A Machine Learning Based Dose Prediction of Lutein Supplements for Individuals With Eye Fatigue

Juntao Kan\textsuperscript{1†}, Ao Li\textsuperscript{2†}, Hong Zou\textsuperscript{2}, Liang Chen\textsuperscript{1} and Jun Du\textsuperscript{1*}

\textsuperscript{1} Nutrilite Health Institute, Shanghai, China, \textsuperscript{2} Department of Bioinformatics, WuXi NextCODE Genomics, Shanghai, China

**Purpose:** Nutritional intervention was always implemented based on “one-size-fits-all” recommendation instead of personalized strategy. We aimed to develop a machine learning based model to predict the optimal dose of a botanical combination of lutein ester, zeaxanthin, extracts of black currant, chrysanthemum, and goji berry for individuals with eye fatigue.

**Methods:** 504 features, including demographic, anthropometrics, eye-related indexes, blood biomarkers, and dietary habits, were collected at baseline from 303 subjects in a randomized controlled trial. An aggregated score of visual health (VHS) was developed from total score of eye fatigue symptoms, visuoagnosis persistence, macular pigment optical density, and Schirmer test to represent an overall eye fatigue level. VHS at 45 days after intervention was predicted by XGBoost algorithm using all features at baseline to show the eye fatigue improvement. Optimal dose of the combination was chosen based on the predicted VHS.

**Results:** After feature selection and parameter optimization, a model was trained and optimized with a Pearson’s correlation coefficient of 0.649, 0.638, and 0.685 in training, test and validation set, respectively. After removing the features collected by invasive blood test and costly optical coherence tomography, the model remained good performance. Among 58 subjects in test and validation sets, 39 should take the highest dose as the optimal option, 17 might take a lower dose, while 2 could not benefit from the combination.

**Conclusion:** We applied XGBoost algorithm to develop a model which could predict optimized dose of the combination to provide personalized nutrition solution for individuals with eye fatigue.

**Keywords:** dose prediction, machine learning, XGBoost, lutein supplements, eye fatigue

**INTRODUCTION**

Eye fatigue also known as asthenopia, is a common condition in both adults and children, which can be caused by various reasons, especially the intensive use of electronic products e.g., computers, cell phones and iPads (1). Nutritional intervention that provides certain benefit to ocular health has been researched for decades (2). Lutein and zeaxanthin are known to protect retina because of their anti-oxidant nature and ability to absorb high-energy blue light produced by visual display units (3). Moreover, the two ingredients, when combined with other botanical ingredients with
rich anthocyanin, were reported to relieve eye fatigue (4). So far, however, most strategies for preventing or reducing the incidence of the symptoms are based on “one size fits all” public health recommendations to the whole population.

Based on anthropometrics, blood biomarkers, dietary habits and physical activities, the solution of personalized nutrition has been tailored to meet specific nutritional needs these years (5). Machine learning as a field of computer science adopts computer algorithms to identify patterns in large datasets with numerous variables, which can be used to predict data-based outcomes (6). Machine learning algorithms, including random forest (RF), extremely randomized trees (ET), extreme gradient boosting (XGBoost), and gradient boosting decision tree (GBDT), usually establish a model from test inputs to make predictions or decisions based on the data (7). Nowadays, machine learning techniques have proven to be highly effective for prediction of response to methotrexate and antidepressant medication, and diagnoses of pediatric diseases and upper gastrointestinal cancer (7–10).

A novel combination of lutein ester, zeaxanthin, extracts of black currant, chrysanthemum and goji berry was previously developed, showing protective effects on eye fatigue, dry eye, and macular function in a randomized controlled study (RCT) (11). Here using machine learning technology, we tried to, by leveraging several algorithms, predict the optimal dose of the combination based on the features collected in the RCT, in order to provide personalized nutrition solution for the future in real world.

METHODS

Clinical Trial

The RCT was conducted in Aier Eye Hospital, Shanghai, China. A total of 360 subjects with eye fatigue were initially enrolled in the study at baseline, and randomized into 4 arms (arm 1: placebo; arm 2, test product containing 6 mg of lutein; arm 3, test product containing 10 mg of lutein; arm 4, test product containing 14 mg of lutein) to receive either test products or placebo orally once daily for 90 days. Each subject had a total of 3 visits at baseline (visit 1, V1), 45 days (visit 2, V2), and 90 days (visit 3, V3) throughout the study. During the study, 42 subjects withdrew after the first visit and 15 more subjects withdrew from the study after the second visit, leaving 303 subjects in the statistical analysis. Scores of eye fatigue, dry eye, and macular function were aggregated to a score named visual health score (VHS).

Feature Selection and Model Evaluation

XGBoost, AdaBoost, and ElasticNet algorithms were applied in the model building. All baseline features were used for training, while VHS of post-intervention was used as the target for prediction. Pearson’s correlation coefficient (PCC) between real and predicted VHS was used to evaluate model performance. Features were selected using stability-selection method (12). In brief, the method took different subset of samples to train model with different complexity (lambda parameter) before outputting a set of stability scores for each feature. Tree depth from 2 to 10 of XGBoost was used as lambda parameter for stability-selection, and the threshold was determined by: (1) keeping reasonably less features; (2) not lowering average PCC of 10× cross validation (10×CV) in training set. Finally, the model was retrained using selected features with the optimized parameters before it was used to predict VHS of post-intervention in test and validation sets where PCC was calculated for model evaluation.

Dose Prediction

Twenty eight subjects in test set and 30 in validation set were used for dose prediction. It was supposed that each subject took 3 different doses (containing 6, 10, 14 mg of lutein). The validated model was used to predict VHS at V2 of each subject that underwent three separate doses. The dose with the highest VHS was recognized as the optimal option.
### TABLE 1 | Abbreviation list.

| Abbreviation | Full Name                                | Abbreviation | Full Name                                |
|--------------|------------------------------------------|--------------|------------------------------------------|
| BFR          | Body fat rate                            | RDWCV        | Red cell distribution width-coefficient of variation |
| CRE          | Creatinine                               | RDWSD        | Red cell distribution width-standard deviation |
| Dairy3       | Whole milk                               | RTLC         | Retinal thickness of left eye-central average |
| DBP          | Diastolic blood pressure                 | RTLCMn       | Retinal thickness of left eyeentral minimum |
| FFM          | Fat free mass                            | RTLU6        | Retinal thickness of left eye-up 6 mm to center |
| Fruit3       | Banana                                   | RTRD3        | Retinal thickness of right eye-down 3 mm to center |
| Fruit8       | Pear                                     | SBP          | Systolic blood pressure                  |
| HDL          | High density lipoprotein cholesterol     | ST           | Schirmer test                            |
| LDL          | Low density lipoprotein cholesterol      | TG           | Triglyceride                             |
| LYMHC        | Lymphocyte count                         | TSEFS        | Total score of eye fatigue symptoms      |
| LYMHP        | Lymphocyte proportion                    | Veg4         | Cruciferous vegetables                   |
| MCHC         | Mean corpuscular hemoglobin concentration | Veg7         | Other vegetables                         |
| MCV          | Mean corpuscular volume                  | VF           | Visceral fat                             |
| MPOD         | Macular pigment optical density          | VHS          | Visual health score                      |
| PDW          | Platelet distribution width              | VP           | Visuognosis persistence                  |
| PLT          | Blood platelet count                     | WBC          | White blood cell count                   |

**FIGURE 1 | Workflow overview.**

**RESULTS**

The Botanical Formula Improved Eye Fatigue

Baseline characteristics, including demographic, anthropometrics, and eye-related indexes, were shown in **Table 2.** An aggregated score, VHS, was developed from 4 eye-related features, including TSEFS, VP, MPOD and ST, to represent an overall eye fatigue level, following a normal distribution (Figures 2A, B). Formula intervention with different doses significantly improved VHS at V2 and V3 compared to placebo arm (Figure 2C). Since there was a slight, but not
significant, increase of VHS at V3 compared to that at V1 in placebo arm, while the values of VHS at V2 and V1 were comparable, the VHS at V2 was chosen as a quantitative measurement of eye fatigue improvement.

**XGBoost Successfully Selected Key Features and Predicted Eye Fatigue Improvement**

All features at baseline were used as training features to construct a baseline regression model to predict VHS at V2 by applying different algorithms. Among them, XGBoost (PCC of training set =0.618) outperformed others such as AdaBoost (0.604) and ElasticNet (0.412), and therefore, was selected for the model building. Next, stability-selection was applied to select key features with getting better performance. A total number of 25 key features (out of 504) were selected with a stability score threshold being 0.8 (Figure 3A). The features were composed of 5 parts: (1) dose; (2) VHS; (3) demographic and anthropometrics features; (4) blood test features and (5) OCT features (Figure 3B).

Based on the selected 25 features, a model was trained and optimized with a PCC of 0.649 in training set, which could predict test and validation set with a PCC of 0.638 and 0.685, respectively (Table 3). Thus, there was no obvious overfitting in our model.

**Our Model Remained Good Performance Without Blood Test and OCT Features**

Considering the feasibility of sample collection, we tried to train our model without blood test features. By redoing stability-selection, another 17 features (out of 447) were selected with the stability score threshold being 0.85. The model was trained and optimized, with a PCC of 0.652, 0.667, and 0.683 in training, test and validation set, respectively (Table 3). Considering the cost of sample collection, we additionally removed OCT features. Another set of 19 features (out of 413) was selected with the stability score threshold being 0.65 to train a new model, in which the PCC was 0.648, 0.623, and 0.679 in training, test and validation set, respectively (Table 3, Figure 4A). After the blood test and OCT features were excluded, both were replaced by dietary features (Figure 4B).

**Our Model Could Predict Personalized Dose for Each Subject**

Among 28 subjects in test set and 30 in validation set, 56 subjects (96.6%) showed significant eye fatigue improvement with VHS elevated by more than 0.1 in 45 days, while other 2 subjects (3.4%) could not benefit from our botanical combination since they already had a relatively high VHS at baseline. 39 subjects (67.2%) should take 14 mg as the optimal dose, and other 17 subjects (29.3%) might take the combination at a lower dose level as the difference value of predicted VHS at V2 was <0.05 among 3 different doses (Figure 5).

**DISCUSSION**

Machine learning techniques have been well applied in medicine these years. Compared with traditional statistical models, they have many advantages including high power and accuracy, ability to model non-linear effects with more complex, high-dimensional and interactive variables, interpretation of large genomic data sets, robustness to parameter assumptions and dispense with normal distribution test (7, 13). Machine learning can be used for diagnoses of pediatric diseases, detection of upper gastrointestinal cancer combined with endoscopy, and prediction of mortality in patients with suspected coronary artery disease or rheumatoid arthritis (9, 10, 14, 15). For drug therapy, it is widely used to predict clinical response to antidepressant and optimal dose of warfarin (8, 16, 17). However, investigation on nutrition intervention is limited, especially lutein-related phytoneutrients supplement.

Personalized nutrition received more and more attention these years. By utilizing machine learning algorithm of stochastic gradient boosting regression, glycemic responses to different types of food was predicted by Zeevi’s group (18). In our study, hundreds of features, including anthropometrics, dietary habits, blood biomarkers, and eye-related indexes, were collected, and eye fatigue improvement was successfully predicted using XGBoost algorithm with PCC of training set 10xCV=0.618, which showed better performance than AdaBoost and ElasticNet algorithms. Upon feature selection and parameter optimization, the PCC was elevated to 0.649, which was comparable to that (0.68) in Zeevi’s study. More importantly, using this validated model, we could predict the optimal dose of the botanical combination of lutein supplements for subjects with eye fatigue symptom and provide them with personalized nutrition solution. In our cohort, a certain number of subjects (67.2%) could benefit from the highest dose, some (29.3%) need lower doses to receive efficient supplement, and a small proportion (3.4%) could not benefit from the intervention. It suggested the rationality and necessity of personalized nutrition solution for phytoneutrient supplementation.

To predict the eye fatigue improvement, VHS was developed from 4 eye-related features, including TSEFS, VP, MPOD, and ST to represent an overall eye fatigue level. Therefore, the
FIGURE 2 | Aggregation of visual health score. (A) Distribution of 4 eye-related features. TSEFS, VP and MPOD were Z-score transformed as they followed normal distribution approximately. ST was not normally distributed, then it was min-max scaling transformed. (B) Distribution of VHS. VHS was taken as mean of these 4 transformed features, which followed a normal distribution of N (0.064,0.485) with P-value of Kolmogorov–Smirnov test = 0.557. (C) Violin plot of aggregated VHS. Red: VHS at V1 (0 day); green: VHS at V2 (45 days); blue: VHS at V3 (90 days).

TABLE 3 | PCC of different models.

|                          | Training set (10 × CV) | Test set | Validation set |
|--------------------------|------------------------|----------|----------------|
| Original model           | 0.649                  | 0.638    | 0.685          |
| Model without blood test features | 0.652                  | 0.667    | 0.683          |
| Model without blood test or OCT features | 0.648                  | 0.623    | 0.679          |

top 2 features in our selected feature list were VHS and dose regardless of blood test and OCT features. Age was reported to be significantly correlated with eye fatigue in both children and white-collar worker with visual display units (19, 20). In particular, age-related macular degeneration could be prevented by lutein supplement, potentially attributed to the elevation in MPOD (21, 22). Age was another selected feature in our study. It was reported that there was a significant inverse relationship between the percentage of body fat and MPOD (23, 24). Body fat loss was positively related with the increased serum concentrations of lutein and zeaxanthin (25). In addition, HDL was significantly related to MPOD, serum lutein and zeaxanthin (26). So anthropometric features, such as BFR, BFM, FFM, and weight, and blood biomarkers, such as HDL, LDL, and TG, were selected for model development. Lutein and zeaxanthin are found in relatively high concentrations in leafy-green vegetables and brightly colored fruits. Cruciferous vegetables, such as kale and brocoli, and spinach are good dietary source (27). In addition, consumption of one egg per day increased serum lutein and zeaxanthin concentrations due to high bioavailability of lutein in the yolk (28). After blood test and OCT features were excluded, both were replaced by dietary features in the selected feature list.

Drawing blood is an invasive method of sample collection, and always leads to incompliance of the individuals. OCT is a non-invasive, cross-sectional imaging technique to evaluate retinal structure but with high cost and limited application (29). Considering the feasibility and cost of sample collection, we excluded blood test and OCT features to train another model, and finally got comparable performance with the original one. This optimization would make the personalized nutrition solution much easier to be accepted by the consumers. Genotype
should be considered as one of the most important features in the personalized nutrition (30). But to date, little is known about the relationship between eye fatigue and genotype. In addition, genotyping increases the cost and time, which is not as convenient, efficient and cheap as the collection of phenotypes, such as anthropometrics, questionnaires and point-of-care.
testing (POCT). Therefore, in this study, we did not collect genotype features. Even so, there was still a report showing that SNP variants in BCMO1 and CD36 were associated with plasma lutein concentration and MPOD in humans (31). It suggests potential application of genotype in personalized nutrition for eye health, which will be considered in our future studies.

Besides the deficiency of genotype features, there are still several limitations in our study. In this study, subjects were administered only 3 doses, which limited the predicted dose to three. And this is the flaw of the algorithm itself. The study is additionally limited by relatively small sample size, limited ophthalmic examination findings in the enrollment of the subjects, and lack of external validation for model development.

In spite of these limitations, the study has its superiority. First, we predicted exact dose instead of response to the intervention. Second, to our knowledge, this is the first study to utilize machine learning technologies to investigate nutrition intervention, particularly the phytonutrients supplementation, for eye health. Third, the feature collection is simple and easy. All features, including anthropometrics, questionnaires and POCT, can be collected in a friendly way by staff instead of experienced doctor. Last but not the least, with the established prediction model, we can provide an integrated personalized nutrition solution in combination with the launched botanical product to the consumers, shedding new light on the direct-selling business.

In conclusion, by selecting different types of features collected in the RCT and applying XGBoost algorithm, we developed a model for predicting the optimized dose of the combination to provide personalized nutrition solution for consumer with eye fatigue in real world.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IRB of Shanghai Nutrition Society. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JK and JD designed the research. AL and HZ conducted the research. JK, AL, and LC analyzed the data. JK and AL wrote the paper. JD had primary responsibility for final content. All authors read and approved the final manuscript.

FUNDING

This study was funded by National Key R&D Program of China (2018YFC1706800), Nutrition Research Foundation of Chinese Nutrition Society (CNS-DSM2018A35), and Innovation Funding of Amway (China) R&D Center.

ACKNOWLEDGMENTS

We would like to thank Dr. Li Zhang and his colleagues from SPRIM China for their excellent work in coordinating the clinical trial.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnut.2020.577923/full#supplementary-material
REFERENCES

1. Abdi S, Rydelberg A. Asthenopia in schoolchildren, orthoptic and ophthalmological findings and treatment. Doc Ophthalomol. (2005) 111:65–72. doi: 10.1007/s10633-005-4722-4
2. Mccusker MM, Durrani K, Payette MJ, Suchecki J. An eye on nutrition: the role of vitamins, essential fatty acids, and antioxidants in age-related macular degeneration, dry eye syndrome, and cataract. Clin Dermatol. (2016) 34:276–85. doi: 10.1016/j.clindermatol.2015.11.009
3. Johnson EJ. Role of lutein and zeaxanthin in visual and cognitive function throughout the lifespan. Nutr Rev. (2014) 72:605–12. doi: 10.1111/nure.12133
4. Yagi A, Fujimoto K, Michihiko K, Goh B, Tsu D, Nagai H. The effect of lutein supplementation on visual fatigue: a psychophysiological analysis. Appl Ergon. (2009) 40:1047–54. doi: 10.1016/j.apergo.2009.04.013
5. Livingstone KM, Celis-Morales C, Navas-Carretero S, San-Cristobal R, Macready AL, Falaffa R, et al. Effect of an Internet-based, personalized nutrition randomized trial on dietary changes associated with the Mediterranean diet: the Food4Me Study. Am J Clin Nutr. (2014) 104:288–97. doi: 10.3945/ajcn.115.129049
6. Obermeyer Z, Emanuel EJ. Predicting the future-big data, machine learning, and clinical medicine. N Engl J Med. (2016) 375:1216. doi: 10.1056/NEJMmp160181
7. Mo X, Chen X, Li H, Li J, Zeng F, Li J, et al. Early and accurate prediction of clinical response to metformin treatatment in juvenile idiopathic arthritis using machine learning. Front Pharmacol. (2019) 10:1155. doi: 10.3389/fphar.2019.01155
8. Chekroud AM, Zotti RJ, Shehzad Z, Guercguevria R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry. (2016) 3:243–50. doi: 10.1016/S2215-0366(15)00471-X
9. Liang H, Tsui BY, Ni H, Valentin CC, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. (2019) 25:433–8. doi: 10.1038/s41591-018-0335-9
10. Loo H, Xu G, Li C, He L, Luo L, Wang Z, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. Lancet Oncol. (2019) 20:1645–54. doi: 10.1016/S1470-2045(19)30637-0
11. Kan J, Wang M, Yao Y, Liu H, Chen L, Zhang X, et al. A novel botanical formula improves eye fatigue and dry eye: a randomized, double-blind, placebo-controlled study. Am J Clin Nutr. (2020) 112:334–42. doi: 10.1093/ajcn/nqa139
12. Meinschansen N, Bühlmann P. Stability selection. J R Stat Soc B. (2010) 72:417–73. doi: 10.1111/j.1467-9868.2010.00740.x
13. Tang J, Liu R, Zhang Y-L, Liu M-Z, Hu Y-F, Shao M-J, et al. Application of machine-learning models to predict tacrolimus stable dose in renal transplant recipients. Sci Rep. (2017) 7:42192. doi: 10.1038/srep42192
14. Lezcano-Valverde JM, Salazar F, León L, Tolezano E, Jover JA, Fernandez-Gutierrez B, et al. Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach. Sci Rep. (2017) 7:10189. doi: 10.1038/s41598-017-10558-w
15. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Lutein and zeaxanthin carotenoids and their role in eye health. Nutrients. (2013) 5:1169–85. doi: 10.3390/nu5041169
16. Chung H-Y, Rasmussen HM, Johnson EJ. Lutein bioavailability is higher from lutein-enriched eggs than from supplements and spinach in men. J Nutr. (2004) 134:1887–93. doi: 10.1093/jn/134.8.1887
17. Menke MN, Dabov S, Knecht P, Sturm V. Reproducibility of retinal thickness measurements in healthy subjects using spectralis optical coherence tomography. Am J Ophthalmol. (2009) 147:467–72. doi: 10.1016/j.ajo.2008.09.005
18. Görmü K, Mathers JC, Grimvald KA, Ahlgren J, Nordström K. Do we know enough? A scientific and ethical analysis of the basis for genetic-based personalized nutrition. Genes Nutr. (2013) 8:373–81. doi: 10.1007/s12263-013-0338-6
19. Borel P, De Edelenyi FS, Vincent-Baudry S, Malezet-Desmoulin C, Marrotta A, Lian B, et al. Genetic variants in BCMO1 and CD36 are associated with plasma lutein concentrations and macular pigment optical density in humans. Ann Med. (2011) 43:47–59. doi: 10.3109/07853890.2010.531757

Conflict of Interest: The study design, implementation, analysis, and interpretation were carried out jointly by the funders. The study products were produced and supplied by Amway (China) R&D Center, JK, LC, and JD are employees of Nutrilite Health Institute, a division of Amway.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kan, Li, Zou, Chen and Du. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.