TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 24, No. 3, June 2026, pp. 904~914
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v24i3.27580

Development of a recommendation system for selecting a formula in cataract surgery

Arseniy Lomakin[1,2], Anastasiya Donskaya[1,2], Alexander Zubkov[1,2]
[1]Department of Software Engineering, Volgograd State Technical University, Volgograd, Russia
[2]Department of Clinical Engineering and Artificial Intelligence Technologies, Volgograd State Medical University, Volgograd, Russia

Article Info

Article history:

Received Sep 30, 2025
Revised Feb 18, 2026
Accepted Mar 29, 2026

Keywords:

Cataract
Intraocular lens
Machine learning
Optical calculations
Recommender system

ABSTRACT

Accurate intraocular lens (IOL) power calculation remains a critical factor for achieving optimal refractive outcomes in cataract surgery. This study analyzes existing methods and software solutions for selecting formulas used to calculate IOL power. To solve this problem, a support medical decision-making recommendation system (SMDRS) was developed to analyze patient biometric data and predict the most suitable calculation formula. Among the evaluated machine learning approaches, the random forest (RF) algorithm demonstrated the highest stability and classification accuracy, leading to its selection as the core predictive engine. The system was validated using retrospective clinical data and evaluated in a functioning ophthalmology clinic. Performance evaluation demonstrated that the system increased the success rate of surgical outcomes in complex cases from 73.5% to 90.5%, thereby confirming its impact on improving the efficiency of optical calculations in clinical practice. By minimizing human error and standardizing decision-making, the proposed solution offers a robust tool for ensuring consistently superior surgical results.

Corresponding Author:

Arseniy Lomakin
Department of Software Engineering, Volgograd State Technical University
28 Lenina Avenue, 400005, Volgograd, Russia
Email: arseny.lomakin@gmail.com

1. INTRODUCTION

In the context of rapid developments in ophthalmology, clinicians must adapt to rapidly evolving diagnostic technologies and intraocular lens (IOL) calculation methods. Cataracts - a condition in which the natural lens becomes clouded, leading to reduced visual acuity and potentially blindness. According to the World Health Organization (WHO), cataracts affect more than 94 million people worldwide and remain the leading cause of visual impairment and blindness. By 2050, this number is expected to exceed 115 million [1].

The main treatment is cataract surgery, where the clouded lens is replaced with an IOL. The success of the operation largely depends on the correct selection of the IOL's optical power [2]. Currently, doctors use one of four main formulas - Barrett, Hoffer, Haigis, or Holladay [3]. However, accuracy depends on patient-specific biometric parameters and the precision of diagnostic devices, which makes the decision highly complex and increases the risk of human error. With more than 20 relevant biometric variables to consider, even experienced ophthalmologists face challenges, especially since measurements from different devices may vary in accuracy. Moreover, existing tools do not provide individualized formula selection across multiple IOL calculation approaches.

Given these challenges, data-driven approaches have emerged as promising tools to support clinicians in selecting IOL formulas. Specifically, machine learning methods are being introduced to reduce decision-making time and process large amounts of biometric and clinical data to identify patterns. These advanced technologies allow for the creation of automated tools that can significantly enhance the precision of preoperative planning.

Thus, the study focused on building a support medical decision-making recommendation system (SMDRS) in ophthalmology using machine learning methods [4]. Based on the patient's biometric data [5], this system provides the clinician with accurate recommendations regarding the choice of formula for calculating the optimal optical power. Implementation of this tool minimizes the human factor and accounts for individual characteristics, ultimately improving the efficiency and accuracy of cataract surgery.

2. METHOD

To build the recommendation system, it was decided to use data accumulated in the scientific repository of anonymized medical data (SRMD) on all clinical cases of cataracts, created specifically for the accumulation of clinical data and scientific research.

2.1. Collecting and conversions of data

When a patient arrives at the clinic, their data is immediately entered into the medical information system, from where it is transferred to the SRMD in the form of a preoperative examination report for patients diagnosed with cataracts. Similarly, data from the surgical protocol and the postoperative examination protocol, used to record the consequences of surgical intervention and evaluate the effectiveness of the treatment, are also entered into the system. The structure of these key medical documents and their constituent data fields is illustrated in Figure 1.

The system architecture is logically divided into two domains: the main clinical processes, where diagnostic data is aggregated, and the research work processes, which rely on a local clinical data repository. This storage component ingests medical documents from the central system to power the SMDRS module without disrupting routine workflows [6]. The resulting separation of environments and the data flow between them are detailed in Figure 2.

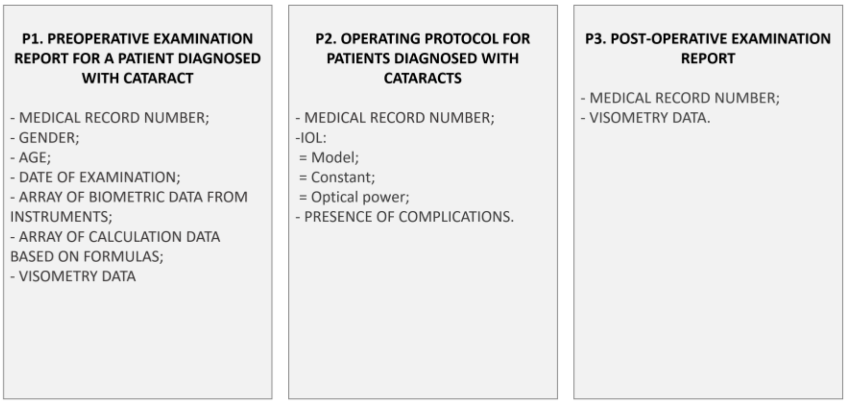


Figure 1. Structure of used medical documents

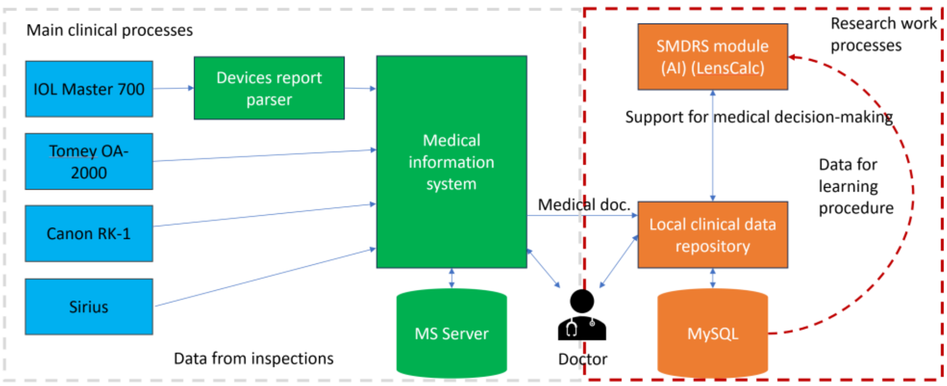


Figure 2. Architectural solution for implementing SMDRS into existing clinical processes

To build a preprocessing pipeline from patient examination reports, a Zeiss IOL Master 500 portable document format (PDF) report parser was initially implemented, which converts key document fields into structured JavaScript object notation (JSON) objects. However, the presence of different branches and other versions of this device required a more adaptable solution to handle format variations. Consequently, the parser was expanded to a more universal one using the Pytesseract computer vision library.

The data obtained by the devices during patient examinations is recorded in a relational database using the entity-attribute-value (EAV) model, which allows for a high degree of scalability but is unsuitable for model training [7]. Therefore, a preliminary transformation was required to convert this raw structure into a usable format. A cascade of structured query language (SQL) queries was developed to convert the EAV tables into third normal form, as illustrated in Figure 3.

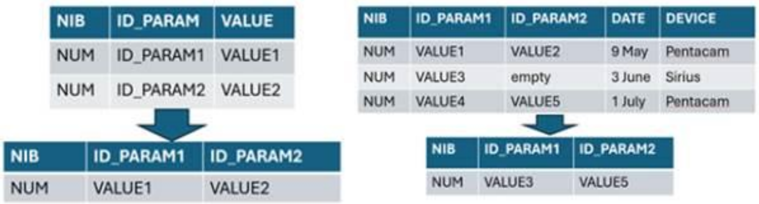


Figure 3. Data conversions

This structural conversion, a dedicated preprocessing pipeline was applied to ensure data quality and consistency. This process consists of main steps.:

a. Missing data handling: records with missing values were handled by strict exclusion to prioritize data quality. This ensured that only clinically complete profiles were used for training.
b. Decoding and cleaning: parameter codes are translated into human-readable names, and duplicate entries are identified and removed to ensure data consistency.
c. Aggregation: measurements from different devices are merged into a single record per eye.
d. Prioritization: inconsistent measurements across different devices are resolved using device hierarchy (from highest to lowest priority): Sirius, IOL Master 700, Tomey OA-2000, Canon RK-1, and Pentacam.

This preprocessing ensures a complete, unified set of parameters for each eye. As a result, the final consolidated dataset included 3213 eyes. This collection covered a broad distribution of axial lengths, including hyperopic, emmetropic, and myopic cases, thereby providing the sufficient scale and diversity required for robust model training.

2.2. Statement of the main formal research task

In our case, in clinical practice, one of four formulas (Haigis, Barrett, Hoffer, or Holladay) is used to calculate the optical power of the lens based on the patient's optometry using data from several Visometric devices. Formally, the task of determining the most suitable formula for calculating the optical power [8] of an IOL can be presented as a classification task. We have a set of structured electronic medical documents (SEMD) of primary patient examinations $M_{P1} = \{P1_1, \dots, P1_n\}$, a set of SEMD of preoperative patient examinations $M_{P2} = \{P2_1, \dots, P2_n\}$, a set of SEMD of postoperative patient examinations $M_{P3} = \{P3_1, \dots, P3_n\}$ and a set of formulas $M_{fm} = \{Fm_1, \dots, Fm_4\}$ [9].

So, we have a set of input characteristics of the patient's biometrics $M_{P12} = M_{P1} \cup M_{P2}$ and the fact of the success of the operation performed using the formula selected by the doctor $C(P3i) \rightarrow \{0, 1\}$ based on retrospective data from the set M_{P3}. Moreover, for each element of the set M_{P12}, we have a set of input characteristics of the patient's eye biometrics $A_n = \{a_1, \dots, a_n\}$. This set includes anatomical measurements such as eye length (AL), anterior chamber depth (ACD), and lens thickness (LT), as well as calculated optical parameters. Explanations for each element of the set A_n are presented in Table 1.

Moreover, the function for binary marking of the success of applying the formula for selecting the optical power of the lens used in the operation based on retrospective data can be represented as (1):

$$\mathrm{C}(P3i) = \{\, 0, if\ SEi > 0.5, 1, if\ SEi \leq 0.5 \quad (1)$$

whereby $SEi = SE(SPHi, CYLi) = |SPHi + 0{,}5 * CYLi|$, where $SPHi$ - postoperative sphere value of the eye, $CYLi$ - postoperative cylinder value of the eye, SEi - postoperative sphere equivalent value of the eye. The threshold of ±0.50 D corresponds to clinically acceptable refractive outcomes in modern ophthalmology.

The Spherical equivalent was chosen as the primary outcome measure because it effectively summarizes the total refractive error. Furthermore, achieving outcomes within this range is a standard benchmark for confirming high-precision surgery and patient satisfaction.

So, the objective function F can be written as (2):

$$F(P1i, Fmj, H_m) = \{ 0,\ if\ C(P3i) = 0;\ 1, if\ C(P3i) = 1 \quad (2)$$

where F represents the classifier algorithm, H_m represents the set of hyperparameters of the classifier F. In this case, the task boils down to training a statistical classifier F', which will be as close as possible to the target function F. This optimization is achieved by maximizing the model's performance according to selected accuracy metrics.

Classification in terms of prediction type can be considered in several ways:

a. Binary answer: the result of the classifier can be considered in terms of belonging to the class of correctness or the impossibility of applying this formula to calculate the diopter of the lens. It can take one of two values $\{0, 1\}$. Thus, $F'\colon M_{P1}{\times}Mfm\ \{0,1\}$
b. Degree of similarity (probability): the result of the classifier can be viewed as the degree to which the transmitted parameters satisfy the conditions for belonging to a particular class. It can take values in the range [0; 1]. Thus $F'\colon M_{P1}Mfm[0,1]$
c. Use of secondary interpretation tools: the results of the classifier are transferred (used) in third-party tools (methods) to obtain additional information about the results of the classifier.

Table 1. Final set An of patient biometric input characteristics based on SEMD primary and preoperative examinations of the patient

Name	Description
GENDER	Patient gender (0 or 1)
EYE	Patient eye (1 for right, 2 for left)
SPH	Eye sphere
CYL	Eye cylinder
AL	Eye length
ACD	Anterior chamber depth
LT	Lens thickness
WTW	Pupillary distance (white to white)
K1	Anterior surface keratometry (mm)
AxF1	Anterior surface keratometry (mm)
K2	Anterior surface keratometry (mm)
AxF2	Anterior surface keratometry (mm)
ΔK	Difference between K1 and K2 (mm)
TK1	Total keratometry (mm)
AxC1	Total keratometry (mm)
TK2	Total keratometry (mm)
AxC2	Total keratometry (mm)
ΔTK	Difference between TK1 and TK2 (mm)
PK1	Posterior surface keratometry (mm)
AxB1	Posterior surface keratometry (mm)
PK2	Posterior surface keratometry (mm)
AxB2	Posterior surface keratometry (mm)
ΔPK2	Difference between PK1 and PK2 (mm)
CW-Chord	Kappa angle (rad)
CCT	Corneal thickness (mm)
AxR	Refractometry
SE-Barrett	Total sphere equivalent of the eye according to Barrett's formula (dpt)
IOL Barrett	Optical power of the lens according to Barrett's formula (dpt)
SE Haigis	Total sphere equivalent of the eye according to Haigis' formula (dpt)
IOL Haigis	Optical power of the lens according to Haigis' formula (dpt)
SE Hoffer	Total sphere equivalent of the eye according to Hoffer's formula (dpt)
IOL Hoffer	Optical power of the lens according to the Hoffer formula (dpt)
SE Holladay	Total spherical equivalent of the eye according to the Holladay formula (dpt)
IOL Holladay	Optical power of the lens according to the Holladay formula (dpt)

The research design is based on a labeled dataset [10]. It was divided into a training set L, a validation set V, and a test set T. The training set is necessary for selecting hyperparameters and tuning machine learning models. The test sample is necessary for evaluating the quality of the models. At the same time, mandatory conditions for these samples must be met, namely, they must not have any intersections, i.e., $L \cap T = \emptyset$. Both samples must also be labeled, i.e., each set of patient biometric characteristics must be assigned a success rate for the operation performed using the formula selected by the doctor.

2.3. Choosing an approach

The next step was to choose between classical or deep machine learning for the task of training classifiers. Classical machine learning is based on methods of mathematical statistics and probability theory, seeking to identify patterns between characteristics and the target variable by minimizing an error function. Deep machine learning, in contrast, uses artificial neural networks with input, hidden, and output layers to perform complex hierarchical transformations of data [11], [12]. In deep learning, weights are adjusted iteratively through feedback loops until the correct result is achieved, requiring significant computational resources. However, classical models are often more suitable for structured data where interpretability and stability on smaller datasets are prioritized over raw feature extraction power.

Classical machine learning models were chosen due to the limited number of available clinical cases, which makes deep learning unsuitable. Initial empirical tests using simple neural network architectures such as multi-layer perceptron's revealed signs of overfitting due to this data scarcity relative to the feature space. Specifically, these networks showed a training accuracy of 63% but failed to generalize on the validation set, achieving only 38% accuracy. Furthermore, transfer learning approaches were considered inapplicable, as pre-trained models for tabular ophthalmic biometric data are virtually absent compared to image-based domains. Given the specific nature of the task, classical approaches such as decision trees (DT), random forests (RF), and gradient boosting (GB) appear more interpretable and better suited to the task at hand.

Globally, machine learning tasks are divided into supervised and unsupervised learning. This study exclusively utilizes supervised learning, where the training data contains input vectors paired with corresponding target values, allowing the system to identify patterns and generalize them for future predictions. Within this framework, the problem is strictly defined as a classification task [13], the goal of which is to assign each input vector a single value from a finite set of discrete classes.

Classifiers are generally based either on probabilistic generative models, which explicitly model input distributions, or on discriminant functions that directly assign an input vector X to a class. If an approach uses Bayes' theorem to model posterior probabilities, it falls under the generative category, whereas direct mapping approaches are considered discriminant. In the context of this research, discriminant binary classification was prioritized to efficiently assign each clinical case to one of two mutually exclusive classes.

Depending on the approach chosen, binary classifiers can output either discrete functions or continuous probability scores. To handle continuous outputs, a compression function such as the sigmoid is used to convert the raw value into a probability within the range of 0 to 1. Finally, to rigorously evaluate these classifiers, the confusion matrix framework was adopted, the detailed metrics of which are described in the following sections.

2.4. Choosing an machine-learning model

To create interpretable binary classification models, the selection focused on classical machine learning algorithms implemented in the Python scikit-learn library [14]. These algorithms were chosen specifically for their transparency and the ability to fine-tune hyperparameters for optimal performance. A comparative overview of the selected methods, detailing their operating principles, advantages, and limitations, is presented in Table 2.

Table 2. Comparative table of classical machine learning algorithms for classification tasks

Algorithm	How it works	Advantages	Disadvantages
DT	A DT builds a branching structure based on training data, breaking it down into subgroups until final decisions are reached	Ease of interpretation, ability to detect nonlinear relationships	A tendency to relearn
GB	A collection of weak models (usually DTs) are combined, each correcting the errors of the previous one	High accuracy, resistance to retraining	Careful tuning of hyperparameters is required
K-nearest neighbors (KNN)	Based on the principle of proximity of objects, a new object is classified based on the classes of its closest neighbors	Ease of implementation, adaptation to complex solution boundaries	Sensitivity to emissions, computing resource requirements
RF	An ensemble of DTs, where each tree votes for a class, and the final result is determined by voting	Resistance to retraining, ability to work with heterogeneous data	They may be less interpretable than individual trees
Support vector machine (SVM)	Constructs a hyperplane in feature space that separates classes	Efficiency in high-dimensional spaces, robustness to overfitting	Sensitivity to emissions

For the experimental phase, the preprocessed dataset was stratified into a training set (70%) and an independent test set (30%). A strict splitting protocol based on individual patients was enforced to ensure that data from fellow eyes of the same patient remained within a single partition, preventing data leakage between the samples. Subsequently, each model was initially trained on standard hyperparameters. For the algorithm with the highest performance, optimal hyperparameters were selected using GridSearchCV, which utilized cross-validation with five folds to evaluate performance across different subsamples. To address potential class imbalance during this phase, we employed a cost-sensitive learning approach by setting class weights inversely proportional to class frequencies. This technique penalizes misclassifications of the minority class more heavily, thereby mitigating bias towards the majority class and improving the model's sensitivity across all formula categories.

2.5. Selecting metrics for evaluating the quality of machine learning models.

Model performance evaluation will include the use of several metrics [15], namely accuracy, precision, recall, F1-score, as well as confusion matrix analysis [16], taking into account macro averaging and weighted averaging variations. These metrics have been specifically chosen because they collectively provide a comprehensive and reliable overview of model effectiveness in binary classification. Such a multi-metric approach is particularly crucial in cases of class imbalance or when different types of errors carry different costs.

Accuracy assesses the overall correctness of the model and is the ratio of correctly classified objects to all objects. While accuracy remains the most widely used metric in binary classification, it can be misleading when applied to imbalanced datasets as it does not distinguish between error types [17]. Therefore, it is critically important to supplement accuracy with other, specific metrics to ensure a complete evaluation.

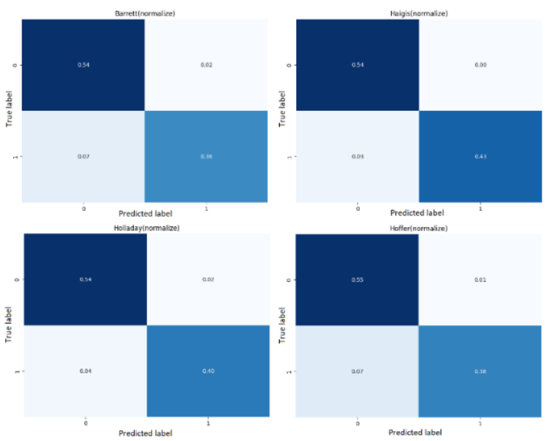
Precision shows what proportion of objects classified as positive are actually positive. This is especially important when false positives are undesirable, such as in medical diagnostics or screening scenarios. However, a high precision [18] can sometimes be accompanied by low recall if many true positives are missed.

Recall (also called sensitivity or completeness) determines what proportion of all actual positive objects were successfully identified by the model. This is critical when it is important to detect all possible positive cases, as in disease detection or fraud detection tasks. High recall ensures that the model does not leave out actual positive cases, even at the expense of some false positives.

F1-score (F-measure) is the harmonic mean between precision and recall [19], combining the strengths of both metrics. It is particularly useful when a balance between precision and recall is needed, or when classes are imbalanced, since F1-score will only be high if both precision and recall are high. It is less sensitive to class imbalance than accuracy and provides a better single-value summary in challenging settings.

Macro averaging calculates the average metric value across all classes, giving equal weight to each class regardless of their frequency. This is useful for evaluating model performance fairly across both major and minor classes in the dataset. Weighted averaging, in contrast, takes class imbalance into account by weighting the metric values by the number of instances of each class, thus providing a more representative average for imbalanced datasets.

Confusion matrix visualizes the classification results by showing the counts of true positive, true negative, false positive, and false negative predictions [20]. This matrix allows for detailed error analysis and helps identify specific cases where the model makes mistakes, which is essential for model improvement and interpretation, especially in sensitive applications. The resulting confusion matrices obtained for each of the four calculation formulas, along with a detailed explanation of the error categories, are illustrated in Figure 4.



Explanation of the error matrices obtained	Class 0 (model response)	Class 1 (model response)
Class 0 (correct answer)	The percentage of cases where the model detected that the formula could not be applied and this is true	The percentage of cases where the model identified the possibility of applying the formula and it was incorrect
Class 1 (correct answer)	The percentage of cases where the model identified the possibility of applying the formula and it is true	The percentage of cases where the model detected that the formula could not be applied and this was an error

Figure 4. Obtained matrices and their explanation

All these metrics allow for the robust evaluation of models in predicting the correct formulas for calculating the optical power of a lens based on eye measurement data. By analyzing multiple performance indicators, we can gain insight into different aspects of classification quality. This comprehensive approach ensures that neither imbalanced data nor specific types of errors are overlooked during the validation process.

3. RESULTS AND DISCUSSION

The following metrics will be used to evaluate the accuracy of the models and compare them: accuracy, precision macro, recall macro, F1-score macro, precision weighted, recall weighted, F1-score weighted, and error matrix. A detailed description of these metrics is provided in section 2.7.

The following models are considered as comparable classical machine learning classification algorithms: DT, GB, k-nearest neighbors (KNN), RF, and SVM. Each of the compared algorithms [21], [22] is an implementation from the scikit-learn library of the Python language [23]. All models considered were compared using standard hyperparameters set by the library that implements them.

Each classifier model was evaluated independently for all four target formulas used to calculate lens diopters. The primary goal of this comparison was to identify the optimal algorithm for each specific calculation method. The results of training on the base models are presented in the Table 3.

Table 3. The results of training on the base models for analyzing the applicability of each formula

Classifier model	Accuracy	Recall macro	Precision macro	F1-score macro	Recall weighted	Precision weighted	F1-score weighted
DT for Barrett formula	0.81	0.82	0.82	0.81	0.81	0.82	0.81
GB for Barrett formula	0.67	0.67	0.68	0.67	0.66	0.67	0.67
KNN for Barrett formula	0.74	0.74	0.74	0.74	0.73	0.74	0.74
RF for Barrett formula	0.88	0.86	0.85	0.86	0.88	0.89	0.89
SVM for Barrett formula	0.61	0.61	0.60	0.61	0.63	0.62	0.61
DT for Haigis formula	0.80	0.81	0.79	0.80	0.80	0.81	0.80
GB for Haigis formula	0.64	0.64	0.65	0.64	0.63	0.64	0.64
KNN for Haigis formula	0.8	0.81	0.79	0.8	0.8	0.81	0.8
RF for Haigis formula	0.88	0.79	0.86	0.82	0.88	0.88	0.87
SVM for Haigis formula	0.59	0.57	0.58	0.59	0.6	0.6	0.59
DT for Holladay formula	0.79	0.77	0.77	0.77	0.79	0.79	0.79
GB for Holladay formula	0.5	0.48	0.49	0.49	0.51	0.51	0.5
KNN for Holladay formula	0.81	0.82	0.8	0.81	0.81	0.82	0.81
RF for Holladay formula	0.90	0.88	0.86	0.87	0.9	0.9	0.9
SVM for Holladay formula	0.53	0.5	0.5	0.53	0.51	0.53	0.53
DT for Hoffer formula	0.87	0.86	0.86	0.86	0.87	0.87	0.87
GB for Hoffer formula	0.68	0.65	0.68	0.66	0.68	0.68	0.68
KNN for Hoffer formula	0.79	0.76	0.77	0.77	0.77	0.81	0.79
RF for Hoffer formula	0.88	0.83	0.87	0.85	0.88	0.88	0.88
SVM for Hoffer formula	0.71	0.69	0.72	0.7	0.72	0.7	0.71

The analysis of the results shows that the highest metric values are achieved using ensemble methods, particularly the RF algorithm. For all four formulas, this model demonstrates the best performance across accuracy, precision, and F1-score (up to 0.90), highlighting its robustness and strong generalization ability.

Statistical analysis confirmed that this performance advantage over other models was significant ($p < 0.05$), with 95% confidence intervals indicating high reliability. The KNN and DT models achieve moderate results, with accuracy and recall remaining at acceptable levels (around 0.74-0.81). In contrast, the SVM and GB algorithms show the least stable performance, with metrics ranging between 0.5 and 0.68, making them less suitable for this task. Therefore, it can be concluded that applying ensemble tree-based methods ensures the highest classification quality and can serve as the foundation for building the recommendation system.

Next, the best base models were tuned using the GridSearchCV [24] method. Based on the comparison of models trained on basic hyperparameters, the RF base algorithm achieved the best result for each of the formulas. As a result, after selecting the hyperparameters, the result shown in Table 4 was achieved for the models.

Table 4. Results for the models that showed the best result numerical metrics

Classifier model	Accuracy	Recall macro	Precision macro	F1-score macro	Recall weighted	Precision weighted	F1-score weighted
Barrett formula	0.90	0.90	0.87	0.88	0.90	0.91	0.91
Haigis formula	0.92	0.89	0.90	0.89	0.92	0.92	0.92
Holladay formula	0.90	0.88	0.86	0.87	0.90	0.90	0.90
Hoffer formula	0.95	0.93	0.95	0.94	0.95	0.95	0.95

All trained models met the requirements for metric threshold values [25], [26]. Crucially, achieving accuracy levels of 0.90-0.95 has direct clinical implications: high precision minimizes the risk of significant postoperative refractive errors, thereby reducing the likelihood of secondary surgical interventions and improving overall patient safety. This confirms that the resulting system is suitable for use in real clinical practice.

As a result, we obtained a system [27] with a robust architectural design that significantly facilitated the work of doctors and improved the accuracy of decision-making when selecting the optical power of IOL for patients with cataracts. The core of this architecture features a scalable node of independent binary classifiers, which process verified medical data in parallel before aggregating the results for final validation. The complete schematic representation of this pipeline is illustrated in Figure 5.

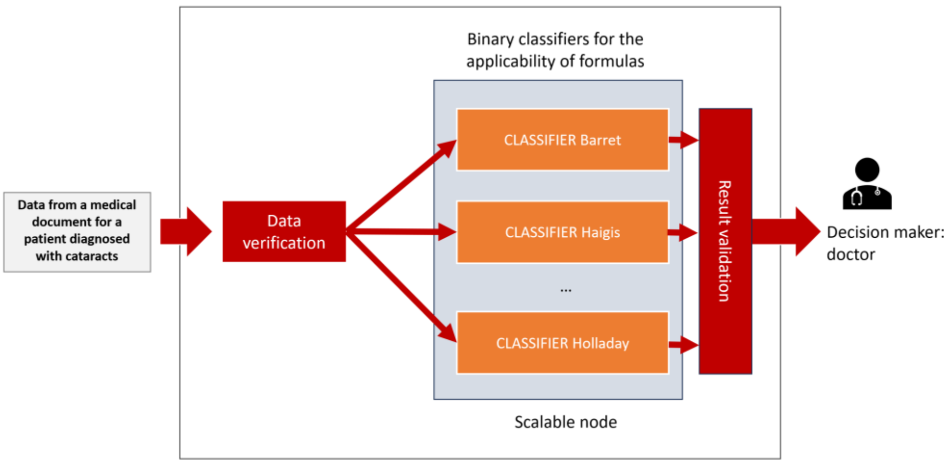


Figure 5. Recommendation system scheme

The developed system is of particular value for novice ophthalmologists, as the automated formula selection significantly reduces the risk of errors associated with lack of experience. The architecture, based on the use of multiple independent classifiers, not only verifies the applicability of different formulas but also generates parallel results that undergo validation before being presented to the doctor. As a result, the specialist receives a reliable set of recommendations grounded in objective data analysis rather than solely on subjective judgment. This not only accelerates the decision-making process but also provides an additional educational effect for inexperienced doctors, helping them more quickly understand the patterns of formula selection.

The clinical validation of the system was conducted over a six-month period from 08-01-2025 to 01-07-2025. The pilot study included a total sample of 674 clinical cases involving patients diagnosed with cataract. In the post-operative context, surgical success was defined as achieving an SE within ±0.50 D of the target refraction one month after surgery.

During this validation period, ophthalmologists interacted with SMDRS through a seamless interface integrated into the electronic health record system, where biometric data was automatically uploaded for instant analysis. Pilot users reported that this automated workflow significantly reduced average preoperative planning time, allowing them to focus more on patient counseling. Qualitative feedback praised the system's high usability, noting that the intuitive visual comparison provided a clear rationale for decisions and increased confidence in complex cases.

However, the integration of automated recommendations into clinical practice entails certain limitations and ethical considerations. While the system demonstrates high accuracy, it relies on retrospective data and may not fully account for rare anatomical anomalies. Therefore, SMDRS is designed strictly as a decision support tool, where the final ethical responsibility remains with the ophthalmologist to avoid automation bias and ensure patient safety.

4. CONCLUSION

The developed system provides clear, data-driven recommendations based on the analysis of multiple factors. This analysis incorporates a wide range of inputs, including biometric eye measurements and clinical characteristics of patients. Specifically, its use minimizes the risk of formula selection errors, ensuring optimal surgical outcomes.

Retrospective data analysis at the clinic was conducted over a three-month testing period to evaluate surgical outcomes in complex clinical cases. In this group, 90.5% of surgeries using SMDRS achieved the target postoperative visual acuity, compared to only 73.5% without it. Thus, the application of SMDRS demonstrated a substantial 17% improvement in the rate of successful postoperative outcomes.

Future research will focus on developing a hybrid AI cascade that integrates both regression and classification algorithms. This system will not only recommend the optimal calculation formula but also directly predict the precise IOL power needed to achieve target refraction. Furthermore, the model will be trained to forecast potential postoperative refractive deviations, thereby reducing the likelihood of refractive surprises in high-risk patients.

Ultimately, the decision support system shows strong potential for integration into ophthalmic clinical practice. Its application standardizes the process of selecting a formula for IOL calculation, reduces dependence on the subjective experience of the surgeon, and incorporates multiple factors that were previously overlooked. As a result, surgical planning becomes more accurate, the risk of complications decreases, and the overall prognosis for visual recovery in patients improves significantly.

ACKNOWLEDGMENTS

The authors would also like to express their sincere gratitude to A. R. Vinogradov for his contribution to this research.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Arseniy Lomakin	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Anastasiya Donskaya				✓		✓				✓		✓		
Alexander Zubkov						✓		✓		✓		✓		

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT
Authors state no conflict of interest.

INFORMED CONSENT
We used a de-identified clinical dataset and therefore obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL
The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee.

DATA AVAILABILITY
The data that support the findings of this study are available from the corresponding author, Arseniy Lomakin, upon reasonable request.

REFERENCES

[1] World Health Organization, Aug. 10, 2023, "Blindness and vision impairment," world health organization. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment (Accessed: Dec. 04, 2025).

[2] Y. Zhou *et al.*, "The accuracy of intraocular lens power calculation formulas based on artificial intelligence in highly myopic eyes: a systematic review and network meta-analysis," *Frontiers in Public Health*, vol. 11, p. 1279718, Nov. 2023, doi: 10.3389/fpubh.2023.1279718.

[3] W. Stopyra, D. L. Cooke, and A. Grzybowski, "A review of intraocular lens power calculation formulas based on Artificial Intelligence," *Journal of Clinical Medicine*, vol. 13, no. 2, p. 498, Jan. 2024, doi: 10.3390/jcm13020498.

[4] F. Funer *et al.*, "Impacts of clinical decision support systems on the relationship, communication, and shared Decision-Making between health care professionals and patients: Multistakeholder interview study," *Journal of Medical Internet Research*, vol. 26, p. e55717, Jun. 2024, doi: 10.2196/55717.

[5] S. Nadeem, "Cataract surgery: historical devices, modern innovations, and future perspectives," *Expert Review of Medical Devices*, vol. 21, no. 11, pp. 991–994, Oct. 2024, doi: 10.1080/17434440.2024.2419477.

[6] M. R. Khan, "Advancements in Deep Learning Architectures: A Comprehensive review of current trends," *Journal of Artificial Intelligence General Science (JAIGS)*, vol. 1, no. 1, Jan. 2024, doi: 10.60087/jaigs.v1i1.29.

[7] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Computer Science: Artificial Intelligence*, May 2017, doi: 10.48550/arxiv.1705.07874.

[8] Z. Chen *et al.*, "Harnessing the power of clinical decision support systems: challenges and opportunities," *Open Heart*, vol. 10, no. 2, p. e002432, Nov. 2023, doi: 10.1136/openhrt-2023-002432.

[9] H.-J. Hu, Y.-L. Chang, Y.-Y. Dai, Y.-L. Yu, and X. Zhang, "Magnetic resonance imaging assessment of the lateral geniculate nucleus volume and height in patients with glaucoma: a Meta-analysis," *International Journal of Ophthalmology*, vol. 16, no. 4, pp. 638–644, Apr. 2023, doi: 10.18240/ijo.2023.04.19.

[10] S. G. Finlayson *et al.*, "The clinician and dataset shift in artificial intelligence," *New England Journal of Medicine*, vol. 385, no. 3, pp. 283–286, Jul. 2021, doi: 10.1056/nejmc2104626.

[11] R. S. Abdulsadig and E. Rodriguez-Villegas, "A comparative study in class imbalance mitigation when working with physiological signals," *Frontiers in Digital Health*, vol. 6, Mar. 2024, doi: 10.3389/fdgth.2024.1377165.

[12] A. Mosavi, S. Shamshirband, E. Salwana, K. wing Chau, and J. H. M. Tah, "Prediction of multi-inputs bubble column reactor using a novel hybrid model of computational fluid dynamics and machine learning," *Engineering Applications of Computational Fluid Mechanics*, vol. 13, no. 1, pp. 482–492, 2019, doi: 10.1080/19942060.2019.1613448.

[13] K. P. Murphy, *Probabilistic Machine Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2022.

[14] N. Purtova and G. v. Maanen, "Data as an economic good, data as a commons, and data governance," *Law Innovation and Technology*, vol. 16, no. 1, pp. 1–42, Nov. 2023, doi: 10.1080/17579961.2023.2265270.

[15] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, no. 1, p. 6086, Mar. 2024, doi: 10.1038/s41598-024-56706-x.

[16] M. Shirdel, M. D. Mauro, and A. Liotta, "Worthiness Benchmark: A novel concept for analyzing binary classification evaluation metrics," *Information Sciences*, vol. 678, p. 120882, Jun. 2024, doi: 10.1016/j.ins.2024.120882.

[17] S. A. Hicks *et al.*, "On evaluation metrics for medical applications of artificial intelligence," *Scientific Reports*, vol. 12, no. 1, p. 5979, Apr. 2022, doi: 10.1038/s41598-022-09954-8.

[18] G. Flores, A. Schiff, A. H. Smith, J. A. Fukuyama, and A. C. Wilson, "A consequentialist critique of binary classification evaluation practices," *Computer Science: Machine Learning*, Jun. 2025, doi: 10.48550/arXiv.2504.04528.

[19] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary

classification evaluation," *BMC Genomics*, vol. 21, no. 6, Jan. 2020, doi: 10.1186/s12864-019-6413-7.
[20] J. Opitz, "A closer look at classification evaluation metrics and a critical reflection of common evaluation practice," *Transactions of the Association for Computational Linguistics*, vol. 12, pp. 820–836, Jan. 2024, doi: 10.1162/tacl_a_00675.
[21] D. Bugl, *Modern Full-Stack React Projects: Build, maintain, and deploy modern web apps using MongoDB, Express, React, and Node.js*, Birmingham, UK: Packt Publishing, 2023.
[22] R. Nazir, A. Bucaioni, and P. Pelliccione, "Architecting ML-enabled systems: Challenges, best practices, and design decisions," *Journal of Systems and Software*, vol. 207, p. 111860, Oct. 2024, doi: 10.1016/j.jss.2023.111860.
[23] Scikit-learn Developers, "API Reference scikit-learn 1.5 documentation," *scikit-learn*, 2026. [Online]. Available: https://scikit-learn.org/stable/api/index.html. (Accessed: May 10, 2024).
[24] W. Xu, Y.-L. Fu, and D. Zhu, "ResNet and its application to medical image processing: Research progress and challenges," *Computer Methods and Programs in Biomedicine*, vol. 240, p. 107660, Jun. 2023, doi: 10.1016/j.cmpb.2023.107660.
[25] C. Ekström and K. Hårleman, "Risk factors for incident open-angle glaucoma in clinical practice in Sweden: A population-based case–control study," *Acta Ophthalmologica*, vol. 101, no. 5, pp. 530–535, Feb. 2023, doi: 10.1111/aos.15644.
[26] D. M. Wright *et al.*, "Validating and Updating the OHTS-EGPS Model Predicting 5-year Glaucoma Risk among Ocular Hypertension Patients Using Electronic Records," *Ophthalmology Glaucoma*, vol. 8, no. 2, pp. 143–151, Nov. 2025, doi: 10.1016/j.ogla.2024.10.009.
[27] L. F. Nakayama, L. Z. Ribeiro, F. K. Malerbi, and C. V. S. Regatieri, "Ophthalmology and Artificial intelligence: Present or Future? A Diabetic Retinopathy Screening Perspective of the Pursuit for Fairness," *Frontiers in Ophthalmology*, vol. 2, p. 898181, May 2022, doi: 10.3389/fopht.2022.898181.

BIOGRAPHIES OF AUTHORS

Arseniy Lomakin received the Bachelor's degree in Software Engineering and in Management from the Volgogtrad State Technical University, Volgograd, Russia, in 2024. He obtains the Master's degree in Software Engineering and Computer Science. He is currently a lecturer in Volgograd State Medical University and head of medical projects. He leads the development and management of innovative medical projects, ensuring alignment with institutional goals, improving MIS in Volgograd region. Supervises technical teams and oversee project milestones from conception to implementation. His current research interests include ophthalmology, machine learning, system analysis, and convolutional neural networks. He can be contacted at email: arseny.lomakin@gmail.com.

Anastasiya Donskaya received the Bachelor's degree in Fundamental Physics from Volgograd State Technical University, Volgograd, Russia, in 2020. In 2022, she obtained the Master's degree in Software Engineering and Computer Science from the same university. In 2025, she completed her Ph.D. studies in Systems Analysis. She is currently a senior lecturer at the Department of Software for Automated Systems, Volgograd State Technical University, and a senior lecturer at the Department of Clinical Engineering and Artificial Intelligence Technologies, Volgograd State Medical University, Ministry of Health of the Russian Federation. Her research interests include machine learning, data analysis, systems analysis, decision support systems, and the application of mathematical methods in medicine. She can be contacted at email: donsckaia.anastasiya@yandex.ru.

Alexander Zubkov received the Bachelor's and Master's degree in Software Engineering, completed Ph.D. studies in System Analysis and defended his degree in Technical Sciences, head of the Department of Information Development at Volgograd State Medical University Ministry of Health of Russia, Associate Professor of the Department of Software for Automated Systems at Volgograd State Technical University, Volgograd, Russia. His current research interests include medicine, machine learning, system analysis, convolutional neural networks and algorithms. He can be contacted at email: zubkov.alexander.v@gmail.com.