Original Article

Leveraging big data for pattern recognition of socio-demographic and climatic factors in correlation with eye disorders in Telangana State, India

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

Purpose: Big data is the new gold, especially in health care. Advances in collecting and processing electronic medical records (EMR) coupled with increasing computer capabilities have resulted in an increased interest in the use of big data in health care. Ophthalmology has been an area of focus where results have shown to be promising. The objective of this study was to determine whether the EMR at a multi-tier ophthalmology network 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. **Methods:** The study was designed by merging a dataset obtained from the Telangana State Development Society to an existing EMR of approximately 1 million patients, who presented themselves with different eye symptoms and diagnosed with several diseases from the years (2011–2019). The dataset obtained included weather and climatic variables to be tested alongside eye disorders. AI creative featuring techniques have been used to narrow down the variables most affected by climatic and demographic factors, with the application of the Cynefin Framework as a guide to simplify and structure the dataset for analysis. **Results:** Our findings revealed a high presence of cataract in the state of Telangana, mostly in rural areas and throughout the different weather seasons in India. Males tend to be the most affected as per the number of visits to the clinic, while home makers make the most visit to the hospital, in addition to employees, students, and laborers. While cataract is most dominant in the older age population, diseases such as astigmatism, conjunctivitis, and emmetropia, are more present in the younger age population. **Conclusion:** The study appeared useful for taking preventive measures in the future to manage the treatment of patients who present themselves with eye disorders in Telangana. The use of clinical big datasets helps to identify the burden of ocular disorders in the population. The overlaying of meteorological data on the clinical presentation of patients from a geographic region lends insight into the complex interaction of environmental factors on the prevalence of ocular disorders in them.

Key words: Artificial intelligence, big data, ocular disease, ophthalmology

India is home to one of the most visually impaired countries in the world, and the state of Telangana ranks as one of the highest states that deals with visual impairment and blindness.[1] 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% by 2020, the prevalence of blindness still stands at 1.99%, according to the National Blindness and Visual Impairment Survey, released in October 2019[1] by the Union Ministry of Health and Family. The prevalence of blindness and visual impairment is one of the highest in Telangana, a state in Southern India, as inferred from the survey.[1] The significant reasons indicated in the survey were due to cataract and refractive error.[2]

This study uses Artificial Intelligence (AI) and machine learning techniques to explore a dataset containing information on 873,448 patients who visited a multi-tier ophthalmology hospital network based in India. The data was extracted from EyeSmart, the hospital's Electronic Medical Record (EMR) and health management system, and then merged with climatic factors to test the correlation between climatic variables and ocular diseases presented by the patients.

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 promising as it offers the benefit of transforming large volumes of clinical data into a more structured data for quality improvement

School of Professional and Continuing Studies, DMgmt in Strategic Leadership Program, Thomas Jefferson University, Philadelphia, PA, USA, [1]Kanbar College of Design, Engineering, and Commerce, Thomas Jefferson University, Philadelphia, PA, USA, [2]Office of Global Affairs and Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, USA, [3]Department of eyeSmart EMR and AEye, L. V. Prasad Eye Institute, Hyderabad, Telangana, India, [4]Department of Obstetrics and Gynaecology, Thomas Jefferson University, Philadelphia, PA, USA

Correspondence to: Dr. Anthony Vipin Das, Associate Director and Consultant Ophthalmologist, L. V. Prasad Eye Institute, Hyderabad, Telangana - 500 034, India. E-mail: vipin@lvpei.org

to help answer 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.[3]

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. 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.[4] This has shed light on recent research in medicine and ophthalmology.

In Section 2 of this paper, we discuss and highlight different statistical tools and methods used to investigate the study of several demographic and climatic factors that impact individuals in Telangana. While Section 3 focuses on providing a thorough analysis of the data findings, Section 4 highlights key findings and trends, and Section 5 includes 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. The study also intends to discover how socio-demographic and climatic factors correlate to the number of individuals affected using Artificial Intelligence tools.

Methods

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. Python was used to merge the datasets using Pandas library through a column mutation process. It was an effective tool as it handled the large and complex dataset, however, the process was 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.

The dataset, which covers clinical visits are from the year 2011-2019, includes demographic information of the patient, including age, gender, profession, date of visit, district of resident, among other variables. To look further into this issue, we searched for other variables that may have a relationship with eye disorders and focused on climate.

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.[5] The findings that relate to temperature and its effect on cataract in older age was consistent in high and low temperatures.

In addition, this research was undertaken to gain insight into how data merging techniques can transform a complex data set into a simpler format for analysis to gain new insight about the development of cataract in patients in Telangana, and hence we used the Cynefin framework as a guide to achieving a more structured dataset. "The Cynefin framework helps leaders determine the prevailing operative context so that they can make appropriate choices".[6] Thus, in healthcare, it can offer practitioners and researchers more structured environments by managing information to help in decision making.[6] The Cynefin Framework [Fig. 1] was developed with an aim to offer clarity within complex projects to support the right direction in the process of decision making for situations where the existing intricacy within the outcomes affect the nature of knowledge, forecast, and choice.[7] In the case of the research being conducted in this paper, we use a complex dataset as an example of a complex project where information is hard to read and make sense of given the volume of information in the EMR system. Organizing the information in the dataset using AI tools can guide researchers or healthcare practitioners navigate through the unordered information and simplify it to read patterns that relate to the cause and effect of eye diseases. Even though the Cynefin Framework was "developed by Snowden for business problems but applies well to medicine".[8] It has varied domains which necessitate different actions, for instance, the straight forward and complex context is considered equivalent to an "ordered state" of the universe which can be interpreted based on the causal and effect association of the facts or findings, and therefore the right direction or pattern can be decoded.[8] However, in the case of "complicated or chaotic" data, where researchers or healthcare practitioners are unable to formulate a definitive cause and effect association, there is no such immediate conceivable relationship, thus, the Cynefin framework guides professionals to choose the right direction based on the "emerging patterns".[8] This means that the chaotic or unordered state of the world requires pattern-dependent management for an effective decision-making in the practice of healthcare.[8]

According to scientific evidence the parts or the components of the complex system may not show direct association in a linear pattern. These factors or parameters are openly prevailing within the environment due to which the interactions between


Figure 1: Cynefin framework and its application in medical complexities

those factors can occur at varying levels via recursive feedback loops.[7]

The above Fig. 2 refers to the cause and effect association between the parameters decoded with the aid of Cynefin Framework. In the figure, the square block indicates the "causal agent", and the dots indicate the "effect agent". The proper lines refer to the direct association in between the two agents whereas the dotted lines refer to the weak or probable association in between the two agents. When the relation or the conduct of the components of the complex adaptive systems cannot be perceived in direct terms, it is said to be emergent and active in nature. Moreover, these factors again show alterations with time and pressure of the surrounding environment due to which it develops into a new form.[9]

The concept of medicine is complex and the complexity that arises at the time of general practice varies widely.[9] The approach of general practice formulates expertise based on the overall profile of the patients. In this regard, the theory of complexity demonstrates that these associations between varied confounding factors are complementary in nature.[9] The clinical specialist has to deal with the complex domain frequently during the practice. It is evident that detailed analysis using statistical or analytical tools (AI) along with investigation and specialized knowledge of the specialist helps to provide excellent solutions to complicated problems, for instance, cataract problems and factors that lead to the development of cataracts.[8] The clinical specialist often finds problems within the disordered or chaotic zone as the data do not show any direct or known association between the varied agents, however, with the help of taking the Cynefin Framework as a tool to convert complex data to simpler data, practitioners or researchers can recognize the probable underlying cause which would be of immense help in the field of healthcare.
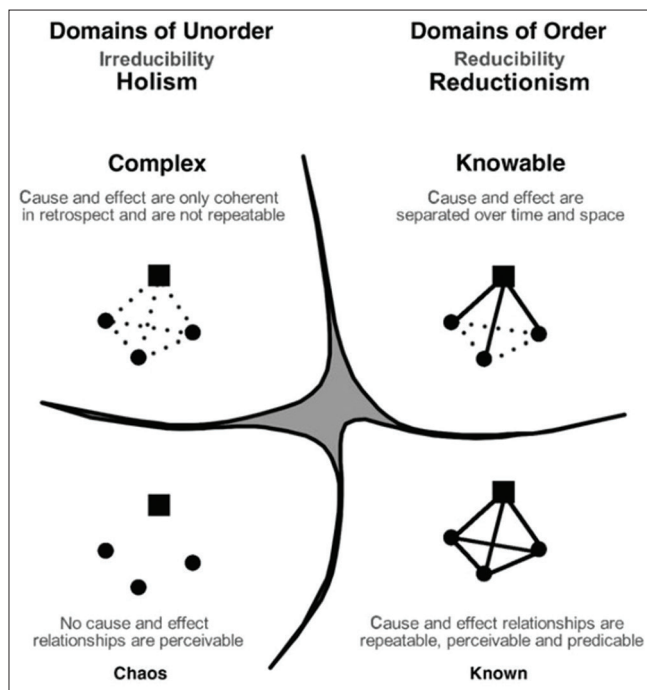
Snowden, the developer of the Cynefin Framework, highlighted that the concept of "best practice" is attributable to identifiable or known problems; for the complicated situations the "good practice" is advisable and for the complex problems the "emergent practice" is considered to be the most suitable.[6] Therefore, it is scientifically evident that with the use of Cynefin Framework, the complex or chaotic data which the research started with can be successfully converted to simpler data to find an emergent association between the co-factors that cause the development of cataract, referred to as "wisdom" in the DIKW pyramid, and ultimately will bring awareness among the community.[6]

Results

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 economic variables examined. 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.

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 to minimize the number of variables have been helpful throughout the analysis as it can shed light on which patterns are more significant.

Fig. 3 indicates that between the years 2011–2019, 53% of the patients were male patients who were seen for eye disorders,

**Figure 2:** Cause and effect association of Cynefin framework in medical complexities
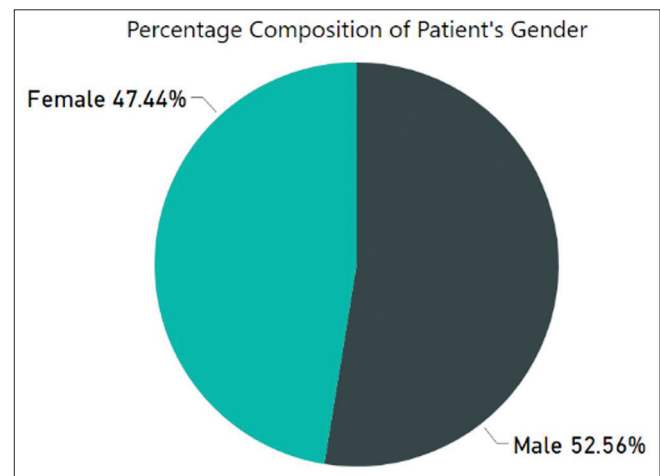
Figure 3: Gender and eye disorders. Clinical visits by gender (2011-2019)

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 the multi-tier ophthalmology network from the years 2011–2019.[4] Globally, one of the social determinants of health that has been universally identified is gender.[10] 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 for having higher rates of blindness in India and being less likely to access appropriate eye services[11,12] 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 where access to healthcare is available and appreciated by both male and female.[13]

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.[10]

In addition, when studying the correlation between profession and clinical visits, it appeared that homemakers, employees in the government and private sectors, and students make the top three categories of those who are most affected. Fig. 4 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.[10] 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.[10]

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
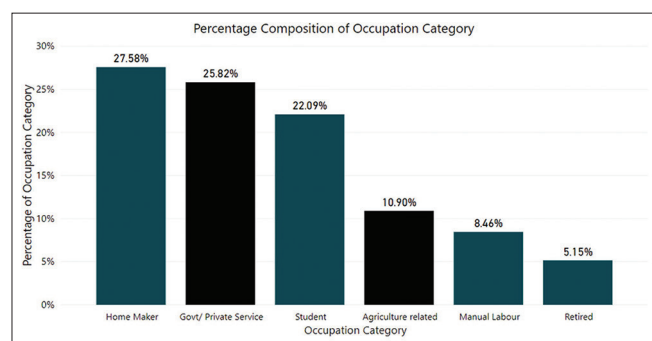
heart diseases than working women, and that is due to lack of education, lifestyle that is based on obesity and cultural myths that do not focus on women's health. Having a similar study related to eye disorders and visual impairment, as per the study based on the sample of the population from Telangana, the same pattern can be seen and it can potentially be from these similar reasons.[14]

Cataract seems to be the leading ocular 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 [Fig. 5].

Astigmatism, which is an irregularity of the shape of the cornea was present in the younger age population. Astigmatism has been linked to being a hereditary condition in ophthalmology [Fig. 6].

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.[15] 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.[15]

Cataract appeared to be the most disease prevalent in regions with higher rainfall. 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 regions with higher rainfall [Fig. 7].

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 pervasive at over 46% of the total global radiation value. [Fig. 8] 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.

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. Pterygium also was also the most present among patients and concentrated at average maximum windspeed of between 10.2 and 10.9. [Fig. 9]

Identifying cataract patients came to 102,509 patients from the years (2011-2019). Therefore, the sample study of patients is based on 102, 509 patients who presented themselves to LVPEI during that timeframe from the state of Telangana. The Fig. 10 above shows the gender composition of cataract patients. There is a total of 48,219 (47.04%) who are male, while 54,290 (52.96%) are female.

The above Fig. 11 shows the breakdown of the number of cataract patients across the districts – urban, rural, and metropolitan. It is clear that cataract is most prevalent in rural

Figure 4: Occupation and clinical visits. Clinical visits by profession (2011-2019)

areas with 58,123 cases of cataract recorded. This is followed by metropolitan with 27,150 cases of cataract recorded while the urban district has 17,236 cases of cataract recorded. Eye diseases worsen the quality of life and satisfaction of an individual.[16] It has been observed that the spatial location,

the cultural and financial condition of the individuals and most importantly the access rate to healthcare organization are considered to be the perceived barriers. It is scientifically evident that cataract is considered to be one of the potential causes behind the visual impairment or blindness mostly
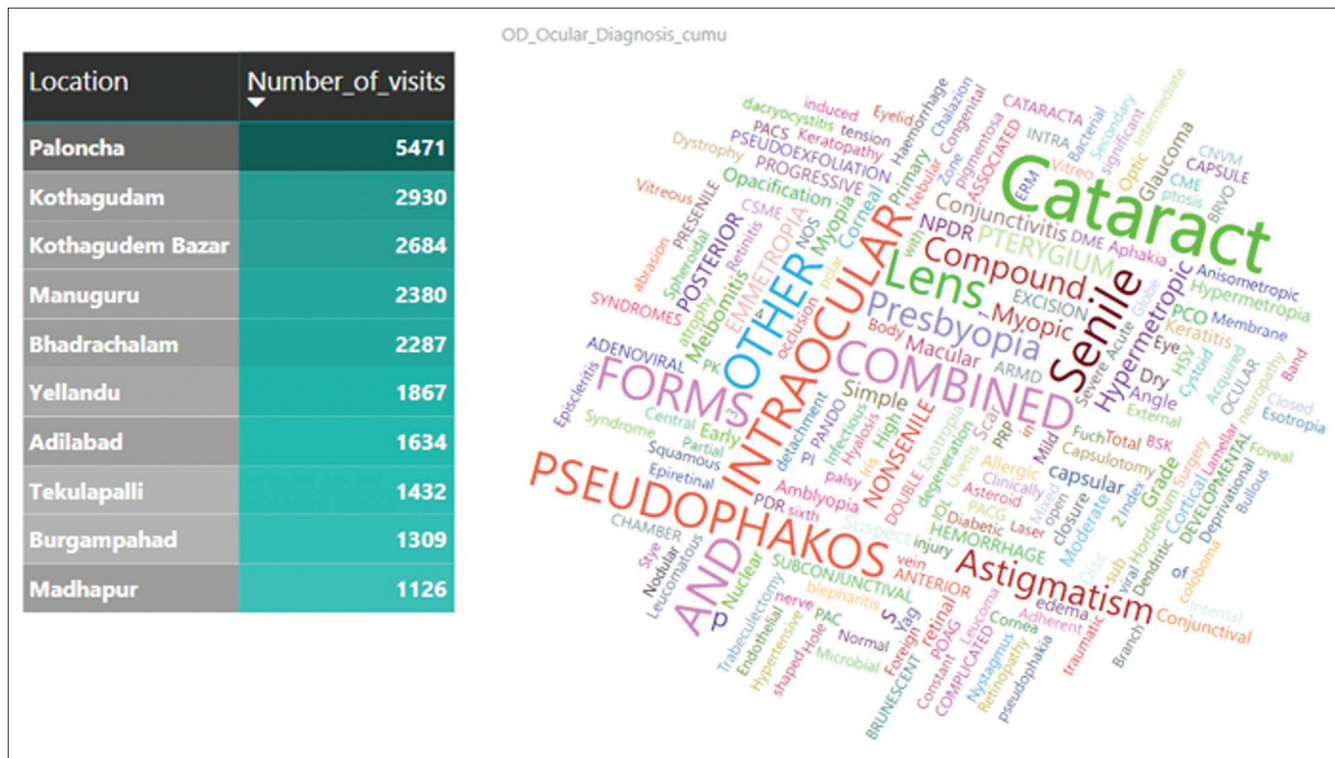

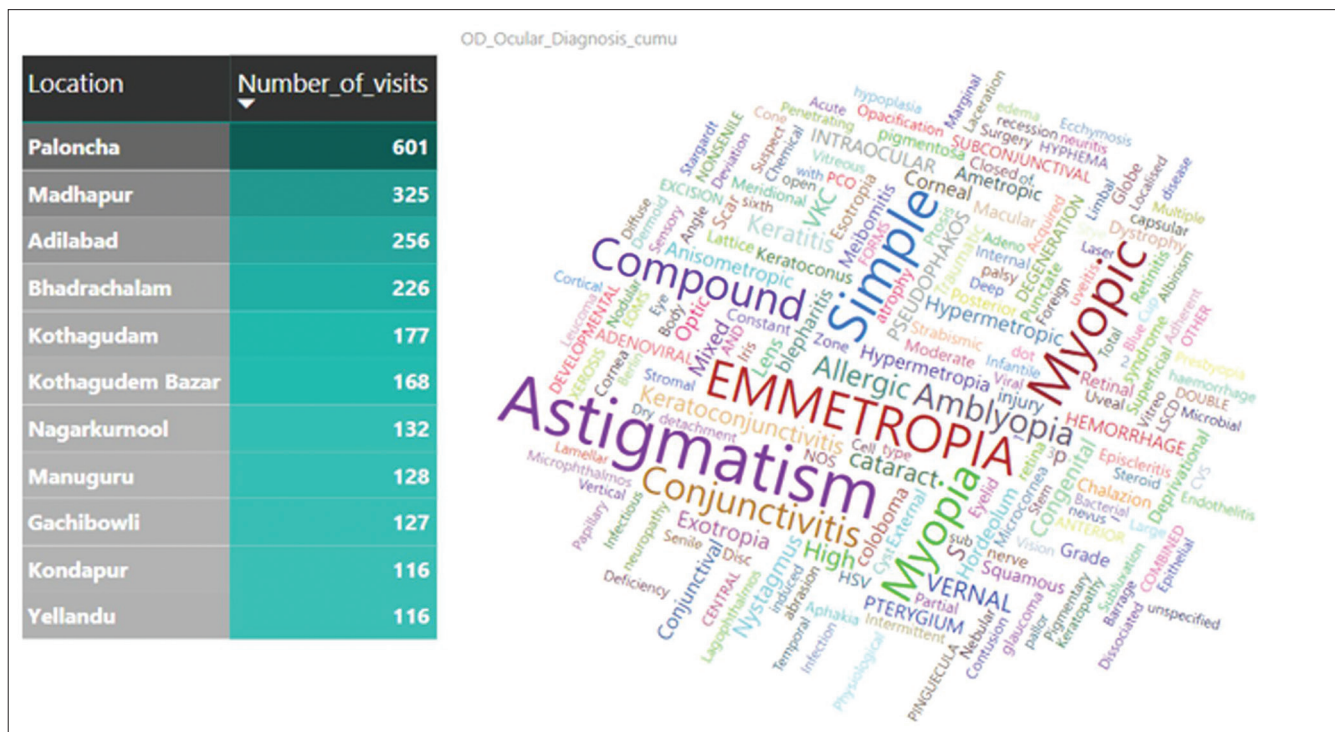**Figure 5:** Location and eye disorders in older age (41-70)

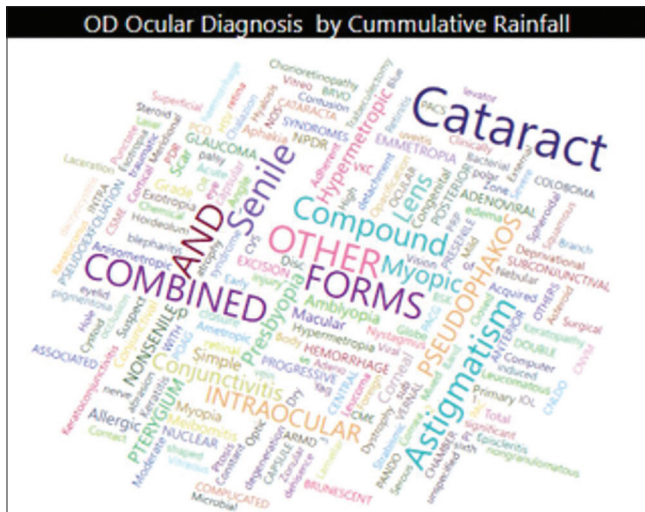

Figure 6: Location and eye disorders in younger age (11-20)

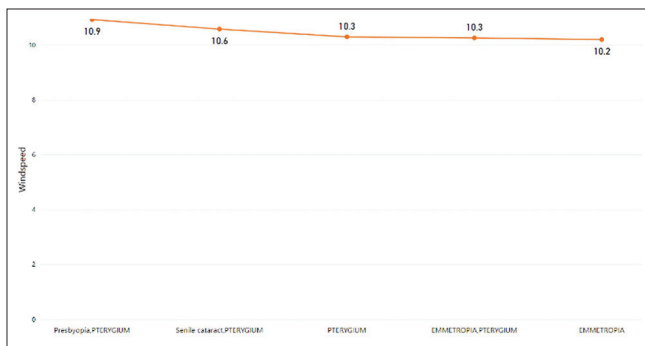**Figure 7:** Cataract most prevalent in regions with higher rainfall

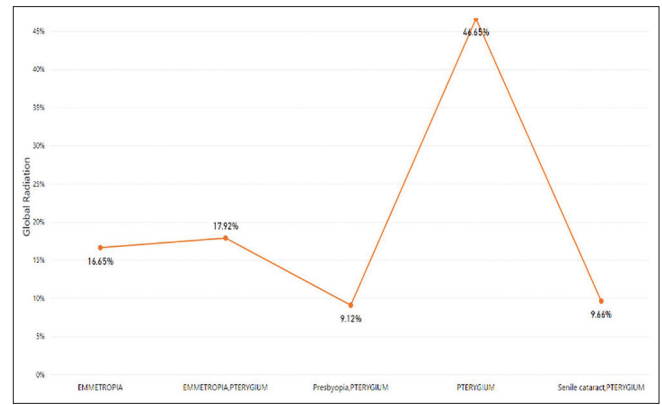**Figure 8:** Consistent prevalence of pterygium in relation to global radiation

Figure 9: Consistent prevalence of pterygium in relation to windspeed
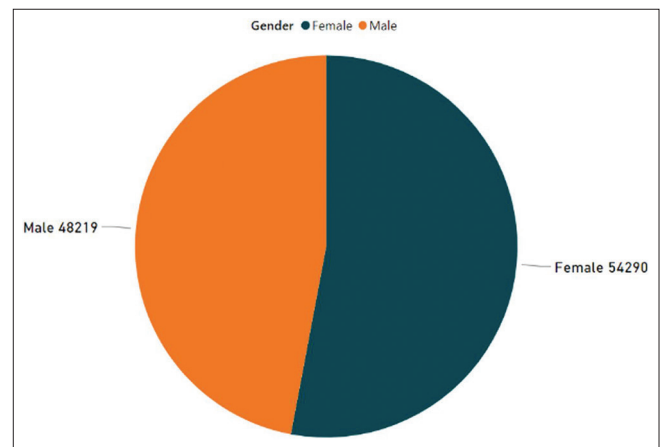

Figure 10: Breakdown of cataract patients by gender

among the middle and poor resourced families.[17] The varied reasons behind these variations might be attributed to the spatial location which creates troubles to access the distant health care centers especially for female patients.[17] Moreover, the awareness about eye diseases and the financial crisis plays the most significant roles for the worsening condition of cataract among the rural populace. The studies conducted by both Joshi and Murthy *et al.* have supported this opinion and highlighted the expanded charges of the modernized strategies involved for cataracts such as intraocular lens implantation[16,17]

Discussion

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 of gender's relation to eye disorders in Telangana. We found the tools we used to be very useful for 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 fewer variables to overcome the challenges in the first phase of data merging.

We recommend that the authorities spend more time and funds on creating awareness to educate individuals and families
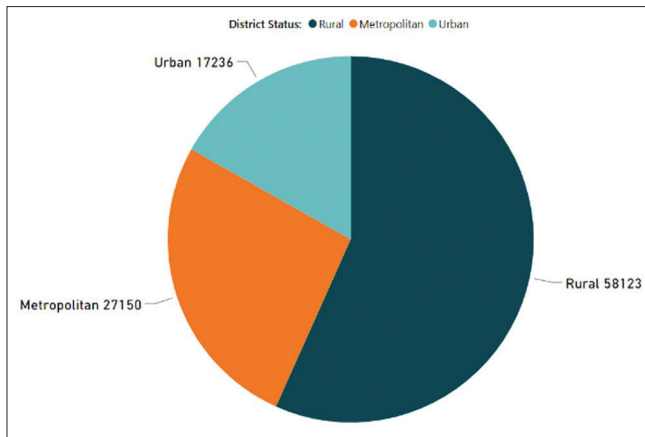
Figure 11: Breakdown of cataract patients by district

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 the 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.

Conclusion

In conclusion, this study aimed to profile the ocular disorders in a large cohort of patients presenting from the state of Telangana in India. Cataract accounted for the majority of the patients who accessed eye care services and were more commonly males. The use of big datasets derived from electronic medical records helps to identify the burden of ocular diseases in the population. The overlaying of meteorological data on the clinical profile helps to understand the complex interaction of the environmental factors on the prevalence of ocular disorders and can assist in a more holistic approach in designing prevention and treatment strategies.

Conflicts of interest

There are no conflicts of interest.

References

1. Banjot Kaur, 2019. Available from: https://www.downtoearth.org.in/news/health/cataract-top-cause-of-blindness-in-india-finds-survey-67187. [Last accessed on 2020 Oct 15].

2. Express News Service, October 2019. Prevalence of Blindness, Visual Impairment High in Telangana. https://www.newindianexpress.com/cities/hyderabad/2019/oct/11/prevalence-of-blindness-visual-impairment-high-in-telangana-2045886.html#:~:text=From%20Telangana%2C%20the%20erstwhile%20Warangal,visual%20impairment%20was%2020.31%20pc. [Last accessed on 2020 Oct 20].

3. Boland MV. Big data, big challenges. Ophthalmology 2016;123:7-8.

4. Das AV, Kammari P, Vadapalli R, Basu S. Big data and the eyeSmart electronic medical record system-An 8-year experience from a three-tier eye care network in India. Indian J Ophthalmol 2020;68:427-32.

5. Telangana State Development Planning Society. [online] Available from: https://tsdps.telangana.gov.in/. [Last accessed on 2020 Aug 11].

6. Snowden DJ, Boone ME. A leader's framework for decision making. Harv Bus Rev 2007;85:68-76, 149.

7. Kempermann G. Cynefin as reference framework to facilitate insight and decision-making in complex contexts of biomedical research. Front Neurosci 2017;11:634.

8. Gray B. The Cynefin framework: Applying an understanding of complexity to medicine. J Prim Health Care 2017;9:258-61.

9. Sturmberg JP, Martin CM. Complexity and health–yesterday's traditions, tomorrow's future. J Eval Clin Pract 2009;15:543-8.

10. Mathews D. How gender influences health inequalities. Nurs Times 2015;111:21–3.

11. Kaur K. Cataract blindness: Socioeconomic factors associated with treatment barriers and high blindness rates for women in rural regions of Andhra Pradesh, 2018. Available from: https://scholarscompass.vcu.edu/cgi/viewcontent.cgi?article=1061&context=auctus. [Last accessed on 2020 Oct 20].

12. Khanna R, Raman U, Rao GN. Blindness and poverty in India: The way forward. Clin Exp Optom 2007;90:406-14.

13. Gorka B.2019. Available from: https://www.jagranjosh.com/current-affairs/india-innovation-index-2019-karnataka-tops-the-list-followed-by-tamil-nadu-and-maharashtra-1571375106-1 [Last accessed on 2020 Oct 20].

14. Marmamula S, Khanna RC, Rao GN. Unilateral visual impairment in rural south India-Andhra Pradesh Eye Disease Study (APEDS). Int J Ophthalmol 2016;9:763-7.

15. Suchitra M. Stream of Ash. Down to Earth,2015. Available from: https://www.downtoearth.org.in/coverage/environment/stream-of-ash-44036. [Last accessed on 2020 Oct 20].

16. Joshi MV. Epidemiological study of patients availing free cataract services of national programme of control of blindness. J Clin Ophthalmol Res 2015;3:9-13.

17. Vs Murthy G, Jain B, Shamanna B, Subramanyam D. Improving cataract services in the Indian context. Community Eye Health 2014;27:4-5.