Data Merging Technique in Cataract Patients in Telangana for Enhancing Public Awareness of Visual Impairment

Amna Alalawi
Strategic Leadership Program
Thomas Jefferson University
Philadelphia, PA, USA
Email: alalawiaj@gmail.com

Les Sztandera
Kanbar College of Design, Engineering, and Commerce
Thomas Jefferson University
Philadelphia, PA, USA
Email: les.sztandera@jefferson.edu

***Abstract*—Data merging is a creative technique used in big data analysis and is considered a model in strategic leadership thinking, one which is used in integrating and applying strategy to resolve complex problems. In addition, data merging can draw a blueprint for effective decisions, especially when harnessing data in healthcare to clarify ambiguity on complex issues. The objective of this study was to examine the data merging technique using the Electronic Medical Record (EMR) in LV Prasad Eye Institute (LVPEI) in India, for the aim of contributing to the management of patient care, and ultimately spreading public awareness of cataracts. Our findings revealed that there is a high presence of cataract in the state of Telangana, mostly in rural areas and throughout the different weather seasons in India. Men tend to be the most affected, while home makers make the most visits to the hospital, in addition to employees, students, and laborers. While cataract is most dominant in the older age population, diseases such as astigmatism and conjunctivitis are more present in the younger age population. 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.**

Keywords –data merging; visual impairment; data analysis; public awareness.

I. INTRODUCTION

India is home to over 8.3 million people with Vision Impairment (VI), the highest in the world [1]. Even though, in 1976, India became the first country in the world to start a national program for control of blindness with 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 [2] 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 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, which is a clouding of the lens – turning the lens from clear to yellow, brown, or even milky white, 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. LVPEI operates out of 106 locations, 86 of them being primary eye care centers located in remote rural villages [10]. For the past 24 years, it has served over 14 million people, over 50 percent of them entirely free of cost, irrespective of the complexity of care needed. To date, LVPEI has trained over 13,000 eye care professionals; its faculty has been awarded 22 PhDs with over 1,000 research paper publications, its sight enhancement and visual rehabilitation services served over 100,000 people, and its eye bank services have harvested about 34,000 donor corneas, and it has transplanted more than 17,000 of them to needy patients [10].

The data used in this study was extracted from EyeSmart, the hospital's 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 [1]. 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].

The goal of an EMR in general is to enable electronic documentation of patients for faster retrieval and research purposes, as well as to transform the entire network into a paperless eco-friendly environment. LVPEI states that Eye Smart is an effective EMR that is enabled for viewing on various digital platforms, such as iPads, iPhones and other tablets. It has also evolved into an effective educational tool

 ISBN: 978-1-61208-891-4

for students and fellows who train at the institute. The standard procedures, classifications, evidence-based medicine protocols integrated into the system help to deliver more effective care, and to also aid in teaching.

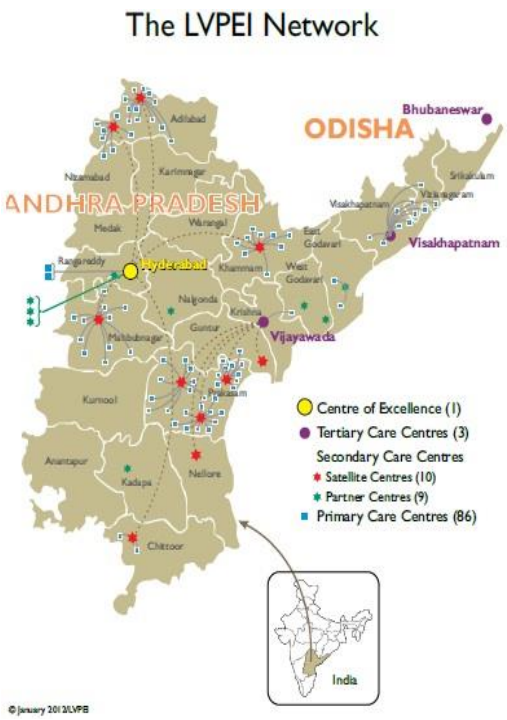

Figure 1. LVPEI Area Map

Figure 1 shows the different centers of LVPEI in the state of Telangana and the total area that it covers in offering eye care to all patients. The map is important to view as the different districts of where patients come from have been examined alongside cataract development.

In Section II of the paper, we explain the data merging technique that was used to merge weather variables with datasets from Eye Smart, to enhance the findings that relate to the development of cataracts in patients, and to potentially offer preventative measures of the development of the disease. The aim is to use knowledge management to gain insight into socio-demographic and environmental factors to shed light on the causes, diagnosis, and treatment of cataracts to enhance ophthalmology practice. In this case, the knowledge management technique consists of applying computational intelligence software to patient data and environmental factors such as race, culture, and climate.

The need for an interconnected health network has reached its peak. Using electronic health records dramatically increases the quality of care for patients and the efficiency of the health care systems. Looking at electronic health systems independently can show limited information about patients. The methodology, which was based on a design thinking approach, offers ways in which EMRs can be studied collectively and holistically to bridge the gap that currently exists between knowledge and practice, and to enhance and improve public awareness of visual impairment.

The rest of the paper is organized into discussing methodology of the research in Section 2, analysis and findings of the data trends in relation to the development of cataracts in Sections 3 and 4, and conclusion including recommendations for preventative measures in Section 5.

II. METHODOLOGY

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.

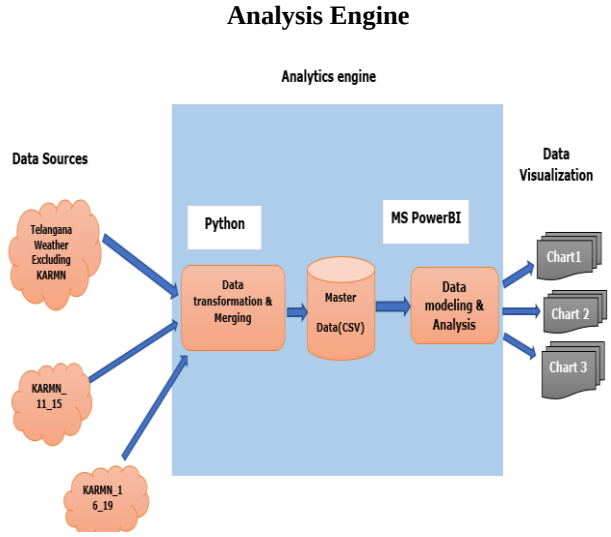

Figure 2. Data Analytics Engine

The purpose of Figure 2 is to explain the process of the data analytics engine that was applied in the research, and that achieved enhanced visibility of patterns and findings. The environment for any data analytics application creation should provide for the following: storing data, processing data, and data analysis alongside visualization. To this end,

 ISBN: 978-1-61208-891-4

the process contributed to knowledge expansion of the causes that develop cataract in patients and showed the relationships between all the co-factors studied. Researchers and practitioners in the big analytics sector can use this diagram to understand how data merging can be processed and can lead to potential strategies for greater impact of result findings. The main advantage to data merging is that, in addition to storing big data, it provides a process through which large pools of data can lead to intelligent insight and perform informed decisions based on variables that are merged alongside each other. The outcome led to decisions that can drive recommendations, growth, planning and prevention, which can be defined as the "wisdom" that is created.

Trends in the IT industry have been transformational in the way data is collected and analyzed through AI techniques. There was an era when people were moving from manual computation to automated, computerized applications, then moved into an era of enterprise level applications, which ultimately gave birth to architectural models such as Software as a Service (SaaS) and Platform as a Service (PaaS). Now, we are in the big data era, which can be processed and analyzed in cost-effective ways [5], such as through Microsoft Power BI. The world is moving towards open source to get the benefits of reduced license fees, data storage, and computation costs. It has really made it lucrative and affordable for all sectors and segments to harness the power of data. This is making Big Data synonymous with low cost, scalable, highly available, and reliable solutions that can churn huge amounts of data at incredible speed and generate intelligent insights.

III. ANALYSIS

The scheme of Cynefin Framework acts as a guidance to healthcare practitioners and researchers because of its foundation in the management of information [6]. This particular tool was developed with an aim to offer support and 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 [6]. It has varied domains which necessitate different actions, for instance, the straightforward and complex context is considered equivalent to an "ordered state" of universe which can be interpreted based on the causal and effect association of the facts or findings, and therefore the right orientation or pattern can be decoded [7]. 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 orientation based on the "emerging patterns" [7]. This means that the chaotic or unordered state of the world requires pattern dependent management for proper orientation and right decision making [7].

This framework proved very useful in this study as it served as a guide to simplify the complex data in a format that could be studied and explored for suggesting reasons that are associated with the development of cataracts, with the hope of taking preventative measures in the future in patient treatment.

Cynefin Framework

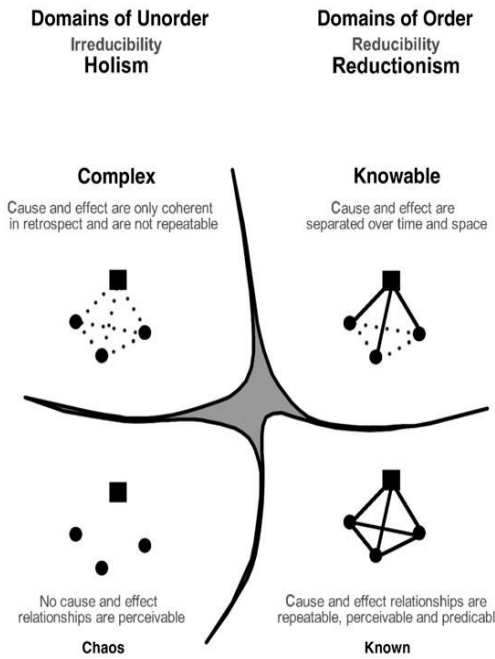

Figure 3. Cause Effect Association Interpretation Using Cynefin Framework

Figure 3 refers to the cause and effect association between the parameters decoded with the aid of the Cynefin Framework [8]. In the figure, the square block indicates the "causal agent", and the dots indicate the "effect agent" [8]. 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 adaptable 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 [8].

The master dataset or the big data, which was explored and analyzed, covered clinical visits between 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 [9].

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.

We followed the model of the Cynefin Framework as a blueprint for analysis to take complicated data into a simplistic form for exploration. 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.

IV. SUMMARY OF FINDINGS

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, which as discussed in the previous section were 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, and then merged with the EMR of patient records using data merging as the main technique for analysis.

The analysis shows that astigmatism (irregularity in the shape of the cornea) and conjunctivitis (inflammation or infection of the conjunctiva) are the most prevalent eye diseases among the youth population (ages 21-40) in Telangana. For older adults, (ages 41-70), the analysis shows that cataract and pseudophakos are the most prevalent eye diseases. Paloncha, Kothagudam, Kothagudam Bazar, Bhadrachalam, Mauguru, Madhapur, Kondapur, Adilabad, Yellandu, Kondapur and Tekulapalli are the top ten locations with the highest number of hospital visits, with Paloncha being identified as a high-risk location because of the presence of the state-run thermal power plant. In addition, the analysis shows a consistent pattern for high prevalence of cataract within the minimum ranges of temperature (20ºC - ~30ºC). The analysis also shows that cataract is the most prevalent eye disease in the rainfall season.

Influence Diagram - Taking Knowledge to Wisdom

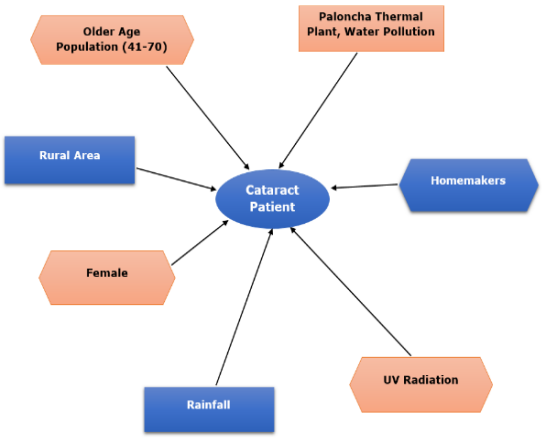

Figure 4. Influence Diagram of Cataract Patients and Co-Factors Affecting Cataract

Figure 4 depicts the summary of findings in a form of an influence diagram. 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). Looking at the data analysis holistically through Figure 4 creates the ease to depict the main causes of the development of cataracts.

V. CONCLUSION

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. In the case of studying the patients' data in Eye Smart, and summarizing the co-factors that play a role in the development of cataracts in Telangana, an influence diagram was created to show the different co-factors that lead to cataract according to the findings generated. This was further analyzed with a systems thinking approach, a method that allows consideration of the whole rather than individual elements of representation of the related co-factors.

The mission of data merging is to try to improve societal outcomes by developing, integrating, and using appropriate analytical tools in these new approaches, to offer healthcare practitioners the opportunity to design better patient-care policies and well-targeted regulations to lower the burdens of the economic and social problems associated with diseases, such as cataracts. An important starting point is to understand how data analysis and its results can be incorporated into the actual decision-making process. decision-making process.

 ISBN: 978-1-61208-891-4

Visual impairment imposes substantial costs on society, particularly to individuals with visual impairment and their families. Eliminating or reducing disabilities from visual impairment through public awareness of preventive care, early diagnosis, more intensive disease treatment, and new medical technologies could significantly improve the quality of life for people with visual impairment and their families, while also potentially reducing national health care expenditure and increasing productivity in the underdeveloped world. 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 individuals with vision impairment.

REFERENCES

[1] 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.

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

[5] S. Gupta, Real-Time Big Data Analytics. Packt Publishing Ltd. 2016.

[6] D.J Snowden and M.E Boone, A leader's framework for decision making. *Harvard Business Review*, 85(11), p.68. 2007.

[7] G. Kempermann, Cynefin as reference framework to facilitate insight and decision-making in complex contexts of biomedical research. *Frontiers in Neuroscience*, 11, p.634. 2017.

[8] J.P. Sturmberg and C.M. Martin, Complexity and health–yesterday's traditions, tomorrow's future. *J Eval Clin Pract*, 15(3), pp.543-548. 2009.

[9] R. Ackoff,. "From Data to Wisdom." *Journal of Applied Systems Analysis*, 16.1 pp. 3-9. 1989.

[10] A Das, P Kammari, R Vadapalli, and S. Basu. Big Data and the Eyesmart electronic medical record system. *Indian Journal of Ophthalmology*. pp. 427-432. 2020.

 ISBN: 978-1-61208-891-4