Article

Dry Eye Disease in the Digital Age: Causes and Diagnostics Using Machine Learning

Tamerlan Rakhmanberdiev [*] and Zhenishbek Orozakhunov

Article

Dry Eye Disease in the Digital Age: Causes and Diagnostics Using Machine learning

Rakhmanberdiev Tamerlan and Zhenishbek Orozakhunov

Department of Computer Science, Ala-Too International University; zhenishbek.orozakhunov@alatoo.edu.kg
*   Correspondence: topkek164@gmail.com

Abstract: Dry Eye Disease (DED) is a widespread condition made worse by modern lifestyle factors such as extended screen time and reduced physical activity. This study focuses on developing a machine learning model, specifically a Random Forest classifier to predict the likelihood of DED based on user data. The dataset included features like screen time, eye discomfort, physical activity, weight, height, blood pressure and other symptoms.The model achieved an accuracy of 71.7% and a log loss of 0.6501, showing moderatel performance in classifying DED. Feature importance shows that the most important features in prediction were eye strain, redness, irritation, screen time, and physical activity. These results align with clinical expectations and demonstrate that lifestyle-related data can be effectively used to detect DED risk.This research not only supports the idea that behavior affects eye health, but also shows how machine learning can be applied to assist in early detection and personalized recommendations for preventing DED.

Keywords: dry eye disease, eye health, machine learning, artificial intelligence

MSC: 616.9 (UDC) - Diseases of the Eye

1. Introduction

With the growing popularity of social media such as Twitter, Facebook and Instagram, and the growing number of people who have access to the Internet, in today's world more and more people are using phones, tablets and computers.However, despite the fact that most people have access to all the information in the world and the ability to communicate with all people from far away, the issue of eye health is raised.More and more people are looking at screens of phones, computers, etc., which make them blink less and thus provoke dry eye disease.This study will be about the impact digital screens have on the eyes, how much of an impact screens have on the development of dry eye disease, disease prevention, and modern treatment practices.By studying all the causes and consequences, prevention and treatment of dry eye disease, I aim to contribute to the understanding of the processes occurring around dry eye disease and the dissemination of awareness, prevention and treatment of this disease. **Hypothesis:** Long exposure to digital screens increases the risk of Dry Eye Disease because of reduced blinks and tear film instability.

2. Literature Review

In this Chapter we will talk about Machine learning and dry eye disease, what causes it and how it affects organism.

2.1. Dry Eye Disease

Dry eye disease (DED) is disease of the eye surface, that is usually caused by tear firm not working properly. Tear firm protects eye and making eye smooth by creating lubrication for eye. If eye firm is not properly functioning, it can lead to eye dryness, irritation and blur vision. Untreated DED causes permanent damage to the eye surface. [1]
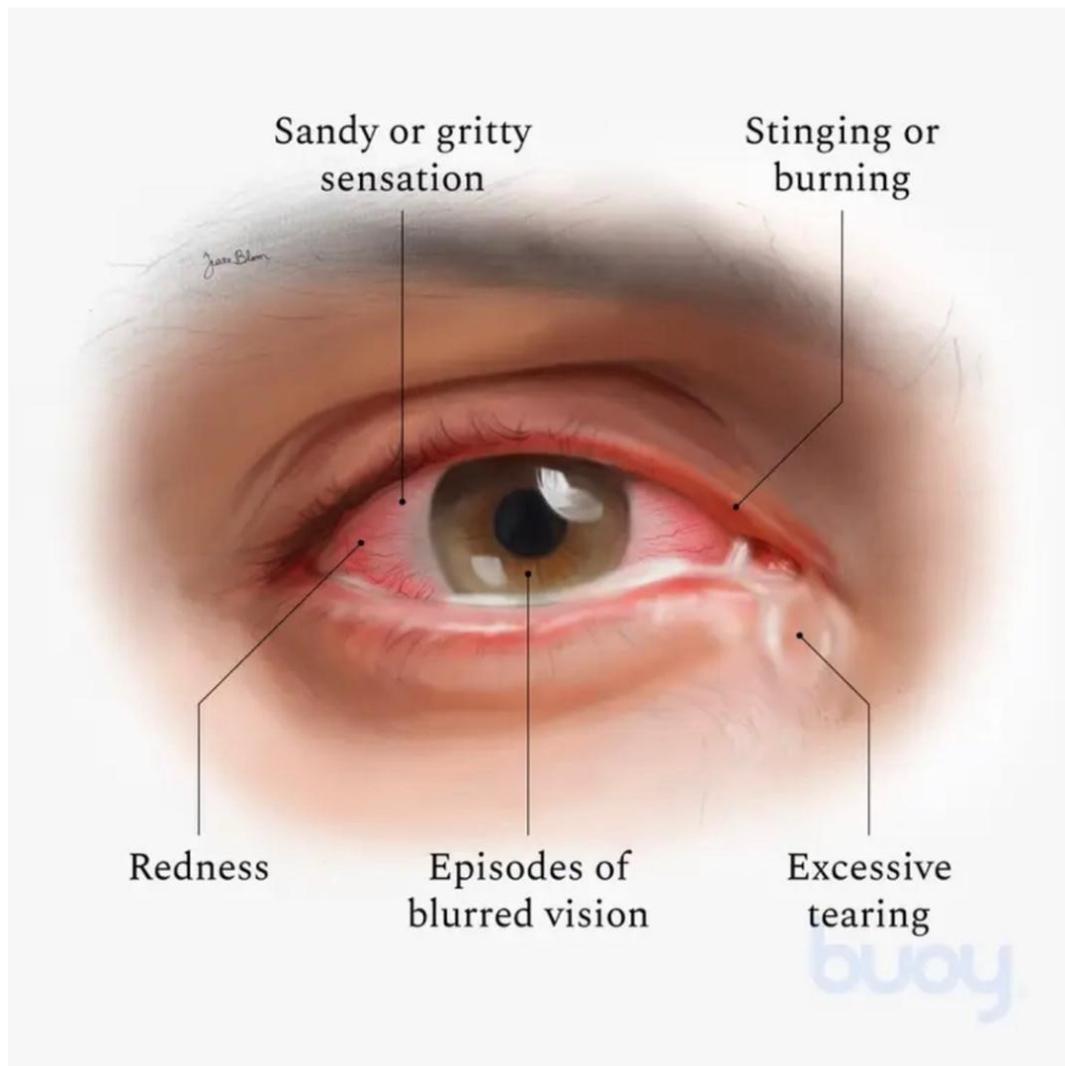
Figure 1. Symptoms of the Dry eye disease.

2.2. Causes

This subsection is listing of main causes of DED.

2.2.1. Tear Firm Produces Less Tears

The Lacrimal Functional Unit (LFU) is a complex, integrated system essential for maintaining the health and stability of the ocular surface. It encompasses various anatomical and neurological components that work synergistically to regulate tear production, distribution, and clearance. Disruption in any part of this unit can lead to dry eye disease (DED).

Tear Firm produces less tear and because of that eye surface becomes drier.

1. Other diseases and conditions like Sjogren's syndrome, allergy, arthritis, lupus, scleroderma, sarcoidosis and thyroid disorders. [2]
2. Medications like antihistamines, decongestants, hormone replacement therapy, antidepressants, and medications for high blood pressure, acne problems, birth control and Parkinson's disease. [2]
3. Outer causes like lenses or eye surgeries. [2]
4. Lack of vitamin A causes dry eye. [3]

2.2.2. Tear Evaporates Faster

Tear Firm produces less normal amount of tears but tears evaporates faster making eye dry.

1. Tears are mix of water, mucus and oils, that are produced by meibomian glands. If the mix is imbalanced, tears can evaporate faster. [4]
2. Eyelid problems. Inward or outward turning of eyelid affects the way we blink, causing faster evaporation of tears. [5]
3. Blinking less can cause faster evaporation of tears. Less amount of blinking can be caused by reading and screen use. [2]
4. Environmental factors such as dry, hot or windy weather, heating or using air conditioner, high altitude and smoke. [5]

2.2.3. Age and Gender

Dry eye disease more common in women than in men. Women who experienced pregnancy, menopause or used birth control are more at risk of DED.

People who older than 50 years old have more risk of DED, because of tear production tends be less as person get older. [2]

2.3. Machine Learning

2.3.1. What is Machine Learning

Machine Learning is a discipline that studies computer programs that automatically improve itself through experience and everything around it. From theoretical laws that defines learning to practical computational laws. The study of machine learning is important because it allows not only make jobs easier by handling them to AI but it also helps humans by studying better laws of learning. [6]

2.3.2. Why Machine learning

Reasons why we chose Machine learning in this study.

1. Machine learning is better at understanding complex data and finding correlations and patterns in the dataset.
2. Machine Learning is more objective if it is done right. The real doctor can be unexperienced or biased towards patients.
3. Machine learning can help people who do not have access towards eye treatment.

2.4. Random Forest Algorithm

Random forest is a machine learning algorithm, that uses prediction of multiple decision trees to reach a single answer.

2.4.1. Decision Tree

A decision tree is an algorithm that breaks down every decision into smaller decisions. Each smaller decision is represented as a node and possible outcome are represented as branches. The final answer is represented as leaves.The overall shape looks like a tree, hence the name. [7]

2.4.2. What is Random Forest Algorithm

Random Forest algorithm as an algorithm that combines several decision trees to reach result. This algorithm is a flexible method used for classification. It is efficient and scalable, making it perfect for handling large datasets, making it perfect algorithm for our study [8].

2.4.3. How does Random Forest Algorithm works

Random Forest Algorithm creates several decision trees and uses randomly chosen data samples. Because of random choosing the algorithm reduces overfitting and improves prediction accuracy [8].
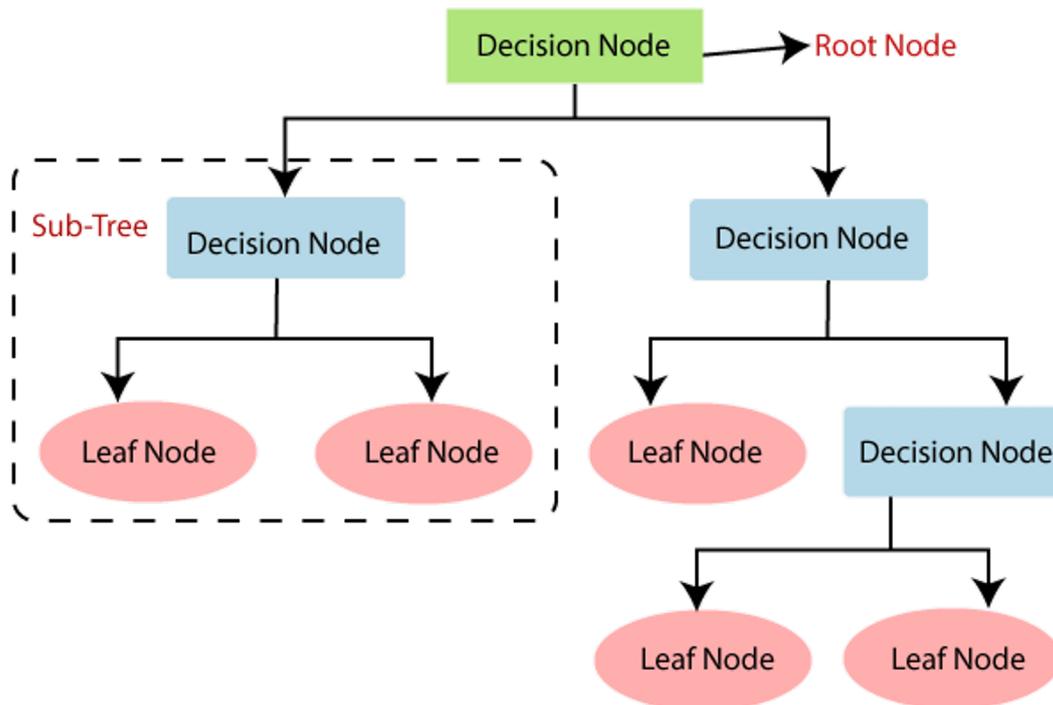
Figure 2. Structure of decision tree

2.4.4. Pros of Random Forest Algorithm

1. Fast to train
2. Handles large datasets
3. Supports classification
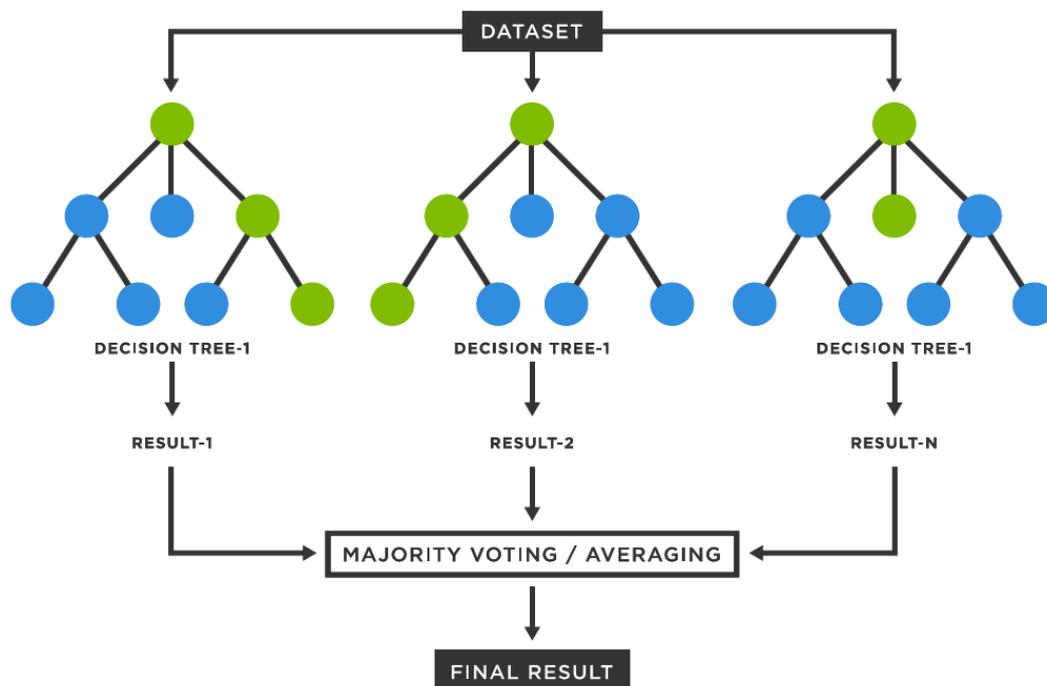4. High performance [8]

Figure 3. Structure of Random Forest

3. Dataset

The dataset used in this study was derived from the publicly available site Kaggle. It has 20000 samples with 25 metrics such as itchiness, screen time, screen filters and others. This section presents all metrics and explains why they are important.

1. Gender. Women have higher chance of having DED than Men. [2]
2. Age. People who 50 years or older have higher chance of having DED. [2]
3. Sleep duration. People who sleep less than 5 hours have higher chance of having DED. [9]
4. Sleep quality. People who have low-quality sleep have more chances of having ED. [9]
5. Stress level. People with stress, anxiety, or other mental health factors can affect the development of DED. [9]
6. Blood pressure. High blood pressure is associated with an increased risk of DED. [9]
7. Heart rate. High heart rate can be sign of DED. [10]
8. Daily steps. Physical activity of walking is considered to be prevention for DED.
9. Physical activity. Physical activity or exercise was associated with relief of DED associated symptoms. [11]
10. Height. There is no correlation between height and DED, but this metric is still useful because with weight metric it allows us to find if patient is obese.
11. Weight. There was a moderate correlation between body fat percentage and dry eye symptoms. [12]
12. Sleep disorder. There is a significant correlation between sleep disorder and DED. [9]
13. Wake up during night. People who waking up during night with dry eyes can have various reason for it. DED is one of the reason. [13] This metric is supporting one.
14. Feel sleepy during day. Daytime sleepiness is a symptom of DED. [14]
15. Caffeine consumption. There is correlation bertween caffeine and increasing tear production. [15]
16. Alcohol consumption.Alcohol consumption can disrupt the tear film, leading to DED symptoms. [16]
17. Smoking. Smoke from cigarettes can irritate eye surface leading to DED symptoms. [17]
18. Medical issue. Often medical issues like arthritis, lupus, Graves disease etc can lead to DED. [2]
19. Ongoing medication. Several medication can contribute to or exacerbate DED. [2]
20. Smart device before bed. Extensive usage of smart devices before bed is able to increase risk of DED. [18]
21. Average screen time. Large amount of screen time contribute to or exacerbate DED. [19]
22. Blue-light filter. Usage of the blue-light filter can negate some of the problems caused by screens and help people with DED to make symptoms easier to handle. [20]
23. Discomfort eye strain. Eye strain is one of the most common symptoms of DED. [21]
24. Redness in eye and Itchiness/Irritation in eye. Eye redness and irritation are common symptoms of DED. [22]
25. Dry Eye Disease. Metric that will be used in the training and testing of the model.

4. Methods

4.1. Instruments for Building

The base programming language that was used in the building of CNN was Python. We chose Python because of rich ecosystem that supports any Neural Network building with ease. [23] The library for ML model building is Scikit learn. Scikit learn is one of the most popular Python libraries for ML algorithm. The reasons why we chose Scikit learn

1. The first reason we chose Scikit learn is popularity. The popularity of library means that there is a lot of support, forums where we can discuss errors and methods to build better ML algorithm.
2. The second reason is that Scikit learn is scalable. That allows us to build a more efficient and scalable ML algorithm with 20000 rows of data in our dataset.

4.2. Data Preprocessing

Prior to training the model, the data were preprocessed using standard techniques.Columns such as Sleep disorder,Wake up during night,Feel sleepy during day,Caffeine consumption,Alcohol consumption,Smoking,Medical issue,Ongoing medication,Smart device before bed,Blue-light filter,Discomfort Eye-strain,Redness in eye,Itchiness/Irritation in eye,Dry Eye Disease have either Y or N, that stands for Yes or No. Y or NO were turned into 1 or 0 for easier training. Same process were done with the Gender column that have either M or F, that stands for Male or Female. The Blood pressure column were divided into Systolic and Diastolic columns for easier training. After turning every data into numerical type we turn all of them into float type for easier training.

Table 1. Data of the same patient before and after preprocessing.

Columns	Before	After
Gender	F	0.0
Age	24	24.0
Sleep Duration	9.5 hours	9.5 hours
Sleep Quality	2	2.0
Stress Level	1	1.0
Blood Pressure	137/89	divided
Systolic	None	137.0
Diastolic	None	89.0
Heart Rate	67 bpm	67.0 bpm
Daily Steps	3000 steps	3000.0 steps
Physical Activity	31	31.0
Height	161 cm	161.0 cm
Weight	69 kg	69.0 kg
Sleep Disorder	Y	1.0
Wake Up During Night	N	0.0
Feel Sleepy During Day	N	0.0
Caffeine Consumption	N	0.0
Alcohol Consumption	N	0.0
Smoking	N	0.0
Medical Issue	Y	1.0
Ongoing Medication	Y	1.0
Smart Device Before Bed	N	0.0
Average Screen Time	8.7 hours	8.7 hours
Blue-light Filter	N	0.0
Discomfort Eye-strain	Y	1.0
Redness in Eye	Y	1.0
Itchiness/Irritation in Eye	N	0.0
Dry Eye Disease	Y	1.0

4.3. Machine Learning Models

Random Forest Classifier: Scikit Learn has RandomForestClassifier function that allows us to use Random Forest Algorithm on our dataset. The model is using 4 decision trees, with maximum depth equal to 30 and random state equal to 42.
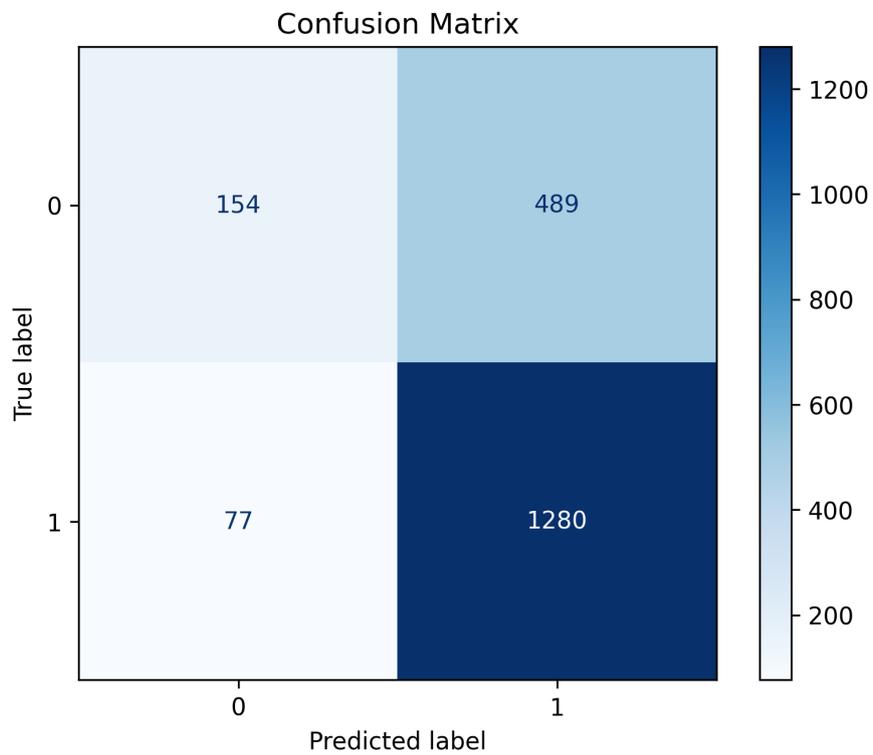
5. Results

5.1. Numerical Score

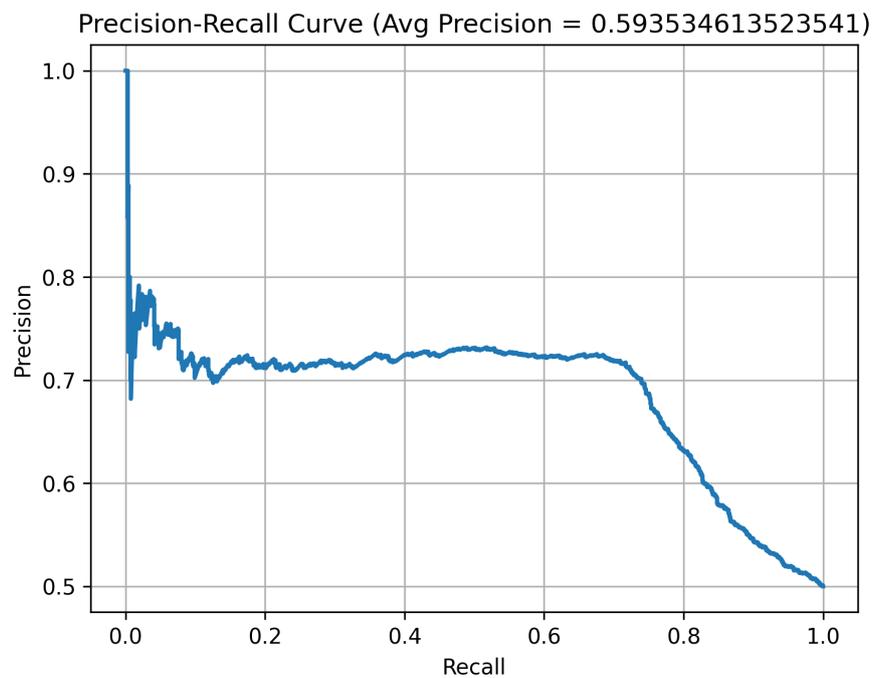Model showed Accuracy that equals to 71.7 percent with Log Loss equls to 0.6501.

5.2. Plots

5.2.1. Confusion Matrix

Figure 4. Confusion Matrix

5.2.2. Precision

Figure 5. Precision
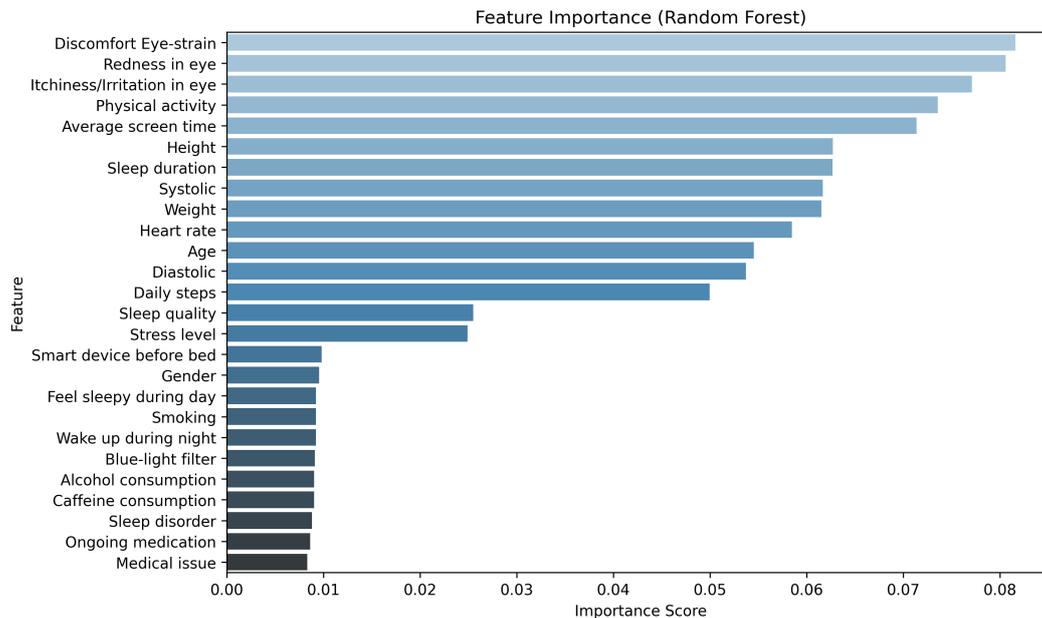
5.2.3. Features Importance

Figure 6. Feature Importance
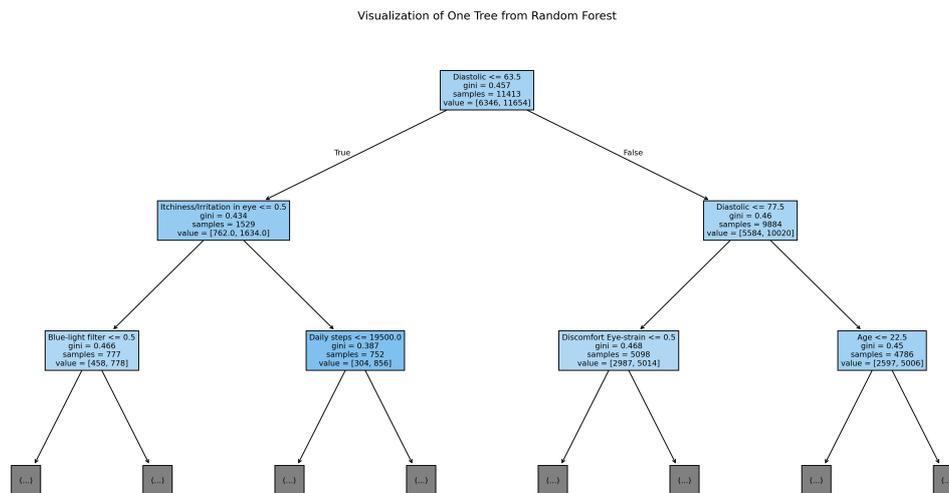
5.2.4. Visualization of Forest

Figure 7. Visualization of One Tree in the Forest

6. Discussion

The Random Forest classifier achieved an accuracy of 71.7 percent. Performance is not perfect but fairly reasonable for classification of Dry Eye Disease (DED). While not perfect, this percentage of accuracy allows, in my opinion, to serve as a supportive tool, for non-clinical decisions or early intervention.

The model's log loss was 0.6501. That is a moderate level of loss. The lower loss is hardly achievable because of subjective nature of the dataset.

The feature importance plot revealed that the most influential features were:

1. Discomfort Eye-Strain
2. Redness in eye
3. Itchiness/Irritation in eye

4.   Physical Activity
5.   Average Screen Time
6.   Height

7. Conclusions

This study' goal was developing a machine learning model,specifically a Random Forest classifier to predict Dry Eye Disease (DED). The model achieved an accuracy of 71.7 percent and a log loss of 0.6501, a reasonable ability to classify DED cases, although some prediction uncertainty remains.

The feature importance analysis revealed that the most influential predictors of DED were Discomfort Eye-Strain, Redness in Eye, Itchiness/Irritation in Eye, Physical Activity, Average Screen Time, and even Height. Most of these results are understandable. Eye strain, redness, and irritation are common symptoms of DED, while physical activity and screen time are factors from outside that are known to impact eye surface health. The most unexpected outcome in this study was impact of patient's height on decision. It can be influenced by other factors that are not in the dataset.

Conclusion: this work shows the potential of machine learning models as support tool for medical field as an early identification of Dry Eye Disease. With access to more objective and more diverse datasets, such models could play a big role not only in ophthalmology, but in all medical field.

References

1.   Craig, J.P.; Nichols, K.K.; Akpek, E.K.; Caffery, B.; Dua, H.S.; Joo, C.K.; Liu, Z.; Nelson, J.D.; Nichols, J.J.; Tsubota, K.; et al. TFOS DEWS II Definition and Classification Report. *The Ocular Surface* **2017**, *15*, 276–283. https://doi.org/10.1016/j.jtos.2017.05.008.
2.   Mayo Clinic Staff. Dry Eyes - Symptoms and Causes. https://www.mayoclinic.org/diseases-conditions/dry-eyes/symptoms-causes/syc-20371863, 2023. Accessed: 2025-05-05.
3.   American Academy of Ophthalmology. Vitamin Deficiency and the Eye. https://www.aao.org/eye-health/diseases/vitamin-deficiency, 2022. Accessed: 2025-05-05.
4.   Healthline Editorial Team. Dry Eyes: Causes, Symptoms, and Treatments. https://www.healthline.com/health/dry-eyes, 2023. Accessed: 2025-05-05.
5.   Medical News Today Editorial Team. Dry Eyes: Symptoms, Causes, and Treatments. https://www.medicalnewstoday.com/articles/170743, 2023. Accessed: 2025-05-05.
6.   Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260.
7.   Rokach, L.; Maimon, O. Decision Trees. In *The Data Mining and Knowledge Discovery Handbook*; Maimon, O.; Rokach, L., Eds.; Springer: Boston, MA, 2005; pp. 165–192. https://doi.org/10.1007/0-387-25465-X_9.
8.   Salman, H.A.; Kalakech, A.; Steiti, A. Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning* **2024**, pp. 69–79. https://doi.org/10.58496/BJML/2024/007.
9.   Zheng, e.a. The Association Between Sleep Disorders and Dry Eye. *Frontiers in Medicine* **2022**, *9*, 832851. https://doi.org/10.3389/fmed.2022.832851.
10.   Shokr, H.; Wolffsohn, J.S.; Trave Huarte, S.; Scarpello, E.; Gherghel, D. Dry eye disease is associated with retinal microvascular dysfunction and possible risk for cardiovascular disease. *Acta Ophthalmologica* **2021**, *99*, e1236–e1242. Epub 2021 Feb 11, https://doi.org/10.1111/aos.14782.
11.   Navarro-Lopez, S.; Moya-Ramón, M.; Gallar, J.; Carracedo, G.; Aracil-Marco, A. Effects of physical activity/exercise on tear film characteristics and dry eye associated symptoms: A literature review. *Contact Lens and Anterior Eye* **2023**, *46*, 101854. Epub 2023 May 10, https://doi.org/10.1016/j.clae.2023.101854.
12.   Ho, K.C.; Jalbert, I.; Watt, K.; Golebiowski, B. A Possible Association Between Dry Eye Symptoms and Body Fat: A Prospective, Cross-Sectional Preliminary Study. *Eye Contact Lens* **2017**, *43*, 245–252. https://doi.org/10.1097/ICL.0000000000000275.
13.   Crum, Jon. Why Am I Waking Up with Dry Eyes? https://drjoncrum.com/why-am-i-waking-up-with-dry-eyes/, 2023. Accessed: 2025-05-06.
14.   Gu, Y.; Cao, K.; Li, A.; et al. Association between sleep quality and dry eye disease: a literature review and meta-analysis. *BMC Ophthalmology* **2024**, *24*, 152. https://doi.org/10.1186/s12886-024-03416-7.
15.   WebMD Editorial Staff. Can Caffeine Help with Dry Eye? https://www.webmd.com/eye-health/caffeine-dry-eye, 2023. Accessed: 2025-05-06.

16.   Assil Eye Institute. Alcohol and Dry Eye Syndrome. https://assileye.com/blog/alcohol-and-dry-eye-syndrome/, 2023. Accessed: 2025-05-06.

17.   WebMD Editorial Staff. Smoking and Dry Eyes: What to Know. https://www.webmd.com/eye-health/smoking-dry-eyes, 2023. Accessed: 2025-05-06.

18.   Zaina Al-Mohtaseb and Scott Schachter and Bridgitte Shen Lee and Jaclyn Garlich and William Trattler. The Relationship Between Dry Eye Disease and Digital Screen Use. *Clinical Ophthalmology* **2021**. https://doi.org/https://doi.org/10.2147/OPTH.S321591.

19.   Jansen, J.A.; Kuswidyati, C.; Chriestya, F. Association between screen time and dry eye symptoms. *Indonesian Journal of Medicine and Health* **2021**. https://doi.org/10.20885/JKKI.Vol12.Iss2.art7.

20.   Cheng, H.M.; Chen, S.T.; Liu, H.J.; Cheng, C.Y. Does blue light filter improve computer vision syndrome in patients with dry eye? *Life Science Journal* **2014**, *11*, 612–615.

21.   Kaur, K.; Gurnani, B.; Nayak, S.; Deori, N.; Kaur, S.; Jethani, J.; Singh, D.; Agarkar, S.; Hussaindeen, J.R.; Sukhija, J.; et al. Digital Eye Strain - A Comprehensive Review. *Ophthalmology Therapy* **2022**, *11*, 1655–1680. Epub 2022 Jul 9, https://doi.org/10.1007/s40123-022-00540-9.

22.   Jenkins Eye Care. Understanding the Link Between Dry Eyes and Eye Irritation. https://jenkinseyecare.com/understanding-the-link-between-dry-eyes-and-eye-irritation/, 2023. Accessed: 2025-05-06.

23.   IEEE Spectrum. Top Programming Languages 2024. https://spectrum.ieee.org/top-programming-languages-2024, 2024. Accessed: 2025-05-06.