Intelligent Manufacturing and Cloud Computing
I.S. Jesus and K. Wang (Eds.)

doi:10.3233/ATDE241398

Enhancing Dry Eye Disease Detection Through the Application of an Ophthalmic Knowledge Graph

Jiazheng Cui[a,b,c,d], Zhiyuan Lin[d,e], Haoyang Liu[b,f], Yunxiao Liu[b,c,f] and Mini Han Wang[g,h,1]

[a] *Beijing Normal University - Hong Kong Baptist University United International College, Zhuhai China*
[b] *Digital Healthcare and Artificial Intelligence Association, Macao China*
[c] *Zhuhai Institute of Advanced Technology Chinese Academy of Sciences, Zhuhai, China*
[d] *Perspective Technology Group, Zhuhai, China*
[e] *Beijing Institute of Technology, Zhuhai, China*
[f] *Macao Polytechnic University, Macao, China*
[g] *The Faculty of Medicine, Chinese University of Hong Kong, Hong Kong, China*
[h] *Zhuhai People's Hospital (Zhuhai Clinical Medical College of Jinan University), Zhuhai, China*

Abstract. Dry eye disease is a multifactorial condition that significantly impacts the quality of life for millions worldwide. Accurate diagnosis and treatment are often challenging due to the complex interplay of symptoms, risk factors, and underlying causes. This study presents an innovative approach to enhance dry eye detection through the development of an ophthalmic knowledge graph. By curating and organizing data from peer-reviewed literature, the knowledge graph captures key entities such as symptoms, treatments, and diagnostic methods, and systematically maps their relationships. The study employs natural language processing (NLP) and deep learning techniques to extract entities and relationships, which are then integrated into the knowledge graph using the Neo4j platform. This structured representation provides clinicians with an intuitive and comprehensive tool for improving diagnostic accuracy and developing personalized treatment strategies. The knowledge graph not only strengthens the connections between diverse clinical data but also facilitates the integration of AI-driven diagnostic methods. This research demonstrates the potential of knowledge graphs in advancing the field of ophthalmology, specifically in the diagnosis and management of dry eye disease.

Keywords. Dry eye disease, Knowledge graph, Artificial Intelligence, Ophthalmic disease detection, Natural language processing

1. Introduction

Dry eye disease is a prevalent and multifactorial condition that affects millions of people worldwide[1], significantly impacting patients' quality of life[2]. It is

[1] Corresponding Author: Mini Han Wang,1155187855@link.cuhk.edu.hk

characterized by symptoms such as discomfort, visual disturbances, and tear film instability, often associated with inflammation of the ocular surface. The increasing incidence of dry eye disease[3], coupled with the complexity of its diagnosis and treatment, presents ongoing challenges for clinicians. Traditional diagnostic methods rely heavily on clinical experience and standard tests[4, 5], which may not always capture the nuanced variations in symptoms and underlying causes across diverse patient populations[6].

Recent advancements in artificial intelligence (AI) and big data analytics have created new opportunities for improving the diagnosis[7] and management of ocular diseases[8] . In particular, the development of knowledge graphs—structured representations of relationships between entities—has gained traction as a powerful tool in medical research and clinical decision-making[9]. By organizing large volumes of medical data into interconnected nodes of information, knowledge graphs[10] provide a comprehensive, intuitive framework for understanding the relationships between symptoms, risk factors, treatments, and outcomes[11].

This study seeks to enhance dry eye detection through the construction of an ophthalmic knowledge graph. The research involves systematically extracting and organizing knowledge from peer-reviewed literature related to dry eye disease, including key entities such as symptoms, diagnostic methods, and risk factors. By leveraging AI-driven natural language processing (NLP)[10, 12] techniques and deep learning models, this knowledge graph is designed to provide clinicians with a more accurate and comprehensive tool for diagnosing and treating dry eye disease. The study also aims to address several critical gaps in current diagnostic methods, including the integration of multi-source data and the ability to identify complex interrelationships between clinical indicators. The knowledge graph is intended to not only improve diagnostic accuracy but also provide insights into personalized treatment strategies, ultimately contributing to better patient outcomes.

2. Materials and Methods

A.Materials

The dataset employed in this study was meticulously curated from a range of scientific publications related to dry eye disease. A comprehensive screening of numerous peer-reviewed articles was conducted to ensure that the data collected was both scientifically robust and current. The selected literature encompassed the most recent advancements in the diagnosis, treatment methods, and associated factors influencing dry eye disease. These papers provided a solid theoretical foundation for the construction of the ophthalmic knowledge graph, as they contained essential concepts pertinent to dry eye diagnosis and treatment. This dataset thus served as a rich source of information for the extraction of entities and relationships, which were instrumental in the subsequent development of the knowledge graph.

B.Methods

The methodology of this study centered on the development of an ophthalmic knowledge graph aimed at enhancing the detection of dry eye disease. The process commenced with a thorough manual review of authoritative literature on dry eye, ensuring both scientific rigor and relevance. The selected papers were chosen based on their focus on the latest advancements in the diagnosis and treatment of dry eye disease. Subsequently, Python and the GPT API were employed for triple extraction, allowing for the identification of key entities and their relationships—such as symptoms, treatment methods, and risk factors—from the curated documents. By designing tailored prompts, the API was able to efficiently extract valuable information from the complex medical texts. Following the extraction phase, manual integration and screening of the data were conducted to ensure both the accuracy and consistency of the triples, with redundant or irrelevant information being systematically removed. The refined data were then processed using a deep learning model, which categorized the entities and their relationships into thematic categories such as "diagnosis," "treatment," and "research focus." Finally, the processed data was organized into a structured format and exported into an Excel spreadsheet. This structured dataset, containing both entities and their relationships, provided the foundation for constructing the ophthalmic knowledge graph, designed to offer enhanced insights into the detection and treatment of dry eye disease.

1) Triple Extraction

To extract pertinent information from the selected literature, we utilized a combination of Python scripting and the GPT API. Tailored prompts were developed to extract triples, which consist of key entities and their corresponding relationships. This approach specifically targeted the identification of associations between entities such as symptoms, treatments, and risk factors relevant to dry eye disease. By employing carefully designed prompt engineering techniques, we were able to systematically retrieve critical insights from complex medical terminology and scholarly data, thus facilitating the generation of structured triples (entity-relationship-entity).

2) Integration and Classification of Entities and Relationships

Upon the initial extraction of triples, manual integration was conducted to validate the accuracy and completeness of the derived data. This step involved eliminating redundant or irrelevant information to maintain the integrity and consistency of the dataset. Subsequently, we employed the latest deep learning model from Zhupu to automate the classification of both entities and their relationships. This model was adept at distinguishing among various categories of entities, including those related to "dry eye disease," "traditional Chinese medicine," and "immunity and inflammation." Furthermore, the relationships between these entities were categorized into specific types, such as "diagnosis," "characteristics," "interactions," and "research hotspots."

3) Data Structuring and Knowledge Graph Construction

Following classification, the processed dataset, comprising entities and their relationships, was organized into a structured format using Excel. This structured data table allowed for further refinement and served as the foundation for constructing the ophthalmic knowledge graph. The table provided a detailed representation of various entity types and their relationships, offering a clear and intuitive visualization of the interconnections between entities. This structured dataset was integral to the development of a comprehensive knowledge graph aimed at improving the detection and understanding of dry eye disease.

3. Result

The classification results are shown in Tables 1 and 2. The first column of the vertical axis lists the names of various entities and relationships, while the second column shows the number of entities in each category. This classification provides a clear structural framework for the subsequent construction of the knowledge graph, making the relationships between nodes in the graph more clear and precise.

Table 1. Entity Classification.

Category	Count
Related to Dry Eye	1485
Management and Prevention	18
Related to Immunity and Inflammation	88
Others	13
Other Diseases	106
AI-Based Diagnostic Methods	118
Related to Traditional Chinese Medicine	180

Table 2. Three Scheme comparing.

Category	Count
Biomarkers	22
Symptoms	40
Definition	3
Related	17
Diagnosis	150
Applied Technology	85
Characteristics	141
Risk Factors	26
Increased Risk	10
Negative Impact	1
Negative Correlation	2
Cause	5
Research Hotspots	376
Treatment & Intervention	6
Interaction	510
Spend More Money	1
Patient Feeling	1
Effect	3
Research Units	39
Research Outcomes	4
Influencing Factors	6
Pain Points	2
Symptoms	2
Principle	18
Reasons	8
Positive Correlation	2
Lead to	9
Marker	19

Finally, we used the Neo4j graph database to store and visualize the processed data. The powerful capabilities of Neo4j allowed us to build a complete knowledge graph by constructing nodes and edges representing the entities and relationships extracted from the triples. Through the visualization of nodes and relationships, the knowledge graph intuitively reflects the complex interconnections between symptoms, treatments, risk factors, and other entities related to dry eye disease. This provides comprehensive knowledge support for enhancing the diagnostic capability of dry eye. The final visualization of the knowledge graph is shown in Figure 1, displaying a network of various entities and their relationships.

During the process of triple extraction, manual integration, and classification using Zhupu's latest large model, a large number of entities and relationships related to dry eye disease were extracted and classified. A total of 1,485 dry eye-related entities were identified, including several important categories, such as 88 entities related to immunity and inflammation, 118 entities for AI-based diagnostic methods, and 180 entities related to traditional Chinese medicine. In addition, 510 "interaction" relationships, 376 "research hotspot" relationships, and 150 "diagnosis" relationships were identified. These relationships further strengthened the connections between entities, forming the core structure of the knowledge graph for dry eye diagnosis. The results of the entity classification reveal multidimensional knowledge information related to dry eye diagnosis and treatment, which was subsequently used to construct the knowledge graph. The distribution of various entities and relationships clearly highlights key areas of focus in current dry eye research and clinical practice. The full ophthalmic knowledge graph is shown in Figure 1:

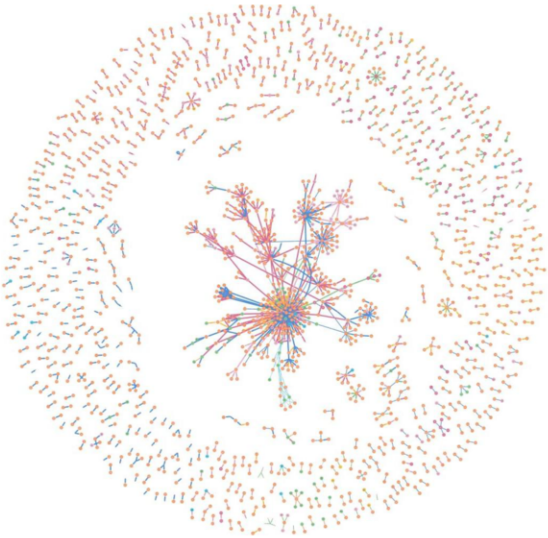

Figure 1. Full Ophthalmic Knowledge Graph.

Using the Neo4j database, we successfully organized the processed entities and relationships into a complete knowledge graph. This knowledge graph, represented through nodes and edges, visually displays information related to dry eye disease, such as symptoms, treatment methods, diagnostic techniques, and risk factors, forming a complex and systematic relationship network. Figure 1 shows the complete ophthalmic knowledge graph. This graph can be applied in various scenarios, including quickly querying diagnostic and treatment information related to dry eye, analyzing the interrelationships between different symptoms and treatment methods, and more. By visualizing the nodes and their relationships, the knowledge graph not only improves the accuracy of dry eye diagnosis but also provides decision support for clinicians, effectively enhancing diagnostic and treatment efficiency.

The constructed ophthalmic knowledge graph aids clinicians in quickly accessing comprehensive information related to patient symptoms, diagnoses, and treatment plans. Through association analysis in the graph, clinicians can rapidly identify cases with similar symptoms to their patients, providing strong support for developing personalized treatment plans. Notably, the multiple connections between "risk factors" and "treatment & intervention" in the graph help refine diagnostic strategies and treatment approaches for patients at different stages of disease progression. Additionally, the AI-based diagnostic methods are validated and strengthened through the knowledge graph, laying a solid foundation for further applying AI technology to dry eye diagnosis. The multi-dimensional information within the graph enhances data analysis, clinical decision support, and future research development.

Thus, the study successfully constructed an ophthalmic knowledge graph aimed at enhancing the detection and treatment of dry eye disease. Through meticulous manual screening of relevant scientific literature and the application of advanced natural language processing techniques, a vast dataset of entities and relationships related to dry eye was curated and classified. (1) A total of 1,485 dry eye-related entities were identified, spanning key areas such as dry eye symptoms, immunity and inflammation, AI-based diagnostic methods, and traditional Chinese medicine. Important relationships were also captured, including 510 interactions, 376 research hotspots, and 150 diagnostic relationships. These relationships provided a robust framework for understanding the interconnections between various aspects of dry eye disease. (2) The classification process revealed multiple dimensions of knowledge related to dry eye diagnosis and treatment. This included the identification of entities within critical categories like "diagnosis," "treatment," and "research focus," offering deeper insights into the most relevant areas of study and clinical practice. These classifications provided a clear structure for building a detailed knowledge graph. (3) Using Neo4j, a comprehensive knowledge graph was developed, representing the complex relationships between symptoms, treatment methods, diagnostic techniques, and risk factors associated with dry eye disease. The visualization of this graph highlighted the intricate connections between various entities, enabling a more precise and systematic understanding of the disease. (4) The constructed knowledge graph demonstrated practical clinical value by enabling faster access to comprehensive information related to patient symptoms and treatment options. The graph's ability to display relationships between risk factors and treatment methods, along with its support for AI-based diagnostic techniques, offers significant benefits for personalized treatment planning and improved diagnostic accuracy.

4. Discussion

The development of an ophthalmic knowledge graph for dry eye detection presents significant implications for both clinical practice and research. By organizing complex relationships between symptoms, treatment methods, risk factors, and diagnostic technologies, the knowledge graph provides an intuitive and comprehensive framework that enhances the diagnostic process. Clinicians can quickly access critical insights, identify similar cases, and develop personalized treatment plans, improving the efficiency and accuracy of patient care. Moreover, the integration of AI-based diagnostic methods into the knowledge graph strengthens the potential for automation and scalability in dry eye diagnosis. The graph provides a foundation for incorporating real-time clinical data, which could lead to the continuous improvement of AI-driven diagnostics. This integration offers new opportunities for telemedicine and remote healthcare, particularly for eye care specialists in underserved areas. The knowledge graph also has broader implications for the standardization of ophthalmic data and cross-institutional collaboration. By establishing a structured repository of dry eye-related information, it could serve as a valuable tool for global researchers working on dry eye and other ocular diseases, providing a shared platform for data exchange and analysis.

While the study demonstrates the potential of using a knowledge graph to improve dry eye detection, several limitations should be acknowledged. First, the dataset used to construct the knowledge graph was based on a curated selection of scientific literature, which may not cover the full spectrum of clinical cases and variations of dry eye disease. There is a possibility of bias in the literature selection, as the focus was on papers covering recent advances, potentially omitting foundational research. Second, although the Neo4j platform enabled effective visualization and interaction with the knowledge graph, its performance may be limited when scaling the system for larger datasets or more complex relationships. Further optimization of the graph for real-time applications in clinical settings may be required. Lastly, the reliance on AI-driven triple extraction for identifying entities and relationships, while efficient, may introduce inaccuracies or miss subtle relationships in the medical literature. Manual curation is necessary to ensure the accuracy of the extracted data, but this approach can be time-consuming and subject to human error.

Future research directions could focus on addressing these limitations and extending the scope of the knowledge graph. First, expanding the dataset by incorporating real-world clinical data from multiple medical centers could enrich the graph with diverse patient cases, ensuring a more comprehensive representation of dry eye disease. By integrating electronic health records (EHR) and patient-reported outcomes, the graph could become more dynamic and reflective of everyday clinical practice. Another promising area for future research involves enhancing the AI-driven extraction process through the use of more sophisticated natural language processing models. Exploring methods such as fine-tuning large language models specifically for medical text could improve the precision and depth of entity and relationship extraction, reducing the need for manual intervention. Moreover, the application of this knowledge graph can be extended beyond dry eye to other ocular diseases or even systemic diseases where ocular indicators play a role. Investigating the use of the graph in detecting early signs of systemic diseases, such as diabetes or cardiovascular conditions, would open new avenues for cross-disciplinary research and diagnostic innovation. Finally, developing user-friendly interfaces and clinical decision support systems based on the knowledge graph would enhance its practical application. Integrating the graph into telemedicine platforms, mobile health apps, and AI-driven diagnostic tools could improve accessibility for clinicians and patients alike, making it a vital asset in the global fight against eye diseases.

5. Conclusion

This study successfully developed an ophthalmic knowledge graph to enhance the detection and diagnosis of dry eye disease by systematically organizing and visualizing complex relationships between symptoms, treatment methods, diagnostic techniques, and risk factors. The integration of AI-driven methods for entity and relationship extraction, combined with manual curation, allowed for the creation of a comprehensive knowledge framework that significantly improves clinical decision-making and diagnostic precision. The knowledge graph provides a powerful tool for clinicians, enabling more efficient access to relevant information and supporting personalized treatment strategies. Additionally, it offers a foundation for integrating real-world clinical data, which can lead to continuous advancements in the diagnosis and treatment of dry eye disease. Despite its promising applications, the study acknowledges certain limitations, including potential biases in the literature-based dataset and challenges in scaling the graph for broader clinical use. Future research should focus on expanding the dataset with real-world clinical data, improving AI extraction methods, and extending the knowledge graph to other ocular and systemic diseases. In conclusion, this study not only demonstrates the potential of knowledge graphs in improving the diagnosis of dry eye disease but also sets the stage for future developments in AI-driven medical diagnostics and the standardization of ophthalmic data, offering significant benefits to both clinical practice and research.

Acknowledgments

This research was funded by the National Science Foundation of China under Grant U22A2041, 82071915 and 62372047, Shenzhen Key Laboratory of Intelligent Bioinformatics under Grant ZDSYS20220422103800001, Shenzhen Science and Technology Program under Grant KQTD20200820113106007, GuangDong Basic and Applied Basic Research Foundation under Grant 2022A1515220015, Zhuhai Technology and Research Foundation under Grant ZH22036201210034PWC, 2220004000131 and 2220004002412.

References

[1] WANG M H, XING L, PAN Y, et al. AI-Based Advanced Approaches and Dry Eye Disease Detection Based on Multi-Source Evidence: Cases, Applications, Issues, and Future Directions [J]. Big Data Mining and Analytics, 2024, 7(2): 445-84.

[2] ZHMUD T, MALACHKOVA N, REDJAK R, et al. Dry eye disease severity and impact on quality of life in type II diabetes mellitus [J]. Front Med, 2023, 10: 1103400.

[3] WANG M H, LIN Z, ZENG P, et al. Optimizing dry eye disease diagnosis: integrating deep learning insights from OSDI questionnaires and ocular blinking videos, F, 2024 [C]. SPIE.

[4] WANG H, ZHOU R. Advanced Diagnosis Method Analysis of the Dry Eye Disease: A Bibliographic Study (2012-2022) [J]. Clinical Medicine Frontiers, 2022, 1.

[5] SHEPPARD J D, NICHOLS K K. Dry eye disease associated with meibomian gland dysfunction: focus on tear film characteristics and the therapeutic landscape [J]. Ophthalmol Ther, 2023, 12(3): 1397-418 %@ 2193-8245.

[6] WANG H, XING L. A Fusion-based AI Approach for Dry Eye Disease Diagnosis using Multiple Sources of Digital Ophthalmic Data: A Bibliographic Study [J]. 2022.

[7] WANG M H, CHONG K K-L, LIN Z, et al. An Explainable Artificial Intelligence-Based Robustness Optimization Approach for Age-Related Macular Degeneration Detection Based on Medical IOT Systems [J]. Electronics, 2023, 12(12): 2697.

[8] ABDELMOTAAL H, HAZARBASANOV R, TANERI S, et al. Detecting dry eye from ocular surface videos based on deep learning [J]. The Ocular Surface, 2023.

[9] WANG M H, ZHOU R, LIN Z, et al. Can Explainable Artificial Intelligence Optimize the Data Quality of Machine Learning Model? Taking Meibomian Gland Dysfunction Detections as a Case Study; proceedings of the Journal of Physics: Conference Series, F, 2023 [C]. IOP Publishing.

[10] RAJABI E, ETMINANI K. Knowledge-graph-based explainable AI: A systematic review [J]. Journal of Information Science, 2024, 50(4): 1019-29 %@ 0165-5515.

[11] NI P, OKHRATI R, GUAN S, et al. Knowledge graph and deep learning-based text-to-GraphQL model for intelligent medical consultation chatbot [J]. Information Systems Frontiers, 2024, 26(1): 137-56 %@ 1387-3326.

[12] ZHU Y. A knowledge graph and BiLSTM-CRF-enabled intelligent adaptive learning model and its potential application [J]. Alexandria Engineering Journal, 2024, 91: 305-20 %@ 1110-0168.