**General**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID[[1]](#endnote-1) |  | | | |
| Use case name | AI solution to predict Post-Operative Visual Acuity for LASIK Surgeries | | | |
| Application domain | Healthcare | | | |
| Deployment  model | Cloud services | | | |
| Status | In operation | | | |
| Scope[[2]](#endnote-2) | Predicting Post-Operative Visual Acuity for LASIK Surgeries from retrospective LASIK surgery data with patient follow-ups. | | | |
| Objective(s) | Given: Pre-operative examination results and demography information about a patient. Predict: Post-operative UCVA after one day, one week and one month of the surgery. | | | |
| Narrative | Short description (not more than 150 words) | LASIK (Laser-Assisted in SItu Keratomileusis) surgeries have been quite popular for treatment of myopia, hyperopia and astigmatism over the past two decades. In the past decade, over 10 million LASIK procedures had been performed in the United States alone with an average cost of approximately $2000 USD per surgery. While 99% of such surgeries are successful, the commonest side effect is a residual refractive error and poor uncorrected visual acuity (UCVA). In this work, we aim at predicting the UCVA post LASIK surgery. We model the task as a regression problem and use the patient demography and pre-operative examination details as features. To the best of our knowledge, this is the first work to systematically explore this critical problem using machine learning methods. Further, LASIK surgery settings are often determined by practitioners using manually designed rules. We explore the possibility of determining such settings automatically to optimize for the best post-operative UCVA by including such settings as features in our regression model. Our experiments on a dataset of 791 surgeries provides an RMSE (root mean square error) of 0.102, 0.094 and 0.074 for the predicted post-operative UCVA after one day, one week and one month of the surgery respectively. | | |
| Complete description | **Introduction to LASIK surgeries**  Refractive surgeries for eye are performed to correct (normalize) the refractive state of the eye, to decrease or eliminate dependency on glasses or contact lenses. This can include various methods of surgical remodeling of the cornea or cataract surgery. LASIK is a refractive eye surgery that uses a laser to correct nearsightedness, farsightedness, and/or astigmatism. In LASIK, a thin flap in the cornea is created using either a microkeratome blade or a femto-second laser. The surgeon folds back the flap, then removes some corneal tissue underneath using a laser. The flap is then laid back in place, covering the area where the corneal tissue was removed. With nearsighted people, the goal of LASIK is to flatten the steep cornea; with farsighted people, a steeper cornea is desired. LASIK can also correct astigmatism by smoothing an irregular cornea into a more normal shape. LASIK surgeries are highly popular; over 10 million LASIK procedures have been performed in the United States alone in the past decade.  **Motivation**  While overall patient satisfaction rates after primary LASIK surgery have been around 95%, it may not be recommended for everybody for two reasons: (1) high cost with potentially no significant improvement for certain types of patients, and (2) possible eye complications after the surgery. LASIK surgeries cost approximately $2000 USD per surgery. An ability to predict post-operative UCVA can help patients make an informed decision about investing their money in undergoing a LASIK surgery or not. It can also help surgeons recommend the most promising type of laser surgery to the patients. How can we perform this prediction? Further, while performing such surgeries, surgeons need to set multiple parameters like suction time, flap and hinge details, etc. These are often set using manually designed rules. Can we design a data driven automated method to suggest the best settings for a patient undergoing a laser surgery of a certain type?  **Problem Definition**  In this paper, we address the following problem.  Given: Pre-operative examination results and demography information about a patient  Predict: Post-operative UCVA after one day, one week and one month of the surgery.  Challenges  The problem is challenging because (1) large amount of data about such surgeries is not easily available; (2) there are a lot of pre-operative measurements that can be used as signals; and (3) data is sparse, i.e., there are a lot of missing values.  **Brief Overview of our Approach**  We model the task as a regression problem. We use domain knowledge to preprocess data by transforming a few categorical features into binary features.We also use average values to impute missing values for numeric features. For categorical features, we impute missing values using the most frequent value for the feature. We evaluate multiple regression approaches. Our experiments on a dataset of 791 surgeries provides an RMSE of 0.102, 0.094 and 0.074 for the predicted post-operative UCVA after one day, one week and one month of the surgery respectively.  **Summary**  – We described a critical problem of predicting post-operative UCVA for patients undergoing LASIK surgeries.  – We modeled the task as a regression problem. We explored the effectiveness of demographic, pre-operative features and surgery settings for the prediction task.  – Using a dataset of 791 LASIK surgeries performed on 404 patients from 2013 and 2014, we tested the effectiveness of the machine learning methods. | | |
| Stakeholders[[3]](#endnote-3) | Hospitals, Patients undergoing LASIK surgeries. | | | |
| Stakeholders’  assets, values[[4]](#endnote-4) |  | | | |
| System’s threats and vulnerabilities[[5]](#endnote-5) | different sources of bias; incorrect AI system use | | | |
| Key performance indicators (KPIs) | ID | Name | Description | Reference to mentioned use case objectives |
| 1 | Recommendation | The system can be used to automatically recommend the right LASIK surgery to the patient. | New use-case in healthcare |
| 2 | Improve accuracy | We found the accuracy of the model to be reasonably good to be practically useful. | Improve accuracy |
|  |  |  |  |
| AI features | Task(s) | Prediction | | |
| Method(s)[[6]](#endnote-6) | Machine Learning, Gradient Boosted Decision Trees Based Regression | | |
| Hardware[[7]](#endnote-7) | Machine with 1 CPU and 2 GB RAM. Any Operating system. | | |
| Topology[[8]](#endnote-8) | LASIK surgeries, UCVA, Uncorrected visual acuity, Regression | | |
| Terms and concepts used[[9]](#endnote-9) |  | | |
| Standardization  opportunities/ requirements |  | | | |
| Challenges and issues | The problem is challenging because (1) large amount of data about such surgeries is not easily available; (2) there are a lot of pre-operative measurements that can be used as signals; and (3) data is sparse, i.e., there are a lot of missing values. | | | |
| Societal  concerns | Description |  | | |
| SDGs[[10]](#endnote-10) | Good health and well-being for people | | |

**Data (optional)**

|  |  |
| --- | --- |
| Data characteristics | |
| Description | The dataset contains information for 404 patients in the age range of 18 to 47 years. 215 of these patients are females, and the rest are males. The 791 LASIK surgeries were done in 2013 and 2014. 397 of the surgeries were performed on the left eye and remaining ones on the right eye. Most of the surgeries are either of the Wavefrontguided- LASIK type or of the Plano-scan-LASIK type. Orbscan is the most popular topography machine used; Oculyzer being the second most popular one. Pre-operative UCVA values vary between 0.15 and 2. Post-operative UCVA values vary between - 0.2 and 1 for day 1, -0.3 and 1 for week 1 and -0.2 and 0.95 for month 1 after the operation. Although usually large datasets improve accuracy of the learned machine learning models, it is difficult to obtain large datasets in this domain. |
| Source[[11]](#endnote-11) | Measured using various medical machines at the LVPEI Eye Institute, Hyderabad, India. |
| Type[[12]](#endnote-12) | Structured Data |
| Volume (size) | 791 instances from 404 patients. |
| Velocity (e.g. real time)[[13]](#endnote-13) | Batch. |
| Variety (multiple datasets)[[14]](#endnote-14) | Single source. Data from multiple centers of the hospital. |
| Variability  (rate of change)[[15]](#endnote-15) | Static. |
| Quality[[16]](#endnote-16) | Contains some noise. High quality after pre-processing. |

**Process scenario (optional)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scenario conditions | | | | | |
| No. | Scenario name | Scenario description | Triggering event | Pre-condition[[17]](#endnote-17) | Post-condition[[18]](#endnote-18) |
| 1 | Pre-processing | Remove unnecessary, noisy, redundant columns. Impute missing values. Remove outliers. | As soon as raw dataset arrives |  | Pre-processed clean data is ready. |
| 2 | Training | Train a model with training samples | Pre-processed clean data is ready. | Pre-processing | Trained regression model |
| 3 | Evaluation | Evaluate whether the trained model is of good accuracy | Completion of training/re-training | Training/re-training | Accuracy values |
| 4 | Prediction/ Deployment | Test new instances using the trained model | When a new patient visits the hospital for LASIK surgery | Training/re-training | Prediction of post-LASIK surgery outcomes |
| 5 | Retraining | Retrain model with more training samples. | Certain period of time has passed since last training/retraining and more training samples are available | Pre-processing | Retrained regression model. |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**Training (optional)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scenario name | Training | | | | |
| Step No. | Event[[19]](#endnote-19) | Name of process/Activity[[20]](#endnote-20) | Primary actor | Description of process/activity | Requirement |
| 1 | Sample Raw data is ready | Pre-processing | AI Cloud Service Provider | Outlier detection, feature selection, missing value imputation | API to perform pre-processsing |
| 2 | Completion of step 1 | Training sample creation | AI Cloud Service Provider | Create training samples by clearly recognizing relevant features and training label for data from step 1 |  |
| 3 | Completion of step 2 | Model training | AI Cloud Service Provider | Train a gradient boosted trees based regression model using training samples from step 2. |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| Specification of training data[[21]](#endnote-21) | |  | | | |

**Evaluation (optional)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scenario name | Evaluation | | | | |
| Step No. | Event[[22]](#endnote-22) | Name of process/Activity[[23]](#endnote-23) | Primary actor | Description of process/activity | Requirement |
| 1 | New patient visits hospital for LASIK surgery | Pre-processing | AI Cloud Service Provider | Get relevant data from various machines based on patient registration form, and do pre-processing. |  |
| 2 | Completion of Step 1 | Prediction | AI Cloud Service Provider | Given pre-processed instances from step 1 and the trained model, compute predictions for the current patient. |  |
| 3 | Completion of Step 2 | Evaluation | AI Cloud Service Provider | Compare the result of Step 2 with that of the results after surgery. |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| Input of evaluation[[24]](#endnote-24) | |  | | | |
| Output of evaluation[[25]](#endnote-25) | |  | | | |

**Execution (optional)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scenario name | Execution | | | | |
| Step No. | Event[[26]](#endnote-26) | Name of process/Activity[[27]](#endnote-27) | Primary actor | Description of process/activity | Requirement |
| 1 | New patient comes in | Pre-processing | Hospital | Pre-process input data from patient |  |
| 2 | Completion of step 1 | Prediction | AI Cloud Service Provider | Hospital uses the model hosted on the cloud to predict post-surgery results for the patient based on input from step 1 |  |
| 3 | Completion of step 2 | Consultation and surgery recommendation | Hospital | Based on results for various types of LASIK surgeries from step 2, suggest the best suitable surgery to patient. |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| Input of Execution[[28]](#endnote-28) | |  | | | |
| Output of Execution[[29]](#endnote-29) | |  | | | |

**Retraining (optional)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scenario name | Retraining | | | | |
| Step No. | Event[[30]](#endnote-30) | Name of process/Activity[[31]](#endnote-31) | Primary actor | Description of process/activity | Requirement |
| 1 | Certain period of time has passed since the last training/retraining | Pre-processing | AI Cloud Service Provider | Outlier detection, feature selection, missing value imputation | API/software to perform pre-processsing |
| 2 | Completion of step 1 | Training sample creation | AI Cloud Service Provider | Create training samples by clearly recognizing relevant features and training label for data from step 1 |  |
| 3 | Completion of step 2 | Model training | AI Cloud Service Provider | Train a gradient boosted trees based regression model using training samples from step 2. |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| Specification of retraining data[[32]](#endnote-32) | | Retraining data has to include recent data | | | |

**References**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| References | | | | | | |
| No. | Type | Reference | Status | Impact on use case | Originator/organization | Link |
| 1 | Resea rch Paper | LASIK  surger y predict ion | Publis hed | High | Microsoft, LVPEI | https://link.springer.com/chapter/10.1007/ 978-3-319-31753-3\_39 |
| 2 | Keyno te video snip | LASIK  surger y predict ion | Availa ble Online | High | Microsoft | https://[www.youtube.com/watch?v=mmD](http://www.youtube.com/watch?v=mmD) z7cwC7CE&t=128s |
| 3 | Relate d Paper | Visual Acuity Predict ion | Publis hed | Medi um | Visx Inc, Sunnyvale, Calif. | https://[www.ncbi.nlm.nih.gov/pubmed/14](http://www.ncbi.nlm.nih.gov/pubmed/14) 50116 |
| 4 | Relate d Paper | Visual Acuity Predict ion for Childre n | Publis hed | Medi um | Department of Ophthalmology, University of Minnesota, Minneapolis, USA. | https[://www.ncbi.nlm.nih.gov/pubmed/89](http://www.ncbi.nlm.nih.gov/pubmed/89) 65225 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**Footnote**

1. Leave this cell blank. [↑](#endnote-ref-1)
2. The scope defines the limits of the use case. [↑](#endnote-ref-2)
3. Stakeholder involved in the scenario - examples are: type of organization; customers, 3rd parties; end users; humans; environment; negative stakeholders (attackers, criminals, etc). [↑](#endnote-ref-3)
4. Assets and values that are valuable to the stakeholders and at the risk of being compromised by the AI system deployment – examples can include competitiveness; reputation or trust; fairness; safety; privacy; stability; etc. [↑](#endnote-ref-4)
5. Threats and vulnerabilities can compromise the assets and values above. Examples are: different sources of bias; incorrect AI system use; new security threats; challenges to accountability; new privacy threats (hidden patterns). [↑](#endnote-ref-5)
6. AI method(s)/framework(s) used. [↑](#endnote-ref-6)
7. Hardware system used. [↑](#endnote-ref-7)
8. Topology is the study of geometric forms differentiated by intersection and bifurcation. The term is used for the graphic aspects network architectures. [↑](#endnote-ref-8)
9. Terms and concepts listed here can be used to extend the work of WG 1 (AWI 22989 and AWI 23053) as necessary. [↑](#endnote-ref-9)
10. The Sustainable Development Goals (SDGs), otherwise known as the Global Goals, are a collection of 17 global goals set by the United Nations General Assembly. SDGs are a universal call to action to end poverty, protect the planet and ensure that all people enjoy peace and prosperity.

    See URL for more details: <http://www.undp.org/content/undp/en/home/sustainable-development-goals.html> [↑](#endnote-ref-10)
11. Origin of data, which could be from instruments, IoT, web, surveys, commercial activity, or from simulations. [↑](#endnote-ref-11)
12. Structured/unstructured Images, voices, text, gene sequences, and numerical. Composite: time-series, graph-structured [↑](#endnote-ref-12)
13. The rate of flow at which the data is created, stored, analysed, or visualized. [↑](#endnote-ref-13)
14. Data from a number of domains and a number of data types. The wider range of data formats, logical models, timescales, and semantics complicates the integration of the variety of data. [↑](#endnote-ref-14)
15. Changes in data rate, format/structure, semantics, and/or quality. [↑](#endnote-ref-15)
16. Completeness and accuracy of the data with respect to semantic content as well as syntactical of the data (such as presence of missing fields or incorrect values) [↑](#endnote-ref-16)
17. Describe which condition(s) should have been met before this scenario happens. [↑](#endnote-ref-17)
18. Describe which condition(s) should prevail after this scenario happens. The post-condition may also define "success" or "failure" conditions. [↑](#endnote-ref-18)
19. The event that triggers the step. This might be completion of the previous event. [↑](#endnote-ref-19)
20. Action verbs should be used when naming activity. [↑](#endnote-ref-20)
21. Training data can be further specified. [↑](#endnote-ref-21)
22. The event that triggers the step. This might be completion of the previous event. [↑](#endnote-ref-22)
23. Action verbs should be used when naming activity. [↑](#endnote-ref-23)
24. Specify input of evaluation. [↑](#endnote-ref-24)
25. Specify output of evaluation. [↑](#endnote-ref-25)
26. The event that triggers the step. This might be completion of the previous event. [↑](#endnote-ref-26)
27. Action verbs should be used when naming activity. [↑](#endnote-ref-27)
28. Specify input of evaluation. [↑](#endnote-ref-28)
29. Specify output of evaluation. [↑](#endnote-ref-29)
30. The event that triggers the step. This might be completion of the previous event. [↑](#endnote-ref-30)
31. Action verbs should be used when naming activity. [↑](#endnote-ref-31)
32. Retraining data can be further specified. [↑](#endnote-ref-32)