Data Merging for the Study of Eye Diseases in Relation to Demographic and Weather Factors in Telangana State, India

Amna Alalawi[1], Les Sztandera[2], Dr. Richard Derman[3], Parth Lalakia[4], Anthony Vipin Das[5], Gumpili Sai Prashanthi[6]

[1]DMgmt in Strategic Leadership Program, Thomas Jefferson University, Philadelphia, PA, USA
Email: Amna.Alalawi@Jefferson.edu

[2]Kanbar College of Design, Engineering, and Commerce, Thomas Jefferson University, Philadelphia, PA, USA
Email: Les.Sztandera@Jefferson.edu

[3]Global Health Research, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, USA
Email: Richard.Derman@jefferson.edu

[4]Global Affairs, Thomas Jefferson University, Philadelphia, PA, USA
Email: Parth.Lalakia@jefferson.edu

[5] Department of EyeSmart EMR & AEye, Indian Health Outcomes Public Health and Economics Research Centre (IHOPE), LV Prasad Eye Institute, Hyderabad, Telangana, India
Email: vipin@lvpei.org

[6] Department of EyeSmart EMR & AEye, Indian Health Outcomes Public Health and Economics Research Centre (IHOPE), LV Prasad Eye Institute, Hyderabad, Telangana, India
Email: saiprashanthi.g@lvpei.org

***Abstract* – Big data is the new gold in today's VUCA world. VUCA is a systems-derived acronym used to refer to a volatile, uncertain, complex, and ambiguous system. Taking a holistic approach in studying and examining big data can yield new insights that can benefit healthcare practitioners and decision-makers in a VUCA system. This study explores advancing the method to collect and process Electronic Medical Records (EMR), to expand the interest in the use of big data in health care. The objective of this study was to determine whether the EMR record in LV Prasad Eye Institute (LVPEI) in India can contribute to the management of patient care, through studying how climatic and socio-demographic factors relate to eye disorders and visual impairment in the State of Telangana, using data merging as the main technique to study the data. Our findings highlighted the presence of cataract in rural areas and throughout the different weather seasons. Men tend to be the most, while home makers make the most visit to the hospital.**

Keywords –big data; ophthalmology; ocular diseases; artificial intelligence.

I. INTRODUCTION

This extended research is based on the critical global healthcare issue of examining the demographic and weather factors that correlate to the development of eye diseases in Telangana, India. A brief presentation of the topic was addressed at the Ninth Data Analytics 2020 Conference in Nice, France [1]. Responding to the major interests from the discussions in Nice, this research includes more statistical analysis of the relationship, as well as a detailed explanation of the systems-approach method used to study the data.

India is home to over 8.3 million people with Vision Impairment (VI), the highest in the world [4]. Even though, in 1976, India became the first country in the world to start a national program for control of blindness for the goal to reduce blindness prevalence to 0.3 percent by 2020, the prevalence of blindness still stands at 1.99 percent, according to the National Blindness and Visual Impairment Survey, released in October 2019 by the Union Ministry of Health and Family [2]. The prevalence of blindness and visual impairment is one of the highest in Telangana, a state in Southern India, as inferred from survey [2]. The significant reasons indicated in the survey were due to cataract and refractive error [3].

All surveys in the country have shown that cataract is the most common cause of blindness and all prevention of blindness programs have been "cataract-oriented."

However, it has recently been recognized that the visual outcome of the cataract surgeries, as well as the training of ophthalmologists has been less than ideal.

This study uses Artificial Intelligence (AI) and machine learning techniques to explore a dataset containing information on 873,448 patients who visited LV Prasad Eye Institute (LVPEI), a multi-tier ophthalmology hospital network, based in Hyderabad. The data was extracted from Eye Smart, the name given to the hospital's Electronic Medical Record (EMR) and health management system. The EMR was then merged with climatic factors to test the correlation between climatic variables and ocular diseases presented by the patients [4]. Studying risk factors, primarily associated with climate and the environment can lead to a better understanding of the causes, diagnosis, and treatment of several eye diseases [6].

In healthcare, ophthalmology deals with the diagnosis and treatment of eye disorders. Some known diseases in ophthalmology are cataracts, retinal disorders, macular degeneration, and others. The relatively rapid and recent adoption of EMRs in ophthalmology has been associated with the promise that the accumulation of large volumes of clinical data would facilitate quality improvement and help answer a variety of research questions. Given that EMRs are relatively new in most practices and that clinical data are inherently more complex than other fields that have been altered by the digital revolution, these proposed benefits have yet to be realized [5].

With the rise of big data, it has now become easier to study how culture, race, climate, and other socio-demographic factors correlate to the spread of ocular diseases. This has shed light on recent research in medicine and ophthalmology. An investigation has been conducted with an aim for the application of AI-based hierarchical clustering as a tool to optimize the in total excellence, values, and the security for the Adult Spinal Deformity Surgery (ADS) [17]. It has been observed that prior to this the ADS classification was based on certain radiographic parameters which have been correlated with the patient related outcomes. But the problems faced immensely by the researchers is to separate out the patients and the patterns manually that is in turn was based on hundreds of data parameters and the process was considered to be practically not feasible.

Therefore, as a methodological approach for every probable cluster of patient (N) done on the basis of (M) surgery were normalized for two year enhancement and the massive rate of complication were computed. Thus, this particular study has therefore, highlighted that unsubstantiated hierarchical clustering of the patterns of findings that helps to initiate the prior operative judgment making by formulating a two-year risk benefit grid. In this way the novel AI-ASD pattern of classification and identification have helped to diminish the risk and overall improvement [17]. It should be mentioned that smaller cities face a lot of difficulties to maintain the sustainable welfare of countries in amalgamation with notable standards of living [18].

In another investigation, the emergency bases of hospitalization along with the causes of mortality were utilized as the replacements of frailty. The researchers used to two different models to assign the deteriorating risk score to every subject of the elderly population residing within the Municipality of Bologna, Italy [18]. The study design was a cohort study with of 58 789 subjects as sample size for overall six years with a four-year monitoring period. The study findings reported excellent power of discrimination along with calibration that demonstrated an excellent anticipating ability of the models utilized [18].

In this respect, another study could also be illustrated that had utilized the application of health administrative databases along with authenticated algorithms to show a correlation in between the residential proximity towards foremost roadways along with the prevalence of three major neurological diseases like dementia, multiple sclerosis, and Parkinson's disease [19]. This particular study design was also a cohort study based on two different populations having the age group in the range of 20–50 years with sample size of 4.4 million suffering from multiple sclerosis and the matured adult groups having the age range of 55–85 years with sample population size of 2·2 million suffering from dementia or other Parkinson's disease. The researchers of the study estimated the associations among the following parameters such as the traffic proximity, incidence of dementia, Parkinson's disease, and the multiple sclerosis using the model of Cox proportional hazards which would also take into consideration of certain individualistic contextual parameters such as any injury to brain, diabetes, and the local income [19].

The study demonstrated the successful application of the health record databases along with the specialized analytical tool for the categorization of patterns of large or big data [19]. In today's world, the relationship based on function in between the non-coding RNA (ncRNA) and the varied types of human diseases, is considered to be a central task within the field of modern research for the formulation of effective therapeutic approaches.

In Section 1, we discuss the methodology used in the study, breaking the sections down to explaining the systematic approach used as a framework for studying the date, and an explanation into the computer software used, Microsoft Power BI, to explain the data merging technique. In Section 2, we discuss the analysis and highlight the specific statistical tools and methods used to investigate the study of several demographic and climatic factors that impact upon the individuals in Telangana. Section 3 focuses on providing a thorough analysis of the data findings. Section 4 highlights key findings and trends, and Section 5 includes the conclusion and recommendations pertaining to the use of big data and data merging in EMR to reveal new insights in the study of visual impairment and eye disorders in Telangana.

II. METHODOLOGY

A. Systematic-Approach Using the DIKW Pyramid

Data science, a discipline that has been emerging over the past few years, centers on analyzing data. Since there is no specific definition of data science, we relied on the explicit meaning of the term, and therefore decided to work on complex data sets and determine a way through which it can be evaluated. To make the process simpler, we decided to study the data from a holistic approach. Putting a framework in place, helped us in creating a plan for studying the data to achieve results that can be of use and that can identify and analyze patterns in the data sets, which can be used for future clinical practices.

The following Data, Information, Knowledge, and Wisdom (DIKW) pyramid framework was chosen as the main framework that explains the reasons for transforming the data. By transforming the data into information, we gained certain knowledge about the topic which then was transformed into wisdom that helped us in not only conducting the investigation effectively, but also to gain effective understanding about the research topic. The readers can also rely on this wisdom to gain conceptual clarity and understanding of the subject matter.

The DIKW Pyramid

Figure 1. DIKW Pyramid

Figure 1 depicts the DIKW Pyramid. Dr. Russel Ackoff, wrote that "wisdom is located at the top of a hierarchy of types," suggesting that as the highest level of the hierarchy of types, wisdom is somehow superior to the types below it. Subsequent depictions of the knowledge hierarchy typically exclude the understanding level [22]. Ackoff's hierarchy does not require that data transform into information, information into knowledge, or knowledge into wisdom. Instead, he stated that each category is included in the next. For example, there can be no wisdom without understanding and no understanding without knowledge [22]. Big data analytical tools have a strong potential that allows healthcare professionals to gander upon the clinical data stored within repositories and assist in the process of informed making of decisions. Today, especially with the COVID-19 situation, the healthcare sector is making use of AI for a wide range of findings.

Social researchers are now expected to possess various skills and abilities, in particular, the ability to apply qualitative methods. Such researchers are also expected that they acknowledge as well as explain their personal positioning. This is mainly due to ‘subjectivity’. Every person has their own thoughts and opinions; due to this reason, another researcher exploring the same topic might have a different perspective which will be seen through their investigation. Such differences in perceptions and ‘frames of mind’ influence the way the study is carried out and understood by the readers [24]. As Morison [24] has written: “Sociological research is a complex enterprise involving a dynamic interplay between personal values, theories and practical data gathering skills. Different sociologists, looking at the same community but not starting from the same theoretical viewpoint, may direct their attention to different aspects of the place they are studying and come up with extremely contrasting results.”

B. Data-Merging Using Microsoft Power BI

To gain insight into the climatic and socio-demographic factors that correlate to the risk of ocular diseases in the State of Telangana, we used multiple approaches utilizing AI and statistical software and programming languages, including Microsoft Power BI and Python to explore the dataset, which contained information on 873,448 patients complaining of eye disorder symptoms across multiple categories of ocular diseases. Publicly available climatic variables were obtained and aligned to the dataset through a process called column mutation, and then examined by Microsoft Power BI, which heavily relies on visual illustrations and statistical storytelling to present findings and new insights. It should be noted that Microsoft Power BI is considered an assortment of software or apps which all together works in amalgamation to transform the unrelated sources of data into a visually pattern oriented, continuous and dynamic insights. Column mutation, which is the merging of datasets, was done through Python, an interpreted, object-oriented programming, that codes the columns in a language called Syntax. It works out on the principle of logical and arithmetic computation. This tool has an advantage to handle large and complex datasets.

The process was however timely given the large volume of patients. The process was repeated more than once to ensure minimal error in the merging process. Microsoft Power BI was then used to model the data in order to obtain the visualizations and insights.

III. Analysis

It has proven valuable to first observe which diseases are the most prevalent in the different areas of Telangana, and what age and gender are most affected to get a full understanding of the criticality of the eye disorder epidemic and to provide a baseline against which to compare the climate and patient demographic variables examined.

The use of EMRs in generating new insights has been an increasing trend in the area of ophthalmology. Research in ophthalmology has benefited greatly from the use of EMRs in expanding the breadth of knowledge in areas such as disease surveillance, health services utilizations and outcomes. In addition, the quantity of data available has increased, that it is now highly recommended to work on data linkage systems in eye research, as such data can offer insights into advantages and limitations for future direction in eye research [6].

The timespan of this dataset is between 2011-2019, a total of 8 years. There has been consistency to already known information through the analysis, specifically on gender and age-related eye conditions. Creative featuring techniques have been used to shed the light on the most critical variables through trial and error of testing for relationship in accordance to eye disease.

The master dataset or the big data, which was explored and analyzed, covered clinical visits are from the year 2011-2019, and included demographic information of the patient, including age, gender, profession, data of visit, district of resident, and symptoms and diagnosis of the patient in relation to eye disease. To look further into this issue, we merged climate variables to the dataset to explore the relationship with eye disorders. The AI approach can be of varied types namely conventional symbolic AI, Computational intelligence, and statistical tools or the combination of all of the above. Here in the present assignment the Computational intelligence approach has been adopted for the analytical purpose [23].

The climate variables we examined were average temperature, minimum temperature, maximum temperature, humidity, rainfall, and solar radiation. This data was retrieved from the Telangana State Development Planning Society in the state of Telangana. The findings that relate to temperature and its effect on Cataract in older age was consistent in high and low temperatures.

IV. Findings and Trends

This section highlights key findings of the study, as well as trends in relation to the subject matter as per the demographic and climatic variables tested.

A. Gender and Eye Disorders

First, we broke down the data to understand which gender in Telangana has a higher rate of risk in developing eye diseases. This break- down helped us to understand the cultural background of Indian culture among men verses women.

Gender and Eye Disorders

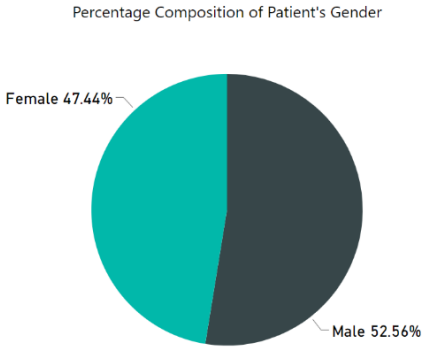

Figure 2. Clinical Visits by Gender (2011-2019)

Figure 2 indicates that between the years 2011-2019, 53% of the patients were male patients who were seen for eye disorders, and 47% were female patients. This finding is in line with the gender study that was conducted on 2.3 million patients of all those who presented to LV Prasad Institute from the years 2011-2019 [4]. Globally, one of the social determinants of health that has been universally identified is gender. In India, health inequalities between men and women have played a pivotal role in disease development, including eye disorders. With respect to eye care, women have been generally cited to have higher rates of blindness in India and are less likely to access appropriate eye services [7][8]. However, as we can see from the study which was focused on Telangana, this is not the case, as male patients exceeded female patients, and this could be for the reason that Telangana has been ranked as one of the top ten innovative and developed states in India according to the India Innovation Index 2019 [15] where access to healthcare is available and appreciated by both male and female.

India has been one of the countries where efforts to strengthen the evidence-base for blindness control has received significant attention from policy planners and program managers. Over the past four decades, a series of population-based blindness and visual impairment surveys have been undertaken in India, using different survey methods. This included detailed eye examination surveys, as well as rapid assessments [8].

B. Occupation and Clinical Visits

Second, to dig deeper into understanding the culture, we studied the occupation of the patients to determine what could be posing a high risk in terms of where someone

spends the day the most. The study showed that home makers, employees in the government and private sectors, and students make the top three categories who visited the clinic the most.

Percentage of Clinical Visits by Occupation

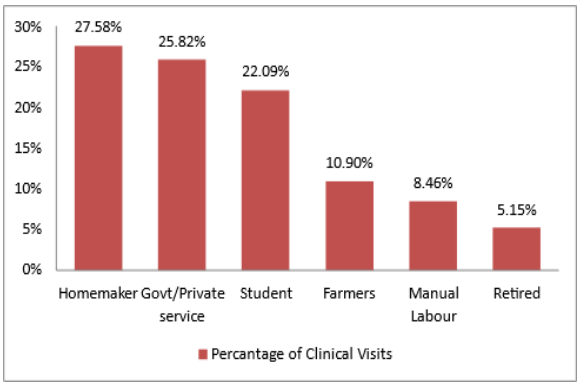

Figure 3. Clinical Visits by Profession (2011-2019)

Figure 3 depicts this analysis and portrays the top six professions taken from the analysis. We can also see that workers in Agriculture and manual laborers tend to present themselves with eye disorders as well, and that could be to the nature of the job, in which they are exposed to certain chemicals, dust, and usually work in heated environments. Recent estimates from the World Health Organization indicate that 90 per cent of all those affected by visual impairment live in the poorest countries of the world [9]. India is home to one-fifth of the world's visually impaired people and therefore, any strategies to combat avoidable blindness must take into account the socio-economic conditions within which people live [9].

Home makers could also translate to housewives, who are at higher-risk of visual disorders, and this is in line with a study that was conducted in 2009 on women in Indian culture, where it showed that housewives are more likely to suffer from heart diseases than working women, and that is due to lack of education, lifestyle that is based on obesityand cultural myths that do not focus on women's health. Having a similar study related to eye disorders and visual impairment, as per the study based on the sample of the population from Telangana, the same pattern can be seen and it can potentially be from these similar reasons [10].

C. Location and Eye Disorders (Older Age 41-70)

Third, we broke down the data to understand the location of the patient, and whether there are underlying reasons in a specific location for a higher rate of risk to developing eye diseases. The study was broken down into examining the location by different age groups.

Number of Visits by Location (Age 41-70)

Location	Number of Visits
Paloncha	5471
kothagudam	2930
kothagudem bazar	2684
Manuguru	2380
Bhadrachalam	2287
Yellandu	1867
Adilabad	1634
Tekulapalli	1432
Burgampahad	1309
Madhapur	1126

Figure 4. Location and Eye Disorders in Older Age Population

Figure 4 shows the locations that people with eye disorders come from and is focused on the older age population. Cataract seemed to be the most disease that has affected older age in Telangana, which is the clouding of the natural human lens. Cataract is a condition known to affect older age, and this study revalidates the information.

D. Location and Eye Disorders (Younger Age 11-20)

The second age group that was studied as per the relation to the location of the patient was the younger population, who we defined from the data as being between the ages 11-20.

Number of Visits by Location (Age 11-20)

Location	Number of Visits
Paloncha	601
Madhapur	325
Adilabad	256
Bhadrachalam	226
kothagudam	177
kothagudem bazar	168
Nagarkurnool	132
Manuguru	128
Gachiowll	127
Kondapur	116
Yellandu	116

Figure 5. Location and Eye Disorders in Younger Age Population

Figure 5 shows the location that people with eye disorders come from and is focused on the younger age population. Astigmatism, which is an irregularity of the shape of the cornea was present in younger age population. Astigmatism has been linked to being a hereditary condition in ophthalmology.

In both contexts, it appeared to be that eye disorders are mostly concentrated in residents from the district of Paloncha, and even though this district has a higher literacy rate than state average is 77%, 10% higher than that of the state average which is at 67%, it has been reported that it has been hit with pollution and contaminated water in 2015. The state-run thermal power plant installed in 2015 caused pollution and health disorders including eye disorders [11]. Residents complained of gray water, and doctors in Paloncha confirmed that the prolonged exposure to air and water pollution has led to higher incidences of respiratory diseases, tuberculosis, skin diseases, blurring of vision and irritation in the eyes, such as Cataract, Cornea, Anterior Segment, Retina, and Glaucoma [11].

E. Consistent Prevalence of Cataract in Rainfall

Fourth, we broke down the data to understand the correlation between the number of visits and rainfall. The sample data was huge, and therefore we limited the study to understand right-eye diseases, as the classifications in Eye Smart were according to right-eye diseases and left-eye diseases separately.

Number of Visits for Right-Eye Diseases by Cumulative Rainfall

Right eye Diseases	Cumulative Rainfall	Number of Visits
Cataract	2.11	4126
Myopia	2.11	8058
Emmetropia	2.10	15253
Pseudophakos	2.12	5837
Presbyopia	2.15	5820
Senile Cataract	2.05	8134
Intraocular Lens	2.03	3146
Hypermetropia	2.13	4164
Simple Myopia	2.12	7686
Astigmatism	2.12	4172

Figure 6. Right Eye Diseases in Relation to Rainfall

Figure 6 shows the diseases that affect the right eye the most when tested alongside rainfall. Globally, cataract is the single most important cause of blindness, and the second most common cause of Moderate and Severe Vision Impairment (MSVI) according to the Global Burden of Disease, Injuries and Risk Factors Study, and it is most predominant in Southeast Asia. Cataract contributed to a worldwide 33.4% of all blindness and 18.4% of all MSVI. Translating the same into actual numbers, cataract caused blindness in 10.8 million of overall 32.4 million blind and visual impairment in 35.1 million of 191 million visually impaired individuals [13].

The close relationship between climate, environment and the development of Cataract is crucial to understand for future preventative measures. In Telangana, it shows that Cataract is the disease most prevalent in rainfall.

F. Consistent Prevalence of Pterygium in Relation to Global Radiation

Fifth, we broke down the data to understand the correlation between right-eye diseases and global radiation. Similar to the previous study of number of visits and rainfall, the sample data was huge, and therefore we limited the study to understand right-eye diseases.

Right-Eye Diseases as Influenced by Global Radiation

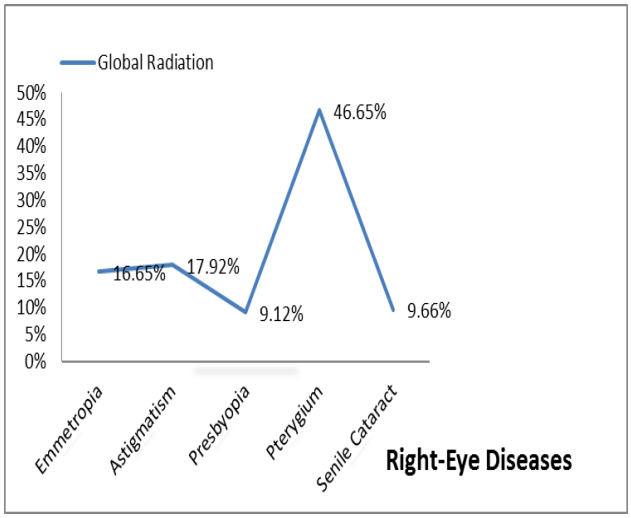

Figure 7. Right Eye Diseases in Relation to Global Radiation

Figure 7 shows the diseases that affect the right eye the most when tested alongside global radiation. We analyzed patients who presented with degeneration symptoms, and correlated the diagnosis to climatic factors, such as humidity, rainfall, temperature and global radiation. The above analysis shows the top 5 most prevalent degeneration right-eye diseases as impacted by global radiation. Pterygium shows to be most prevalent at over 46% of the total global radiation value. The analysis was done on a

patient basis and not a disease basis, as the data showed that one patient can develop more than one disease.

G. Consistent Prevalence of Pterygium in Relation to Windspeed

Finally, we broke down the data to understand the correlation between right-eye diseases and windspeed. Similarly again, the sample data was huge, and therefore we limited the study to understand right-eye diseases.

Right-Eye Diseases as Influenced by Windspeed

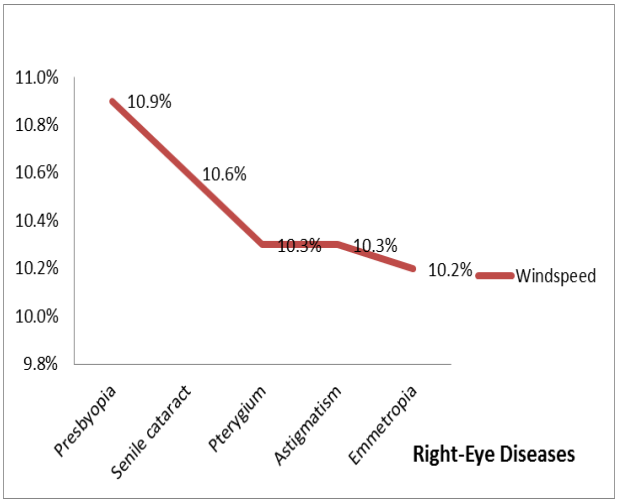

Figure 8. Right Eye Diseases in Relation to Windspeed

Figure 8 shows the diseases that affect the right eye when tested alongside windspeed. The analysis above shows the top 5 most prevalent right-eye diseases with degeneration as a symptom and how the diseases are influenced by maximum windspeed. Presbyopia, Cataract, and Pterygium were the most present among patients and concentrated at average maximum windspeed of between 10.2 and 10.9.

As concluded from the research, the healthcare EMR system is large and complex, one that does not naturally lend itself to easy analysis, design or even understanding. Therefore, the complexity and critical nature of the system beg for the development and use of good, representative models [25]. As per the case study of the research, to analyze the data in Eye Smart, an influence diagram was created to show the different co-factors that lead to cataract according to the findings generated. This was further analyzed using a systems- thinking approach, a method that allows consideration of the whole rather than individual elements of representation of the related co-factors. The influence diagram was useful to showcase the summary of the findings of the research, as it allows for us to see the connection of the variables in a picture format. This format of summary is useful for healthcare practitioners, and in the case of our study, for ophthalmologists, and decision-makers, in the field of eye-care management.

Influence Diagram - Taking Knowledge to Wisdom

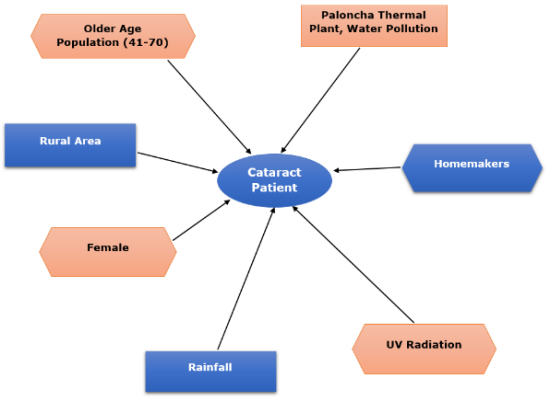

Figure 9. Influence Diagram of Cataract Patients and Co-Factors Affecting Cataract

Figure 9 depicts the influence diagram of the main study. Influence diagrams are closely related to decision trees and often used in conjunction with them. An influence diagram displays a summary of the information contained in a decision tree. It involves four variable types for notation: a decision (a rectangle), chance (an oval), objective (a hexagon), and function (a rounded rectangle). Influence diagrams also use solid lines to denote influence. In the case of the influence diagram generated below, this has been equivalent to the definition of creating “wisdom” in the DIKW pyramid explained earlier in previous chapters. Looking at the data analysis holistically creates the ease to depict the main causes of the development of cataracts. Figure 9 is a drawing of the influence diagram that explains the connection of the weather factors in relation to the development of the most right-eye disease found in the sample population of the study, which is cataract.

V. Conclusion and Future Work

This data analytics study provides an expanded exploration of how socio-demographic and climatic factors affect the prevalence of visual impairment and eye disorders in Telangana. Applying several statistical techniques, including pattern recognition, and generating other data visualizations, we were able to validate previously identified findings about gender’s relation to eye disorders in Telangana. We found the tools we used to be very useful for a discovery research to better understand the sample set of patients and to generate informative and understandable visuals.

Big data can serve to boost the applicability of clinical research studies into real-world scenarios, where population, race, and climate create a challenge. It equally provides the

opportunity to enable effective and precision medicine by performing patient stratification. This is indeed a key task toward personalized healthcare. A better use of medical resources by means of personalization can lead to well-managed health services that can overcome the challenges of a diverse population where poverty is high. Thus, creative featuring and data merging for health management of EMRs can have an impact on future clinical research.

From a systems perspective, we observe that a patient is influenced by several co-factors that result in the development of eye disorders, and that is significant in studying patient care from a holistic standpoint. AI tools create the pathway to merging publicly available data and aligning multiple variables as part of the overall influence. This technique is widely applied in decision-making and outcome assessment for an enhanced healthcare experience, in which modeling knowledge and expert experience are studied more thoroughly for new pattern recognition. However, variables must be minimized in order to capture the underlying knowledge, or otherwise patterns will be harder to spot. Thus, we attempt to apply this in the future with less variables to overcome the challenges in the first phase or data merging.

We recommend that the authorities spend more time and funds on creating awareness to educate individuals and families about the visual impairment crisis in Telangana. Creation of awareness is one of the most comprehensive approaches to sensitize communities concerning the consequences of eye disorders, but also one of the avenues to equip individuals with knowledge, skills, and correct attitudes towards a healthier lifestyle.

Besides creation of awareness, this study also recommends ophthalmologists' understanding of all factors that influence a disease other than medical history, and to look at each patient uniquely in terms of social income, cultural upbringing, and offer a more individualistic approach in educating a patient from the criticality of self-care, to help patients deviate away from high risk situations that can cause eye disorders, and to find ways from an earlier age for more effective preventative results that can reduce the number of affected individuals with vision impairment in Telangana.

Visual impairment has continually exhibited an escalating trend in underdeveloped countries over the past years, and in India, the burden of visual impairment is high in urban and rural areas. Eye-care services should be accessible and affordable to individuals in need. Future work will be on discovery research on how socio-demographic and climatic factors correlate to the development of other diseases. It is important to mention that hardly any investigation had been conducted with the application of AI tools to categorize huge data based on demographic and weather variables. Therefore, this study adds new insights to the field of ophthalmology.

REFERENCES

[1] A. Alalawi, A. Das, R. Derman, P. Gumpili, P. Lalakia, L. Sztandera, "A Data-Driven Approach for Eye Disease Classification in Relation to Demographic and Weather Factors Using Computational Intelligence Software," The Ninth International Conference on Data Analytics, IARIA, Proceedings of 2020. pp. 21-27.

[2] G. Rao, N. Khanna, and A. Payal, "The global burden of cataract," *Current Opinion in Ophthalmology*, *22* no. 1, pp. 4-9, 2011.

[3] A. Das, P. Kammari, and S. Ranganath Vadapalli, "Big data and the eyeSmart electronic medical record system-An 8-year experience from a three-tier eye care network in India," *Indian Journal of Ophthalmology*, *68* no. 3, pp. 427, 2020.

[4] S. Marmamula, R. Khanna, and G. Rao, "Unilateral visual impairment in rural south India–Andhra Pradesh Eye Disease Study (APEDS)", *International Journal of Ophthalmology*, *9* no.5, pp.763, 2016.

[5] M. V. Boland, "Big data, big challenges," Ophthalmology, *123* no.1, pp. 7-8, 2016.

[6] A. Clark, J. Ng, N. Morlet, and J. Semmens, "Big data and ophthalmic research," *Survey of Ophthalmology*, *61* no.4, pp. 443-465, 2016.

[7] E.M. Messmer, "The pathophysiology, diagnosis, and treatment of dry eye disease," *Deutsches Ärzteblatt International*, *112* no.5, pp. 71, 2015.

[8] D. Matthews, "How gender influences health inequalities," *Nursing Times*, *111* no.43, pp. 21-23, 2015.

[9] N. Diaz-Granados, et al. "Monitoring gender equity in health using gender-sensitive indicators: a cross-national study," *Journal of Women's Health*, *20* no.1 pp. 145-153, 2011.

[10] G.V.S Murthy, S.K. Gupta, D. Bachani, R. Jose, and N. John, "Current estimates of blindness in India," *British Journal of Ophthalmology*, *89* no.3, pp.257-260, 2005.

[11] S.K Angra, G.V. Murthy, S.K. Gupta, and V. Angra, "Cataract related blindness in India and its social implications," *The Indian Journal of Medical Research*, *106*, pp. 312-324, 1997.

[12] A. Janati, H. Matlabi, H. Allahverdipour, M. Gholizadeh, and L. Abdollahi, "Socioeconomic status and coronary heart disease," *Health Promotion Perspectives*, *1* no.2, pp. 105, 2011.

[13] S. Kumar, "National security environment," *India's National Security*, pp. 25-236, Routledge India, 2016.

[14] S.G. Honavar, "Eliminating cataract blindness: Are we on target?" *Indian Journal of Ophthalmology*, *65* no.12, pp. 1271, 2017.

[15] G. Bakshy https://www.jagranjosh.com/current-affairs/india-innovation-index-2019-karnataka-tops-the-list-followed-by-tamil-nadu-and-maharashtra-1571375106-1, 2019, [Online]. (Accessed on 07.09.2020).

[16] C.P. Ames, et al. "Artificial intelligence based hierarchical clustering of patient types and intervention categories in adult spinal deformity surgery: towards a new classification scheme that predicts quality and value," *Spine*, *44* no.13, pp. 915-926, 2019.

[17] F. Bertini, G. Bergami, D. Montesi, G. Veronese, G. Marchesini, and P. Pandolfi, "Predicting frailty condition in elderly using multidimensional socioclinical databases," *Proceedings of the IEEE*, *106*(4), 723-737, 2018.

[18] H. Chen, J. Kwong, R. Copes, P.J. Villeneuve, A. Van Donkelaar, and A.S. Wilton, "Living near major roads and the incidence of dementia, Parkinson's disease, and multiple sclerosis: a population-based cohort study," *The Lancet*, *389*(10070), pp. 718-726, 2017.

[19] E.P. Barracchia, G. Pio, D. D'Elia, and M. Ceci, "Prediction of new associations between ncRNAs and diseases exploiting multi-type hierarchical clustering," *BMC bioinformatics*, *21* no.1, pp. 1-24, 2020.

[20] B. Familiar and J. Barnes, "Data Visualizations, Alerts, and Notifications with Power BI. In Business in Real-Time Using Azure IoT and Cortana Intelligence Suite," pp. 397-473, *Apress*, Berkeley, CA, 2017.

[21] M. Lutz, "Programming Python: powerful object-oriented programming," *O'Reilly* Media, 2010.
[22] G. Marcus and E. Davis, "Rebooting AI: Building artificial intelligence we can trust," *Pantheon*, 2019.
[23] Russell L. Ackoff, "From data to wisdom," Journal of applied systems analysis, 16.1. pp. 3-9, 1989.
[24] A.P. Brown, "Qualitative method in applied research," 10 (2), pp. 229, 2010.
[25] M. Longman, London, Methods in Sociology, 1986.
[26] K. Patrick, and M. J. Schwandt. "Health systems: A dynamic system-benefits from system dynamics," Proceedings of the Winter Simulation Conference, 2005, IEEE.

.