SINET: Ethiop. J. Sci., **44(2): 223–233, 2021** **ISSN: 0379–2897 (PRINT)**

eISSN: 2520–7997

Date received: April 26, 2021; Date revised: September 03, 2021; Date accepted: October 23, 2021
DOI: https://dx.doi.org/10.4314/sinet.v44i2.9

An integrated knowledge-based system for early detection of eye refractive error using data mining

Teklay Birhane[1,*], Dr. Dereje Teferi[2] and Tariku Mohammed[3]

[1]Department of Information Science, Mekelle University, Mekelle, Ethiopia.
*E-mail corresponding author: teklaybirhane12@yahoo.com
[2]School of Information Science, Addis Ababa University, Addis Ababa, Ethiopia. E-mail: dereje.teferi@aau.edu.et
[3]Department of Information Science, Haramaya University, Haramaya, Ethiopia. E-mail: tarejimma@gmail.com

ABSTRACT*:* Refractive error is one of optical defect in the human visual system. Refractive error is a very common disease these days in all populations and in all age groups. Uncorrected and undetected refractive error contributes to visual impairment, blindness and places a considerable burden on a person in the world. The long use of technological devices such as smart phones also poses a new burden on the human eye. The intensity and brightness of these digital devices open a new door for high prevalence of eye refractive errors. Early medical diagnosis of the disease may help in avoiding complications and blindness. Data mining algorithms can be applied to help in ophthalmology and detection of an eye disease at an early stage. So mining the ophthalmology data in efficient manner is a critical issue. This research work deals with development of an integrated knowledge-based system that helps to detect eye refractive error early and provides appropriate advice for the patients. In this study, the hybrid knowledge discovery process model of data mining that was developed for academic research is used. About 9000 ophthalmology data from selected eye health centers are used to build the model. The sample data was preprocessed for missing values, outliers, and noise. Then the model is built using decision tree (J48 and REPTree) and rule induction (JRip and PART) algorithms. The PART algorithm has registered better predictive performance with accuracy of 60% and 96.45% for subjective and objective based model evaluation, respectively as compared to J48, REPTree, and JRip. Finally, the knowledge discovered with this algorithm is further used to build the knowledge-based systems. The Java programing language is used to integrate data mining results to knowledge-based system. The performance of the proposed system is evaluated by preparing test cases. Overall, the knowledge based system resulted in 89.2% accuracy. Finally the study concludes that discovering knowledge using data mining techniques could be used as a functional eye refractive error detection system.

Keywords/Phrases: Data Mining Techniques, Eye Refractive Error, Knowledge-Based System

INTRODUCTION

Refractive error is one of optical defect in the visual system, which causes induced blurred vision and blindness. Refractive error nowadays is one of the frequent reasons behind the visual impairment and blindness in the world (Fageeri *et al.*, 2017). Refractive error places a considerable burden and creates a negative impact on individuals and society. People living in low-income countries are majority of those affected by refractive error with minimal access to eye care (Varadarajan *et al.*, 2018). School-age children are particularly vulnerable groups for this disease (Jafer Kedir and Abonesh Girma, 2014). Where uncorrected and undetected, refractive error have an adverse effect on learning capability and educational potential, as well as high economic cost to the family and government (Nebiyat Kassa *et al.*, 2014). It causes a serious blurred vision and blindness. The poor vision and inability to read material on the chalkboard due to refractive error can greatly affect a child's participation in education system and other social interactions (Sheeladevi *et al.*, 2019). Moreover, children often think that their problem is common to all and may not speak

*Author to whom correspondence should be addressed.

about it to their family or teachers. Due to this, many children are forced to leave their school in Ethiopia (Tibebu Kassa and Getu Degu, 2000). This problem is common in Ethiopia since the country has relatively poor health coverage and eye health care system.

The lack of skilled workers in eye healthcare contributes to poor diagnosis and weak treatment of patient having eye refractive error conditions in Ethiopia, mainly due to lack of domain experts and eye health care facilities (Nebiyat Kassa *et al.*, 2014). There are novel technologies such as artificial intelligence (AI) that offer a low-cost method of screening and diagnosing eye disease in the developing world (Varadarajan *et al.*, 2018). Therefore, the use of AI is important in order to solve such problems. Artificial Intelligence (AI) has shown promising results in the diagnosis and interpretation of medical datasets to make intelligent decisions with the design and development of intelligent algorithms (Puaschunder, 2020).

As sub part of AI, data mining techniques enable to extract useful knowledge from large data sources using various analysis techniques that integrate traditional methods for analyzing data with complex algorithms (Cheng *et al.*, 2018).

Therefore, by considering the importance of early medical diagnosis of a disease, data mining techniques can be applied to help in ophthalmology and detection of eye disease at an early stage. Early identification of vision problems and eye refractive errors, including high refractive errors like hyperopia, myopia and astigmatism is critical for optimal treatment (Fageeri *et al.*, 2017).

Data mining extract accurate information from large amount of data generated by medical centers which cannot typically extract by medical experts (Codreanu *et al*, 2011). Such information can be used as key input knowledge in developing effective and efficient knowledge-based system (KBS) (Kedir Eyasu *et al.*, 2020). KBS is part of Artificial Intelligence which contains the knowledge and analytical skills of human experts in a specific problem domain (Blondet *et al.*, 2019). It is a good solution to reduce human and medical errors, functioning in a specific domain to offer wise decisions with better reasoning ability (Tripathi, 2011).

Fageeri *et al*. (2017) and Varadarajan *et al*. (2018) developed predictive model for diagnosis of eye refractive error using data mining algorithm techniques such as decision tree, support vector machine, naïve Bayes and deep learning neural network. Although, there are some developed models that perform well, they lack the provision of advice for the patients and optometrists about what course of action to be taken in accordance to the detected eye refractive error condition. The integration of eye refractive error detection with a KBS system will help ophthalmologists and optometrists to provide a better service to patients. To this end, there is a need to integrate data mining results with a reasoning system to design an intelligent and self-learning knowledge based system.

This research aimed to develop an integrated knowledge-based system that automatically acquires knowledge from ophthalmology dataset by applying data mining techniques. The researchers integrated the data mining model with knowledge base systems to detect eye refractive disease and provide advice for optometrists and ophthalmologists. The knowledge-based system has a learning capability to update the knowledge base as the size of data increases and new rules are generated.

With the aforementioned in mind, this study explores and addresses the following research questions.

1. What are the relevant attributes for the eye refractive error detection model?
2. How do we develop an integrated prototype model for early detection of eye refractive error?
3. To what extent does the integrated prototype model works in detection process?

RESEARCH METHODOLOGY

Research Design

A Hybrid Knowledge Discovery Process (HKDP) model is followed in this research. It was developed based on the more industry based CRISP-DM model by adopting it to academically inspired research (Cios *et al.*, 2019). This model provides a more general and research-oriented description of the steps.

The CRISP-DM model has only three major feedback sources, while the hybrid model has more detailed feedback sources and a modified final step for the application of the discovered knowledge. Subsequently, the hybrid knowledge discovery modeling process is chosen for this study. As mentioned, the HKDP model is a six step process namely: understanding the business problem, understanding the data, data preparation, data mining, evaluation of the discovered knowledge, and use of discovered knowledge.

Problem Understanding

Domain problem understanding is assessed and investigated in depth with domain experts for constructing eye refractive error predictive model to improve the early detection process of eye diseases.Ten-domain experts are selected from Fitsum Birhan Specialized Eye Center and St. Louis Eye Clinics in Mekelle, Ethiopia to define the business problem and determine the data-mining goal. In addition, observations and review of articles is done to understand the problem area in depth.

Based on the suggestion from domain experts, variables age, sex, both parts of right and left eye (spherical, cylindrical and axial) refractive error measure values are chosen to detect and identify refractive error condition.

Data Understanding

In this study, real life medical dataset is employed. The eye diseases dataset was collected from the two Eye Clinics. Data from 2018 and 2019 is considered for the study.

Initially, the dataset collected in paper based table format is converted to MS-Excel format by selecting the attributes chosen to be important for the problem domain. The dataset has 9000 instances and 9 attributes with 508 missing values. Table 1 briefly describes the variables that are being considered in the initial dataset.

Data Preparation and Preprocessing

Data mining uses cleaned and transformed data, searches the data by using different techniques and algorithms, and then outputs patterns and relationships to the evaluation step of the knowledge discovery from data (Li *et al*., 2011 and Mihaela, 2006).

According to Han and Kamber (2012), the data processing task of data mining includes data cleaning, data integration, data reduction, and data transformation. Before feeding the data to Data Mining, the quality of data has to be ensured. Well-accepted multidimensional data quality measures such as accuracy, completeness, consistency, timeliness, and interpretability are used.

Table 1: List of raw data variables in the initial dataset

S. No	Attribute Name	Value Type	Description
1.	uid	Nominal	Patient ID number
2.	age	Numeric	Patient age
3.	od_sph	Numeric	The force required from the lens of the right eye to correct focus values ranging between 0 to +/- 20
4.	od_cyl	Numeric	A vision defect, which collects X-hyphen-to-right eye on many points, so-called astigmatism. values ranging between 0 to +/- 25
5.	od_axis	Numeric	This is linked to CYL and tells where the place of astigmatism.Values ranging between 0 to 180 degrees.
6.	os_sph	Numeric	Left eye sphere, which indicates the power of the lens prescribed to correct the Left eye vision.
7.	os_cyl	Numeric	Left eye cylinder
8	os_axis	Numeric	Left eye axis
9	Class	Nominal	Class of diseases: Hyperopic Astigmatism, Myopia, Myopic Astigmatism, Hyperopia, Mixed Astigmatism or Normal

Attribute Selection

The importance of attributes in eye refractive error predictive modeling is checked by evaluating on all training data. Attributes selection has passed through all possible combinations of attributes in the data to find which subset of attributes works best for refractive error prediction. Accordingly, attributes selected using best first techniques in WEKA are od_sph, od_cyl, od_axis, os_sph, os_cyl, and os_axis. From the eight (8) attributes stated in table 1 six (6) of them are used in the study the remaining two attributes ***uid*** and ***age*** has no effect in this study because the detection process of eye refractive error using refraction based eye examination does not depends on patient id or patient age. Thus, before preceding to the experiments, the accuracy of the model in WEKA with the selected attributes is evaluated.

Weka data mining tool was used for knowledge extraction and data analysis. To map, the knowledge acquired from the data mining classifiers model with the knowledge-based system, Java Net Beans IDE 8.2 with JDK 8 is employed. In addition, SWi-prolog integrated with JPL is applied to develop and implement the system.

EXPERIMENTATION AND RESULTS

The next step in the hybrid knowledge discovery process model, after preparation and preprocessing of the data, is model development followed by model evaluation. A dataset with 9000 records is used for training and testing the predictive model constructed in this study.

Four data mining classification algorithms are experimented to discover knowledge from the data for predicting refractive error and to build the analytical model for diagnosis and treatment of refractive error. Classification is a data mining technique, based on supervised machine learning, which maps data into predefined groups or classes (Sharma *et al*., 2013). It is the process of building a model of classes from a set of records that contain class labels (Liu and Wu, 2012). It often describes these classes by looking at the characteristics of data already known to belong to the classes (Sharma *et al*., 2013).

The four classification algorithms used in this study are J48 pruned, REPTree, PART, and JRip. J48 and REPTree are tree-based classifiers in WEKA whereas JRip and PART are rule-based classifiers. The four algorithms are selected because they use decision tree or rule based structure. These kind of algorithms are known to perform well for numeric data types.

Hence the aim of this study is constructing of rule based KBS and because of their capable of generating powerful rules the four classifiers algorithms are selected. A rule-based classification extracts a set of rules that show relationships between attributes of the dataset and the class label (Kedir Eyasu *et al.*, 2020). The dataset is split into training and test using 10-fold Cross-validation. Thus, for each classifier, 10 experiments are conducted and the average evaluation result is reported.

Experiment 1- J48 classifier

A decision tree produces a sequence of rules that can be used to classify the given data of attributes together with its classes (Milovic, 2012 and Brázdil *et al*., 2018). It is mainly used in classification and prediction. It is a simple and powerful way of representing knowledge (Milovic, 2012).The strengths of J48 are simple to understand and visualize, requires little data preparation, and can handle both numerical and categorical data. Drawn backs of J48 are poor results on very small datasets and overfitting can easily occur.

The experiment shows that, the model developed using J48 classifier has a tree of 63 nodes with 32 leaves. The algorithm has correctly classified 8702 instances and only 298 instances are classified incorrectly taking 0.47 seconds to build the model. In Table 2 and all the following experiments, A = Mixed Astigmatism, B = Hyperopia, C = Normal, D = Myopic Astigmatism, E = Myopia, F = Hyperopic Astigmatisms

Table 2. Confusion Matrix for J48 classification algorithm.

Confusion Matrix						
	A	B	C	D	E	F
A	1396	70	22	6	1	5
B	0	1432	57	1	6	4
C	7	54	1397	27	9	6
D	1	8	0	1490	0	1
E	2	4	4	3	1487	0
F	0	0	0	0	0	1500

Experiment 2- REPTree classier

The second experiment is based on REPTree algorithm. REPTree creates multiple trees in different iterations using the regression tree logic. After that it selects the best one from all generated trees and that will be considered as the representative. Basically REPTree is a fast decision tree learner that builds a decision/regression tree using information gain/variance and prunes predictions made by the tree using the mean square error measure. This algorithm has correctly classified 8703 instances out of 9000. That is, it has incorrectly classified 297 instances taking 0.13 seconds to build the model. In addition, the tree generated has 53 nodes.

Table 3. Confusion Matrix for REPTree classification algorithm.

Confusion Matrix	A	B	C	D	E	F
A	1390	70	22	11	2	5
B	0	1489	0	2	5	4
C	10	97	1350	27	10	6
D	4	8	0	1481	6	1
E	0	4	0	3	1493	0
F	0	0	0	0	0	1500

Experiment 3-PART classier

PART is a rule-based classifier that uses a set of IF-THEN rules for classification. It is an expression of IF condition THEN conclusion. PART, rule induction algorithm generated 25 rules by involving all the attributes of the dataset. The algorithm registered prediction accuracy of 96.47% in which 8682 instances out of 9000 are correctly classified. The algorithm has incorrectly classified only 318 instances and took 0.64 seconds to build.

Table 4. Confusion Matrix for PART classification algorithm.

Confusion Matrix	A	B	C	D	E	F
A	1410	57	23	4	1	5
B	30	1409	50	4	3	4
C	7	68	1390	27	2	6
D	1	7	1	1490	0	1
E	2	3	9	3	1483	0
F	0	0	0	0	0	1500

Experiment 4 – JRip classifier

The other rule induction algorithm selected for this study is JRip. JRip is a rule-based classifier algorithm that uses a set of IF-THEN rules for classification process. JRip correctly classified 8659 instances from 9000. The numbers of incorrectly classified instances are 341. The algorithm has generated 15 rules. See Table 5.

Model evaluation and selection

All the selected classifier algorithms generated rules from the dataset. The performance of algorithms is compared, and the one, which performed better, is selected as prime choice for the knowledge acquisition step. Both objective and subjective based model evaluation methods are used to select the best model for development of the KBS. Objective based model evaluation measurement is generally based upon the statistical and inherent structure of the mined patterns.

The accuracy, sensitivity (TPR), precision, recall, and F-measure of all the four classifiers, obtained from the experiment, are shown in Table 6.

Table 5. Confusion Matrix for JRip classification algorithm.

Confusion Matrix	A	B	C	D	E	F
A	1384	72	25	12	2	5
B	16	1442	30	5	1	6
C	7	88	1363	25	9	8
D	0	6	2	1490	1	1
E	3	5	5	4	1480	3
F	0	0	0	0	0	1500

Table 6. Performance of Classifiers.

Objective model evaluations	Classification algorithms J48 pruned	REPTree	JRip	PART
Correctly Classified	96.69%	96.7 %	96.21%	96.47%
Incorrectly Classified	3.31%	3.3 %	3.79%	3.53%
TP Rate	96.7%	96.7%	96.2%	96.5%
Precision	96.7%	96.9%	96.3%	96.5%
Recall	96.7%	96.7%	96.2%	96.5%
F-Measure	96.7%	96.7%	96.2%	96.5%
Time taken in seconds	0.47	0.13	10.64	0.67

As can be seen from Table 6, REPTree classification algorithm performs better in all objective based model evaluation methods. However, over 3.3% of the test data is still incorrectly classified. The rule acquired from the classifier algorithms will be used for constructing the knowledge base and hence these errors need to be minimized as much as possible. To this end, subjective model evaluation method is used to identify and exclude the incorrect rules from the knowledge base.

Discussion was held with domain experts about the significance and correctness of the rules generated by the four algorithm. The rules provided by the models can be easily understood by human experts without any difficulty. The rules of the selected four algorithms are distributed to ten purposively selected domain experts and their evaluation is collected as shown in Table 7.

Table 7.Subjective model evaluation results.

Classification algorithms Subjective model evaluation	J48 pruned	REP Tree	JRip	**PART**
Domain Expert Knowledge	10%	10%	20%	**60%**
No of rules	32	53	15	**25**

Rule Extraction

From the four models that are built, the model developed with PART rule induction classifier is selected for developing the knowledge base because it registered better performance (average of objective and subjective model evaluation method). Some of the interesting rules generated by PART model are presented below.

RULE 1: IF os_cyl<= -0.2 AND os_sph<= 0.454007 AND od_cyl> 0 AND os_cyl> -0.75 THEN Eye refractive error disorder = Hyperopic Astigmatism (72.0).

RULE 2: IF os_cyl<= -0.2 AND os_sph<= 0.454007 AND od_sph<= 0 AND od_cyl<= -0.25 THEN Eye refractive error disorder = Myopic Astigmatism (741.0).

RULE 3: IF os_sph<= -0.25 AND os_cyl> -0.2 AND od_cyl> -0.25 AND od_sph<= -1.75 THEN Eye refractive error disorder = Myopia (1479.0).

RULE 4: IF os_cyl<= -0.2 AND os_sph<= 0 AND od_sph<= -2 THEN Eye refractive error disorder = Myopic Astigmatism (550.0/2.0).

RULE 5: IF os_cyl<= -0.2 AND od_cyl<= -0.25 AND os_sph> 0.5 AND od_sph> 0.75 THEN Eye refractive error disorder = Mixed Astigmatism (1333.0).

RULE 6: IF os_sph> 0 AND os_cyl<= 0.25 AND od_cyl<= 0.25 AND od_sph> 0 AND os_cyl> -0.2 THEN Eye refractive error disorder = Hyperopia (1089.0/15.0).

RULE 7: IF os_sph<= 0 AND os_axis<= 0 AND od_sph<= 0 AND od_cyl> -0.25 AND os_sph> -0.75 THEN Eye refractive error disorder = Normal (1169.0).

Mapping Predictive Model to Knowledge Based Systems

Java programming language is used to integrate the data-mining model to the knowledge-based systems. An integrator module that translates a Java program from a text file to a Prolog representation is implemented.

The knowledge discovered using data mining techniques (PART classification algorithm) is used in building the knowledge-based systems that predict eye refractive error conditions. Then, the knowledge-based systems are used for early detection of uncorrected refractive error by inferring from the inference engine of the knowledge-based system. Moreover, the developed knowledge based systems have a self-learning capability to generate new rules and update its knowledge base automatically.

Architecture of the Prototype System

Architecture is a blueprint showing how the components of the integrated knowledge-based system interacts and interrelates.

Figure *1* illustrates the architecture of the proposed system.

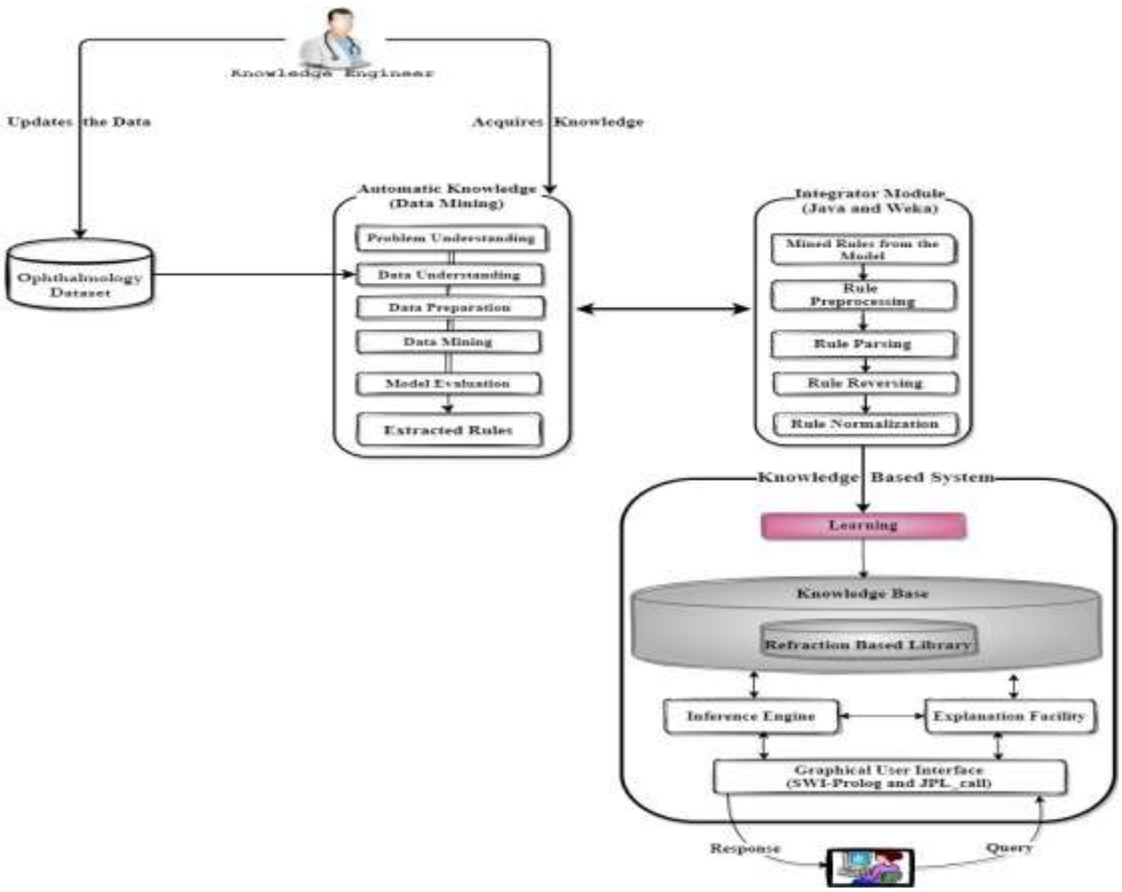

Figure 1. General Architecture of the prototype KBS.

Integrator Module

The results of the data mining are the basis for the development of the knowledge base system through the integrator application. The facts generated from the Data mining can be represented as a rule in knowledge base. The PART model is integrated with knowledge-based system automatically for designing intelligent early detection of eye refractive error as depicted in Figure 2.

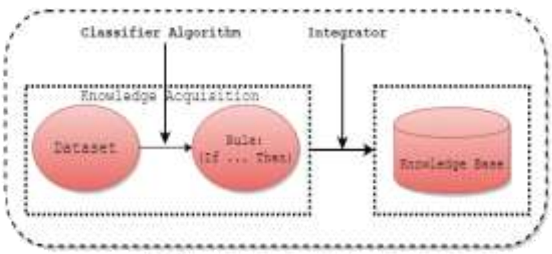

Figure 2. Conceptual Design of the Integration Process.

Overall, the following six steps are used to develop the KBS. Each step is discussed as follows:

1. *Dataset Preparation:* - The ophthalmology dataset is collected from two-eye clinic centers manual database/record named Fitsum Birhan Specialized Eye Center and St. Louis Eye Clinic.

2. *Rule mining*: - at this stage, rules are extracted from the ophthalmology dataset by applying classifier algorithm using Weka. PART classifier algorithm is selected as the best predictive model.

3. *Rule Tokenization*: - at this stage, removal of undesirable characters and replacing some tokens with other equivalent tokens is performed.

For example, comparison operator '<=' (less than or equal to) in PART rules is replaced by its PROLOG equivalent '=<'. The character ":"in PART is replaced by ":-" in equivalent character which means IF in PROLOG. Moreover, the conjunction operator 'AND' is replaced by its PROLOG equivalent ',' used for joining two conditions.

4. *Rules and facts Parsing* - this is the process of analyzing a string of symbols, either in natural language or in computer languages, according to the rules of a formal grammar. In this context, parsing is analyzing the components of the generated PART rules. A given PART rule is composed of: (condition)

implication (conclusion). If condition is evaluated true then conclusion is executed.

5. *Rule Reversing*: - the rule reverse is used to exchange the place of the Left Hand Side (LHS) and the Right Hand Side (RHS) of the rule. This is done because PROLOG uses the reversed rule format. That means PROLOG starts with conclusion and goes for the conditions that made the conclusion. PART algorithm rule is reversed from the format IF … THEN to the format THEN… IF PROLOG understandable format. For example:

Rule: *attribute*$_1$= *value*$_1$ AND *attribute*$_2$ = *value*$_2$ AND *attribute* $_n$ = *value* $_n$ THEN *(conclusion).*

Reversed rule format: *(conclusion):- attribute*$_1$= *value*$_1$, *attribute*$_2$ = *value*$_2$, *attribute* $_n$ = *value* $_n$.

Finally, period (.) is placed at the end of all reversed rules to tell PROLOG end of statement.

6. *Rule Normalization*: - normalization is done to change all tokens in reversed rule into lower case to be understood by PROLOG.

Knowledge based system

A typical knowledge based system consists of the following subsystems: knowledge representation, knowledge base, Inference engine, explanation facility and user interface.

A. Knowledge representation: - The most commonly used methods of knowledge representation are production rule, frame, and network. Knowledge captured from experts and other sources must be organized in such a way that a computer program can access this knowledge whenever needed and draw conclusions. In KBS, production rules are used and represented in the form of rules by rule-based representation technique, since it permits the relationships that make up the knowledge base to be broken down into manageable unit.

B. Knowledge Base: - is a set of rules or encoded knowledge about detection process of eye refractive error disorders. The knowledge collected from domain experts, document analysis and data mining is stored in the knowledge base as set of rules using a rule-based knowledge representation method. The rules are stored as refraction based library.

C. Inference Engine: - An inference engine is the brain of the Knowledge Based System, which directs the system to derive a conclusion by looking for possible solutions from the knowledge base and recommending the best possible one. Since the objective of the proposed Knowledge Based System is diagnosis of eye refractive error and the Prolog's built-in inference mechanism is forward chaining, a goal derived forward inference mechanism that tries to prove or disprove the goal is used.

D. User Interface: - It is the means of by which the user and a computer system interact. The acceptability of a knowledge-based system depends on the quality of the user interface. The user interface is used as the means of interaction between the user and the knowledge based system.

Figure 3. User Interface of the Developed KBS

E. Explanation facility: The KBS system can describe "what", a request to repeat for clarification before it reaches on its conclusions. This ability is usually important because the type of problems to which a KBS is needed requires an explanation of the result presented to the end-users. It has also the ability of justifying "why" a certain problem is being questioned in order to reason out what it means and how it benefits. The system included "what" and "why" explanation facilities in problem solving. Moreover, the explanation facilities included in the system are easily understandable by the end-users.

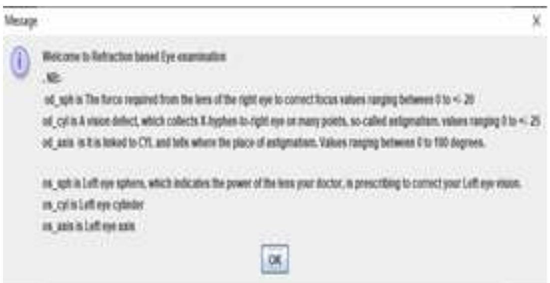

Figure 4. User interface of the explanation facility.

Figure 5. User interface for detected eye disorder result.

System Testing

Before deployment of the system to actual use, it has to be evaluated by domain experts to assure whether the system meets its objectives or not. To evaluate the performance of the KBS, test cases are prepared and given to the system. The outputs of the system are compared against domain experts' judgment.

Table 8. KBS **accuracy test.**

	Prototype System Response/output							
	Diagnoses Type	Normal	Myopia	Hyperopia	Mixed-astigmatism	Hyperopic-astigmatism	Myopic-astigmatism	Total
Domain Experts suggestion	Normal	7	0	0	0	0	0	7
	Myopia	3	5	0	0	0	0	8
	Hyperopia	1	0	10	0	0	0	11
	Mixed-astigmatism	0	0	0	4	0	0	4
	Hyperopic-astigmatism	0	0	0	0	3	0	3
	Myopic-astigmatism	0	0	0	0	0	4	4
	Total	11	5	10	4	3	4	37

The system accuracy test, Table 8, shows that test cases evaluation by the KBS and domain experts' judgment. The column side shows the result of KBS and the row side shows suggestion of domain experts.

The entry under column Normal indicates that out of eleven instances seven of the instances are correctly identified as Normal, three of the instances are incorrectly identified as Myopia and one instance is incorrectly identified as hyperopia. The entries in all the rest columns show that the system has correctly identified the instance. Overall, the system has correctly detected and diagnosed 33 test instances out of 37 to their correct class. This means the system has 89.2% detection accuracy.

DISCUSSION

Eye disease diagnosis is one of real-world medical problem. Detection of eye disease in its early stages can prevent complications and blindness.

In this study, the possibility of integrating data mining models with knowledge-based system is realized and explored. The use of data mining techniques to build the knowledge base of the KBS can be taken as a strong feature of the developed system.

The integration process began by taking initial samples of ophthalmology datasets. The collected datasets was preprocessed and made suitable for mining steps. Since the system focuses on refraction-based eye examination process, automatic knowledge acquisition mechanism is applied in this study to make intelligent decision-making. PART data mining classifier is employed for automatic knowledge acquisition. PART performed best among the selected four-classifier algorithms namely J48, REPTree and JRip with an accuracy of 60% and 96.45% for subjective and objective based model evaluation respectively. Automatic integrator application is built to map the datamining model with knowledge-based system and construct knowledge base. Following the successful integration of induced knowledge with the knowledge-based system, an integrated KBS for detection of eye refractive error conditions is built. System performance testing is undertaken to make sure that the KBS works well. Testing the system registered a promising result of 89.2% accuracy.

In this study, promising result is achieved in integrating data mining induced patterns with knowledge-based system for detecting eye refractive error conditions and providing appropriate advice for ophthalmologists.

This study provides automated medical diagnosis and treatment of eye disorders and diseases. The knowledge-based system improves decision-making process on eye refractive error detection. The system provides advice about detected refractive error conditions to the ophthalmologist as well as which action to take in accordance to the detected eye refractive error condition. The system can also become an expert knowledge-sharing tool to be used by other medical personnel who are not specialists in diagnosis of eye refractive errors, especially for hospitals that do not have an ophthalmologist. Moreover, this research helps to increase awareness on the negative impacts of eye refractive errors and helps to detect them early for better treatment.

In addition, this study will motivate future researchers to work on developing knowledge-based system using data mining techniques in other fields of studies especially in areas where there is shortage of domain experts.

The developed integrated KBS supports specific types of eye refractive error conditions. However, each type of eye refractive error has sub categories. It is possible to increase the number of eye diseases in future work using similar studies. The following are recommended as potential research area for future study.

- The need for additional types of eye refractive disorders solution for users who have multiple eye disorder diagnosis.
- The use of image datasets to improve the performance of the model in detection of eye refractive error conditions.
- Developing hybrid knowledge based system, which is capable of employing rule based reasoning, and case based reasoning with integrated data mining.

ACKNOWLEDGMENTS

The authors acknowledge Mekelle University for allowing this research as well as Haramaya University for writing support letter to the eye hospitals where the study was conducted. Moreover, the authors are grateful to the management of the eye hospitals considered in this study for providing the dataset.

REFERENCES

1. Sheeladevi S, Seelam B, Nukella PB, Borah RR, Ali R and Keay L. (2019). Prevalence of refractive errors, uncorrected refractive error, and presbyopia in adults in India: A systematic review. *Indian Journal of Ophthalmology*, **67**:583-92. https://doi.org/10.4103/ijo.IJO

2. Blondet, G., Duigou, J. Le, and Boudaoud, N. (2019). A knowledge-based system for numerical design of experiments processes in mechanical engineering. *Expert Systems with Applications*, **122(6)**:289–302. https://doi.org/10.1016/j.eswa.2019.01.013

3. Brázdil, T., Chatterjee, K., Křetínský, J., and Toman, V. (2018). Strategy representation by decision trees in reactive synthesis. *Journal of Artificial Intelligence and Lecture Notes in Bioinformatics*, 385–407. https://doi.org/10.1007/978-3-319-89960-2

4. Cheng, Y., Chen, K., Sun, H., Zhang, Y., & Tao, F. (2018). Data and knowledge mining with big data towards smart production. *Journal of Industrial Information Integration*, **9(8)**:1–13. https://doi.org/10.1016/j.jii.2017.08.001

5. Kedir Eyasu,Worku Jimma andTakele Tadesse (2020). Developing a Prototype Knowledge-Based System for Diagnosis and Treatment of Diabetes Using Data Mining Techniques. *Ethiopian Journal of Health Sciences*, **30(1)**:115–124. https://doi.org/10.4314/ejhs.v30i1.15

6. Fageeri, S. O., Ahmed, S. M. M., 1. Almubarak, S. A., and Mu'Azu, A. A. (2017). Eye refractive error classification using machine learning techniques. *Journal of International Conference on Communication, Control, Computing and Electronics Engineering.* https://doi.org/10.1109/ICCCCEE.2017.7867660

7. Cios, K.J., Pedrycz, W., Swiniarski, R.W. and Kurgan, L.A. (2019). Data Mining A Knowledge Discovery Approach. *International Journal of Chemical Information and Modeling*, **53(9)**: 1689-1699.

8. Han, J., Kamber, and Pei, J. (2012). Data mining Concepts and Techniques Preface and Introduction. Morgan Kaufmann, 3rd edtion,1-740.

9. Nebiyat Kassa, Alemayehu Woldeyes and Chalachew Misganaw (2014). Prevalence and Factors Associated With Refractive Error Among Primary School Children in Addis Ababa, Ethiopia. *International Journal of Medical and Health Sciences Research*, **1(9)**:92–104.

10. Jafer Kedir and Abonesh Girma. (2014). Prevalence of refractive error and visual

impairment among rural school-age children of Goro District, Gurage Zone, Ethiopia. *Ethiopian Journal of Health Sciences,* **24(4):** 353–358. https://doi.org/10.4314/ejhs.v24i4.11

11. Codreanu, D.E., Parpandel, D.E. and Popa, I. (2011) Data Mining : Extracting Knowledge from Data. Nicolae Titulescu University Publishing.

12. Mihaela, O. (2006). On the Use of Data-Mining Techniques in Knowledge-Based Systems. 1–4.

13, Li, J.-S., Yu, H.-Y., and Zhang, X.-G. (2011). Data Mining in Hospital Information System. New Fundamental Technologies in Data Mining. https://doi.org/10.5772/14052.

14. Liu, Q., & Wu, Y. (2012). Encyclopedia of the Sciences of Learning. Encyclopedia of the Sciences of Learning. https://doi.org/10.1007/978-1-4419-1428-6

15. Milovic, B. (2012). Prediction and decision making in Health Care using Data Mining. *International Journal of Public Health Science (IJPHS)*, **1(2):** 69–76. https://doi.org/10.11591/ijphs.v1i2.1380

16. Puaschunder, J. M. (2020). The Potential for Artificial Intelligence in Healthcare. *Electronic Journal of Social Science Research Network* (SSRN), *6(2)*:94–98. https://doi.org/10.2139/ssrn.3525037

17. Sharma, S., Agrawal, J., Agarwal, S., & Sharma, S. (2013). Machine learning techniques for data mining: A survey. *Journal of International Conference on Computational Intelligence and Computing Research, I.* https://doi.org/10.1109/ICCIC.2013.6724149

18. Tibebu Kassa, G. D. A. (2000). prevalance of refractive error in preschool and school childern of Debark and Kola Diba towns,Northwest Ethiopia. *Ethiopian Journal of Health Development,* **17(2)**: 117–124.

19. Tripathi, K. P. (2011). A Review on Knowledge-based Expert System : Concept and Architecture. *Journal of Artificial Intelligence Techniques - Novel Approaches & Practical Applications*, **4(4)**:19–23.

20. Varadarajan, A. V., Poplin, R., Blumer, K., Angermueller, C., Ledsam, J., Chopra, R., Keane, P. A., Corrado, G. S., Peng, L., and Webster, D. R. (2018). Deep learning for predicting refractive error from retinal fundus images. *International Journal of Investigative Ophthalmology and Visual Science,* **59(7)**:2861–2868. https://doi.org/10.1167/iovs.18-23887