10-2020

A Systematic Approach to Big Data Analysis in Cataract Patients In Telangana State, India

Amna Alalawi
Thomas Jefferson University, alalawiaj@gmail.com

A SYSTEMATIC APPROACH TO BIG DATA ANALYSIS IN CATARACT PATIENTS IN TELANGANA STATE, INDIA

by

Amna Alalawi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Management
in
Strategic Leadership

at

THOMAS JEFFERSON UNIVERSITY

2020

A SYSTEMATIC APPROACH TO BIG DATA ANALYSIS IN CATARACT PATIENTS IN TELANGANA STATE, INDIA

Amna Alalawi

Doctoral Dissertation Committee Approval:

Signature goes here____

Dr. Les Sztandera, PhD, Chairperson Date

____

Dr. Robbin Durie, PhD, Committee Member, Internal Reader

____

Dr. Steven Herrine, MD, Committee Member, 1st External Reader

____

Dr. Richard Derman, MD, Committee Member, 2nd External Reader

ABSTRACT

Big data is the new gold, especially in healthcare. Advances in collecting and processing Electronic Medical Records (EMRs), coupled with increasing computer capabilities have resulted in an increased interest in the use of big data in healthcare. Big data require collection and analysis of data at an unprecedented scale and represents a paradigm shift in healthcare, offering on one hand the capacity to generate new knowledge more quickly than traditional scientific approaches, and, on the other hand, a holistic understanding of specific illnesses when socio-demographics are incorporated in the analysis. Big data promises more personalized and precision medicine for patients with improved accuracy and earlier diagnosis, and therapy geared to an individual's unique combination of genes, environmental risk, and precise disease phenotype.

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 record in LV Prasad Eye Institute (LVPEI), based in Hyderabad, India, can contribute to the management of patient care, through studying how climatic and socio-demographic factors relate to cataracts, clouding of the lens – turning the lens from clear to yellow, brown or even milky white, which cause visual impairment and blindness if left untreated. 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 were diagnosed with several ocular diseases from the years (2011-2019), a timeframe of 8 years. The dataset obtained included climatic variables to be tested alongside the development of cataracts in patients. Microsoft Power BI was used to analyze the data through prescriptive and descriptive data analysis techniques to read patterns that can dig deeper into high-risk climatic and socio-demographic factors that correlate to the development of cataract.

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. Women 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 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 cataracts in Telangana. In addition, this research created a pathway for new methods in the study of how EMRs contribute to new knowledge in ophthalmology. Results indicated that cultural upbringing, climatic factors, and proximity to the state-run thermal plant play a significant role in the presence of cataracts. Through testing the methodology used, observations indicate that the AI technique used is only effective when variables are minimized. Reflections suggest that studying patients through a more holistic and systematic approach can reveal new insights that can help bridge the gap between existing knowledge and practice for an aim to provide enhanced ophthalmic care in India.

DEDICATION

I dedicate this proposal to my Angel in Heaven, Sami.

Sami, my precious son, and the sweetest boy on the planet, died on August 14th, 2019, at the age of nine at King's College Hospital in London. Sami was diagnosed with a late stage liver cirrhosis because of a rare genetic disease called Wilson's Disease. After two liver transplants were performed to try and save his life, Sami lost the battle of survival. Sami was and still is my everyday inspiration to fight hard for the best in life, and to enjoy the blessing of a new day.

AI in healthcare can shed the light on unanswered medical complexities that are yet to be discovered, and because of that, I chose this topic.

"I love you and miss you every single day of my life 'habeebi' Sami." -- Mama

**habeebi is an Arabic word that means 'my beloved'*

ACKNOWLEDGEMENTS

"There is no treasure more profitable than knowledge"— Imam Ali Ibn Abi Taleb (AS)

First and foremost, I begin by expressing my gratitude to Almighty '*Allah*', God, the Most Gracious, the Most Merciful, for His endless blessings and kindness which enabled me to come to the United States to pursue the journey of Doctorate of Management in Strategic Leadership at Thomas Jefferson University, a door that opened numerous opportunities for my professional development and personal growth.

Second, I am thankful for my Advisor and Chair, Dr. Les Sztandera, PhD, Professor of Computer Science, who has been second to none in his support to me every step of the way. Dr. Sztandera proposed the topic of Artificial Intelligence in Healthcare to me, a remarkable field of study in today's world, which enables IT tools and technology to mimic human thought in big data analysis. His efforts in empowering me to believe in myself is one that I appreciate the most. Throughout working with Dr. Sztandera, I learned that dedication is crucial for one's success, and that nothing comes easy, but that hard work pays off. I was honored to win the 'Best Paper Award' at the Eighth International Data Analytics Conference, which took place in Porto, Portugal. Jefferson has fully sponsored my participation in the conference, which I am grateful for. The paper titled "Leveraging Statistical Methods and AI Tools for Analysis of Demographic Factors of Opioid Overdose Deaths" was published in the journal of *Advances in Life Sciences, Volume 12, July 2020.*

I am also thankful for my Committee members from Thomas Jefferson University, Dr. Richard Derman, Associate Provost of Global Affairs, Director of Global Health Research, and Professor of Obstetrics and Gynecology; Dr. Robbin Durie, Adjunct Professor in Business Policy

and Strategy, and Director of the Medical Electronic Catalog Program at the US Department of Defense Supply Agency; and Dr. Steven Herrine, Professor of Medicine, and Vice Dean of Academic Affairs and UME at the Sidney Kimmel Medical College. I was honored by my Committee's continuous guidance which helped me understand the criticality to think of a medical impact that I can bring into my research. I was honored to obtain a grant for my research from India's Bill and Melinda Gates Foundation, which was supported and approved by Dr. Derman. I enjoyed taking part in a research that is part of Jefferson's global initiatives. Throughout my work, I was introduced to Dr. Anthony Vipin, Consultant Ophthalmologist, based in Hyderbad, India, who generously shared datasets from the LV Prasad Eye Institute's Electronic Medical Record (EMR), which I based my dissertation on examining.

I am also thankful for Dr. Larry Starr, PhD, Director of my program "Doctorate of Management in Strategic Leadership", who provided me with tremendous support throughout my stay in the United States as an international student. I appreciate that he believed in my potential and accepted me in the program. Throughout my learning in the program, I understood the concept of systems thinking and how to combine philosophies of complexity theory into my research.

Finally, I thank my entire academic community, family, and friends, with special appreciation to Ms. Hannah Bar-Giora, Director of International Students at Jefferson East Falls, and Mrs. Colleen Duffey Giambalvo, Associate Director of Applicant Experience, who both supported me very much during my stay in Philadelphia and provided me with 'sisterly love' during a difficult time of my life. The loss of my beloved son, Sami, mid-way throughout my doctorate degree was an event that will remain the hardest in my life. I am fortunate to have been around the best family and friends who pushed me to see the light after loss.

I hereby present my dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Background

India is home to over 8.3 million people with Vision Impairment (VI), the highest in the world. 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 (Survey, 2019). Prevalence of blindness and visual impairment is one of the highest in Telangana, a state in Southern India, as inferred from the survey. The major reasons indicated in the survey were due to cataract and refractive error.

The close relation between climate, environment, and the development of diseases in the developing world is crucial to understand, and by studying these factors that relate to ocular diseases, more information can be revealed for future preventative measures. The impact on climate change and climate trends due to global warming has caused an array of health-related problems (McMichael, et al, 2003; Hales, et al,2003). According to previous epidemiological studies, cataracts, for example, can occur due to the high level of exposure to UVR, and it is a serious public health issue because it is one of the causes of blindness throughout the world. The increased exposure to sunlight is responsible for cortical cataract. The risk of the disease due to the increased level of ozone depletion is estimated to be within 5 – 20%. The findings designated that the rate of cortical cataract would rise to 1.3 – 6.9 % within the year 2050 (Liu, et al, 2009; Katoh, et al, 2001; Schein, et al, 1994). Moreover, enhanced exposure to UVR results in premature aging of the lens, which may contribute to nuclear sclerosis and myopia, ocular diseases that cause cloudiness, hardening, and yellowing of the central region of the lens.

Eye health risk in India has been linked to poor air quality and pollution, and in recent years ophthalmologists reported a dramatic increase in the number of patients complaining of eye problems. Telangana covers 33 districts in India, each with a different climate trend, and a unique socio-demographic factor. Studying the correlation of these trends can help explain the relationship between these factors and the development of ocular diseases in patients who presented themselves to LVPEI from the years 2011-2019. For example, high temperature causes dryness in the eye, a condition known as dry-eye disease (Heussner, et al, 2014). India is on the verge of a dry eye disease epidemic. The prevalence of dry eye disease will be in about 40% of the urban population by 2030, which is around 275 million people. A recent publication by LV Prasad Institute goes further in explaining the effects of dry-eye disease, which subsequently results in ocular surface damage. Several risk-factors have been identified as key players in developing the disease such as age, autoimmune disease, lifestyle influences, contact lens use, and others (Dontheneni et al, 2019). The research, however, fails to identify climate factors which may have a big role in developing the disease in India. Given that India is a sub-continent, and climate factors differ from one to another, early intervention could help hopefully eliminate the development of the disease using Artificial Intelligence (AI) machine learning tools.

Statement of the Problem

In 1976, India launched ‘Vision 2020, The Right to Sight’ program. The first of its kind in the world, the program’s stated aim was to reduce the prevalence of blindness among the population to 0.3% by the year 2020. However, the National Blindness and Visual Impairment Survey published in October 2019 by the Union Ministry of Health and Family indicates that 8.3 million citizens remain vision impaired (VI), representing a prevalence of 1.99%, which is the highest national VI population globally (Survey, 2019). Within the national picture, the situation in Telangana stands out as particularly gloomy as regards to blindness and other types of visual impairment, particularly through the high incidence of cataracts and refractive error. A cataract is a clouding of the lens – turning the lens from clear to yellow, brown, or even milky white. This blocks the light from passing into the eye, and eventually causes vision loss and blindness.

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 (Suchitra, 2015).

Hence, the current study focuses on examining the causes of cataracts using computational intelligence software as this condition remains the single most important cause of blindness, both in India and globally (Vashist et al., 2020). Moreover, the Global Burden of Disease, Injuries, and Risk Factors study indicates that cataracts are the second most frequent cause of moderate and severe VI (MSVI) globally, and the first in Southeast Asia. Statistics reported by Honavar (2017) indicate that cataracts have caused the blindness of 10.8 million of the world's 32.4 million blind population; furthermore, of the 191 million global VI population, 32.4 million individuals suffered from cataracts. A total of $5.73 billion investment was estimated to be required for eliminating blindness due to cataract between 2010 and 2020 (He et al., 2017). Cataract remains a concern for public health, especially in low- and middle-income countries.

Identifying the best preventative measures for future practice requires a careful assessment of the relationship between environmental factors, including climate, and the development of cataracts. Recent research carried out among 12,000 patients in India found that increased sun exposure, tobacco smoking, and exposure to indoor smoke from kitchen fires all play a role in the development of cataract. Thus, the climate change caused by global warming can be expected to impact levels of VI, as well as other health-related issues (McMichael et al., 2003; Hales et al., 2003).

This finding supports earlier epidemiological studies which found that high levels of UVR exposure can be responsible for cortical cataract, a type of cataract which occurs in the lens cortex, which is the part of the lens that surrounds the central nucleus. Given current levels of ozone depletion, estimated at 5-20%, this increased UVA exposure clearly represents a danger to public health and suggests a potential increase in the incidence of cortical cataracts to 1.3–6.9% of the global population by 2050, or an increase of 5-20% in risk (Liu et al., 2009; Katoh et al., 2001; Schein et al., 1994). Further health problems caused by increased UVR exposure include the premature aging of the lens, which can be a contributory factor in nuclear sclerosis, which causes the central region of the lens to harden and turn cloudy and yellowish in color.

Pollution is also a factor in the rising incidence of ocular disease in India, in which the connection is apparent. Past studies have shown that there are high levels of subclinical ocular surface changes in people working and residing in highly polluted areas (Saxena et al., 2013). The 33 districts encompassed by the state of Telangana each have their own climactic peculiarities and socio-demographic profile, offering the opportunity to investigate correlations between a range of environmental factors and the development of cataracts among patients who participated in LVPEI care between 2011 and 2019.

Merging datasets that include demographic and environmental factors can expand on the detection of ocular diseases, when analyzed accurately through computational intelligence softwares that can recognize patterns. Expanding on this area, the research will focus on analyzing big data in *Eyesmart's* health records, regarding patients who presented with symptoms and complaints of eye disorders in LVPEI from the years 2011-2019. The demographic information of these patients alongside climatic factors in the different districts of Telengana, will reveal new insights that have not been studied before in the development of ocular diseases. Therefore, the investment in AI in ophthalmology should be important, to act as a strategic role in clinical practice – one which will depend on early detection of ocular diseases, and perhaps prevention.

Significance of the Study

The era of 'big data' in healthcare, even though provides information about patients worldwide, creates a challenge in terms of transforming that data into information that can be leveraged in research. Practitioners are in need to better understand the causes of diseases and how to offer preventative measures. Since the late 1990s, substantial efforts have been made to ensure that health agencies have the information technology and training needed for easy access of information, communicating health issues and for data exchange to better understand emerging trends in diseases. However, little systematic work has been done to understand the information needs of the health sector. Researches from the Centers for Disease Control and Prevention (CDC), part of the World Health Organization (WHO), state that identifying the gap between knowledge and practice in healthcare and being able to set priorities for the development of information remain a challenge (Ola, 2014). If studied effectively, big data offers huge possibilities for health record management, clinical trials and new preventative measures.

Today, leveraging big data through Artificial Intelligence (AI) software to create new knowledge has become a healthcare pattern. This is because the focus in healthcare research has moved towards a more patient-centered study through using big data. Due to the complex and dynamic factors that cause the health issues around the world, this research advocates the importance of a systematic approach in merging publicly available datasets through AI software. Studying the race, culture and climatic factors that affect the development of diseases are crucial to understand. A case study is focused on revealing socio-demographic and climatic insights that correlate to the development of ocular diseases in the state of Telengana, located in Southern India, for an aim to enhance ophthalmic patient care in LV Prasad Eye Institute (LVPEI), based in Hyderabad, the capital city of Telangana. Currently, LVPEI uses EyeSmart, an Electronic Medical Record (EMR) of more than 1.2 million patients for research studies.

In healthcare, ophthalmology deals with the diagnosis and treatment of eye disorders. Some known diseases in ophthalmology are cataracts, retinal disorders, macular degeneration, dry-eye disease 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 (Boland, 2016).

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 (Clark et al, 2016).

Studying socio-demographic risk factors, such as race, culture and the climate, especially associated with the environment can lead to a better understanding of the causes, diagnosis and treatment of several eye diseases. Using computational intelligence software to analyze patients' data alongside environmental and climatic variables can provide insights which can modernize the way treatment is offered at eye clinics across India, including LVPEI, which is the focus of this joint study between Thomas Jefferson University (TJU) and LVPEI. The proposed research will help build information that will contribute to the existing knowledge in the field of big data in ophthalmology.

Research Questions

The following research questions have been formulated:

1. Does the Cynefin Framework offer potential as a means to transform raw data into a form which can be used by healthcare researchers and practitioners aiming to increase early cataract detection rates in Telangana?
2. Can data merging techniques be applied to EMRs to generate new knowledge shedding light on correlations between the development of cataracts and socio-demographic and environmental factors among residents of the state of Telangana, southern India?

Researcher Positionality

As my dissertation has been discovery research, it is crucial to explain myself, my background, and my understanding of the topic from a data-driven perspective to how I analyzed the data. It was an integral part of the research process, as I was positioned to convert information from the analysis into new knowledge by navigating through complex data sets using AI software and converting it to a simplistic dataset that a non-healthcare practitioner like myself can understand and look for patterns. Here I wanted to explain that a layman can also read and understand the data as well as its analysis to form his or her opinions about the subject matter. Even before starting the study, I had come to understand that this topic is a difficult one, and thus a non-healthcare practitioner cannot understand the study and its results, especially that the study is focused on ophthalmic care in the state of Telangana in India. Therefore, to make the investigation as successful as possible, I scheduled frequent phone calls and video meetings with Dr. Anthony Vipin Das, Consultant Ophthalmologist and Clinical Director of Innovation at L V Prasad Eye Institute, who took me through a typical research process at the Institute and what is expected from a research and how to optimize the use of the EMR record, Eysmart, which I will further explain in my research in the following chapters. Dr. Vipin guided me through the process of organizing the EMR record, transforming it into useful data that can go beyond finding general information that causes the development of cataract in patients in the state of Telangana.

I wanted the data to be easily comprehensible for the readers, while at the same time, understand the technicality of how patient data is used in clinical practice. Here I emphasized on making the data and its analysis as simple and easy to read and understand as possible. Further, I also focused on providing an extensive discussion about the data but made sure that I did not use too many technical jargons. This way, I was able to provide a thorough discussion and analysis of the data.

It is an explicit self-consciousness about the researcher's social, political and value positions in relation to how these might have influenced the design, implementation and interpretation of the theory, data and conclusions (Griffiths, 1998; Greenbank, 2003). Therefore, my positionality as a data analyst plays a central role in the research design and findings of the study. Further, it also influences the way I analyzed the findings and discussed the results. Here I explained the results, findings, and discussion in first-person. This way, I aimed at presenting my position in terms of social and value-related aspects. By doing so, I tried to make the study as relatable to the readers as possible.

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, I relied on the explicit meaning of the term, and thereby I decided to work on complex data sets and determine a way through which it can be evaluated. To make the process simpler, I decided to use AI-based software. It helped me in identifying and analyzing patterns in the data sets, which can be used for future clinical practices. In the context of the current study, I mainly used descriptive modeling analysis to reach the findings. The techniques I used here helped me in transforming the data into such formats that could be easily evaluated and analyzed. In this case, I observed that the process of transforming the data was lengthy and a time-consuming one, but it was extremely rewarding as well.

The following Data, Information, Knowledge, and Wisdom (DIKW) pyramid framework explains the reasons for transforming the data. By transforming the data into information, I gained certain knowledge about the topic which then was transformed into wisdom that helped me in not only conducting the investigation effectively but also 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

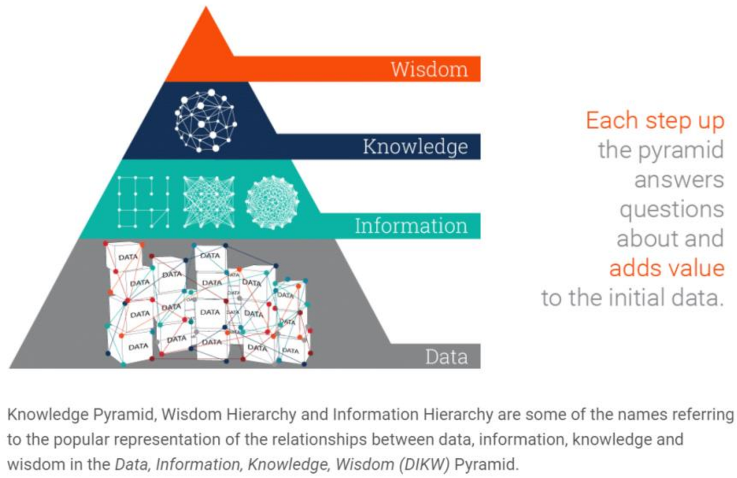


Figure 1: DIKW Pyramid

Dr. Russel Ackoff, wrote that "wisdom is located at the top of a hierarchy of types," (Ackoff, 1989) 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. 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 (Ackoff, 1989).

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 (Brown, 2010: 238). As Morison (1986: 56) 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.”

Self-awareness for me as a researcher was very important. It enabled me to develop my interpretations about the topic and analyze the complex data in a simple manner. I learned a great deal about my skills and learning through the leadership models, while I focused on merging my data analysis skills in developing an effective research design. I was also able to relate the process of design thinking in analyzing big data to my coursework in the Doctor of Strategic Leadership. The methodology chosen for the research helped me understand the modern way of leading through technology, and how AI software can manipulate big data and generate analysis, which when applied on the case study of LVPEI allows for ophthalmologists to better understand their patients and how to use patients’ data to enhance patient care. Strategic leadership in healthcare is an ongoing field where big data, AI, and the approach to clinical care are becoming best practices in public health for establishing new ways to eliminate insidious diseases and epidemics.

CHAPTER 2: LITERATURE REVIEW

Data Analytics

Over the last 45 years, the general activity of making sense of data has evolved from decision support, to executive support, to online analytical processing, to business intelligence, to analytics and now to "big data". One of the terms proposed to define big data is the collection and interpretation of massive data sets, made possible by vast computing power that monitors a variety of digital streams, and analyzes them using "smart" algorithms (Davenport, 2014). Healthcare has become of the key emerging users of big data (Dimitrov, 2016).

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. However, some of the challenges that are being addressed today include issues of privacy and security, which require developing and improving the use of AI. Big data analytics in healthcare is also considered to be at a starting point, however, with the rapid increases of interest in AI, this will soon see a change.

Application of Artificial Intelligence in Healthcare

The term "Artificial Intelligence" has been introduced by John McCarthy within a conference held at Dartmouth in 1956 and explained it as "the science and engineering of making intelligent machines" (Society for the Study of Artificial Intelligence and Simulation of Behavior, 2018). Following that, a period came where there was diminished interest and resource in the field of research with regard to AI had been observed also being referred as AI winter (Crevier, 1993). The concept of AI is a significant part of computer science that seeks to develop intricate models with the attributes of human intelligence. This concept is generally referred to as "General AI" (Copeland, 2016). The concept that operates now at present belongs to "Narrow AI" and here technology can execute assignment superior in comparison to human beings, for instance, the recognition of face and speech (Copeland, 2016). These techniques have advanced facilities from deep learning and AI for the purpose of analysis of speech, identification of face and image and have immense application in the field of natural language processing, autonomous vehicles, and in medicine.

Artificial Intelligence (AI) applications in healthcare has gained attention in recent years. According to B.J Copeland, Professor of Philosophy and Director of the Turing Archive for the History of Computing, University of Canterbury, New Zealand, AI is regarded as the ability of the computer-controlled robot or a digital computer to perform tasks linked to intelligent beings. AI is commonly applicable in project development systems said to be endowed with potential intellectual processes (2019). Eric Topol, cardiologist, author and researcher, argues that AI in medicine is beginning to have an impact on clinicians, predominantly via rapid and accurate image interpretation; for health systems, by improving workflow and reducing medical errors; and for patients, by enabling them to process their own data to promote health (Topol, 2019).

AI in Ophthalmology

On the other hand, ophthalmology is a significant branch of surgery and medicine that focuses on diagnosis as well as treatment of eye disorders (Nelson, 2018). A partial list associated to ophthalmology includes macular degeneration, cataract, dry eyes, proptosis, glaucoma and diabetic retinopathy. The study of AI and ophthalmology is more pertinent and relevant when handling such areas like early disease detection, pattern recognition and data management.

According to Akkara and Kuriakose, ophthalmologists and researchers in Kerala, India, AI is said to have made its entry in healthcare and modern life, and is showing promising results (2019). In addition, ophthalmology is a growing field that is engaging in measurable data and imaging, which makes it ideal for AI and machine learning application.

AI has demonstrated its significance in the field of ophthalmology for the interpretation and recognition of patterns or trends within the clinical images for instance the identification of the 3D retinal scans by Google's DeepMind Health. The field of medicine is constantly evolving with the rapid enhancement and development of the computer based AI tools which had resulted in precision and success in the process of diagnosis across different fields of specialization. With the rapid utilization of AI in radiology, some specialists also have suggested that AI may replace radiology in the future. However, the question arises whether AI will substitute the physicians or boost up their role in their field with the help of AI rather than replacement. Therefore, the main objective of this research is to comprehend the significance of AI in the field of medicine as AI has expanded its role in the field of radiology, cardiology, ophthalmology and pathology. Thus, overall it will enhance the role of the physicians without hampering the physician – patient relationships.

Modern day healthcare research is based on the analysis of large findings, and in the field of ophthalmology, the success of AI tools will depend on the amalgamation of varied aspects of application along with the tool (Patel et al, 2009). With this approach, it is significant to manage the information of the patients stored within the electronic medical records in a holistic approach for analysis. However, the approach creates a problem for the researchers to utilize AI tools in the aspect of modern health research along with the help of healthcare professionals. The tool will allow healthcare providers to acquire a certain approach into the clinical records of patients and to formulate better care and better services to patients. Ophthalmologists are making use of artificial intelligence or machine learning tools, which is making the process of detection automated, and detecting a variety of problems among the patient with enhanced level of supervision. According to Andrew P. Schachat, MD, Editor in Chief of Ophthalmology Retina (2019), the significance of AI, profound learning through computers that would augment the care services for the patients can be clearly comprehended. Moreover, according to published scientific evidence on deep learning models that could perfectly project the succession of age-related macular degeneration through the model named DeepSeeNet. This model analyzed color fundus pictures (58,402), and also examined by itself nearly 900 pictures that has been acquired from longitudinal follow-up investigation of 4,549 sample size participants from AREDS. The findings revealed a comparatively better precise identification of large drusen and changes in the pigmentation which is considered to be as per the diagnosis by a retina specialist. Moreover, the computer could also simplify whether it is fluid or not just like a retina specialist (Schachat, 2019).

Machine Learning, a branch under data science, which combines statistics, computer science, information technology, and data visualization, is a powerful tool with infinite possibilities to enhance clinical practice in identifying certain ocular conditions (Consejo et al, 2019). When combining multiple variables in a study, such as social health factors, environmental factors, and climatic factors, new insights and expanded findings can be generated to help in clinical decision making and in patients care.

One strength that makes AI ideal for ophthalmology is that ophthalmology is a field of medicine with a lot of imaging and measurable data, thus ideal for AI application. AI has been more relevant in the analytical process of the retinal fundus images associated to diabetic retinopathy. This can be followed by what is referred as age-related macular degeneration, retinopathy of prematurity and glaucoma. Based on the findings made by Nelson (2018), major technological companies have taken strides towards AI and ophthalmic use. IBM's AI, for instance, has the capacity of predicting the visual field data associated to OCT scans. In addition, DeepMind Health, Google's artificial intelligence business, helps in the diagnosis of eye disease by analyzing medical images. It analyzes 3D retinal scans for signs of major eye diseases, such as glaucoma or diabetic retinopathy. AI can also analyze the scan immediately while patients would ordinarily have to wait days for a specialist to review the images (Esson, 2018).

A recent study in the *Journal of Ophthalmology*, asserted that healthcare has emerged as a significant area at the center of AI application (Lu et al, 2018). A range of studies have been linked to another branch of Data Science, called Deep Learning (DL). The algorithms in DL are said to be performed at high levels. This has been applicable to breast histopathology analysis as well as skin cancer classification. Other areas of concern as noted by the study include ophthalmology, lung cancer detection and cardiovascular risk prediction. With a vast range of applications, it is paramount to note that AI in ophthalmology is imminent and unstoppable following the development of AI algorithms and accessible data sets such as Messidor, EyePACS and Kaggle's data set. Subsequent observations made in the *International Journal of Ophthalmology*, indicate that AI can be applied in terms of both (DL) and Machine Learning (ML). This has bolstered subsequent diagnosis of ocular diseases which covers the most leading causes of blindness, diabetic retinopathy, cataract, age-related macular degeneration as well as glaucoma (Du et al, 2018). ML approaches, said to have been introduced by Sandrina Nunes and Miguel Caixinha, who have been paramount in monitoring and diagnosing ocular diseases. ML largely attracts small data sets but it can turn out to be cumbersome when it comes to handling visual features. DL, as part of the AI application, is known for having the ability of discovering the most intricate structures across data sets even without specifying the rules.

Further observations made by Lu et al. (2018) noted that current studies are putting more focus on machine learning, which can attain satisfactory outcomes. More focus on AI and diabetic retinopathy has attracted towards retinal microvasculature, which leads to damage. More people are essentially affected by diabetic retinopathy, and this has been turned into a public health problem across the world. Large scale screening of the diabetic retinopathy is on high demand as far as treatment and management is put into consideration. Practitioners have consistently called for early intervention, which taps into diabetic retinopathy automatic identification. Further attention given to neovascularization detection, microaneurysm, cotton wool spot, hemorrhage and exudation has raised hopes of AI application. In this case, computers can receive images which can be labeled as diagnostic lesions before identifying the final judgment and input images.

Besides, Ting et al. (2019) noted that adoption of DL in natural language processing, image recognition and speech recognition has impacted the approach towards healthcare. Apparently, in ophthalmology, the application of DL in visual fields, fundus photographs as well as optical coherence tomography has led to achievement of robust classification performance. The large number of imaging and image processing techniques available nowadays present new opportunities to develop decision-support tools that assist clinicians with the diagnosis of almost any ocular condition.

In the sphere of Data Analytics, another branch in data science, healthcare has shown the potential to become ‘smarter’ and more effective. The use of artificial intelligence has been enabled by big data, along with markedly enhanced computing power and cloud storage, across all sectors (Topol, 2019). The healthcare industry generates a lot of databases including information on patients, diseases, demographic and much more. The potential to improve the quality of healthcare delivery and at the same time reducing cost, is very promising. The massive quantities of data, also referred to as ‘big data’, hold the promise of supporting a wide range of medical and healthcare functions, including clinical decision support, disease surveillance, and population health management (Ragupathi, 2014). Carol McDonald, a developer in health systems and expert in Java, is also a supporter of utilizing big data to reduce cost, in addition to improve health outcomes (2019).

Researchers at the Johns Hopkins School of Medicine discovered they could use data from Google Flu Trends, a novel internet-based influenza surveillance system that uses search engine query data to estimate influenza activity, to predict sudden increases in flu-related emergency room visits at least a week before warnings from the Center of Disease Control. Despite predictive flaws in Google Flu Trends, the analysis of Twitter updates was as accurate as official reports at tracking the spread of cholera in Haiti after the January 2010 earthquake (Ragupathi, 2014). The authors used HealthMap, an automated surveillance platform, to measure the volume of news media generated during the first 100 days of the outbreak, and they also looked at the number of 'cholera' posts on Twitter. The study found that online social media and news feeds were faster than, and broadly as accurate as the official records at detecting the start and early progress of the epidemic, which hit Haiti after the earthquake in January 2010 and has killed more than 6,500 people (Hirschfeld, 2012).

Issa et al (2014), researchers at Georgetown University Medical Center support the use of electronic medical records (EMRs) to create new knowledge in the field of healthcare. EMRs are rich with clinical data that chart patient progression with respect to disease, medications, with other demographic information. Family history, diet, medications and occupational exposures are just some of the documented information that could be invaluable in determining unique treatments for individuals (2014). EMR information is uniquely positioned to aid in the discovery of new findings when coupled with other datasets. In other words, combining biomedical information with environmental and social contributors would provide a holistic system view of a patient, and will highlight new ways to intervene in enhancing patient care either by early detection of a disease, or by a more accurate prevention method. To date, no such platform exists as EMR records have not been transformed yet into the study for clinical medicine, but this is expected to change (Issa et al, 2014). In addition, government agencies worldwide are releasing public datasets about the services provided in healthcare, such as Medicare in the US. These datasets can be queried by multiple classes of users, including hospitals, patients, physicians and policy makers. However, to realize the true value of the information present in these datasets, appropriate analysis, including classification and clustering needs to be performed. Additional insight can be gained by combining the healthcare data with other data sources such as demographics and epidemiology (Rao et al, 2015).

Furthermore, Yang et al (2015), researchers in Information Technology and experts in Computer Science from China, state that since healthcare covers complex processes of the diagnosis, treatment, and prevention of diseases, EMRs can be used in detecting medical problems at an earlier stage, if the data is collected and managed properly. Moreover, they state that technologies are not solely used anymore for therapeutic purposes, but analysis using big data and cloud computing can reveal trends and can be used in predictive medicine. Existing methodologies for the detection and analysis of medical conditions will have to be revised and extended to discover deep knowledge and deliver enhanced patient care. Cloud computing can support the analysis of big data through innovative technologies and softwares. Data mining (DM) is the computing task to discover unsuspected patterns from the observational datasets to help users to make better decisions (Zhou et al, 2010). In general, the discovered patterns are novel understandings and represent something hidden in the available dataset that the users did not know before. “The application of data mining algorithms for medical data analysis and utilization can be classified into two categories, i.e., unsupervised (descriptive) and supervised (predicative) approaches. The unsupervised methods mainly concern data clustering, i.e., grouping data into clusters by measuring the similarity between objects or EMRs to discover unknown patterns or relationships in the available datasets. The typical unsupervised data mining approaches covers data clustering, association rule mining, and sequence discovery” (Yoo et al, 2012).

EMRs have been adopted only relatively recently in ophthalmological practice globally; however, take-up has been swift, largely due to the realization that this electronic record management will enable large amounts of clinical data to be used in research. The practice is, however, in its infancy; moreover, clinical data tend to be more complex than other types of data whose management has been transformed by the digital revolution and thus the expected benefits have not yet been seen in all cases (Boland, 2016).

Nonetheless, EMRs have already proved of significant value within ophthalmological research, offering new insights into matters ranging from disease surveillance through how the health service is utilized, and giving a more detailed picture of outcomes. Furthermore, there has been an increase in the sheer quantity of data available to researchers. It has therefore been recommended that data linkage systems are used to guide future research (Clark et al., 2016).

Significance of AI in Research

The market value of AI in the healthcare sector is constantly rising at about 40% and it will reach up to $6.6 billion by 2021 (Frost & Sullivan, 2016). Large records can be stored within cloud and training or learning on algorithms allows gaining perspectives in the field of diagnostics, treatment management and patient findings (Bresnick, 2018b). AI is well structured to manage the repetitive procedures of work, deals with huge cores of information and helps to make decisions without errors. The study of Frost & Sullivan had highlighted that with the use of AI, patient outcomes can be augmented from 30 to 40% and diminishes the expenditures of treatment by 50% (Hsieh, 2017a). According to specialists, AI have immense application within varied areas of health such as dealing with long term diseases and up taking the right judgment (Bresnick, 2016). AI has made a rapid progress in healthcare because of large chunks of data available in the EMR in the recent years along with raised efficiency of computing and technology (Pratt, 2018). Therefore, AI could be able to evaluate the findings of patients acquired from varied origins for instance the fitness trackers and home monitors that assist physicians to look after patients which would not have been possible without AI (Pratt, 2018).

CHAPTER 3: CASE STUDY AND METHODOLOGY

Electronic Medical Record "Eye Smart"

The current research uses a case study to explore how data merging techniques can generate new knowledge from the Electronic Medical Record (EMR) of LV Prasad Eye Institute (LVPEI), known as "Eye Smart", with the aim of identifying socio-demographic patterns in the development of cataracts among patients in the state of Telangana, southern India. LVPEI was established in 1986 in the city of Hyderabad as a not-for-profit, non-governmental, and comprehensive eye care institution. It states on the website, that LVPEI's service model comprises of three pillars; equity, which translates as treating all patients regarding of economic status equally, efficiency, which is through using world class tools and technology, and excellence, which is the standard of performance that the eye institute ensures to deliver.

LVPEI operates out of 106 locations, 86 of them being primary eye care centers located in remote rural villages. 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 transplanted more than 17,000 of them to needy patients.

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 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, also aids in teaching. Some of the problems that are usually encountered are related to low connectivity and power outage at times, while the cost savings reported include manpower cost with the medical record department, paper printing and storage.

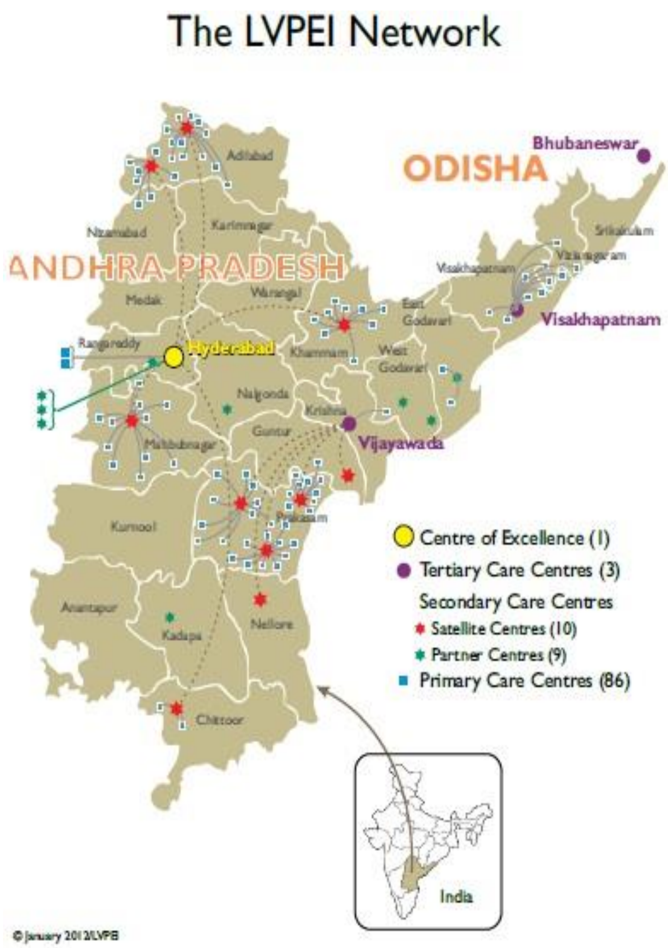


Figure 2: LVPEI Area Map

(Source:lvpei.org)

The figure above 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.

This section explains the data merging technique that was used to merge weather variables with datasets from Eye Smart, to enhance on 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.

Design Thinking and its Application on Eye Smart

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. This paper presents ways in which EMRs can be studied collectively and holistically to bridge the gap that currently exists between knowledge and practice.

Currently, the data of patients in LV Prasad Eye Institute is stored in Eye Smart, a national award winning ophthalmic Electronic Medical Record (EMR) and Hospital Management System developed in-house by the institute. The EMR was established to enable electronic documentation of patients, as well as for faster retrieval of information, as well as for research purposes (Vipin, 2012). The EMR, which was leveraged and combined with other datasets, includes demographic information on 873,450 patients from the state of Telangana, who were diagnosed with different eye disorders between the years 2011-2019. 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. 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. The climatic variables which were the main focus in the research included temperature, humidity, pressure, windspeed, rainfall, and global radiation during the Monsoon period, Post Monsoon, Summer, and Winter.

A total of 176 variables were initially studied, ranging from number of patient visits, type of complaints, ocular diagnosis, gender, age, occupation, among many others. AI creative featuring techniques were used to narrow down the variables most affected by climatic and demographic factors and programmed for analysis. Regarding the size of the original dataset, it is always better to have a larger number of subjects, but one should be cautious with the number of variables included to avoid overfitting. In addition, it is important to apply appropriate data pre-processing to maximize the chance of obtaining a successful classification tool (Ting et al. 2019).

A Systems Approach to Patient Care

Because of the emergence of big data in academic knowledge building, small data, especially publicly available data can be used in further studies through merging them with other data to be utilized in big data analytics. Some data is limited in scope, size, variety, and are used to generate specific answers to questions only. While on the other hand, when merged with other datasets, they produce new information through communication technologies (ICTs). Therefore, big data is being reconceived in the context of knowledge expansion, and is being pooled, linked and scaled into data infrastructure for them to be more open to analysis (Kitchin and Lauriault, 2015). EMRs make it easy to use digital data and reuse it for a low marginal cost. In addition, the data can be manipulated and analyzed by exposing them to computational algorithms. As such, procedures and calculations that would be difficult to undertake by hand or using analogue technologies become possible in just a few microseconds, enabling more and more complex analysis to be undertaken or the replication of objects and results (Kitchin and Lauriault, 2015).

The Analytics Engine

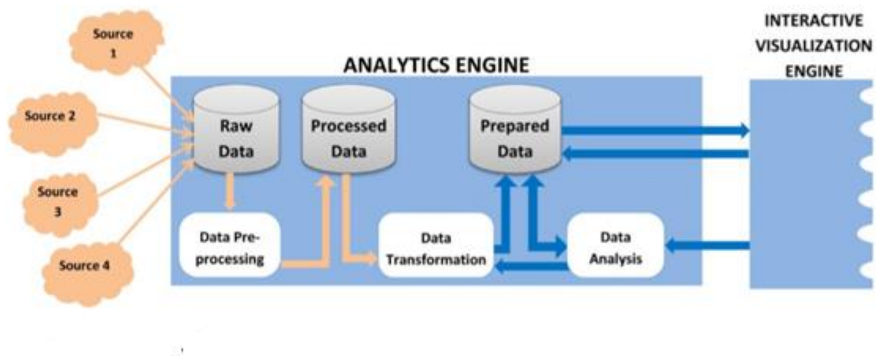


Figure 3: The Analytics Engine Component of VA Tools

Three data sets in excel binary format were received under the following names:

- Telangana Weather Excluding KARMN
- KARMN_11_15
- KARMN_16_19

These different data sets contained the records of patients who were treated at LVPEI. The datasets were highly complex and required a long time to download and be transferred into the AI tool for analysis. Eye symptoms were captured and recorded for each eye and categorized under

OD, which is short for oculus dexter, and is the right eye; OS, which is short for oculus sinister, and is the left eye; left-eye, right-eye, or both eyes.

To start the analysis, a master data set had to be generated to be read on Microsoft Power BI, the AI software that was chosen for the research due its capacity to read big data cells and generate robust and creative visuals. The data sets were first converted from an excel binary format into a standard excel file which can be easily read by the python pandas data library, the language used in programming and data merging.

Python programming language was first used for merging and cleaning the data. A python script was written using Jupyter notebook development environment. The Pandas library was used to read-in and merge the three different data sets into one master data set. Further data extraction/cleaning was carried out to prepare the data for use on Microsoft Power BI. The reason for the choice of Python was because of its strength in manipulating large and complex data sets.

The procedure of data merging is included in the appendix.

The final output of the analysis procedure in Python is a data file named: Master weather data in a Comma Separated Value (CSV) format. This file was further loaded into Microsoft Power BI for final analysis. Microsoft Power BI was further used for modeling, analyzing and visualizing insights from the data. Filtering was done to drill-down to specific insight. The output of the final analysis and visualization on Microsoft Power BI are presented throughout the research document.

Data Analysis Process Diagram

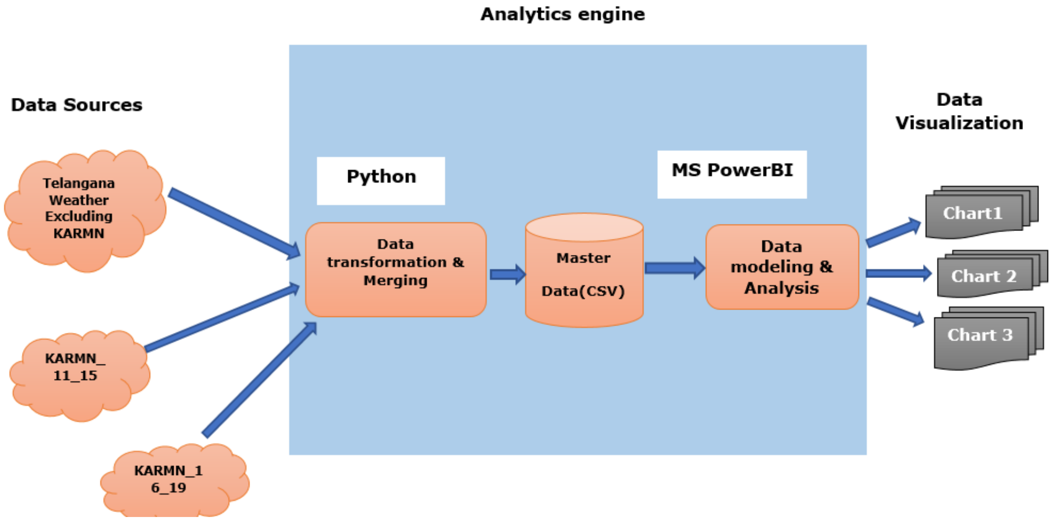


Figure 4: Data Transformation from Python to MS Power BI

The purpose of the model in figure 4 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. A more detailed figure is illustrated and included in the appendix. To this end, 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 together. The outcome lead 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 flavors such as SAAS and PaaS. Now, we are in the big data era, which can be processed and analyzed in cost-effective ways (Gupta and Saxena, 2016), 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.

CHAPTER 4: FINDINGS AND TRENDS

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. The section will also discuss methodology testing and how variables were tested alongside each other on Microsoft Power BI to achieve results that can be used to build on knowledge and create wisdom, as defined in the research earlier.

Initially, before embarking on studying cataract as a specific disease, it was important to gain a glimpse of the composition of the state of Telangana in relation to patients, eye diseases, and other social and demographic factors that are crucial to the study, and this phase is defined as phase I of testing. This has helped in distinguishing different patterns in cataract patients verses non cataract patients, while exploring the correlation of demographic factors to cataract patients in phase II of the study.

The EMR record was tested extensively in phase I of the methodology in order to create a research design that can lead to a simplistic and structured data analysis, again, transforming the EMR from an unstructured to a structured zone for analysis as per the Cynefin Framework discussed in the earlier chapter.

Phase 1- Highlights of Overall Indication to Eye Diseases in Telangana

From the analysis of the overall data, the following trends were identified:

- 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 hospital visits, with Paloncha being identified as a high-risk location because of the presence of the state run thermal power plant.

- 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.

Some of the visual patterns have been captured and identified below for a more in-depth explanation:

A. *Gender and Eye Disorders*

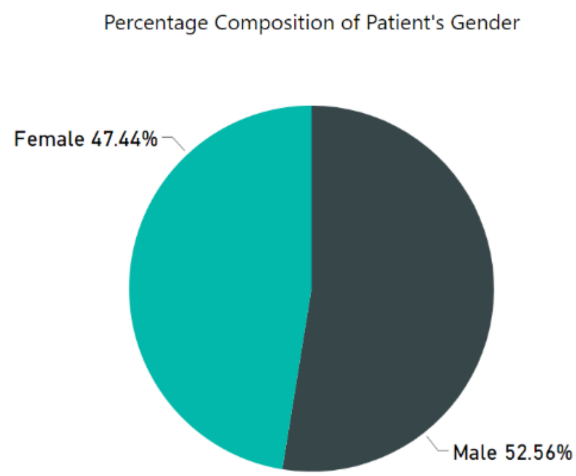


Figure 5. Clinical Visits by Gender (2011-2019)

Figure 1 indicates that between the year 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 (Das et al, 2020). 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 (Clark et al 2016, Messmer 2015). 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 (Gorka, 2019) 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 (Mathews, 2015).

In addition, when studying the correlation between profession and clinical visits, it appeared that home makers, employees in the government and private sectors, and students make the top three categories of those are who most affected. Figure 2 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 (Diaz, 2011). 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 (Diaz, 2011).

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 (Murthy, 2005).

B. *Occupation and Clinical Visits*

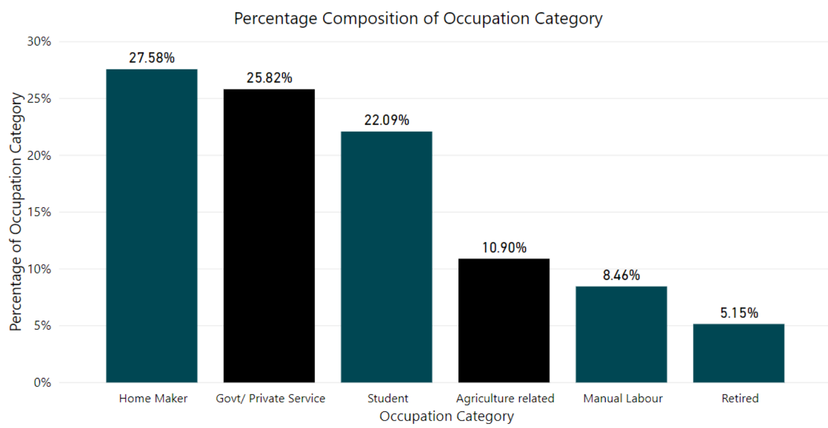


Figure 6. Clinical Visits by Profession (2011-2019)

C. *Location and Eye Disorders in Older 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 7. Location and Eye Disorders in Older Age Population

Cataract seemed to be the most disease that has affected older age in Telangana. Cataract is a condition known to affect older age, and this study revalidates the information.

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

Age Category											
Select All	> 100 yrs	0-10 yrs	11-20 yrs	21-30 yrs	31-40 yrs	41-50 yrs	51-60 yrs	61-70 yrs	71-80 yrs	81-90 yrs	91-100 yrs

Location	Number_of_visits
Paloncha	531
Kondapur	188
Adilabad	181
Madhapur	171
Kukatpally	149
Bhadrachalam	131
Gachibowli	125
Kothagudem Bazar	119
Miyapur	119
Kothagudam	98

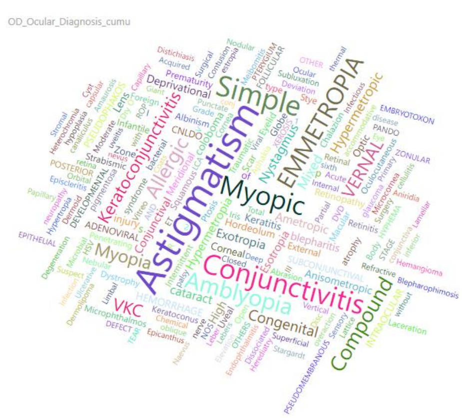


Figure 8. Location and Eye Disorders in Younger Age (0-10) Population

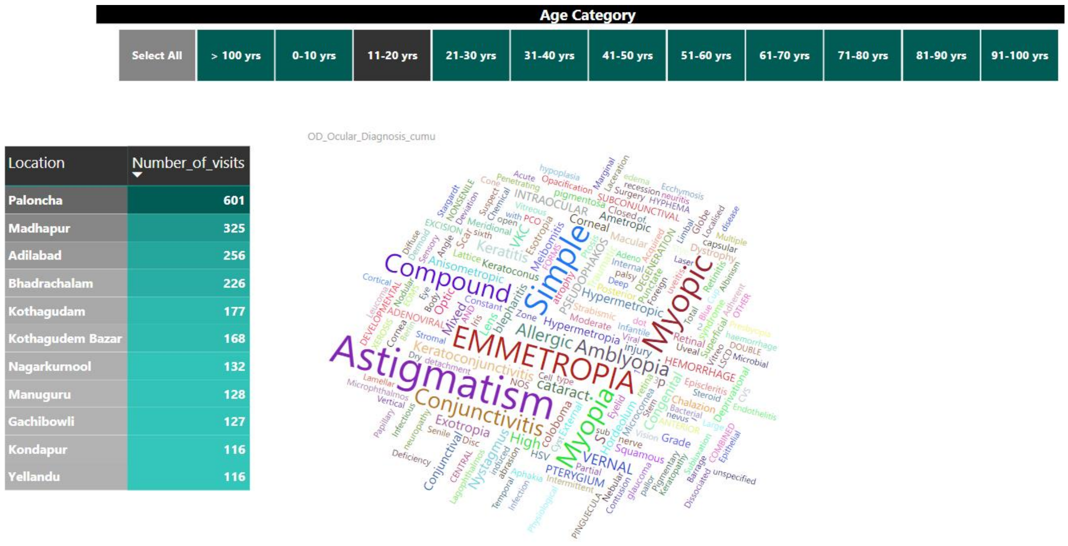


Figure 9. Location and Eye Disorders in Younger Age (11-20) Population

Astigmatism, which is an irregularity of the shape of the cornea was present in younger age population as per Figures 4 and 5. 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 (Kaur, 2019). 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 (Kaur, 2019).

E. Consistent Prevalence of Cataract in Rainfall

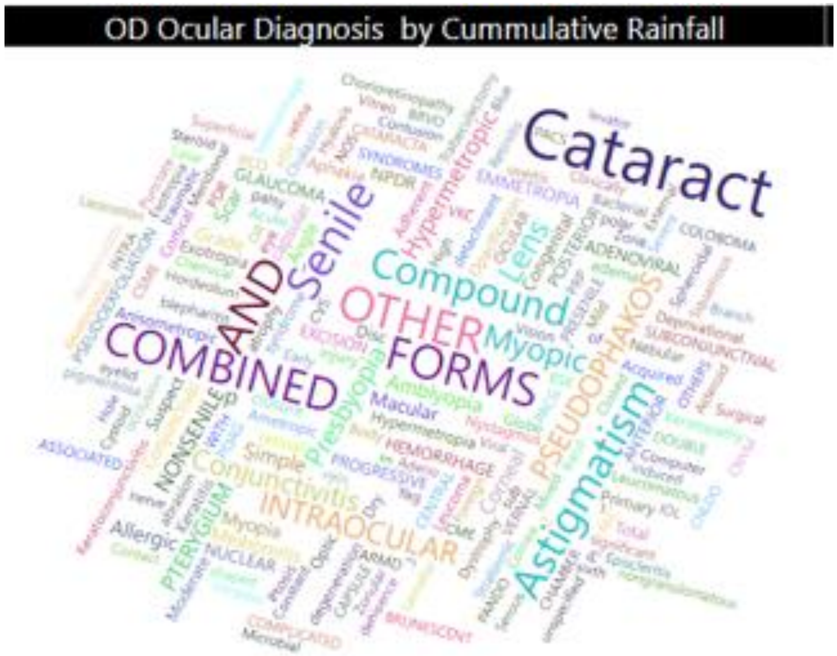


Figure 10. Cataract Most Prevalent Disease in Rainfall

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 rainfall.

Phase 2 - Analysis of Cataract in Telangana

After completing phase 1 of the research, it has been revealed from the data that there is a high cataract rate among patients, and therefore this has led to examining cataract as the main disease in this paper. In addition, cataract is a major cause of vision impairment in many low-income settings (Vashit et al, 2011). The analysis identified demographic correlations to cataract development as well as a potential relationship with weather variables which have been tested in several ways throughout the data analysis process.

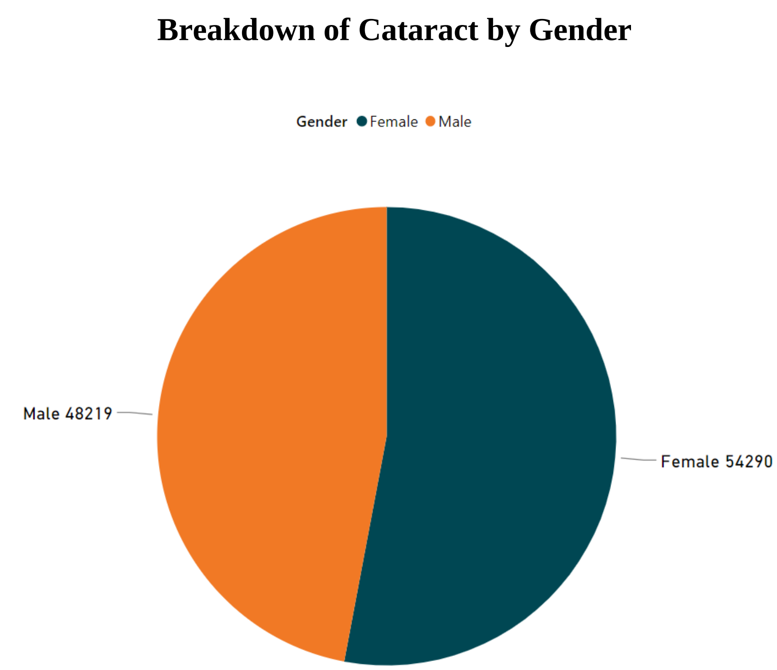


Figure 11. Cataract by Gender

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. The figure 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.

Breakdown of Cataract Patients by District

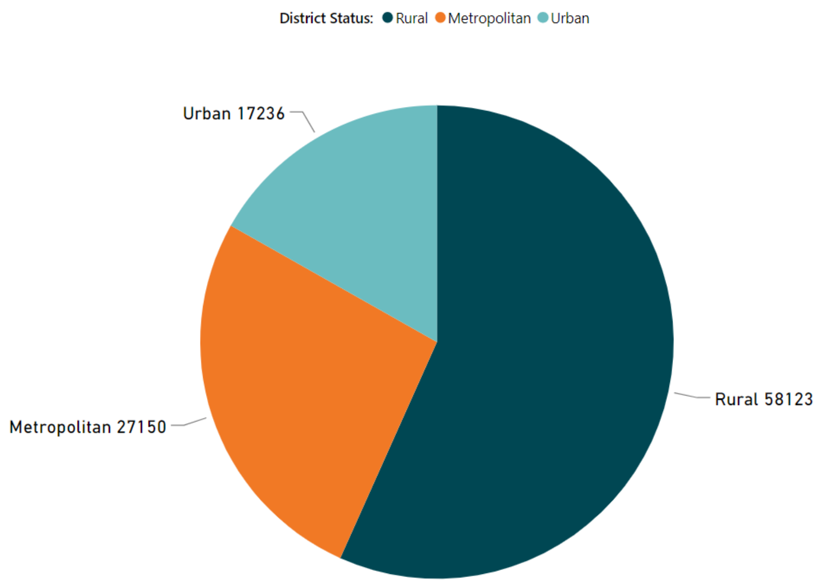


Figure 12. Cataract by District

The above figure 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 the rural 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 (Thevi, et al, 2012). 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 among the middle and poor resourced families (Thevi, et al, 2012). 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 the female patients (Kaur, 2018).

Moreover, the awareness about the 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 (2015) and Murthy et al. (2014) have supported this opinion. Moreover, again as per the affirmations of Kuldeep Dole (2013) who conducted a cross sectional survey highlighted about the expanded charges of the modernized strategies involved for cataract such as intraocular lens implantation and microincision phaco. Another important reason behind the rising number of cases in the rural areas are due to their mindset to utilize the traditional eye medicines which demonstrated contradictory outcomes due to the toxicity and infections caused by the agents.

Cataract Patients by Age

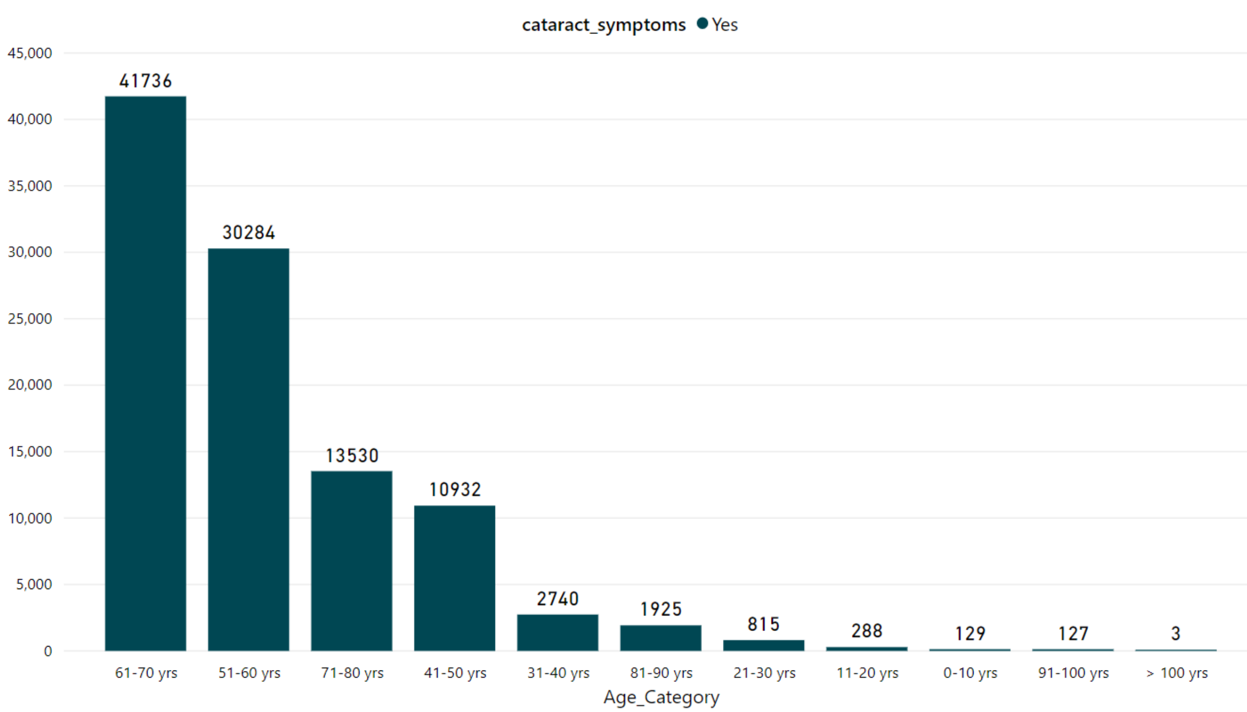


Figure 13. Cataract by Age

Figure 13 shows the distribution of cataract patients among the different age categories. People in the age category of 61-70 have the higest recorded cases of cataract with 41,736 cases of cataract recorded in this age group. This is followed by people within the age bracket of 51-60, with 30,284 cataract cases recorded and 13,530 cataract cases recorded within the age bracket of 71-80. The younger population below the age of 30 do not seem to present themselves with cataract symptoms. Previously in the research, it was explained that Astigmatism and Conjunctivitis are the most prevalent eye diseases among the youth population (ages 21-40) in Telangana.

Cataract Patients as influenced by Occupation

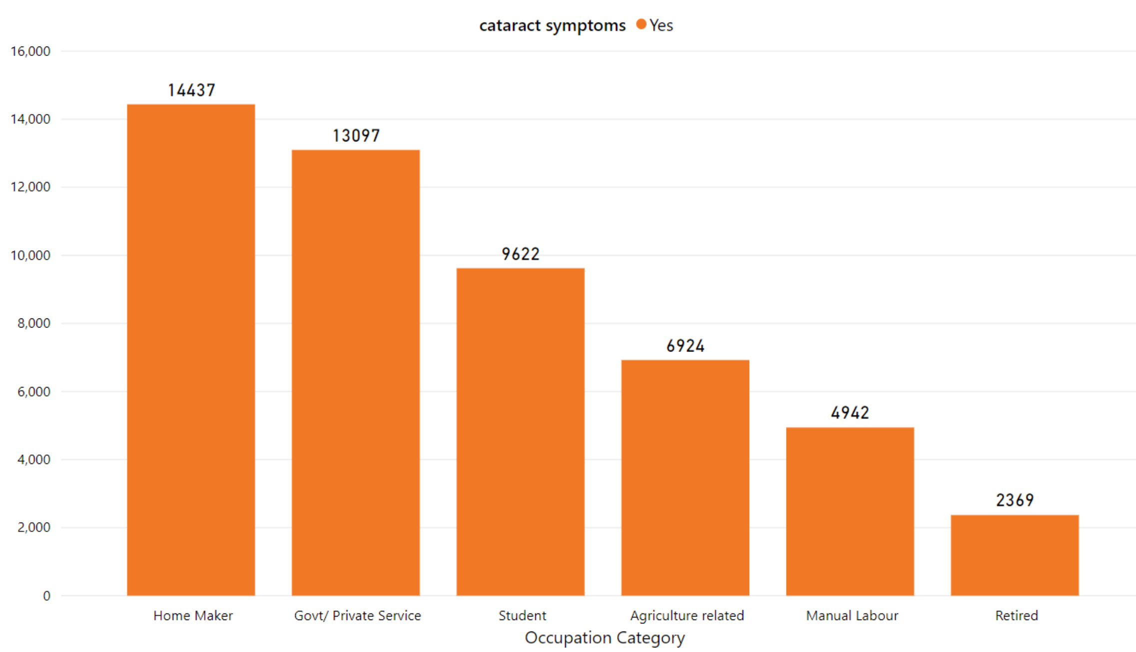


Figure 14. Cataract by Occupation

The figure above shows how the different occupation categories are influenced with cataract as a disease amongst patients in that occupation. Homemakers recorded the highest cases of cataract with 14,437 cases recorded. This is followed by people in Government/Private Service with 13,097 cataract cases recorded, and then students at 9,622. This is in line with the correlation between profession and clinical visits on the overall EMR studied in Telangana. It can be concluded that certain professions are a general reason of why eye diseases in general are present in patients. Homemakers, employees in the government and private sectors, and students make the top three categories of those are who most affected. It is also seen that workers in agriculture and who are defined as 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.

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 (Murthy, 2005).

Cataract Patients tested for Smoking and Diabetes

According to a scientific meta analysis conducted by Ye, et al., 2012 a overall of 13 prospective cohort along with 8 case control investigations were included within the analysis. The findings of the investigation highlighted that smoking status has a positive correlation with age related cataract (ARC) condition. The study showed the statistical validation of (OR 1.41, 95% CI 1.23–1.62) for the cohort group of investigation and (OR 1.57, 95% CI 1.20–2.07) for the case control gathering of the investigation. Moreover, the subgroup analysis of the investigation also focused about the positive association in between smoking and the development of nuclear cataract having the statistical validation of (NC; OR 1.66, 95% CI 1.46–1.89). Same trends of outcomes were observed also for the category of the case controlled studies NC OR 1.86, 95% CI 1.47–2.36 and also for the cases of posterior subcapsular cataract having the statistical validation with the Odds Ratio of 1.60 and 95% CI 0.97–2.65. With regard to the pathogenesis of the condition, the researchers within the study highlighted that the smoke of tobacco possesses varied range of harmful poisonous chemicals for instance nicotine, and obnoxious gas carbon monoxide, free radicals that cumulatively enhances the oxidative induced stress and therefore eventually plays a significant role for the development of cataract (Beebe, et al, 2010; Truscott, 2005).

Another potential factor is the presence of diabetes for the development of cataract among both the populace of developing and developed countries. Though the exact pathogenesis is still not vivid with regard to diabetes however, past scientific confirmations have revealed that the polyol pathway calls for the initiation of the pathological condition (Kiziltoprak, et al, 2019). Therefore, diabetes is considered as potential risk factor for cataract development and also enhances the hazard factor for the intricacies arising after phacoemulsification surgical process among the diabetic populace in comparison to the non diabetic group (Kiziltoprak, et al, 2019). Throughout the world the comprehension of the pathological mechanism for the hindrance or delaying of the development of cataract is still under research and is a matter of challenge among the diabetic populace (Pollreisz, et al, 2010).

Table 1: Diabetes and Smoking Patients in Relation to Cataract

Count of uid	cataract_symptoms	Smoking Symptom	Diabetic Symptom
102479	Yes	No	No
28	Yes	No	Yes
2	Yes	Yes	No
102509			

NB: Count of uid, stands for the count of patients' ID, a unique identifier of each patient.

The table able above shows the number of cataract patients who are either diabetic, have history/symptom of smoking or both.

A total of **102,509** cataract patients were recorded.

102,479 cataract patients are neither diabetic nor have symptom of smoking. This means these patients had only symptom of cataract.

28 of the cataract patients are diabetic, while only **2** of the cataract patients had symptom of smoking.

However, in the study conducted, it seemed there is a very low association between diabetes and cataract. Similarly, hardly any association was available between smoking and cataract within the study population. These may be attributed to the error occurred during the time of assortment of data as not every detail may have been reported for all patients within the EMR. According to scientific evidence, cataract is considered to be a multifactorial disease with varied factors part, such as hereditary material, age, nutrition, status, exposure to ultraviolet radiation, and trace metals can also be the causal factors for the development of cataract (Balasubramanian, et al, 1993) .

Cataract Patients Tested Against Weather Variables

Plotting Only Cataract column against Weather variables:

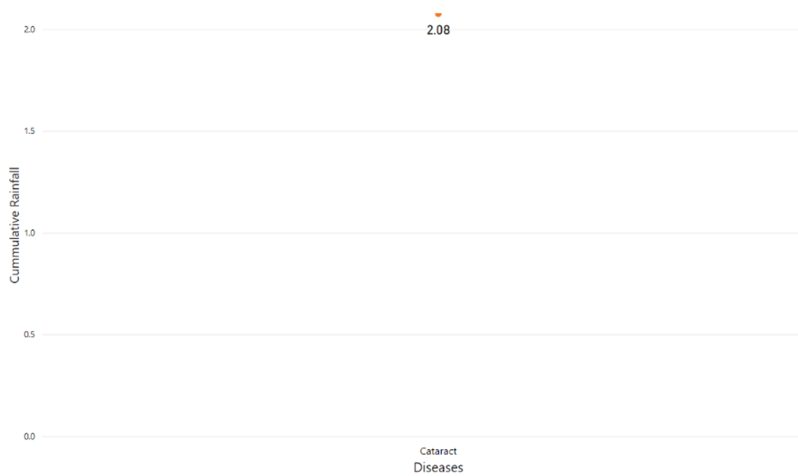


Figure 15. Cataract by Weather Variables

This particular plotting has shown a correlation with continuous variable cumulative rainfall and cataract and this is in accordance with the investigation of Balasubramanian, et al, 1993 as he reported that cataract was found to be more frequent in cloudy weather conditions as pupils of the eye expands more to allow more light to enter and the lens gets enhanced exposure to UV radiations. Moreover, water vapors can filter IR not UV radiation.

The data has been generated with the aid of data analytics tool to explore the association in between specific cataract condition and variables of weather. The groups were categorized as cataract patients and non cataract ones so that the impact of weather condition can be correlated with ocular diseases (Refer Appendix for the reports).

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

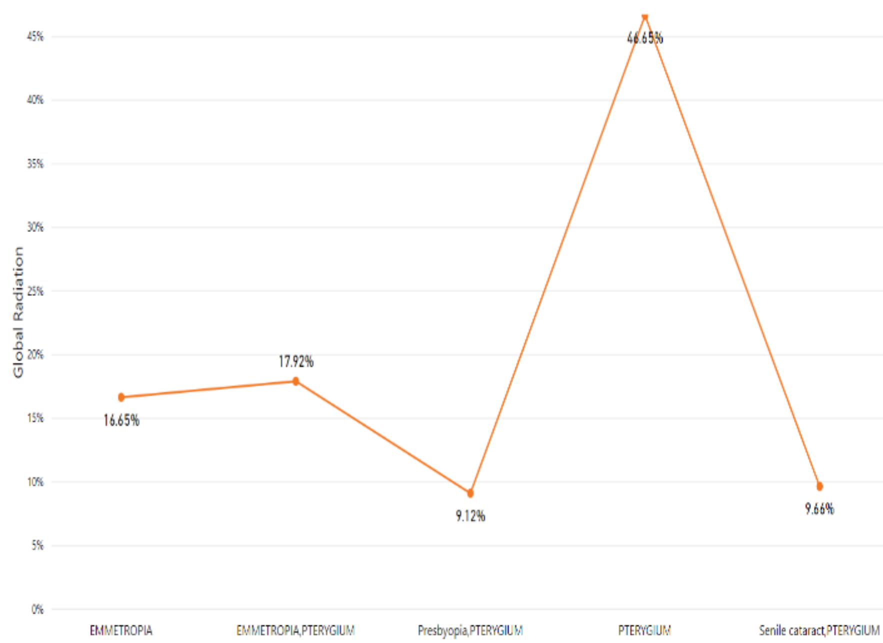


Figure 16. Pterygium in Relation to Global Radiation

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 (pinkish tissue growth on the cornea) 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.

Consistent Prevalence of Pterygium in Relation to Windspeed

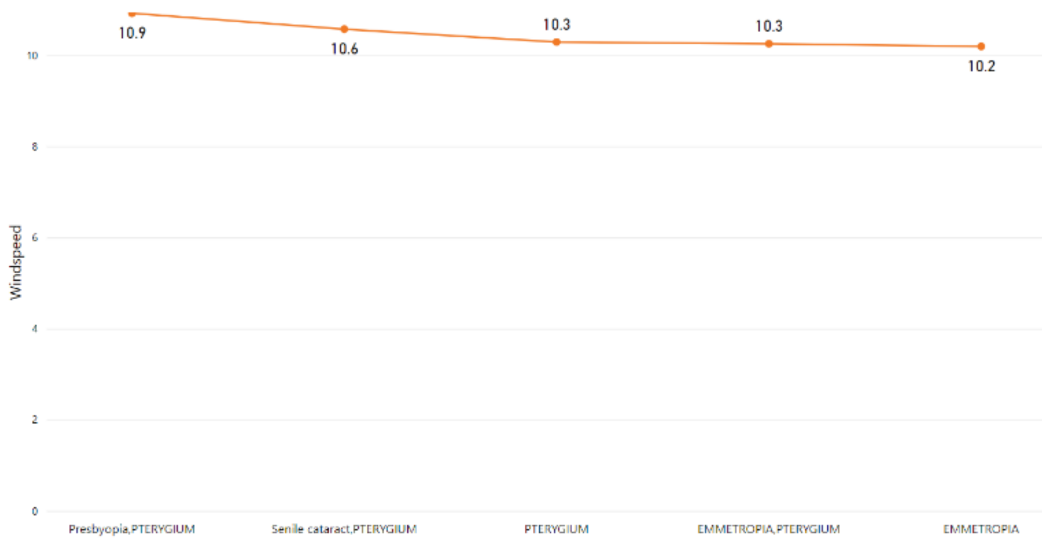


Figure 17. Pterygium in Relation to Windspeed

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

When the outcomes for global radiation exposure of non - cataract or emmetropia populace is compared with cataract or ocular disease populace, a huge difference is observed. The figures demonstrated that predominance of Pterygium is highly associated with global radiation, however with wind speed factor all the conditions demonstrated equivalent effects. A study conducted by Shah, et al, 2016 and Balasubramanian, et al, 1993 highlighted that cataract has a very close association with exposure to radiation. It is evident that in Australia where people were exposed to expanded level of UV radiation showed early initiation of age-related cataract (Hollows, et al, 1981). Moreover, in Nepal, the predominance of cataract was 3.8 times more where people were exposed to more than 12 hours of outdoor exposure (Brilliant, et al, 1983). A study conducted by Nordmann, 1962 also highlighted radiation as causal factor for cataract and ocular diseases (Nordmann, 1962).

All tables and charts generated are included in the appendix.

CHAPTER 5: CONCLUSION AND RECOMMENDATIONS

Responses to Research Questions

This research project 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.

The first research question that guided the study was:

1. **Does the Cynefin Framework offer potential as a means to transform raw data into a form which can be used by healthcare researchers and practitioners aiming to increase early cataract detection rates in Telangana?**

The scheme of Cynefin framework acts as a guidance to healthcare practitioners and researchers because of its foundation in the management of information (Snowden, 2007). 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 (Snowden, 2007). It has varied domains which necessitate different actions, for instance, the straight forward 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 (Kempermann, 2017). 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" (Kempermann, 2017). This means that the chaotic or unordered state of the world requires pattern dependent management for proper orientation and right decision making (Kempermann, 2017).

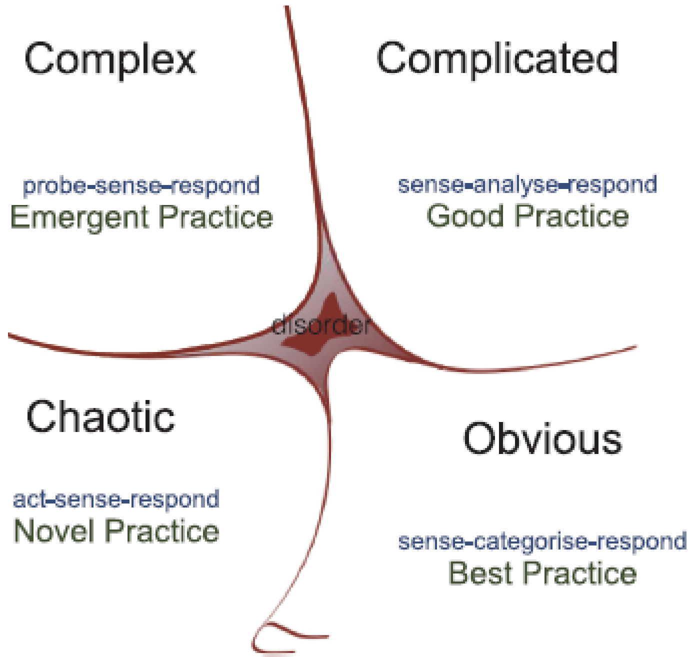


Figure 18: Varied Types of Medical Complexities

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 those factors can occur at varied levels via recursive feedback loops (Gray, B., 2017).

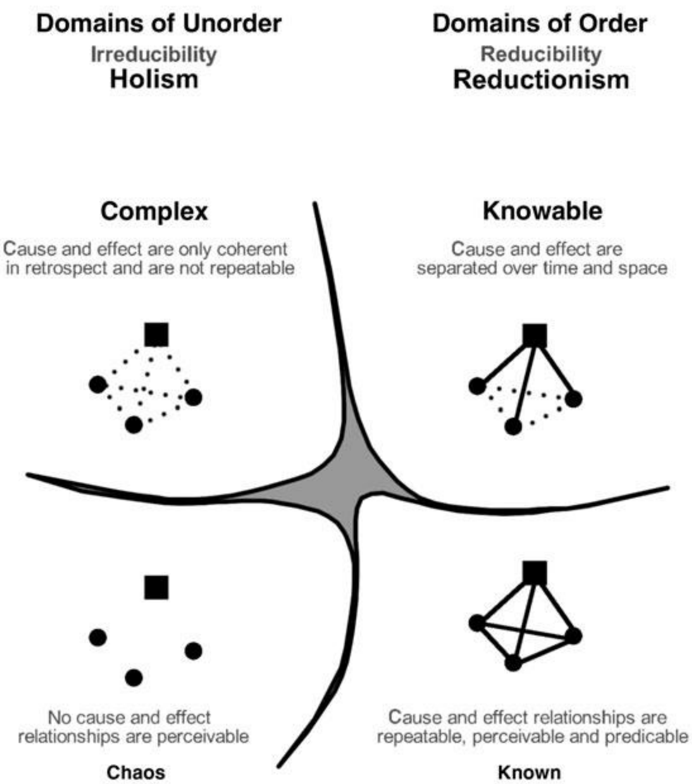


Figure 19: Cause Effect Association Interpretation Using Cynefin Framework

The above figure refers to the cause and effect association between the parameters decoded with the aid of Cynefin framework (Sturmberg, 2009). In the figure, the square block indicates the "causal agent", and the dots indicate the "effect agent" (Sturmberg, 2009). 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 (Sturmberg, 2009).

The concept of medicine is complex and the complexity that arises at the time of general practice varies widely (Gray, B., 2017). 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 complimentary in nature (Gray, B., 2017). 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 (Kempermann, 2017). 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 (Kempermann, 2017). 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 (Kempermann, 2017).

The second research question that guided the study was:

2. **Can data merging techniques be applied to EMRs to generate new knowledge shedding light on correlations between the development of cataracts and socio-demographic and environmental factors among residents of the state of Telangana, southern India?**

Although there is no one specific definition for "data merging", it is inferred that it is a process that brings together various datasets under one cloud to achieve a more structured data for analysis. The structured dataset in this research was the master data set that was created by merging patient information from Eye Smart with weather variables. The process to test the success of the merging technique came from various trials and errors into reading and mapping the data. The merging process was very timely, and also created multiple errors in downloading and converting the file throughout the process, until finally creating a manageable file for analysis on Python.

In addition, what appeared most challenging was that data structuring in healthcare is difficult because of the high level of ambiguity and complexity in the data concepts themselves. Once you get past the simplicity of standard demographic data such as age and gender, which are fairly easy to analyze, you quickly find that clinical data that is dependent on multiple symptoms, diagnosis, and treatments, are not as easy to understand, and therefore it is necessary to have a practitioner in the field to guide you through the process of analysis. In the case of analyzing cataract patients, Dr. Anthony Vipin offered tremendous guidance, which helped in understanding many of the nature of the eye disorders, specifically cataract.

Overall, by discovering associations and understanding patterns and trends within the data, big data analytics has the potential to improve care, save lives and lower costs. Thus, big data analytics applications in healthcare take advantage of the explosion in data to extract insights for making better informed decisions (Raghupathi, 2014). Potential benefits include detecting diseases at earlier stages when they can be treated more easily and effectively; managing specific individual and population health and detecting health care fraud more quickly and efficiently. Numerous questions can be addressed with big data analytics.

Health data volume is expected to grow dramatically in the years ahead [6], and this method, if used effectively, can help in the discovery of associations and understanding of patterns and trends within the data, which can offer the potential to improve care and save lives. Thus, big data analytics applications in healthcare take advantage of the explosion in data to extract insights for making better informed decisions (Raghupathi, 2014).

Influence Diagram - A Systems Approach to Patient Care

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 (Keolling and Schwandt, 2005). 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.

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.

Influence Diagram - Taking Knowledge to Wisdom

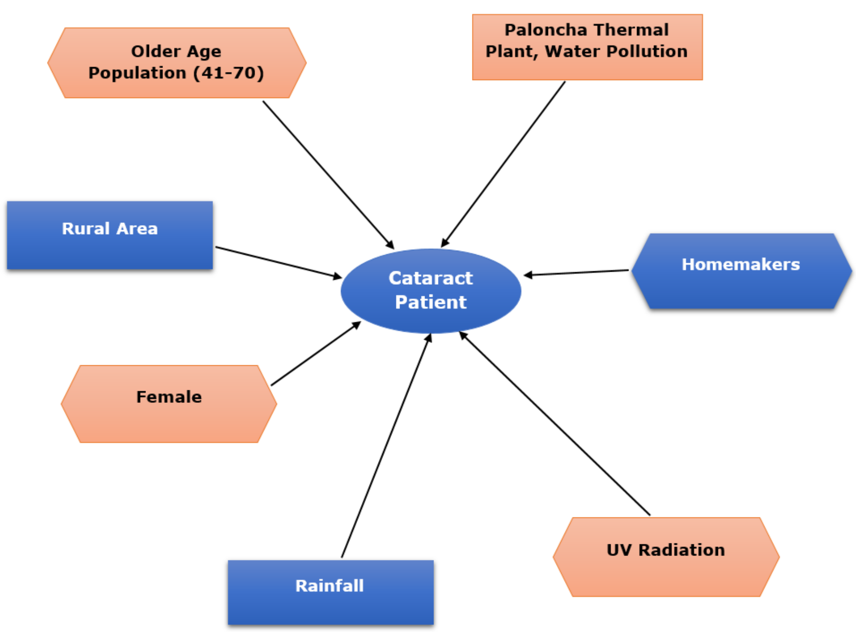


Figure 20: Influence Diagram of Cataract Patients and Co-Factors Affecting Cataract

Summary

This data analytics study provided an expanded exploration of how socio-demographic and climatic factors affect the prevalence of cataracts in Telangana. Applying several statistical techniques, including pattern recognition, and generating other data visualizations, I was able to validate previously identified findings and shed the light on the socio-demographic and weather factors that correlate to the development of cataracts. The machine learning tools that were used throughout the study showed that they are useful for a discovery research to better understand the sample set of patients and to generate informative and understandable visuals.

From a systems perspective, the ‘wisdom’ that was achieved was depicted in an influence diagram that showed the co-factors that result in the development of cataracts. 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, the study showed that variables must be minimized, in order to capture impactful underlying knowledge, or otherwise patterns will be harder to spot.

As a researcher, I recommend that authorities in the state of Telangana 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, I recommend that ophthalmologists look beyond understanding medical symptoms only, and focus on all factors that influence the development of ocular diseases, such as social income, cultural upbringing, and weather and climate. This can lead to ophthalmologists offering 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.

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.

REFERENCES

Ahuja, A.S., 2019. The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ*, *7*, p.e7702.

Akkara, J.D. and Kuriakose, A., 2019. Role of artificial intelligence and machine learning in ophthalmology. *Kerala Journal of Ophthalmology*, *31*(2), p.150.

Balasubramanian, D., Bansal, A.K., Basti, S., Bhatt, K.S., Murthy, J.S. and Rao, C.M., 1993. The biology of cataract. The Hyderabad cataract research group. *Indian journal of ophthalmology*, *41*(4), p.153.

Beebe, D.C., Holekamp, N.M. and Shui, Y.B., 2010. Oxidative damage and the prevention of age-related cataracts. *Ophthalmic research*, *44*(3), pp.155-165.

Boland, M.V., 2016. Big data, big challenges. *Ophthalmology*, *123*(1), pp.7-8.

Bresnick, J 2016, *Big Data, Artificial Intelligence, Iot May Change Healthcare In 2017*. Accessed 1 July 2020, [online] Available at: <https://healthitanalytics.com/news/big-data-artificial-intelligence-iot-may-change-healthcare-in-2017>.

BRILLIANT, L.B., Grasset, N.C., POKHREL, R.P., KOLSTAD, A., LEPKOWSKI, J.M., BRILLIANT, G.E., HAWKS, W.N. and Pararajasegaram, R., 1983. Associations among cataract prevalence, sunlight hours, and altitude in the Himalayas. *American journal of epidemiology*, *118*(2), pp.250-264.

Brown, A.P., 2010. Qualitative method and compromise in applied social research. *Qualitative Research*, *10*(2), pp.229-248.

Centers for Disease Control and Prevention (CDC, 2000. Information needs and uses of the public health workforce--Washington, 1997-1998. *MMWR. Morbidity and mortality weekly report*, *49*(6), p.118.

Clark, A., Ng, J.Q., Morlet, N. and Semmens, J.B., 2016. Big data and ophthalmic research. *survey of ophthalmology*, *61*(4), pp.443-465.

Consejo, A., Melcer, T. and Rozema, J.J., 2019, January. Introduction to machine learning for ophthalmologists. In *Seminars in ophthalmology* (Vol. 34, No. 1, pp. 19-41). Taylor & Francis.

Copeland, M., 2016. What's the Difference Between Artificial Intelligence. *Machine Learning, and Deep Learning*.

Copeland, B.J., 2019. Artificial intelligence. Encyclopaedia Britannica.

Crevier, D., 1993. *AI: the tumultuous history of the search for artificial intelligence*. Basic Books, Inc.

Davenport, T.H., 2014. How strategists use "big data" to support internal business decisions, discovery and production. *Strategy & Leadership*.

Dimitrov, D.V., 2016. Medical internet of things and big data in healthcare. *Healthcare informatics research*, *22*(3), pp.156-163.

Donthineni, P.R., Kammari, P., Shanbhag, S.S., Singh, V., Das, A.V. and Basu, S., 2019. Incidence, demographics, types and risk factors of dry eye disease in India: Electronic medical records driven big data analytics report I. *The ocular surface*, *17*(2), pp.250-256.

Du, X.L., Li, W.B. and Hu, B.J., 2018. Application of artificial intelligence in ophthalmology. *International Journal of Ophthalmology*, *11*(9), p.1555.

Esson, A 2018, *Deep Mind Diagnoses Eye Disease With Artificial Intelligence*. Accessed 7 July 2020. [online] Available at: <http://www.frontlinegenomics.com/news/19080/deepmind-diagnoses-eye-disease-with-artificial-intelligence/>

Foreman, J.W., 2013. *Data smart: Using data science to transform information into insight*. John Wiley & Sons.

Frost & Sullivan 2016, Frost & Sullivan, 2016 From $600 M to $6 billion, artificial intelligence systems poised for dramatic market expansion in healthcare. Accessed 1 July 2020, [online] Available at: <https://ww2.frost.com/news/press-releases/600-m-6-billion-artificial-intelligence-systems-poised-dramatic-market-expansion-healthcare/>

Galor, A., Kumar, N., Feuer, W. and Lee, D.J., 2014. Environmental factors affect the risk of dry eye syndrome in a United States veteran population. *Ophthalmology*, *121*(4), pp.972-973.

Gray, B., 2017. The Cynefin framework: applying an understanding of complexity to medicine. *Journal of Primary Health Care*, *9*(4), pp.258-261.

Gupta, S., 2016. *Real-Time Big Data Analytics*. Packt Publishing Ltd.

Heussner, N., Holl, L., Nowak, T., Beuth, T., Spitzer, M.S. and Stork, W., 2014. Prediction of temperature and damage in an irradiated human eye—Utilization of a detailed computer model which includes a vectorial blood stream in the choroid. *Computers in biology and medicine*, *51*, pp.35-43.

Hirschfeld, D., 2012. Twitter data accurately tracked Haiti cholera outbreak. *Nature*.

Hollows, F. and Moran, D., 1981. Cataract-the ultraviolet risk factor. *The Lancet*, *318*(8258), pp.1249-1250.

Hsieh, P., 2017. AI in medicine: rise of the machines. *Forbes*, *Forbes Magazine*, *30*.

Issa, N.T., Byers, S.W. and Dakshanamurthy, S., 2014. Big data: the next frontier for innovation in therapeutics and healthcare. *Expert review of clinical pharmacology*, *7*(3), pp.293-298.

Joshi, M.V., 2015. Epidemiological study of patients availing free cataract services of national programme of control of blindness. *Journal of Clinical Ophthalmology and Research*, *3*(1), p.9.

Kaur, K., 2018. Cataract Blindness: Socioeconomic Factors Associated with Treatment Barriers and High Blindness Rates for Women in Rural Regions of Andhra Pradesh.

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

Kitchin, R. and Lauriault, T.P., 2015. Small data in the era of big data. *GeoJournal*, *80*(4), pp.463-475.

Kiziltoprak, H., Tekin, K., Inanc, M. and Goker, Y.S., 2019. Cataract in diabetes mellitus. *World journal of diabetes*, *10*(3), p.140.

Kuck Jr, J.F.R., 1983. Effect of Long-Wave Ultraviolet Light on the Lens. *Ophthalmic research*, *15*(4), pp.216-219.

Liu, B., Xu, L., Wang, Y.X. and Jonas, J.B., 2009. Prevalence of cataract surgery and postoperative visual outcome in Greater Beijing: the Beijing Eye Study. *Ophthalmology*, *116*(7), pp.1322-1331.

Lu, P., Chen, X., Liu, X., Yu, L., Kang, Y., Xie, Q., Ke, L. and Wei, X., 2008. Dry eye syndrome in elderly Tibetans at high altitude: a population-based study in China. *Cornea*, *27*(5), pp.545-551.

Lu, W., Tong, Y., Yu, Y., Xing, Y., Chen, C. and Shen, Y., 2018. Applications of artificial intelligence in ophthalmology: general overview. *Journal of ophthalmology*, *2018*.

McDonald, C 2019, *Streaming Machine Learning Pipeline For Sentiment Analysis Using Apache APIs. MAPR*. Accessed 6 November, 2019. [Online]. Available at: <https://mapr.com/blog/author/carol-mcdonald/>

McMichael, A.J., Campbell-Lendrum, D.H., Corvalán, C.F., Ebi, K.L., Githeko, A., Scheraga, J.D. and Woodward, A., 2003. *Climate change and human health: risks and responses*. World Health Organization.

Morison, M., 1986. Methods in sociology.

Murthy, G.V., Jain, B.K., Shamanna, B.R. and Subramanyam, D., 2014. Improving cataract services in the Indian context. *Community Eye Health*, *27*(85), p.4.

Nelson, L., 2018. *Pediatric ophthalmology*. Lippincott Williams & Wilkins.

Nordmann, J., 1962. Cataracts caused by radiations. *L'Annee therapeutique et clinique en ophtalmologie*, *13*, p.35.

Ola, O. and Sedig, K., 2014. The challenge of big data in public health: an opportunity for visual analytics. *Online journal of public health informatics*, *5*(3), p.223.

Patel, V.L., Shortliffe, E.H., Stefanelli, M., Szolovits, P., Berthold, M.R., Bellazzi, R. and Abu-Hanna, A., 2009. The coming of age of artificial intelligence in medicine. *Artificial intelligence in medicine*, *46*(1), pp.5-17.

Patel, V.L., Yoskowitz, N.A. and Arocha, J.F., 2009. Towards effective evaluation and reform in medical education: a cognitive and learning sciences perspective. *Advances in Health Sciences Education*, *14*(5), pp.791-812.

Pollreisz, A. and Schmidt-Erfurth, U., 2010. Diabetic cataract—pathogenesis, epidemiology and treatment. *Journal of ophthalmology*, *2010*.

Pratt, M.K., 2018. Artificial intelligence in primary care. *Medical economics*, *95*.

Quer, G., Muse, E.D., Nikzad, N., Topol, E.J. and Steinhubl, S.R., 2017. Augmenting diagnostic vision with AI. *The Lancet*, *390*(10091), p.221.

Raghupathi, W., & Raghupathi, V.(2014). *Big data analytics in healthcare: Promise and potential. Health Information Science and Systems*, *2*(1), p.3.

Raghupathi, W. and Raghupathi, V., 2014. Big data analytics in healthcare: promise and potential. *Health information science and systems*, *2*(1), p.3.

Rao, A.R., Chhabra, A., Das, R. and Ruhil, V., 2015, October. A framework for analyzing publicly available healthcare data. In *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)* (pp. 653-656). IEEE.

Saxena, R., Srivastava, S., Trivedi, D., Anand, E., Joshi, S. and Gupta, S.K., 2003. Impact of environmental pollution on the eye. *Acta Ophthalmologica Scandinavica*, *81*(5), pp.491-494.

Schachat, A.P., 2019. The year's most intriguing papers from the journal, Ophthalmology. *Presented at: Hawaiian Eye*.

Schachat, A.P., 2019. Ophthalmology Retina is now indexed in MEDLINE/PubMed. *Ophthalmology Retina*, *3*(6), pp.457-458.

Shah, S.I.A., Shah, S.A. and Rai, P., 2016. Factors associated with pterygium based on history and clinical examination of patients in Pakistan. *Journal of current ophthalmology*, *28*(2), pp.91-92.

Snowden, D.J. and Boone, M.E., 2007. A leader's framework for decision making. *Harvard business review*, *85*(11), p.68.

Society for the Study of Artificial Intelligence and Simulation of Behavior 2016. What is artificial intelligence? Accessed 01 July 2020 [online] Available at <https://aisb.org.uk/new_site/?p=73 https://aisb.org.uk/new_site/?p=73>

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

Thevi, T., Basri, M. and Reddy, S.C., 2012. Prevalence of eye diseases and visual impairment among the rural population–a case study of Temerloh hospital. *Malaysian family physician: the official journal of the Academy of Family Physicians of Malaysia*, *7*(1), p.6.

Ting, D.S.W., Pasquale, L.R., Peng, L., Campbell, J.P., Lee, A.Y., Raman, R., Tan, G.S.W., Schmetterer, L., Keane, P.A. and Wong, T.Y., 2019. Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology*, *103*(2), pp.167-175.

Topol, E.J., 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, *25*(1), pp.44-56.

Truscott, R.J., 2005. Age-related nuclear cataract—oxidation is the key. *Experimental eye research*, *80*(5), pp.709-725.

Vashist, P., Talwar, B., Gogoi, M., Maraini, G., Camparini, M., Ravindran, R.D., Murthy, G.V., Fitzpatrick, K.E., John, N., Chakravarthy, U. and Ravilla, T.D., 2011. Prevalence of cataract in an older population in India: the India study of age-related eye disease. *Ophthalmology*, *118*(2), pp.272-278.

Vipin, A., 2012. EyeSmart EMR. *Health Management, 7(1)*.

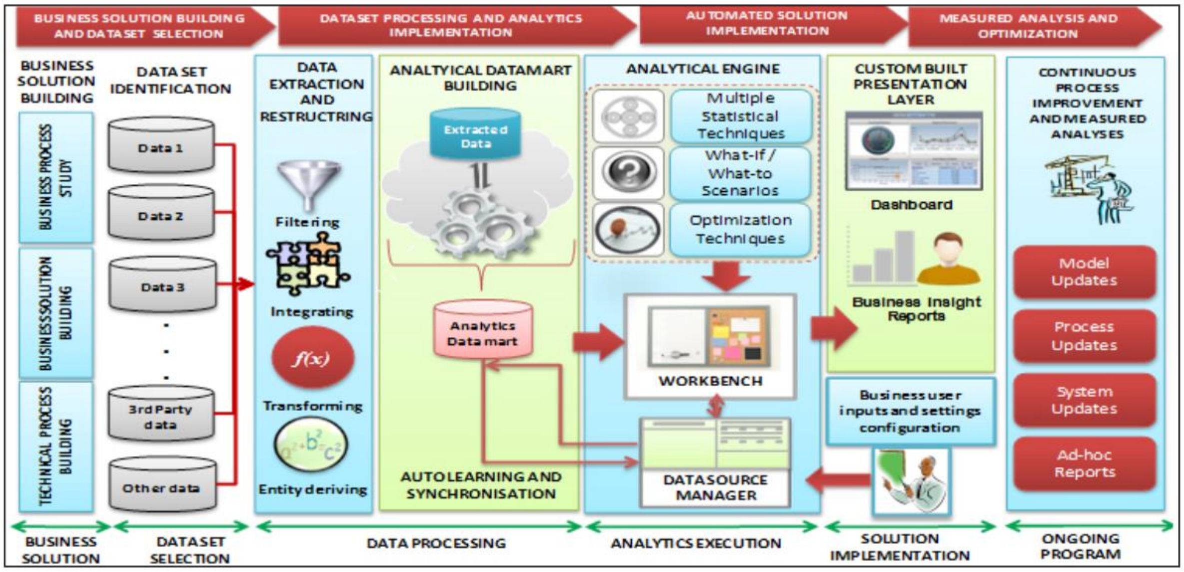
Yang, J.J., Li, J., Mulder, J., Wang, Y., Chen, S., Wu, H., Wang, Q. and Pan, H., 2015. Emerging information technologies for enhanced healthcare. *Computers in industry*, *69*, pp.3-11.

Ye, J., He, J., Wang, C., Wu, H., Shi, X., Zhang, H., Xie, J. and Lee, S.Y., 2012. Smoking and risk of age-related cataract: a meta-analysis. *Investigative ophthalmology & visual science*, *53*(7), pp.3885-3895.

Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J.F. and Hua, L., 2012. Data mining in healthcare and biomedicine: a survey of the literature. *Journal of medical systems*, *36*(4), pp.2431-2448.

Zhou, X., Chen, S., Liu, B., Zhang, R., Wang, Y., Li, P., Guo, Y., Zhang, H., Gao, Z. and Yan, X., 2010. Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. *Artificial Intelligence in medicine*, *48*(2-3), pp.139-152.

Appendix A



Source: http://www.statanalytics.com/images/analytics-services-over.jpg

Appendix B (I): Reading the datasets with Python in Jupyter Notebook

Telangana Weather data analysis Project

Import Python data analysis library

In [2]:
```
import pandas as pd
import seaborn as sns
```

Read in all the 3 data sets using the pandas library

In [5]:
```
data_excluding_Karmn = pd.read_excel('Telangana Excluding KARMN Excel data.xlsx')
Karmn_11_15_data = pd.read_excel('Karmn_11_15_data.xlsx')
Karmn_16_19_data = pd.read_excel('Karmn_16_19_data.xlsx')
```

Appendix B (II): Exploring the read datasets

Check the firt few records of one of the data read (data_excluding_Karmn)

This confirms the data is read in as expected

```
In [6]: data_excluding_Karmn.head()
```

Out[6]:

	Center_Short_Code	th_hospital_type	Center_Category	uid	tp_mrno	Age	Gender	Age_Category	Patient_Category	Patient_Category_Status	...	Advised_
0	CORSC	2	RURAL	15833	CC15028	24	Male	21-30 yrs	GP	Paying	...	
1	CORSC	2	RURAL	15835	CC15030	24	Male	21-30 yrs	GP	Paying	...	
2	CORSC	2	RURAL	15839	CC15034	26	Female	21-30 yrs	GP	Paying	...	
3	CORSC	2	RURAL	15840	CC15035	31	Male	31-40 yrs	GP	Paying	...	
4	CORSC	2	RURAL	15842	CC15037	31	Male	31-40 yrs	GP	Paying	...	

5 rows × 176 columns

Appendix B (III): Merging the two subsets of the datasets

Merge Karmn_11_15 data and Karmn_16_19 data

```
In [10]: karmn_data = Karmn_11_15_data.append(Karmn_16_19_data, ignore_index = True)
```

Check the shape of the combined karmn_data

The karmn_data contains a total of 428736 records of patients

```
In [11]: karmn_data.shape
Out[11]: (428736, 176)
```

Merge the data_excluding_Karmn with the Karmn data to create a combined Master data

```
In [15]: master_data = karmn_data.append(weather_data, ignore_index = True)
```

Appendix B (IV): Merging all data and exporting for analysis

Merge the data_excluding_Karmn with the Karmn data to create a combined Master data

In [15]:
```
master_data = karmn_data.append(weather_data, ignore_index = True)
```

Extraction & cleaning of data relevant for analysis

In [16]:
```
master_data = master_data[['Center_Short_Code', 'th_hospital_type', 'uid', 'Age', 'Gender', 'Age_Category',
              'Age_Category', 'Patient_Category_Status', 'Location', 'DistrictStatus',
              'Cumm_Rainfall', 'Temperature_min', 'Temperature_max', 'Patient_Occupation', 'OD_Ocular_Diagnosis_cumu',
              'OS_Ocular_Diagnosis_cumu', 'Humidity_min', 'Humidity_max', 'Month_First_Visit','SystemicDiagnosis', 'Noe
```

Check the dimension of the final merged master data

The Master data contains a total of 873,447 records of patients

In [16]:
```
master_data.shape
```

Out[16]:
```
(873447, 176)
```

Export the merged data to a csv file for analysis on Microsoft Power BI

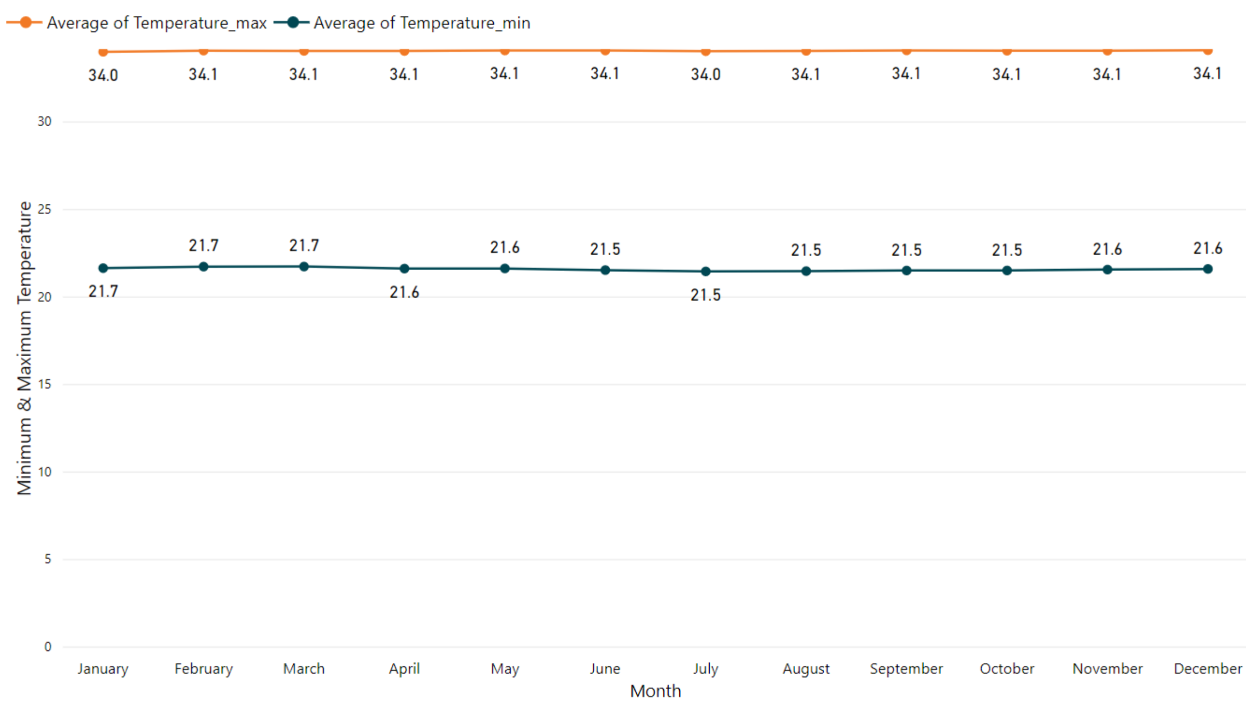
In [18]:
```
master_data.to_csv('Master Weather data.csv')
```

Appendix C: Cataract Patients as influenced by Minimum & Maximum Temperature Conditions

Diseases	cataract symptoms	Average of Temperature_min	Average of Temperature_max
INTRAOCULAR LENS, PSEUDOPHAKOS Nuclear cataract	Yes	21.30	33.73
INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	Yes	21.47	33.92
INTRAOCULAR LENS, PSEUDOPHAKOS Total cataract	Yes	21.45	34.09
Nuclear cataract	Yes	21.37	33.74
Nuclear cataract INTRAOCULAR LENS, PSEUDOPHAKOS	Yes	21.33	33.76
Nuclear cataract,INTRAOCULAR LENS, PSEUDOPHAKOS cataract	Yes	21.31	33.85
Posterior sub-capsular cataract	Yes	21.22	33.95
Posterior sub-capsular cataract - PSC	Yes	21.61	33.72
Senile cataract	Yes	21.48	33.91
Senile cataract cataract,INTRAOCULAR LENS, PSEUDOPHAKOS	Yes	21.34	33.98
Senile cataract INTRAOCULAR LENS, PSEUDOPHAKOS	Yes	21.46	33.91
Senile cataract,INTRAOCULAR LENS, PSEUDOPHAKOS cataract	Yes	21.34	33.99
Total cataract	Yes	21.46	34.17
Total cataract INTRAOCULAR LENS, PSEUDOPHAKOS	Yes	21.53	34.12
Total cataract,INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	Yes	21.33	33.99
TOTAL SENILE CATARACT,INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	Yes	21.25	33.86

The table above shows the prevalent diseases at different minimum & maximum temperature conditions.

Appendix D: Cataract Patients as influenced by Minimum & Maximum Temperature across the months

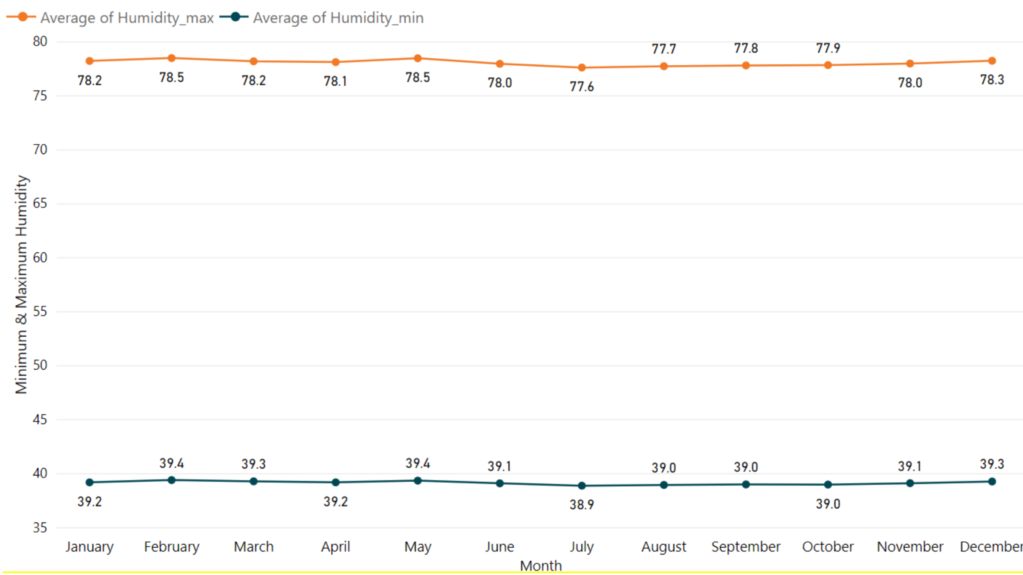


Appendix E: Cataract Patients as influenced by Minimum & Maximum Humidity Conditions:

Diseases	Average of Humidity_min	Average of Humidity_max	cataract symptoms
INTRAOCULAR LENS, PSEUDOPHAKOS Nuclear cataract	37.99	75.95	Yes
INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	38.43	77.25	Yes
INTRAOCULAR LENS, PSEUDOPHAKOS Total cataract	38.77	77.26	Yes
Nuclear cataract	38.13	76.32	Yes
Nuclear cataract,INTRAOCULAR LENS, PSEUDOPHAKOS cataract	38.06	75.85	Yes
Posterior sub-capsular cataract	37.99	75.60	Yes
Posterior sub-capsular cataract - PSC	38.38	77.03	Yes
Senile cataract	38.56	77.41	Yes
Senile cataract cataract,INTRAOCULAR LENS, PSEUDOPHAKOS	38.24	77.07	Yes
Senile cataract INTRAOCULAR LENS, PSEUDOPHAKOS	38.38	77.14	Yes
Senile cataract,INTRAOCULAR LENS, PSEUDOPHAKOS cataract	38.25	77.11	Yes
Total cataract INTRAOCULAR LENS, PSEUDOPHAKOS	38.89	77.48	Yes
Total cataract,INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	38.29	77.31	Yes

The table above shows the prevalent diseases at different minimum & maximum humidity conditions.

Appendix F: Cataract Patients as influenced by Minimum & Maximum Humidity Conditions across the months of the year



Appendix G: Cataract Patients as influenced by the cumulative rainfall

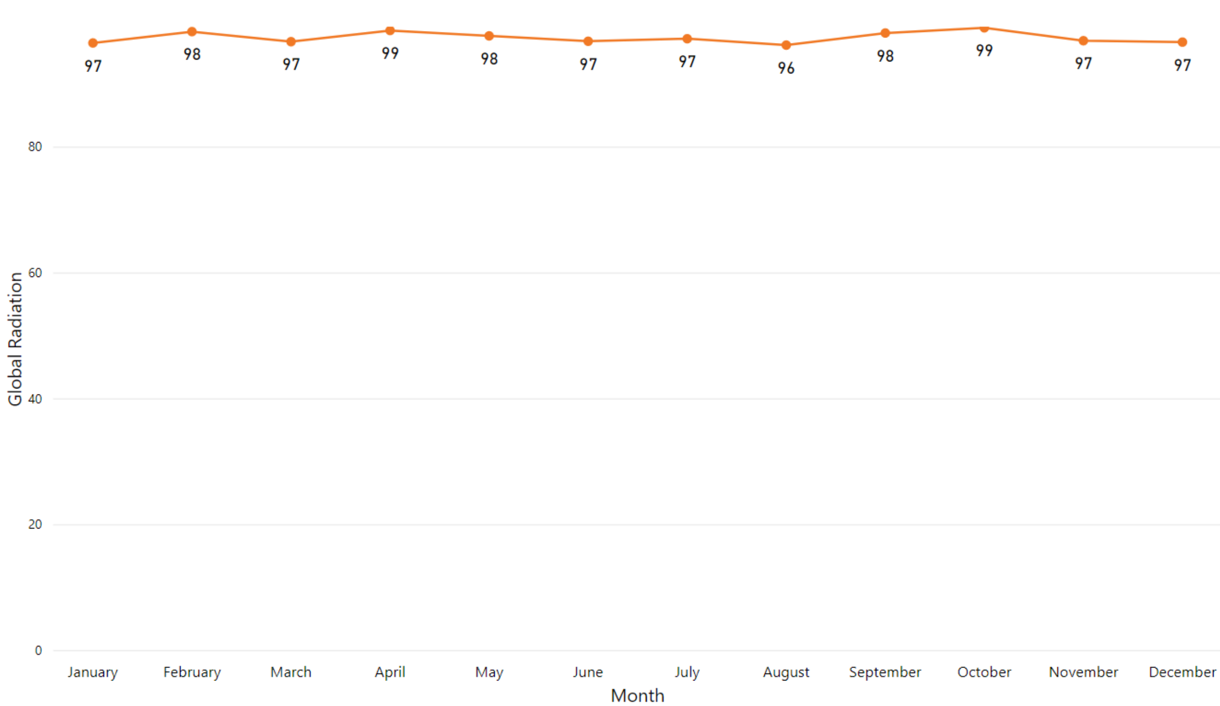
Diseases	Average of cumulative_Rainfall	cataract symptoms
INTRAOCULAR LENS, PSEUDOPHAKOS Nuclear cataract	2.14	Yes
INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	2.05	Yes
INTRAOCULAR LENS, PSEUDOPHAKOS Total cataract	2.21	Yes
Nuclear cataract	2.12	Yes
Nuclear cataract INTRAOCULAR LENS, PSEUDOPHAKOS	2.16	Yes
Nuclear cataract,INTRAOCULAR LENS, PSEUDOPHAKOS cataract	2.20	Yes
Posterior sub-capsular cataract	2.23	Yes
Posterior sub-capsular cataract - PSC	2.06	Yes
Senile cataract	2.06	Yes
Senile cataract cataract,INTRAOCULAR LENS, PSEUDOPHAKOS	2.05	Yes
Senile cataract INTRAOCULAR LENS, PSEUDOPHAKOS	2.06	Yes
Senile cataract,INTRAOCULAR LENS, PSEUDOPHAKOS cataract	2.05	Yes
Total cataract	2.25	Yes
Total cataract INTRAOCULAR LENS, PSEUDOPHAKOS	2.19	Yes
Total cataract,INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	1.98	Yes
TOTAL SENILE CATARACT,INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	1.87	Yes

Appendix H: Cataract Patients as influenced by Global radiation

Diseases	Average of Global_Radiation	cataract symptoms
INTRAOCULAR LENS, PSEUDOPHAKOS Nuclear cataract	102.69	Yes
INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	97.52	Yes
INTRAOCULAR LENS, PSEUDOPHAKOS Total cataract	99.04	Yes
Nuclear cataract	101.28	Yes
Nuclear cataract INTRAOCULAR LENS, PSEUDOPHAKOS	101.19	Yes
Nuclear cataract,INTRAOCULAR LENS, PSEUDOPHAKOS cataract	101.30	Yes
Posterior sub-capsular cataract	102.43	Yes
Posterior sub-capsular cataract - PSC	96.34	Yes
Senile cataract	98.01	Yes
Senile cataract cataract,INTRAOCULAR LENS, PSEUDOPHAKOS	97.79	Yes
Senile cataract INTRAOCULAR LENS, PSEUDOPHAKOS	97.87	Yes
Senile cataract,INTRAOCULAR LENS, PSEUDOPHAKOS cataract	97.73	Yes
Total cataract	98.25	Yes
Total cataract INTRAOCULAR LENS, PSEUDOPHAKOS	98.34	Yes
Total cataract,INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	97.61	Yes
TOTAL SENILE CATARACT,INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	96.99	Yes

The table above shows the prevalent diseases at different Global radiation conditions.

Appendix I: Cataract Patients as influenced by Global radiation across the months of the year



Appendix J: Cataract Patients as influenced by Windspeed:

Diseases	Average of Windspeed_max	cataract symptoms
TOTAL SENILE CATARACT,INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	11.13	Yes
Total cataract,INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	10.75	Yes
Total cataract INTRAOCULAR LENS, PSEUDOPHAKOS	10.05	Yes
Total cataract	10.02	Yes
Senile cataract,INTRAOCULAR LENS, PSEUDOPHAKOS cataract	10.36	Yes
Senile cataract INTRAOCULAR LENS, PSEUDOPHAKOS	10.01	Yes
Senile cataract cataract,INTRAOCULAR LENS, PSEUDOPHAKOS	10.38	Yes
Senile cataract	9.93	Yes
Posterior sub-capsular cataract - PSC	9.27	Yes
Posterior sub-capsular cataract	10.76	Yes
Nuclear cataract,INTRAOCULAR LENS, PSEUDOPHAKOS cataract	10.17	Yes
Nuclear cataract INTRAOCULAR LENS, PSEUDOPHAKOS	10.00	Yes
Nuclear cataract	9.87	Yes
INTRAOCULAR LENS, PSEUDOPHAKOS Total cataract	10.09	Yes
INTRAOCULAR LENS, PSEUDOPHAKOS Senile cataract	10.04	Yes
INTRAOCULAR LENS, PSEUDOPHAKOS Nuclear cataract	10.09	Yes

The table above shows the prevalent diseases at different Global Windspeed conditions.

Appendix K: Non-Cataract Patients with Weather Variable

Appendix K (I): Non-Cataract Patients as influenced by Cumulative Rainfall

Diseases	Average of cumulative_Rainfall	cataract symptoms
CATARACT, SENILE, OTHER AND COMBINED FORMS	2.32	No
Compound Hypermetropic Astigmatism	2.05	No
Compound Myopic Astigmatism	2.07	No
Compound Myopic Astigmatism Simple Myopia	2.06	No
EMMETROPIA	2.15	No
EMMETROPIA,Presbyopia	2.37	No
INTRAOCULAR LENS, PSEUDOPHAKOS	2.15	No
nan	2.13	No
Presbyopia	2.17	No
Simple Hypermetropia	2.10	No
Simple Hypermetropia,Presbyopia	2.25	No
Simple Myopia	2.12	No
Simple Myopia Compound Myopic Astigmatism	2.05	No
Simple Myopic Astigmatism	2.10	No
VKC - Vernal Keratoconjunctivitis	2.03	No

The table above shows the prevalent diseases affecting patients without cataract symptoms (Non-cataract patients) at different cumulative rainfall conditions.

Appendix K (II): Non-Cataract Patients as influenced by Minimum & Temperature conditions

Diseases	cataract symptoms	Average of Temperature_min	Average of Temperature_max
VKC - Vernal Keratoconjunctivitis	No	21.46	33.91
Simple Myopic Astigmatism	No	21.67	33.75
Simple Myopia Compound Myopic Astigmatism	No	21.59	33.71
Simple Myopia	No	21.69	33.83
Simple Hypermetropia,Presbyopia	No	21.83	34.29
Simple Hypermetropia	No	21.77	33.70
Presbyopia	No	21.46	33.98
nan	No	21.66	33.99
INTRAOCULAR LENS, PSEUDOPHAKOS	No	21.66	33.97
EMMETROPIA,Presbyopia	No	22.29	34.79
EMMETROPIA	No	21.71	33.98
Compound Myopic Astigmatism Simple Myopia	No	21.59	33.72
Compound Myopic Astigmatism	No	21.60	33.70
Compound Hypermetropic Astigmatism	No	21.60	33.63
CATARACT, SENILE, OTHER AND COMBINED FORMS	No	22.21	34.62

The table above shows the prevalent diseases affecting patients without cataract symptoms (Non-cataract patients) at different minimum & maximum temperature conditions.

Appendix K (III): Non-Cataract Patients as influenced by Minimum & Humidity conditions

Diseases	Average of Humidity_min	Average of Humidity_max	cataract symptoms
CATARACT, SENILE, OTHER AND COMBINED FORMS	41.69	82.08	No
Compound Hypermetropic Astigmatism	38.29	76.73	No
Compound Myopic Astigmatism	38.49	77.03	No
EMMETROPIA	39.00	77.70	No
EMMETROPIA,Presbyopia	42.38	83.22	No
INTRAOCULAR LENS, PSEUDOPHAKOS	38.87	77.55	No
nan	39.06	78.03	No
Presbyopia	38.46	76.78	No
Simple Hypermetropia	38.26	76.45	No
Simple Hypermetropia,Presbyopia	40.06	79.38	No
Simple Myopia	38.65	77.11	No
Simple Myopic Astigmatism	38.49	76.90	No
VKC - Vernal Keratoconjunctivitis	38.50	77.44	No

The table above shows the prevalent diseases affecting patients without cataract symptoms (Non-cataract patients) at different minimum & maximum humidity conditions.

Appendix K (IV): Non-Cataract Patients as influenced by Global Radiation

Diseases	Average of Global_Radiation	cataract symptoms
VKC - Vernal Keratoconjunctivitis	97.15	No
Simple Myopic Astigmatism	100.05	No
Simple Myopia Compound Myopic Astigmatism	100.79	No
Simple Myopia	99.75	No
Simple Hypermetropia,Presbyopia	95.61	No
Simple Hypermetropia	98.90	No
Presbyopia	99.92	No
nan	96.88	No
INTRAOCULAR LENS, PSEUDOPHAKOS	97.69	No
EMMETROPIA,Presbyopia	90.92	No
EMMETROPIA	98.35	No
Compound Myopic Astigmatism Simple Myopia	100.67	No
Compound Myopic Astigmatism	101.05	No
Compound Hypermetropic Astigmatism	100.92	No
CATARACT, SENILE, OTHER AND COMBINED FORMS	91.61	No

The table above shows the prevalent diseases affecting patients without cataract symptoms (Non-cataract patients) at different global radiation conditions.

Appendix K (V): Non-Cataract Patients as influenced by Windspeed:

Diseases	Average of Windspeed_max	cataract symptoms
VKC - Vernal Keratoconjunctivitis	10.10	No
Simple Myopic Astigmatism	9.21	No
Simple Myopia Compound Myopic Astigmatism	9.43	No
Simple Myopia	9.27	No
Simple Hypermetropia,Presbyopia	9.22	No
Simple Hypermetropia	8.85	No
Presbyopia	10.03	No
nan	9.47	No
INTRAOCULAR LENS, PSEUDOPHAKOS	9.40	No
EMMETROPIA,Presbyopia	8.23	No
EMMETROPIA	9.38	No
Compound Myopic Astigmatism Simple Myopia	9.42	No
Compound Myopic Astigmatism	9.33	No
Compound Hypermetropic Astigmatism	9.24	No
CATARACT, SENILE, OTHER AND COMBINED FORMS	8.41	No

The table above shows the prevalent diseases affecting patients without cataract symptoms (Non-cataract patients) at different windspeed conditions.