Clinical & Refractive Optometry CRO
ONLINE CE A CE JOURNAL FOR OPTOMETRY

CRO Journal. Published online February, 2026.
doi: 10.57204/001c.154657

CE Credit - Topic Review

Unveiling Common LASIK Complications: A Machine-Learning Based Analysis of FDA Reported Cases

Anantha Krishnan[1]*, Salman Sarkar[2], Syed Ishaq[3], Meena Hasaji[3], Laxman Badavath[4]

CRO (Clinical & Refractive Optometry) Journal

ABSTRACT

Purpose: This study explores complications associated with Laser-Assisted In Situ Keratomileusis (LASIK) surgery by applying natural language processing (NLP) and machine learning to a dataset of FDA-reported adverse events. The goal is to identify common complication patterns and predictors to support patient safety, preoperative counseling, and clinical decision-making.

Methods: A retrospective analysis was performed using the publicly available "LASIK Complications" dataset from Kaggle, which includes 1,368 reports submitted to the U.S. Food and Drug Administration (FDA) between June 1997 and April 2019. Text data were cleaned, tokenized, and vectorized using term frequency-inverse document frequency (TF-IDF). A Random Forest classifier was developed and tuned through three-fold cross-validation. Model performance was assessed using accuracy, kappa statistic, sensitivity, specificity, and feature importance metrics.

Results: The model achieved a cross-validation accuracy of 48.76% and a test accuracy of 28.21% ($\kappa = 0.1988$, $p = 0.2661$). Among 19 complication types, the Visual Symptoms category showed the highest sensitivity (76.39%) and specificity (85.57%). Feature importance analysis revealed that terms like "halos," "dry," and "night" were the most influential descriptors, aligning with well-known post-LASIK experiences. In contrast, complications such as infection and corneal flap issues showed no sensitivity.

Conclusion: While overall accuracy was limited, the model successfully identified clinically relevant descriptors, highlighting the potential of NLP and machine learning in analyzing patient-reported outcomes. Future work should address class imbalance, enrich contextual features (e.g., n-grams), and apply explainable AI tools like SHAP or LIME to improve clinical adoption and interpretability.

Keywords: Laser eye surgery; complications; natural language processing; Machine Learning; Laser Assisted In Situ Keratomileusis.

1 INTRODUCTION

Refractive surgery provides a permanent, aesthetic, and effective alternative to conventional refractive error correction methods such as glasses and contact lenses. During the past three decades, procedures such as laser-assisted in situ keratomileusis (LASIK) and photorefractive keratectomy (PRK) have gained widespread acceptance due to advances in surgical techniques, laser technology, and a better understanding of corneal biomechanics[1–3]. Despite high satisfaction rates and consistent visual outcomes, the success of refractive surgery is not universal[3]. A subset of patients continues to experience unsatisfactory outcomes due to factors such as under or overcorrection, regression, or postoperative complications[4].

Understanding the predictive factors that influence the efficacy and safety of refractive surgeries is essential for optimizing patient outcomes. Gomel et al identified younger age, lower spherical equivalent (SE), and male gender as significant predictors of successful surgical outcomes in a large retrospective cohort[5]. Their findings underscored the value of thorough preoperative assessments and highlighted that safety indices have improved over time, reflecting ongoing advancements in surgical techniques and laser technology. Similarly, Deng et al examined factors influencing dominant eye selection in refractive surgery candidates[6]. They reported that variables such as best-corrected visual acuity (BCVA), pupil diameter, Kappa angle displacement, and astigmatism vector J45 play a critical role in surgical planning and postoperative outcomes. These insights are particularly relevant for personalized treatment strategies, including monovision correction, and contribute to maximizing patient satisfaction with visual performance.

Recent advances in artificial intelligence (AI) have shown promise for addressing persistent gaps in refractive surgery outcomes. For example, a study successfully employed machine learning models to improve diagnostic accuracy in retinopathy of prematurity and Meibomian gland dysfunctions[7,8]. These techniques can be adapted to refractive surgery, where AI-powered algorithms could integrate patient demographics, clinical data, and surgical parameters to predict both safety and efficacy. By systematically identifying risk factors and guiding customized enhancements (for example, flap relifts, and top-up ablations), AI can streamline clinical decision making, optimize surgical workflows, and reduce postoperative complications.

While significant progress has been made, gaps remain in understanding the interplay of various preoperative factors affecting refractive surgery outcomes[5]. The lack of comprehensive studies addressing predictive factors for both safety and efficacy necessitates further exploration. Moreover, enhancements in refractive surgery, including flap relifts and top-up ablations, are critical for patients with residual refractive errors[9–11]. However, robust meta-analyses establishing preferred enhancement protocols are limited, and current decision-making algorithms vary across practices[12–14].

This study aims to investigate the predictive factors that influence refractive surgery outcomes by combining machine learning analysis of FDA-reported complications with insights from existing literature, including[5,6]. Specifically, it seeks to identify the most informative textual descriptors extracted from patient-reported data using a Random Forest model, and to contextualize these features considering established clinical predictors of surgical efficacy and dominant eye selection. In doing so, the study also explores the broader implications of these findings for optimizing surgical planning, enhancing patient satisfaction, and informing the development of standardized enhancement protocols in refractive surgery practice.

2 METHODS

Dataset Description

The dataset used in this study, titled LASIK Complications, was obtained from Kaggle and compiled from publicly available reports by the United States Food and Drug Administration (FDA)[15]. It documents historically reported complications associated with laser eye surgeries from June 1997 to April 2019. The dataset comprises 1,368 unique records, each consisting of four primary fields: (1) a unique report identifier (id), (2) the report date, (3) a free-text description of the complication, and (4) keywords summarizing the condition. Commonly reported issues include impaired vision, chronic pain, astigmatism, dryness, halos, and starbursts. The data reflects the growing adoption of laser eye procedures following FDA approval in the late 1990s, with the global personalized LASIK surgery market projected to reach $3.69 billion by 2022. Owing to its unstructured nature, the dataset is well suited for natural language processing (NLP) tasks aimed at extracting patterns and identifying key clinical complications.

Data Pre-processing

A comprehensive pre-processing pipeline was employed to standardize the dataset and improve the quality of its features for model training. The textual data underwent normalization, including conversion to lowercase, removal of punctuation and numerals, and elimination of standard English stop-words. Subsequently, stemming and lemmatization were applied to reduce words to their base forms. Tokenization was performed, and the resulting tokens were transformed into numerical representations using Term Frequency–Inverse Document Frequency (TF-IDF) vectorization, yielding a sparse feature matrix with 8,644 predictors. To address missing values, absent keywords were labeled as "Unknown". The complications keywords were then converted into categorical variables and encoded into numeric factors with syntactically valid R variable names. This pre-processing ensured consistency across records and facilitated compatibility with downstream machine learning algorithms.

Model Development and Evaluation

The dataset was divided into training (80%) and testing (20%) subsets, with stratified sampling to preserve class distribution. A Random Forest classifier was implemented using the ranger package in R, selected for its scalability and efficiency in handling high-dimensional data. Model hyperparameters were tuned using three-fold cross-validation. The hyperparameter grid included three values for *mtry* (no of features randomly tested at each split: 2, 5, and 10) and three values for *min.node.size* (minimum samples a node must have to be split: 1, 5, and 10), with the classification split rule set to gini (choosing splits that minimise *Gini* impurity). The cross-validation objective was to maximize classification accuracy. The optimal configuration was identified at *mtry* = 10 and *min.node.size* = 5. The final model was evaluated on the test set using overall accuracy, Cohen's kappa (κ) statistic, and confusion matrix analysis. Additionally, feature importance scores were computed to identify the most influential predictors contributing to classification performance.

3 RESULTS

Model Performance and Accuracy Trends

The performance of the Random Forest model demonstrated a consistent increase in classification accuracy with higher values of *mtry*. The highest cross-validation accuracy of 48.76% was observed at *mtry* = 10 and *min.node.size* = 5 (Figure 1). This pattern suggests that selecting a larger number of predictors (i.e. symptoms) per split provides richer contextual information, thereby improving predictive performance. In contrast, variations in *min.node.size* had a relatively modest effect, with a value of 5 offering the best trade-off between underfitting and overfitting tendencies.

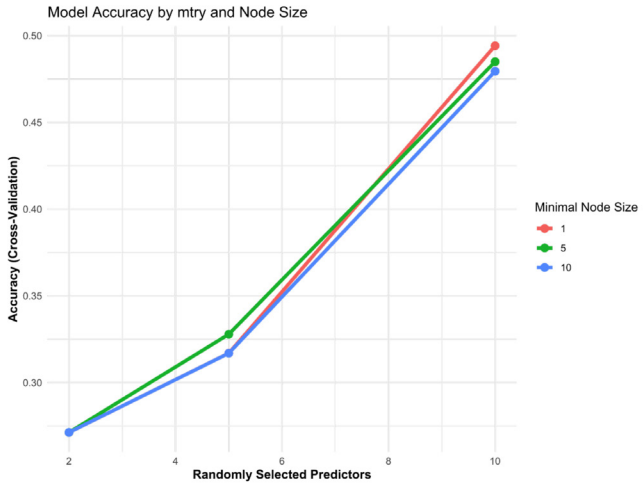


Fig. 1: Cross-validation accuracy trends across varying values of randomly selected predictors (mtry) and min.node.size. Accuracy improves consistently with higher mtry, reaching its peak at mtry = 1 and min.node.size = 5. While variations in minimal node size show limited impact, a value of 5 offers a balanced trade-off between model complexity and predictive performance.

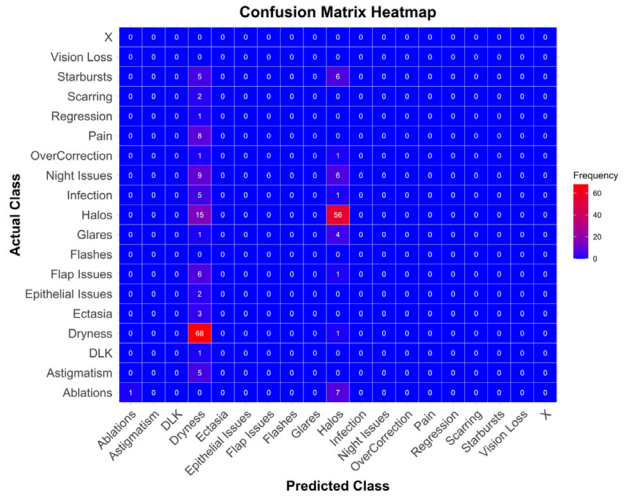


Fig. 2: Confusion matrix heatmap of the Random Forest classifier across 19 complication categories. The y axis shows actual labels, and the x-axis shows predicted labels. Diagonal cells indicate correct classifications, with color intensity reflecting prediction frequency (red = high, blue = low). "X" denotes cases where the model defaulted to an unspecified or ambiguous class.

Confusion Matrix and Classification Performance

The confusion matrix demonstrates that the Random Forest model achieved an overall accuracy of 28.21% (Cohen's κ = 0.1988, p = 0.2661), indicating modest predictive ability beyond random chance (Figure 2). Notably, the Visual Symptoms complication category showed high sensitivity (76.39%) and specificity (85.57%), making it reliably distinguishable from others. In contrast, multiple complication categories such as Infection, Corneal Flap Issues, and Irregular Astigmatism exhibited zero sensitivity, indicating no correct predictions for these complications and highlighting the model's limitations due to class imbalance.

In contrast, several complication categories such as Infection, Corneal Flap Issues, and Irregular Astigmatism recorded zero sensitivity and positive predictive value (PPV), indicating that the model failed to identify any true positives in these categories. The overall classification accuracy on the test set was 28.21%, accompanied by a Cohen's kappa statistic of 0.1988, reflecting only fair agreement between predicted and true labels. The no-information rate (NIR) was 26.37%, and the corresponding p-value of 0.2661 suggests that, for certain classes, the model's performance did not significantly exceed random chance.

Feature Importance Analysis

The feature importance analysis highlights the most influential predictors contributing to the model's classification decisions (Figure 3). The top three features halos, dry, and night achieved the highest importance scores, emphasizing their central role in identifying postoperative complications. Additional prominent features included ectasia, starbursts, severe, glare, and pain, all of which represent clinically relevant symptoms commonly reported after laser eye surgery.

Notably, ectasia emerged as one of the prominent features, which aligns with clinical expectations given its severity and the urgency it typically demands. This suggests that the model is capable of prioritizing diagnostically and semantically relevant terms, reinforcing its potential utility in clinical screening and triage workflows.

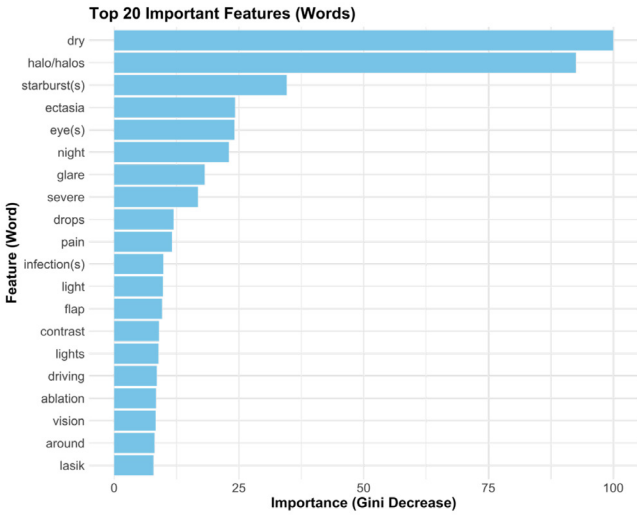


Fig. 3: Bar plot of the top 20 most important words influencing Random Forest classification, ranked by Gini impurity reduction. Words like "halos," "dry," and "starbursts" had the highest impact on model decisions.

4 DISCUSSION

Model Findings

Although the model's quantitative performance was limited (accuracy = 28.21%; Cohen's κ = 0.1988), its ability to identify clinically meaningful symptoms from unstructured patient reports demonstrates the promise of artificial intelligence in refractive-surgery care. The complication category such as Visual Symptoms was consistently identified with relatively high sensitivity (76.39%) and specificity (85.57%).

In contrast, several complications, notably those related to severe corneal and infection issues, were never correctly predicted by the model. The model's difficulty in recognizing corneal infections or flap irregularities, is caused primary due to insufficient reports as there are far fewer reports describing these problems. In clinical terms, this is similar to trying to identify rare adverse events from a limited case series. Similar challenges have been documented in prior studies on natural language processing (NLP) for healthcare applications, where sparse data and label imbalance often compromise model performance[16–18]. Consistent with the findings of[18], who emphasized the importance of dataset balance in improving classification accuracy, future work should explore strategies such as data augmentation, oversampling, or cost-sensitive learning to enhance model robustness.

Strategies for Model Improvement

While the current study prioritized interpretability and model transparency using a baseline Random Forest approach with TF-IDF features, several complementary strategies could enhance performance in future iterations. One such strategy involves consolidating semantically similar terms e.g., halo and halos, or light and lights to reduce redundancy and improve lexical representation. Although basic stemming was applied during preprocessing, more nuanced semantic grouping was deliberately avoided to maintain the traceability of individual descriptors. This trade-off was made to preserve clarity in the model's feature importance outputs.

Incorporating *n-gram* features (e.g., dry eyes, night driving) could also enrich contextual representation by capturing adjacent word relationships. However, due to the relatively short and sparsely worded nature of many FDA reports, unigram-based models were initially favored to mitigate overfitting risks. As demonstrated by[19], future work may benefit from experimenting with *n-gram* embeddings once sufficient contextual density is ensured.

To improve model interpretability and clinical alignment, mapping textual features to standardized diagnostic categories such as linking ectasia to ectatic corneal disorders remains a promising direction. Although this was not performed in the current pipeline, such mappings could align the feature space with medical ontologies and enhance clinical relevance. This is consistent with the recommendations of[20] and[21], who emphasize the value of embedding domain knowledge into predictive healthcare models.

Furthermore, incorporating explainable artificial intelligence (XAI) techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) would support local interpretability by elucidating how individual features contribute to specific predictions. As shown by[22] and[23], these methods help translate complex model behavior into clinically actionable insights, thereby increasing trust and transparency in real-world applications.

Comparative Insights with Existing Literature

Recent studies have demonstrated the utility of machine learning techniques in ophthalmology, optometry and related healthcare domains. For instance, Kermany et al applied deep learning for the detection of diabetic retinopathy, achieving high sensitivity and specificity[24]. Their model leveraged a large, well-balanced image dataset, whereas the present study highlights the unique challenges associated with textual data from patient reports challenges that necessitate specialized preprocessing and feature engineering strategies.

Similarly,[25] developed AI-based screening tools for multiple retinal diseases, further illustrating the potential of machine learning in ophthalmic diagnostics. However, unlike our text-based approach, their work focused exclusively on image-based inputs, thereby requiring different methodologies for feature representation and interpretation.

In addition,[26] and[20] emphasized the importance of interpretability in healthcare-related AI systems. Their call for integrating domain expertise (such as optometrists and ophthalmologists) into model design aligns with our recommendation to map textual features to clinically recognized diagnostic categories. Such integration not only improves model transparency but also enhances its practical relevance in real-world clinical settings.

The findings from this study extend these earlier works by demonstrating how text-based AI can also generate insights that are clinically meaningful to optometrists. Integrating such AI tools into routine practice could strengthen collaborative care between optometrists and surgeons, improving both patient satisfaction and long-term visual outcomes.

Implications for Clinical Practice

The findings of this study offer several meaningful implications for clinical practice. The identification of features such as halos, dry, and night as top-ranked predictors aligns with well-established postoperative patient experiences, including night vision disturbances and dry eye syndrome. This concordance suggests that NLP-based models can reliably provide clinically relevant descriptors and may aid in the early detection and monitoring of LASIK-related complications.

Future iterations incorporating richer contextual information and explainable AI (XAI) frameworks could serve as effective decision-support tools, offering real-time insights derived from unstructured textual data. In particular, the integration of temporal analysis such as that proposed by[27] may further enhance clinical utility by revealing patterns and trends in complication reports over time. These capabilities could help clinicians proactively anticipate and mitigate risks associated with refractive eye surgeries.

The finding in this study highlights patterns of symptoms such as dryness, night-vision disturbances, and halos as the most frequent surgical complication to refractive surgeries. Recognizing these potential complications can help optometrists provided targeted guidance and counseling to patients at higher risk. Initiate management strategies in timely manner by routinely assessing tear-film and monitor visual-quality in post-op patients as a precautionary measure.

5 CONCLUSIONS AND FUTURE PERSPECTIVES

While the Random Forest model demonstrated moderate predictive performance, its ability to highlight clinically relevant features among the most influential terms suggests strong potential for further development. For optometrists who are involved in pre- and post-LASIK co-management, AI-driven insights could assist in early detection of postoperative discomfort, guide individualized patient counseling, and strengthen collaborative care pathways with refractive surgeons. Future, research should focus on enhancing feature representation through advanced natural language processing (NLP) techniques, addressing class imbalance, and employing explainable AI methodologies to improve interpretability and clinical relevance. By building upon these foundational insights, subsequent models could achieve higher predictive accuracy and offer tangible benefits for clinical practice in ophthalmology.

6 ACKNOWLEDGEMENTS

The authors appreciate the Food and Drug Administration of the United States for making LASIK-related adverse event reports publicly available, and Kaggle[15] for hosting the "LASIK Complications" dataset used in this analysis. Their open data provision was essential for conducting this study.

7 DECLARATION

This study did not receive any funding. The authors received no financial support for the research, authorship, or publication of this article. The authors declare no conflict of interest or competing interests. This study did not involve human participants, animals, or patient-identifiable data and therefore did not require ethics approval. Consent for publication is not applicable. The dataset analyzed is publicly available on Kaggle under the title "LASIK Complications"[15]. No new materials were generated in this study. The code used for data processing and model development is available upon reasonable request from the corresponding author. The author was solely responsible for the study's conception, data analysis, interpretation, and manuscript preparation.

REFERENCES

1. Fernández-Sánchez V, Ponce ME, Lara F, Montés-Micó R, Castejón-Mochón JF, López-Gil N. Effect of 3rd-order aberrations on human vision. *J Cataract Refract Surg*. 2008;34(8):1339-1344. doi:10.1016/j.jcrs.2008.04.017
2. Huhtala A, Pietilä J, Mäkinen P, Uusitalo H. Femtosecond lasers for laser in situ keratomileusis: a systematic review and meta-analysis. *Clin Ophthalmol Auckl NZ*. 2016;10:393-404. doi:10.2147/OPTH.S99394
3. Sandoval HP, Donnenfeld ED, Kohnen T, et al. Modern laser in situ keratomileusis outcomes. *J Cataract Refract Surg*. 2016;42(8):1224-1234. doi:10.1016/j.jcrs.2016.07.012
4. Moshirfar M, Basharat NF, Bundogji N, et al. Laser-Assisted In Situ Keratomileusis (LASIK) Enhancement for Residual Refractive Error after Primary LASIK. *J Clin Med*. 2022;11(16):4832. doi:10.3390/jcm11164832
5. Gomel N, Negari S, Frucht-Pery J, Wajnsztajn D, Strassman E, Solomon A. Predictive factors for efficacy and safety in refractive surgery for myopia. *PLOS ONE*. 2018;13(12):e0208608. doi:10.1371/journal.pone.0208608
6. Deng B, Luo L, Zhou M, et al. Factors influencing dominant eye selection in refractive surgery patients: A correlation analysis. *Contact Lens Anterior Eye*. 2024;47(1). doi:10.1016/j.clae.2023.102101
7. Krishnan A, Salman Sarkar Md, Badavath L. Machine Learning Applications in Retinopathy of Prematurity Diagnosis Using the ROP Retinal Image Dataset. *J Neonatal Surg*. 2025;14(1S):820-827. Accessed February 18, 2025. https://www.jneonatalsurg.com/index.php/jns/article/view/1607
8. Krishnan A, Sarkar S, Badavath L. Clustering-Based Analysis of Meibomian Gland Morphology for Automated Assessment of Meibomian Gland Dysfunction Using the MGD-1k Dataset. *J Neonatal Surg*. 2025;14(4S):231-236. doi:10.52783/jns.v14.1777
9. Alió JL, Montés-Mico R. Wavefront-guided versus standard LASIK enhancement for residual refractive errors. *Ophthalmology*. 2006;113(2):191-197. doi:10.1016/j.ophtha.2005.10.004
10. Davis EA, Hardten DR, Lindstrom M, Samuelson TW, Lindstrom RL. Lasik enhancements: a comparison of lifting to recutting the flap. *Ophthalmology*. 2002;109(12):2308-2313; discussion 2313-2314. doi:10.1016/s0161-6420(02)01245-9
11. Lee BS, Gupta PK, Davis EA, Hardten DR. Outcomes of photorefractive keratectomy enhancement after LASIK. *J Refract Surg Thorofare NJ 1995*. 2014;30(8):549-556. doi:10.3928/1081597X-20140711-08
12. Jacobs DS, Lee JK, Shen TT, et al. Refractive Surgery Preferred Practice Pattern®. *Ophthalmology*. 2023;130(3):P61-P135. doi:10.1016/j.ophtha.2022.10.032
13. Peñarrocha-Oltra S, Soto-Peñaloza R, Alonso-Arroyo A, Vidal-Infer A, Pascual-Segarra J. Laser-based refractive surgery techniques to treat myopia in adults. An overview of systematic reviews and meta-analyses. *Acta Ophthalmol (Copenh)*. 2022;100(8):878-893. doi:10.1111/aos.15160
14. Wen D, McAlinden C, Flitcroft I, et al. Postoperative Efficacy, Predictability, Safety, and Visual Quality of Laser Corneal Refractive Surgery: A Network Meta-analysis. *Am J Ophthalmol*. 2017;178:65-78. doi:10.1016/j.ajo.2017.03.013
15. Bansal S. Lasik (Laser Eye Surgery) Complications. 2021. Accessed July 23, 2025. https://www.kaggle.com/datasets/shivamb/lasik-complications-dataset

16. Chapman WW, Nadkarni PM, Hirschman L, D'Avolio LW, Savova GK, Uzuner O. Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. *J Am Med Inform Assoc*. 2011;18(5):540-543. doi:10.1136/amiajnl-2011-000465

17. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A Review of Challenges and Opportunities in Machine Learning for Health. Published online December 5, 2019. doi:10.48550/arXiv.1806.00388

18. Hairani H, Widiyaningtyas T, Prasetya DD. Addressing Class Imbalance of Health Data: A Systematic Literature Review on Modified Synthetic Minority Oversampling Technique (SMOTE) Strategies. *JOIV Int J Inform Vis*. 2024;8(3):1310-1318. doi:10.62527/joiv.8.3.2283

19. Robertson AM, Willett P. Applications of n-grams in textual information systems. *J Doc*. 1998;54(1):48-67. doi:10.1108/EUM0000000007161

20. Sirocchi C, Bogliolo A, Montagna S. Medical-informed machine learning: integrating prior knowledge into medical decision systems. *BMC Med Inform Decis Mak*. 2024;24(Suppl 4):186. doi:10.1186/s12911-024-02582-4

21. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. *BMC Med Inform Decis Mak*. 2020;20(1):280. doi:10.1186/s12911-020-01297-6

22. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. Published online November 25, 2017. doi:10.48550/arXiv.1705.07874

23. Ribeiro M, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In: DeNero J, Finlayson M, Reddy S, eds. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. Association for Computational Linguistics; 2016:97-101. doi:10.18653/v1/N16-3020

24. Kermany DS, Goldbaum M, Cai W, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell. 2018;172(5):1122-1131.e9. doi:10.1016/j.cell.2018.02.010

25. Ting DSW, Cheung CYL, Lim G, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA*. 2017;318(22):2211-2223. doi:10.1001/jama.2017.18152

26. Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci U S A*. 2019;116(44):22071-22080. doi:10.1073/pnas.1900654116

27. Shickel B, Loftus TJ, Ruppert M, et al. Dynamic predictions of postoperative complications from explainable, uncertainty-aware, and multi-task deep neural networks. *Sci Rep*. 2023;13(1):1224. doi:10.1038/s41598-023-27418-5