africa J Ophthalmol”, 2013.

[18] Simunovic MP, Regan BC, Mollon JD, “Is color vision deficiency an advantage under scotopic conditions?, Invest Ophthalmologic Vis Scince 2010.

[19] Alemayehu W and Bedri A, “ A review on Prevalence and causes of blindness and low vision in horn of africa”, Health Develoment, 2015.

[20] Berhane Y, Worku A, Bejiga A, Adamu L, Alemayehu W, Bedri A, Haile Z, Ayalew A, Adamu Y, Gebre T and Kebede T.D, “Prevalence and causes of blindness and low vision in Ethiopia”, Ethiopian Journal of Health Development, 2007.

[21] Koberlein J, “The economic burden of visual impairment and blindness: a systematic review” ,BMJ Open, 2013.

[22] Wittenborn J.S. and D.B. Rein. “Cost-effectiveness of glaucoma interventions in Barbados and Ghan”, Optom Vis Sci, 2011.

[23] Kuper H, “Does cataract surgery alleviate poverty? Evidence from a multi-centre intervention study conducted in Kenya, Philippines and Bangladesh”, PLoS One, 2010.

[24] “Special issue on national blindness, low vision and trachoma 2005–2006” Ethiopian Journal of Health Development, 2007.

[25] Steven L.Mansberger, “Are you compliant with addressing Glaucoma adherence?” Am J Ophthalmology, 2010.

[26] Tham, Y.-C, “Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis”, Ophthalmology, 2014.

[27] Pascolini D and M. SP, “Global estimates of visual impairment”, Br J Ophthalmol, 2012.

[28] An, G., Omodaka, K.; Tsuda, S., Shiga, Y., Takada, N., Kikawa, T., Nakazawa, T., Yokota, H., Akiba, M., “Comparison of machine-learning classification models for glaucoma management.” Journal of healthcare engineering, 2018.

[29] J. Meier, R. Bock, G. Michelson, L. G. Nyl, and J. Hornegger, “Effects of preprocessing eye fundus images on appearance based glaucoma classification,” Proc. CAIP, 2007.

[30] Gupta N, Congdon N, Dada T, Lerner F, Olawoye S, Resnikoff S, Wang N, Wormald R. “Guidelines for glaucoma eye care”, International council of ophthalmology, 2015.

[31] Chauhan, B. C., House, P. H., McCormick, T. A. and LeBlanc, R. P. (1999), “Comparison of conventional and high-pass resolution perimetry in a prospective study of patients with glaucoma and healthy controls”, Archives of Ophthalmology.

[32] [Online].Available:https://en.rspca.org/RU/Knowledge representation and reasoning [Accessed date 03/ 30/ 2020].

[33] Natalie Schellac and William Stewart, Glaucoma: a brief review, USA: Department of pharmacy, Sefako Makgatho Health Sciences University, 2016.

[34] Albert S Khouri “Primary Open Angle Glaucoma,primary closure angle glaucoma, 2018.

[35] Anil K Mandal and Debasis Chakrabarti, “Update on congenital glaucoma”, India, 2018.

[36] Andre Marais and Elzbieta Osuch, “The medical management of glaucoma”, Sefako Makghato Health Sciences University,south Africa,2017.

[37] Fayyad, Usama, Gregory Piatetsky-shapiro, and Padhraic Smyth. “From Data Mining to Knowledge Discovery in Database”, 1996.

[38] M. C. Azevedo, “Data Mining Concept” ,an overview,2014.

[39] M. C. Azevedo, Data Mining Descoberta de Conhecimento em bBase de Dados, FCA, 2005.

[40] K. Charly, "Data Mining for the Enterprise", 31st Annual Hawaii Int. Conf. on System Sciences, 1998.

[41] S. Institute, “SAS Enterprise Minor SEMMA”, 2016.

[42] A. B. Michael J. and S. L. Gordon, “Data Mining Techniques for Marketing, Sales, and Customer Relationship Management”, Second ed., Indianapolis, Indiana: Wiley Publishing, 2004.

[43] Guoqing Chena, Qiang Weia, De Liub, Geert Wetsc, “Simple association rules (SAR) and the SAR-based rule discovery”, china, 2002.

[44] Mahamed G.H. Omran, Andries P Engelbrecht1 and Ayed Salman “An Overview of Clustering Methods”, 2014.

[45] S. V.Sathiya and N. Dr.Sai Satya, "Data Mining Tasks Performed By Temporal Sequential Pattern," International Journal of Research and Computational Technology, vol. 2, 2012.

[46] A. B. Michael J. and S. L. Gordon, “Data Mining Techniques For Marketing, Sales, and Customer Relationship Management”, Second ed., Indianapolis, Indiana: Wiley Publishing, 2004.

[47] Beniwal, Sunita, and Jitender Arora. "Classification and feature selection techniques in data mining", International journal of engineering research & technology, 2012.

[48] S. V.Sathiya and N. Dr.Sai Satya, "Data Mining Tasks Performed By Temporal Sequential Pattern," International Journal of Research and Computational Technology, vol. 2, 2012.

[49] R. Raman and K. Prasad, "Applications of Knowledge Based Systems in Mining", APCOM Proceedings of the Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries, Vols. 167- 180, pp. 1-11, 1987.

[50] Avram, Gabriela, "Empirical study on knowledge based systems", Electronic Journal of Information Systems Evaluation, 2005.

[51] Forkan, A., et al., Bdcam: Big data for context-aware monitoring-apersonalized knowledge discovery framework for assisted healthcare.IEEE transactions on cloud computing, 2018.

[52] Ma´ıla Claro, Leonardo Santos, Wallinson Silva Fl´avio Arau´jo, Nayara Moura, “Automatic Glaucoma Detection Based on Optic Disc Segmentation and Texture Feature Extraction”, clei electronic journal, volume 19, number 2, august 2016.

[53] Kinjan Chauhan and Dr. Ravi Gulati, “A Proposed Framework for diagnosis of Glaucoma - A Data mining Approach” International Journal of Engineering Research and Development, Volume 3, Issue 5 ,August 2015.

[54] Sadaf Malik , Nadia Kanwal and Mamoona Naveed Asghar, “Data Driven Approach for Eye Disease Classification with Machine Learning”, International Journal of Engineering Research and Development, July 2019.

[55] Archana L. Rane and P. D. Mahajan, “Classifying Chief Complaint in Eye Diseases using Data Mining Techniques” International Journal of Engineering Research and Applications, murch 2012,India.

[56] Kinjan Chauhan, Prashant Chauhan and Anand Sudhalkar, “Data Mining Techniques for Diagnostic Support of Glaucoma using Stratus OCT and Perimetric data” International Journal of Computer Applications, oct 2016.

[57] Samina Kahalid, Tehmina Khalil and Adeel M. Syed “Machine Learning Techniques for Glaucoma Detection and Prediction”, Science and Information Conference, Aug 2014.

[58] Ritu Sindhu, “Data Mining Techniques for Glaucoma Detection”, International Journal of Advanced Research in Electronics and Communication Engineering, June. 2018.

[59] [Online]. Available: https://rapidminer/products › feature-list .com. [Accessed 05/ 02/ 2020].

[60] Marwan Salman, “Data mining based pediatric eye diseases classification among children attending outpatient eye department of Tikrit teaching hospital”, university of Tikrit, Tikrit, Iraq.

[61] Falco Nogatz & Philipp K¨orner,” Prolog Coding Guidelines: Status and Tool Support”, International Journal of Advance Engineering and Research Development, Niederrhein University, Germany, 2018.

[62] Dong, Y. and X. Wang, a New Over-Sampling Approach: “Random-SMOTE for Learning from Imbalanced Data Sets.” KSEM'11Proceedings of the 5th international conference on Knowledge Science, Engineering and Management, 2011.

[63] Pouria Kaviani and Mrs. Sunita Dhotre, “Short Survey on Naive Bayes Algorithm”, International Journal of Advance Engineering and Research Development, November 2017.

[64] Eibe Frank and Ian H. Witten, “Generating Accurate Rule Sets Without Global Optimization”, International Journal of Darshan Institute on Engineering Research and Emerging Technology, December 2014, India .

[65] Korting, Thales Sehn. "C4. 5 algorithm and Multivariate Decision Trees and Image Processing Division”, National Institute for Space Research, Canada, 2017.

[66] Ramesh Prasad Aharwal, Meghna Dubey, and S.P. Saxena “ JRIP Rules for E-Governance Data, 2011.

[67] Chee-Fai TAN, "A Prototype of Knowledge-Based System for Fault Diagnosis in Automatic Wire Bonding Machine," Turkish J. Eng. Env. Sci, 2008.

[68] Abdulkerim M. “Towards Integrating Data Mining with Knowledge Based System: The Case of Network Intrusion detection”, unpublished, M.Sc. Thesis, Addis Ababa University, Ethiopia, 2013.

[69] Weichbroth, Paweł, and Marcin Sikorski. "User interface prototyping. Techniques, methods and tools." Studia Ekonomiczne 234, 2015.

[70] M. C. Azevedo, Data Mining Descoberta de Conhecimento em bBase de Dados, FCA, 2005.

APPEDIX I

User acceptance testing Evaluation							
S No	**Criteria's for evaluating the proposed model**	**poor**	**fair**	**Good**	**Very good**	**excellent**	**Average**
1	Simplicity of the model						
2	Does the model contains the necessary symptoms to diagnosis the glaucoma type?						
3	Efficiency of the model relative to its response time						
4	Does the model recommend the appropriate treatment ordering for the diagnosed type of glaucoma type?						
5	Is the user interface suitable for the user?						
6	Would you like to use the predictive model frequently?						
7	Is the model effective in diagnosis and treatment of glaucoma ailment type?						

APPENDIX II

Expert interview questions

1. What is glaucoma ailment?
2. What procedures do you follow to identify the glaucoma ailment?
3. How to treat each glaucoma in a medical and manual system?
4. How doctors can order treatment for glaucomatous patient based on laboratorial experiment result?
5. How the doctors can be differentiate those types of glaucoma ailment?
6. How to order a treatment for each type of glaucoma ailment?
7. What type of treatment is ordered for each type of glaucoma?

APPEDX II
SAMPLE CODES

Sample java codes

```
package mglaucoma;
import java.io.*;
import javax.swing.BoxLayout;
import javax.swing.JButton;
import javax.swing.JLabel;
public class homepage extends javax.swing.JFrame {
    public homepage() {
        initComponents(); }
    @SuppressWarnings("unchecked")
    // <editor-fold defaultstate="collapsed" desc="Generated Code">
    private void initComponents() {
        jButton1 = new javax.swing.JButton();
        jButton2 = new javax.swing.JButton();
        jLabel4 = new javax.swing.JLabel();
        jLabel5 = new javax.swing.JLabel();
        jButton3 = new javax.swing.JButton();
        jLabel6 = new javax.swing.JLabel();
        jLabel7 = new javax.swing.JLabel();
        setDefaultCloseOperation(javax.swing.WindowConstants.EXIT_ON_CLOSE);
        setTitle("Well come to GKBS of  Homepage");
        setBackground(new java.awt.Color(0, 51, 51));
        setBounds(new java.awt.Rectangle(1, 1, 1, 1));
        setForeground(new java.awt.Color(0, 153, 204));
        jButton1.setBackground(new java.awt.Color(0, 204, 255));
        jButton1.setFont(new java.awt.Font("Times New Roman", 1, 12)); // NOI18N
        jButton1.setForeground(new java.awt.Color(51, 51, 255));
        jButton1.setText(" IDENTIFICATION GLAUCOMA  TYPE");
        jButton1.setName(""); // NOI18N
```

```
        jButton1.addActionListener(new java.awt.event.ActionListener() {
            public void actionPerformed(java.awt.event.ActionEvent evt) {
                jButton1ActionPerformed(evt);            }        });
        jButton2.setBackground(new java.awt.Color(0, 204, 255));
        jButton2.setFont(new java.awt.Font("Tahoma", 1, 12)); // NOI18N
        jButton2.setForeground(new java.awt.Color(255, 0, 51));
        jButton2.setText("EXIT SYSTEM");
        jButton2.addActionListener(new java.awt.event.ActionListener() {
            public void actionPerformed(java.awt.event.ActionEvent evt) {
                jButton2ActionPerformed(evt);            }        });
        jLabel4.setBackground(new java.awt.Color(0, 204, 255));
        jLabel4.setFont(new java.awt.Font("Times New Roman", 2, 14)); // NOI18N
        jLabel8setText("Well come to  type Identification and  Treatment of Glaucoma Ailment @ Borumeda
Hospital");
jLabel5.setBorder(javax.swing.BorderFactory.createLineBorder(new java.awt.Color(0, 0, 0)));
        jButton3.setBackground(new java.awt.Color(0, 204, 255));
        jButton3.setFont(new java.awt.Font("Tahoma", 1, 12)); // NOI18N
        jButton3.setForeground(new java.awt.Color(255, 0, 51));
        jButton3.setText("ABOUT GLAUCOMA");…………
```

Sample prolog codes

```
/* %system: Prolog
% KBS for TYPE IDENTIFICATION AND TREATMENT GLAUCOMA AILMENT
To run, type go.*/
domains:-
   glaucomaailmenttype,indication = symbol.
   Patient,name = string.
 go:-
write('========================================================'), nl,
write('*Well Come Borumeda Hospital Center of eye Treatment*'),nl,
write('*KBS for type identification and treatment of glaucoma ailment *'), nl,
write('*Developed By Belete M. *'), nl,
write('*======================================================'),nl.
start:-
```

```
write("who is the patient's name? "),nl.
read(patient),nl,
write(patient,"Do you have the follwing symptom (y/n)?"),nl,
  write(" does the glaucoma patient " symptom?).
  predicates:-
 symptom (patient,indication),
 response (char),
 hypothesis(patient,glaucomaailmenttype).
  /*glaucomaailmenttype_and_symptom*/
 primar_open_angle_glucoma :- poag.
symptom(patient_have_a_symptom_of_vomiting):-verify(patient,"have_a_symptom_of_vomiting (y/n) ?").
symptom(is_the_patient_live_in_lowland):-verify(is_the_patient,"live_in_lowland (y/n) ?").
symptom(is_the_age_of_patient_young):-verify(is_the_age_of_patient,"child (y/n) ?").
symptom(is_the_age_of_patient_young):-verify(is_the_age_of_patient,"yeard (y/n) ?").
symptom(is_the_age_of_patient_young):-verify(is_the_age_of_patient,"young (y/n) ?").
symptom(is_the_patient_live_in_lowland):-verify(is_the_patient,"live_in_highland(y/n) ?")……………
```

Sample tools Integration code

```
* @author Belete Mamo
String prologfilelink="consult('Glaucoma.pl')";
Query = new query ();
System.out.println(query+ (query.hasSolution()?"successfully connected":"fail to connect")); .
String[] name = new String[]{"PACG ","POAG ","CG",” SG”};
String[] results = new String[]{“PACG ","POAG ","CG",” SG”};
for(int k=0 ; k <= 4 ; k++){ System.out.println(name[k]+querytext);
if(query.hasSolution(name[k]+querytext)){if(results[k] == "POAG") jLabel18.setForeground(Color.PINK);
else if(results[k] == "PACG") jLabel18.setForeground(Color.BROWN);
else jLabel18.setForeground(Color.PINK);
jLabel18.setText("Result : "+jTextField1.getText().toString()+" is "+results[k]);
k = 3;
} else if( K == 3){ jLabel18.setForeground(Color.RED);
jLabel18.setText("Result : Not enough input to identify glaucoma types");
```