Use of data mining technology to identify factors affecting cataracts (Applying to Makkah Eye Hospital 2016-2017)

Nabila Tariq Abdul Nabi Muhammad

- Al-Neelain University - College of Graduate Studies - Department of Information Systems

Abstract

The problem of the study is the high incidence of factors affecting the disease (Cataract) in Khartoum state. Because of insufficient information on the disease helping in the prediction of the approximate number of the expected cases. This study aims to discover the important relationships between eye diseases and the factors influencing their causes. And providing future informationon eye diseases to help in decision making. And provision adequate resources to fight these diseases through building a suitable model.The importance of this theory is that Data mining is an important shape of decision supporting which helps in decision making. And also allows for accurate queries and continuously develop it without having a specific goal idea at first. The scientific importance is to build future predictions and extracting new information determines whether there is a virtual relationship reflecting the reality of the data. The descriptive approach was followed because it is a way to describe the subject that you want to study. And depict the results that have been reached on digital forms that can be interpreted. Besides learning the phenomenon as it is actually.The analysis and design systems approach was also followed because it helps in the analysis inputs, outputs, and processes. Weka software was used to build a model to know the future prediction of the disease and an Excel worksheet was used to review the results. It was reached that there is a direct correlation between the average temperature and the number of cases, strong positive relationship between the disease and the age

of the patient, and a poor relationship concerning genetics. It is recommended to uses seasonal data to help in an accurate prediction of Cataract cases. Besides examining the other factors of the infection. It is also recommended to add other attributes related to individuals such as gender and living area.
Key Words:data mining/cataracts/ model/Linear regression

المستخلص

مشكلة الدراسة هي ارتفاع نسبة العوامل المؤثرة على مرض (الساد) بولاية الخرطوم. بسبب عدم كفاية المعلومات عن المرض مما يساعد في التنبؤ بالعدد التقريبي للحالات المتوقعة. يهدف هذا البحث إلى اكتشاف العلاقات المهمة بين أمراض العيون والعوامل التي تؤثر على أسبابها. وتوفير المعلومات المستقبلية عن أمراض العيون للمساعدة في اتخاذ القرار. وتوفير الموارد الكافية لمكافحة هذه الأمراض من خلال بناء نموذج مناسب. تكمن أهمية هذه النظرية في أن التنقيب عن البيانات هو شكل مهم لدعم القرار مما يساعد في صنع القرار. ويتيح أيضًا إجراء استعلامات دقيقة وتطويرها باستمرار دون أن يكون لديك فكرة هدف محددة في البداية. تكمن الأهمية العلمية في بناء تنبؤات مستقبلية واستخراج معلومات جديدة يحدد ما إذا كانت هناك علاقة افتراضية تعكس حقيقة البيانات. تم اتباع النهج الوصفي لأنه طريقة لوصف الموضوع الذي تريد دراسته. وتصور النتائج التي تم التوصل إليها على الأشكال الرقمية التي يمكن تفسيرها. إلى جانب تعلم الظاهرة كما هي في الواقع. كما تم اتباع نهج أنظمة التحليل والتصميم لأنه يساعد في تحليل المدخلات والمخرجات والعمليات. تم استخدام برنامج Weka لبناء نموذج لمعرفة التنبؤ المستقبلي للمرض وتم استخدام ورقة عمل Excel لمراجعة النتائج.

تم التوصل إلى أن هناك علاقة طردية بين متوسط درجة الحرارة وعدد الحالات، وعلاقة موجبة قوية بين الموضوع المريض، وعلاقة ضعيفة فيمايتعلق بالوراثة. يوصى باستخدام البيانات الموسمية للمساعدة في التنبؤ الدقيق بحالات الساد. إلى جانب فحص العوامل الأخرى للعدوى. يوصى أيضًا بإضافة سما ت أخرى متعلقة بالأفراد مثل الجنس ومنطقة المعيشة.

الكلمات المفتاحية:تنقيب البيانات / مرض الكاتاراكت/ نموذج/ الانحدار الخطي

1.0 Introduction:

The Advances in scientific and the widespread use of technology in various aspects of daily life has increased the ability to generate and collect data quickly in this era and this has contributed to the computerization of most of the works, sciences and services that have been provided daily everywhere around the world so that most products of all kinds have become A digital code that distinguishes them from each other.

Data mining is a process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use [1]. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. Aside from the raw analysis step, it also involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.

The term "data mining" is a misnomer, because the goal is the extraction of patterns and knowledge from large amounts of data, not the extraction (mining) of data itself. It also is a buzzwordand is frequently applied to any form of large-scale data or information processing (collection, extraction, warehousing, analysis, and statistics) as well as any application of computer decision support system, including artificial intelligence (e.g., machine learning) and business intelligence. The book Data mining[2]: Practical machine learning tools and techniques with Java(which covers mostly machine learning material) was originally to be named just Practical machine learning, and the term data mining was only added for marketing reasons. Often the more general terms (large scale) data analysis and analytics—or, when referring to actual methods, artificial intelligence and machine learning—are more appropriate.

The actual data mining task is the semi-automatic or automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining, sequential pattern mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting is part of the data mining step, but do belong to the overall KDD process as additional steps(3). The difference between data analysis and data mining is that data analysis is used to test models and hypotheses on the dataset, e.g., analyzing the effectiveness of a marketing campaign, regardless of the amount of data; in contrast, data mining uses machine learning and statistical models to uncover clandestine or hidden patterns in a large volume of data.

1.1 Statement of the problem.

The problem of this study lies in the:

lack of sufficient information on the disease-causing factors, and the lack of forecasts approximate numbers expected casualties.

1.2 Objectives of the Study

This study aims to fulfill the following objectives:

1. Discover the important relationships between eye disease (cataract) and the factors affecting its etiology.
2. Providing future information about white water disease to help take the necessary measures to prevent it.
3. Providing predictions of future disease numbers.

1.3 Hypotheses of the study

The followings are the hypotheses of the study which will be tasted on the basis of the results:

1. There is a direct relationship between high temperature and increased incidence of disease.
2. There is a strong direct relationship between increased age and disease.
3. There is a weak relationship between the genetic factor and the disease.

1.4Significance of the study

Theoretical importance: The importance of research lies in the existence of data to predict the etiology of the disease, and to provide information for disease prevention.

Scientific importance: Building future predictions and extracting new data to determine whether there is any apparent relationship that reflects a reality in the nature of the data.

2.0Literature Review:

2.1 Data mining:

Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. Data mining depends on effective data collection, warehousing, and computer processing. are used to build machine learning models that power applications including search engine technology and website recommendation programs(4).

Data mining has become very popular in the information technology industry and is also emerging in the field of chemo metrics. This has coincided with the evolution of the field of chemo metrics into the wider field of cheminformatics. Data mining is necessary because of the increasing availability of very large amounts of data and the pressing need for converting such data into useful information and knowledge. Data mining is essentially the science of extracting information from large data sets and databases.

As Han and Kamber1 point out, the term 'data mining' is a misnomer. In analogy, if a geologist talks about gold mining, he is looking for gold among rocks and sands. Thus, data mining should be better named 'knowledge mining from data' or simply 'knowledge mining'(5). Apart from finding an appropriate name for a new discipline, it is also a controversial task to define the discipline, as researchers tend to disagree about the extent and limitations of a particular field of science. In this context, the definition of Hand et al(6).fits well with the field of chemo metrics where data mining is not usually applied to huge databases. They state: "Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner". These relationships and summaries present themselves as models and patterns such as linear relationships, tree structures, clusters, recurrent patterns in images, and spectral data. It is clear that we are rather referring here to observational data than to experimental data, which means that we are looking for additional or 'secondary information' in the data unrelated to the main purpose for which the data were collected. The definition also refers to 'large data sets' and essentially also 'complex data sets'. If we were only interested in small data sets with a few variables, then we would be discussing 'traditional exploratory data analysis'. This brings us to the point that there is a strong connection between 'data exploration techniques' and 'data mining methods'. Essentially, data exploration techniques belong to the arsenal of data mining tools, but should not be completely confused. Data mining of large databases involves more stages and more complex algorithms than simple data exploration. As such, data mining requires the integration of techniques from multiple disciplines including statistics, mathematics, machine learning, database technology, data visualization, pattern recognition, signal processing, information retrieval, and high-performance computing. Based on this view, an efficient data mining system consists of different components. Some goals of data mining include supervised classification and regression. These topics have been extensively presented in other sections of 'Comprehensive Chemo metrics' and will not be repeated here. However, it needs to be stated that when dealing with very large and complex data sets, additional issues arise even with traditional methods of regression analysis and supervised classification(7).

2.1.1Data Mining Objectives and Outcomes:

Data Mining is concerned with the search for new knowledge in data. Such knowledge is usually obtained in the form of rules which were previously unknown to the user and may well prove useful in the future. These rules might take the form of specific rules induced by means of a rule induction algorithm or may be more general statistical rules such as those found in predictive modeling. The derivation of such rules is specified in terms of Data Mining tasks where typical tasks might involve classifying or clustering the data(8).

A highly desirable feature of Data Mining is that there be some high-level user interface that allows the end-user to specify problems and obtain results in as friendly as matter as possible. Although it is possible, and in fact common, for Data Mining to be carried out by an expert and the results then explained to the user, it is also highly desirable that the user be empowered to carry out his own Data Mining and draw his own conclusions from the new knowledge. An appropriate user interface is therefore of great importance.

Another secondary objective is the use of efficient data access and data processing methods. Since Data Mining is increasingly being applied to large and complex databases, we are rapidly approaching the situation where efficient methods become a sine qua non. Such methods include Distributed and Parallel Processing, the employment of Data Warehousing and accompanying technologies, and the use of Open Database Connectivity (ODBC) to facilitate access to multi-databases(9).

2.1.2The importance of data mining:

Data mining involves exploring and analyzing large blocks of information to glean meaningful patterns and trends. It can be used in a variety of ways, such as database marketing, credit risk management, fraud detection, spam Email filtering, or even to discern the sentiment or opinion of users.

The data mining process breaks down into five steps(10):

1. organizing collect data and load it into their data warehouses.
2. They store and manage the data, either on in-house servers or the cloud.
3. Business analysts, management teams and information technology professionals access the data and determine how they want to organize it.
4. Application software sorts the data based on the user's results.
5. User presents the data in an easy-to-share format, such as a graph or table.

2.1.3 Data Mining Technologies:

The analytical techniques used in data mining are often well-known mathematical algorithms and techniques. What is new is the application of those techniques to general business problems made possible by the increased availability of data and inexpensive storage and processing power. Also, the use of graphical interfaces has led to tools becoming available that business experts can easily use.

Some of the tools used for data mining are:

Artificial neural networks - Non-linear predictive models that learn through training and resemble biological neural networks in structure.

Decision trees - Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset.

Rule induction - The extraction of useful if-then rules from data based on statistical significance.

Genetic algorithms - Optimization techniques based on the concepts of genetic combination, mutation, and natural selection.

Nearest neighbor - A classification technique that classifies each record based on the records most similar to it in an historical database.

2.1.4 The Future of Data Mining:

In the short-term, the results of data mining will be in profitable, if mundane, business related areas. Micro-marketing campaigns will explore new niches. Advertising will target potential customers with new precision. In the medium term, data mining may be as common and easy to use as e-mail. We may use these tools to find the best airfare to New York, root out a phone number of a long-lost classmate, or find the best prices on lawn mowers.

The long-term prospects are truly exciting. Imagine intelligent agents turned loose on medical research data or on sub-atomic particle data. Computers may reveal new treatments for diseases or new insights into the nature of the universe. There are potential dangers.

2.2 Eyes disease A cataract:

A cataract is a cloudy area in the lens of the eye that leads to a decrease in vision. Cataracts often develop slowly and can affect one or both eyes[(11)]. Symptoms may include faded colors, blurry or double vision, halos around light, trouble with bright lights, and trouble seeing at night. This may result in trouble driving, reading, or recognizing faces[(12)].Poor vision caused by cataracts may also result in an increased risk of falling and depression. Cataracts

cause half of all cases of blindness and 33% of visual impairment worldwide. Cataracts are most commonly due to aging but may also occur due to trauma or radiation exposure, be present from birth, or occur following eye surgery for other problems(13).Risk factors include diabetes, longstanding use of corticosteroid medication, smoking tobacco, prolonged exposure to sunlight, and alcohol(14).The underlying mechanism involves accumulation of clumps of protein or yellow-brown pigment in the lens that reduces transmission of light to the retina at the back of the eye. Diagnosis is by an eye examination(15).
Prevention includes wearing sunglasses, a wide brimmed hat, eating leafy vegetables and fruits, and avoiding smoking(16).Early on the symptoms may be improved with glasses. If this does not help, surgery to remove the cloudy lens and replace it with an artificial lens is the only effective treatment. Cataract surgery is not readily available in many countries, and surgery is needed only if the cataracts are causing problems and generally results in an improved quality of life(17).

2.3 Previous Studies

2.3.1Nehal M. Samy El Gendy, Ahmed A. Abdel-Kader (2018):

Purpose. To highlight the prevalence of selected ophthalmic diseases accidentally discovered at first-time screening of a large sample of patients from the Middle East and North Africa visiting a large referral university hospital checkup unit based in Cairo. Material and Methods. A cross-sectional study of two thousand and thirteen subjects coming for routine ophthalmic medical checkups from different Middle East countries (mainly Egypt, Sudan, and Yemen). Patients were evaluated for prevalence of diabetic retinopathy, glaucoma, ocular hypertension, cataract, and amblyopia. Patients' demographic data and medical history were collected. Complete ophthalmic examination was performed. Investigations were done when needed to confirm suspected conditions. Results. The study included 1149 males and 864 females. 652 Sudanese patients, 568 Yemeni patients, 713 Egyptian patients, and 63 patients from different Gulf and North African countries like Saudi Arabia, Qatar, Libya, and Jordan. Sudanese patients showed a higher percentage of glaucoma (13.3%) and ocular hypertension (8.3%). Yemeni patients showed the highest prevalence of amblyopia (6.7%), diabetic retinopathy (8.6%), and cataract (4.2%). The group of relatively higher economic classification seemed to show fewer prevalence's of these ophthalmic conditions. Yemeni patients tended to have a high percentage of persistent myelinated nerve fibers. Conclusion. Different ophthalmic conditions were discovered for the first time at the general checkup clinic. Certain conditions were more common than others in certain countries. The lack of regular checkups and the unavailability of medical services due to low to moderate socioeconomic status as well as political turbulence may account for the delay in initial diagnosis of many treatable conditions[18].

2.3.2 Judith Kouassi Nzoughet, Khadidja Guehlouz (2020):

Glaucoma is an age related disease characterized by the progressive loss of retinal ganglion cells, which are the neurons that transduce the visual information from the retina to the brain. It is the leading cause of irreversible blindness worldwide. To gain further insights into primary open-angle glaucoma (POAG) pathophysiology, we performed a non-targeted metabolomics analysis on the plasma from POAG patients (n = 34) and age- and sex-matched controls (n = 30). We investigated the differential signature of POAG plasma compared to controls, using liquid chromatography coupled to high resolution mass spectrometry (LC-HRMS). A data mining strategy, combining a filtering method with threshold criterion, a wrapper method with iterative selection, and an embedded method with penalization constraint, was used. These strategies are most often used separately in metabolomics studies, with each of them having their own limitations. We opted for a synergistic approach as a mean to unravel the most relevant metabolomics signature. We identified a set of nine metabolites, namely: nicotinamide, hypoxanthine, xanthine, and 1-methyl-6,7-dihydroxy-1,2,3,4-tetrahydroisoquinoline with decreased concentrations and N-Acetyl-L-Leucine, arginine, RAC-glycerol 1-myristate, 1-oleoyl-RAC-glycerol, cystathionine with increased concentrations in POAG; the modification of nicotinamide, N-Acetyl-L-Leucine, and arginine concentrations being the most discriminant. Our findings open up therapeutic perspectives for the diagnosis and treatment of POAG[19].

3.0 Data analysis and discussion of results

In this topic the researcher deals with a precise description of the method and procedures that I follow in carrying out this study, and this includes a description of the study community, the procedures that were taken to ensure its effectiveness and impact, the method followed to apply it, and the statistical treatments by which the data were analyzed. And extract the results.

Methodology

The study will use the descriptive analytical method and the systems analysis and design approach was used in the analysis of outputs, inputs and processing. about the Investigating Use of data mining technology to identify factors affecting cataracts applying to Makkah Eye Hospital.

Limits of the study

This study will be limited to Makkah Eye Hospital, in the period between January 2016 - October 2017.

Tools of the study:

Weka: Java-based machine learning software developed at the University of Waikato in New Zealand, from open source software, containing a set of presentation tools and algorithms for data analysis and predictive modeling, in addition to graphical user interfaces for easy access to these functions. Wicca supports several data mining tasks, specifically, the Regression Clustering, Association, Data pre-processing, Classification, Visualization Selection feature and. All weka technologies are based on the assumption that data is available as an appendix, each data point is described with a fixed number of attributes or numeric, normally (nominal attributes) and also supports some other attributes. Wiccan provides access to SQL databases using Database java connectivity

Study population and Sample size:

The study population constitutes of 3222 individuals for 22 months to determine the number of individuals with cataract disease from the total community.

Looking the result model at Figure (1) displays the output form results: The weka only uses columns that have a statistical contribution occurring within the model, and ignores those that are not conducive to creating a good model. The regression model showed that genetics had no influence on disease incidence. As for the other factors, 21,653 were added to the average temperature factor, and 2.8558 were added to the average age factor. The output also shows the data predicted by the model, the real data, and the amount of each error in each of the training data. The correlation coefficient shows the predictions that are related to the outputs, the closer to the correct one, the better. Results The correlation coefficient when evaluating the test data is 0.9905. mean after the model's predictions from the actual data points returned a result of 8.3776.

The sample also illustrates other different ways to calculate the error rate (11.2524: Root mean squared error).

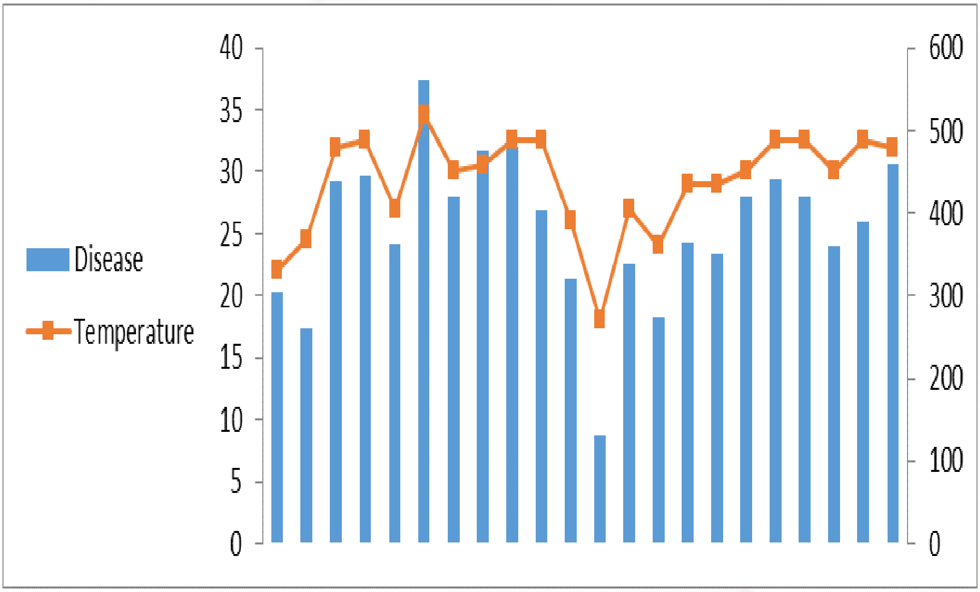

Figure (2): shows the relationship between temperature and disease rate

There is a direct relationship between temperature and the number of cataract infections, due to the effect of high temperature on eye pressure, where we note that the highest temperature (34.5) represents the optimal degree of high incidence of the disease.

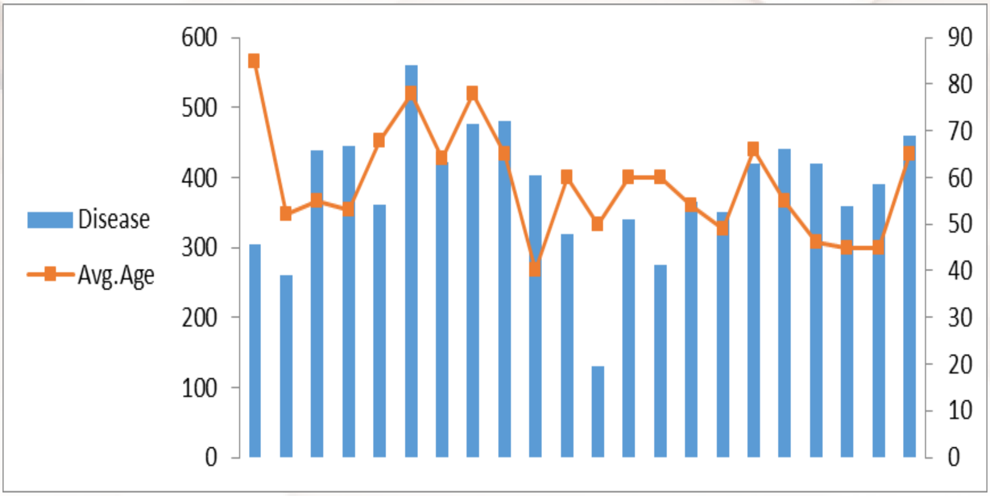

Figure (3): shows the relationship between life expectancy and disease rate

We note that the average age is one of the effective factors in the spread of the disease.The relationship between the number of injuries and age is a direct one, as a rise in the average age increases the number of disease infection. This may be attributed to the influence of other factors on age that increases the chance of contracting the disease.

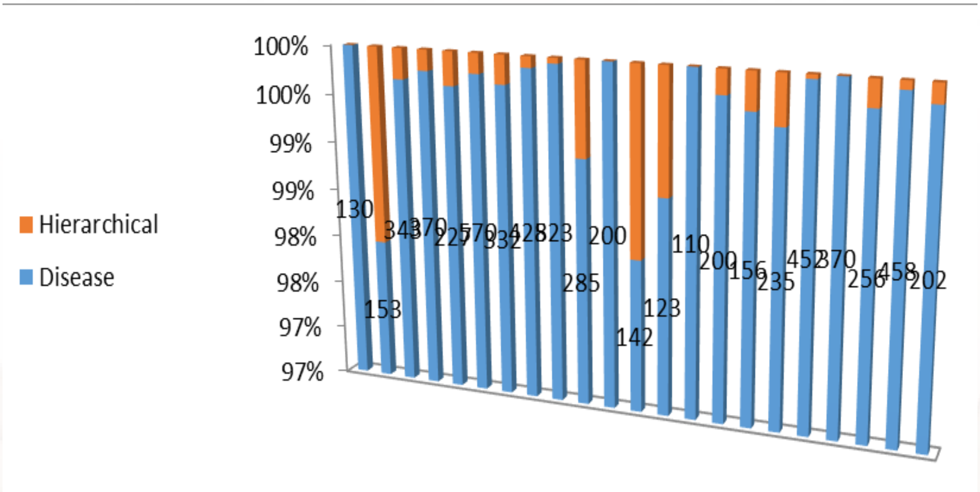

Figure (4): shows the percentage of the heredity factor in developing the disease

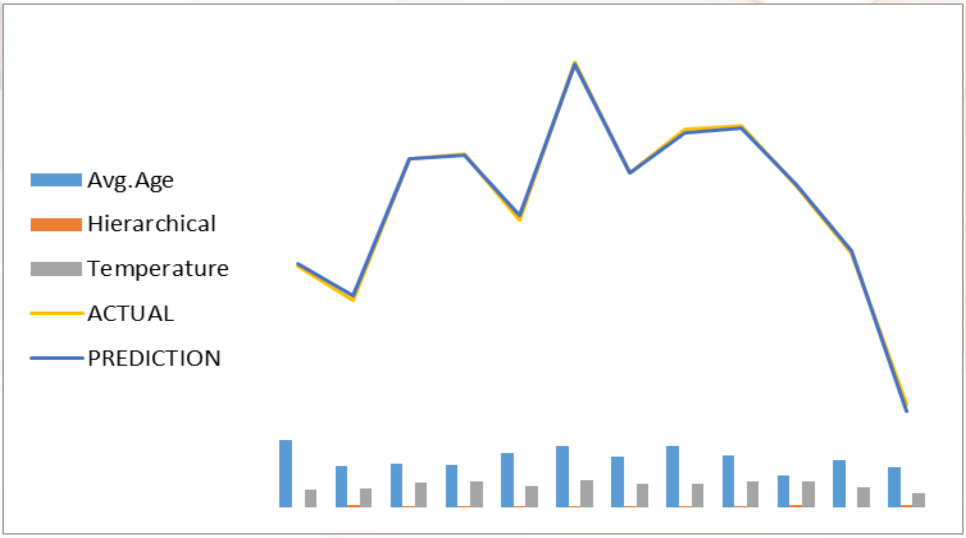

Figure (5) showing the predicted disease injuries using the model with real injuries.

The disease model based on influencing factors that was constructed through data mining techniques had an ability to interpret hospital cataract injury data (2016-2017) through factors (age, temperature, genetics).

Conclusion:

The researcher reviewed the related literature and the previous studies as conceptual, theoretical frame work of the study.

The researcher used the descriptive, analytical method for the study, the subject of the study Use of data mining technology to identify factors affecting cataracts (Applying to Makkah Eye Hospital), then the researcher used the (weka) TO analyze the data.

According to discussion of the results of the study approached the following results:

1. There is a direct relationship between high temperature and increased incidence of disease.According to statistics, the rate of injuries increases with the increase in temperature, especially in the outskirts of the state.
2. There is a strong direct relationship between increased age and disease.
3. There is a weak relationship between the genetic factor and the disease.

Recommendations:

In light of the results of this study the researcher recommends the following:

1. Adding other characteristics related to community members, such as gender and region.
2. Study the factors affecting other eye diseases.
3. Using existing data of the seasons of the year to more accurately predict cataract disease, while examining the contribution of other factors to the disease incidence rate.
4. Applying the study to different rural areas and determining whether the factors have the same effect

Model results

```
=== Run information ===

Scheme:weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8
Relation:     SAD
Instances:    12
Attributes:   4
              Temperature
              Hierarchical
              Age
              Disease
Test mode:evaluate on training data

=== Classifier model (full training set) ===


Linear Regression Model

Disease =

     21.653  * Temperature +
      2.8558 * Age +
   -411.7065

Time taken to build model: 0.03 seconds
```

inst#,	actual,	predicted,	error
1	304	307.404	3.404
2	260	267.295	7.295
3	439	438.259	-0.741
4	445	443.374	-1.626
5	362	367.12	5.12
6	560	558.076	-1.924
7	421	420.656	-0.344
8	476	471.464	-4.536
9	480	477.644	-2.356
10	404	406.249	2.249
11	320	322.621	2.621
12	130	120.838	-9.162

```
=== Evaluation on test set ===
=== Summary ===

Correlation coefficient                  0.9905
Mean absolute error                      8.3776
Root mean squared error                 11.2524
Relative absolute error                 19.0833 %
Root relative squared error             21.5131 %
Total Number of Instances               10
```

Figure (1) displays the output form results

References:

(1) M. F. El-Bab, N. Shawky, A. Al-Sisi, and M. Akhtar, "Retinopathy and risk factors in diabetic patients from Al-Madinah Al-Munawarah in the Kingdom of Saudi Arabia," Clinical Ophthalmology, vol. 6, pp. 269–276, 2012.

(2) F. Al-Maskari and M. El-Sadig, "Prevalence of diabetic retinopathy in the United Arab Emirates: a cross-sectional survey," BMC Ophthalmology, vol. 7, no. 1, p. 11, 2007.

(3) H. A. Osman, N. Elsadek, and M. A. Abdullah, "Type 2 diabetes in Sudanese children and adolescents," Sudanese Journal of Paediatrics, vol. 13, no. 2, pp. 17–23, 2013.

(4) R. B. Khandekar, M. A. Bamashmus, and A. A. Gunaid, "Diabetic retinopathy, visual impairment, and ocular status among patients with diabetes mellitus in Yemen: a hospital based study," Indian Journal of Ophthalmology, vol. 57, no. 4, pp. 293–298, 2009.

(5) T. A. Macky, N. Khater, M. A. Al-Zamil, H. El Fishawy, and M. M. Soliman, "Epidemiology of diabetic retinopathy in Egypt: a hospital-based study," Ophthalmic Research, vol. 45, no. 2, pp. 73–78, 2011.

(6) M. Kohlhaas, A. G. Boehm, E. Spoerl, A. Pursten, H. J. Grein, and L. E. Pillunat, "Effect of central corneal thickness, corneal curvature, and axial length on applanation tonometry," Archives of Ophthalmology, vol. 124, no. 4, pp. 471–476, 2006.

(7) M. A. Kass, D. K. Heuer, E. J. Higginbotham et al., "The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma," Archives of Ophthalmology, vol. 120, no. 6, p. 701, 2002.

(8) P. J. Foster, R. Buhrmann, H. A. Quigley, and G. J. Johnson, "The definition and classification of glaucoma in prevalence surveys," The British Journal of Ophthalmology, vol. 86, no. 2, pp. 238–242, 2002.

(9) J. C. Mwanza, J. D. Oakley, D. L. Budenz, D. R. Anderson, and Cirrus Optical Coherence Tomography Normative Database Study Group, "Ability of cirrus HD-OCT optic nerve head parameters to discriminate normal from glaucomatous eyes," Ophthalmology, vol. 118, no. 2, pp. 241–248.e1, 2011.

(10) F. Kyari, M. M. Abdull, F. B. Sallo et al., "Nigeria normative data for defining glaucoma in prevalence surveys," Ophthalmic Epidemiology, vol. 22, no. 2, pp. 98–108, 2015.

(11) Gimbel, HV; Dardzhikova, AA (January 2011). "Consequences of

waiting for cataract surgery". Current Opinion in Ophthalmology. 22 (1): 28–30.

(12) Chan, WH; Biswas, S; Ashworth, JL; Lloyd, IC (April 2012). "Congenital and infantile cataract: aetiology and management". European Journal of Pediatrics. **171** (4): 625–30

(13) Christen WG, Manson JE, Seddon JM, Glynn RJ, Buring JE, Rosner B, Hennekens CH (August 1992). "A prospective study of cigarette smoking and risk of cataract in men". JAMA. **268** (8): 989–993.

(14) Hodge, WG; Whitcher, JP; Satariano, W (1995). "Risk factors for age-related cataracts". Epidemiologic Reviews. **17** (2): 336–346.

(15) he global burden of disease : 2004 update. Geneva, Switzerland: World Health Organization. 2008. p. 35.

(16) Kahn HA, Leibowitz HM, Ganley JP, Kini MM, Colton T, Nickerson RS, Dawber TR (Jul 1977). "The Framingham Eye Study. I. Outline and major prevalence findings". Am. J. Epidemiol. **106** (1): 17–32.

(17) Guo Y, Yan H (2006). "Preventive effect of carnosine on cataract development". Yan Ke Xue Bao. **22** (2): 85–88.

(18) Nehal M. Samy El Gendy , Ahmed A. Abdel-Kader, " Prevalence of Selected Eye Diseases Using Data Harvested from Ophthalmic Checkup Examination of a Cohort of Two Thousand Middle Eastern and North African Subjects" ,Ophthalmology Department, Kasr Al Ainy Medical School, Cairo University, Cairo, Egypt,2018,pp47-91

(19) Judith Kouassi Nzoughet, Khadidja Guehlouz,"A Data Mining Metabolomics Exploration of Glaucoma",Singapore Eye Research Institute, Singapore National Eye Centre, Duke-NUS, the Special Issue Metabolomics in Human Tissues and Materials,Published: 28 January 2020,pp217-239.