IEEE International Conference on Biomedical and Health Informatics (BHI), Ioannina, Greece, September 2022.

Explainable Machine Learning for Vitamin A Deficiency Classification in Schoolchildren

Jayroop Ramesh
Computer Science and Engineering
American University of Sharjah
Sharjah, United Arab Emirates
b00057412@aus.edu

Donthi Sankalpa
Computer Science and Engineering
American University of Sharjah
Sharjah, United Arab Emirates
g00062902@aus.edu

Amar Khamis
College of Medicine
Mohammed Bin Rashid University
Dubai, United Arab Emirates
amar.hassan@mbru.ac.ae

Assim Sagahyroon
Computer Science and Engineering
American University of Sharjah
Sharjah, United Arab Emirates
asagahyroon@aus.edu

Fadi Aloul
Computer Science and Engineering
American University of Sharjah
Sharjah, United Arab Emirates
faloul@aus.edu

***Abstract*—Vitamin A deficiency is one of the leading causes of visual impairment globally. While blood tests are common approaches in developed countries, various socioeconomic and public perspectives render this a challenge in developing countries. In Africa and Southeast Asia, the alarming rise of preventable childhood blindness and delayed growth rates has been dubbed as an ”epidemic”. With the proliferation of machine learning in clinical support systems and the relative availability of electronic health records, there is the potential promise of early detection, and curbing ocular complication progression. In this work, different machine learning methods are applied to a sparse dataset of ocular symptomatology and diagnoses acquired from Maradi, Nigeria collected during routine eye examinations conducted within a school setting. The goal is to develop a screening system for Vitamin A deficiency in children without requiring retinol serum blood tests, but rather by utilizing existing health records. The SVC model achieved the best scores of accuracy: 75.7%, sensitivity:83.7%, and specificity: 74.9%. Additionally, Shapley values are employed to provide post-hoc clinical explainability (XAI) in terms of relative feature contributions with each classification decision. This is a vital step towards augmenting domain expert reasoning, and ensuring clinical consistency of shallow machine learning models.**

***Index Terms*—Electronic Health Records, Decision Making, Machine Learning, Ophthalmology, Optometry, Refractive Errors, Vitamin A, XAI**

I. Introduction

The World Health Organization reports the inordinate prevalence of Vitamin A deficiency in developing countries, particularly those in Africa and South-East Asia. The most severe effects of this illness is observed in young children, and is primarily associated with poor dietary nutrition. While it is a preventable condition, an estimated 250,000 to 500,000 children become blind, with increase in the risks for respiratory and diarrhoeal infections, decrease in growth rates and bone development, and worsened likelihood of survival. [1].

It is becoming imperative to identify patients with Vitamin A deficiency, for targeted supplementation by doctors because of the association between this condition and major ocular degenerative conditions as well as delayed growth. For example, one of the most common co-occurring condition is that of refractive errors. Refractive errors are aberrations causing poor focus of light onto the retina, resulting in impaired vision. More specifically, it is the mismatch between the optical refractive determinants of the eye in terms of the corneal curvature, lens power, lens location and axial length. Refractive errors can be categorized as myopia, hyperopia, or astigmatism, and stem from a multitude of potential causes including environmental, genetic and even lifestyle factors [2]. Electronic real-world clinical data consisting of patients' health, medical history and treatment outcomes and train intelligent, predictive models to aid diagnosticians and facilitate early screening [3]. As found by the studies [4] and [5], Vitamin A awareness is extremely low, and so is the general scarcity of clinical preventative measures owing to socio-economic reasons. Relatively speaking, ocular afflictions, such as visual acuity, night blindness, etc. are physically noticeable in children and therefore exists more chances on an individual levels for consultation and intervention with an optometrist/ophthalmologist. A retinol serum test measures the level of vitamin A in the blood through venipuncture, and while this is a routine procedure in developed countries, the costs of blood tests and persisting stigma around needles as a consequence of the widespread safety issues of HIV [6] dissuade the undertaking of regular tests. As such, alternatives enabled by technical advancements can offer promising auxiliary approaches for screening for Vitamin A disorders, in a resource-effective and cost-effective manner.

Prior studies in the domain intersection of opthalmology, optometry and machine learning can be divided into two categories; 1) predicting refractive error risk (myopia, hyperopia or astigmatism), and 2) assessing vision outcomes after treatment. The following paragraph provide an overview of the recent advances in terms of their key results and the machine learning approaches used. Notably, Vitamin A is not considered as predictor variables in analysis of ocular diseases across these works, and commonly utilize axial length measurements, spherical equivalent readings and age. This is likely due to the lower prevalence of childhood blindness and lesser severity of Vitamin A deficiency in developed countries where the studies are conducted.

[7] proposes logistic regression and multilayer perceptron to identify people at risk of developing myopia. Their dataset and was derived from the Orinda Longitudinal Study of Myopia database. After feature selection, the logistic regression model performed better than the multilayer perceptron at classification accuracy, of 92.2% with 5 features. This study shows the influence of lifestyle activities on the prognosis of myopia in the context of machine learning. [8] proposes an SVM for the prediction of myopia (diopters) in grade 6 children based on historical ocular measurements and behavior data. The dataset used is collected from several primary schools belonging to the Henan province in China The SVM model achieved accuracy, precision, sensitivity, specificity, F1-score and AUC scores of 92%, 95%, 94%, 94%, 94% and 98% respectively through a 10-fold cross-validation approach. This study is of importance in consolidating the impact of lifestyle factors over time (5 years) on myopia prediction. [9] leverages Big Data and the random forest algorithm to predict myopia (diopters) in terms of the right eye spherical equivalent parameter up to 8 years in the future. The dataset was compiled using electronic health records of children aged 6 to 20 years from 8 different ophthalmic centers across a 10 year period. Additionally, two population based datasets named Guangzhou Outdoor Activity Longitudinal Trial, and Refractive Error Longitudinal Study. The algorithm when applied to the internal validation subset, 7 external validation subsets and 2 population based datasets achieved MAE ranging from 0.503 to 0.799 for 8 years, as well as AUC ranging from 85.2% to 88.8% for 8 years, 80.2% to 88.6% for 8 years. [10] sought to estimate the physiological elongation of ocular axial length in right eyes using linear regression based on routine clinical data to better understand the amount of myopia progression in children undergoing ortho-K treatment. The dataset was constructed using electronic health records from the Peking University People's Hospital, Beijing, China. The linear regression model obtained R^2 score of 0.87, and the primary contributing factors were age, spherical equivalent and mean K reading. This is beneficial for detecting children who can become more myopic at an early age by defining the physiological component in the progression of axial length elongation. [11] adopts machine learning to identify ideal candidates to undergo corneal refractive surgery for refractive error correction (in both eyes). The dataset was retrospectively collected from the BVIIT Eye Center in South Korea. The best performing model was an ensemble classifier composed of support vector machines, multilayer perceptron, random forest, AdaBoost, and LASSO, which obtained an external validation score of AUC: 97.2% and accuracy: 93.4% through a 10-fold cross validation approach.

For the successful adoption of models in critical domains such as healthcare, medical experts must have a clear understanding of the model's behavior, their potential biases and diagnostic capability [12]. The ever increasing complexity of new machine learning models makes it difficult for experts to discern their nature, thereby introducing the need for methods to provide explainability for artificial intelligence (XAI) and transparency as to which features were most important for the trained models.

The contributions of this study are listed as follows.

1) Present machine learning methods utilizing data symptoms and refractive error diagnoses derived from electronic health records during routine eye checkups.
2) Provide explainable Vitamin A deficiency screening as a surrogate method without requiring relatively expensive or invasive blood tests in developing countries.

This paper is organized as follows: section III discusses the methodology, section III presents the results and discussion, and section VI concludes the work.

II. Materials and Methods

The end-to-end sequential pipeline of i) dataset preprocessing, ii) model training, iii) hyperparameter optimization culminating in iv) evaluation is depicted in Figure 1.

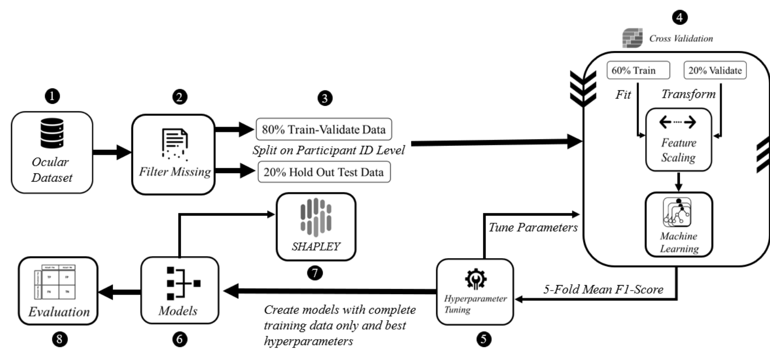

Fig. 1: End-to-end model creation pipeline

A. Dataset

This data was collected from the town of Maradi in Niger, Africa from school children aged 6-15, studying in grades 1 to 9 during government mandated annual eye checkups. Initially there are 125 variables, across 86,216 records. Primarily, patient-specific information such as name, ID, family occupation and school name were removed for de-identification. Aligning with the results of Little's MCAR test, filtering was performed for rows where columns had missing data higher than 70%. Redundant columns as per Spearman's correlation higher than 0.5 were also removed separately. The pre-processed dataset had 24 categorical variables, across 18,423 unique patient records. The final set of variables chosen for constructing models include age, gender, presence/absence of itching, blepharitis, conjunctivitis, cataracts, blunt force trauma, ptosis, pthisis bulbi, myopia, hyperopia, astigmatism, exotropia, esotropia, alternative squint, amblyopia, nystugmus, megalocornea, and opaque cornea disorders. Examining the features in light of medical knowledge, it is found that frequent itching, and corneal impairments like cataracts have a strong association with the target variable [13].

To overcome bias due to the relative abundance of patients with eye disease but no vitamin A deficiency versus patients with the former and the latter, random under-sampling and random over-sampling was utilized. This yielded an equal distribution of 1466 in each class, for a total of 2932 instances. The dataset instances after separation by patient ID were divided into a standard 60-20-20 training-valid-test, across five-fold cross-validation.

B. Models

In concordance with recent literature [3], the frequently used traditional and ensemble machine learning methods were employed. For the former, Support Vector classifier (SVC), Logistic Regression (LR) and K-Nearest Neighbors (KNN) were used, and for latter, Light Gradient Boosting (LGB), eXtreme Gradient Boosting (XGB), Categorical Boosting (CB), and Random Forest (RF) were used.

C. Shapley Values

For XAI, Shapley values, which is a concept from cooperative game theory that fairly distributes a payout generated by the grand coalition in a game to each of its players is utilized. The classic form of the Shapley value estimation for a feature i is presented in Equation 1 [14]:

$$\phi_y(x_i; x) = \sum_{i \not\ni S} \frac{|S|!(n-1-|S|)!}{n!}\left[f_y(x_S \cup x_i) - f_y(x_S)\right] \quad (1)$$

To estimate Shapley value for feature x_i of a single data instance x in our ocular dataset, a set of all possible feature unions subsets with n features excluding i is generated. The value of feature i is obtained by the difference between the results of the characteristic function f_S applied on the set of all features, and the set of features excluding i. The Shapley value is then estimated by averaging the marginal contributions of x_i across all generated feature union orderings [14]. When a feature x_i contributes highly to the classification of data instance x, it will have a relatively larger Shapley value score. When a feature x_i has negligible contribution to a particular classification outcome, its corresponding Shapley value score is lower than the other features.

III. Results

The classification metrics of accuracy, sensitivity, specificity and F1-score are employed for the quantitative evaluation of the algorithms. The results are reported in Table I. The rationale for testing multiple algorithms is to provide a systematic evaluation benchmark in the realm of machine learning for the purpose outlined in this work. Empirically, random under-sampling yielded quantitatively better results than random over-sampling after experimentation with different stratification of data subsets. Regardless, the difference in performances were only marginally higher in favor of random under-sampling, by a factor of 2-3% on sensitivity and F1-score. To improve model performance, hyper-parameter optimization method Random Search was leveraged for tuning a search space of parameters during cross-validation. The search space is a bounded domain of hyperparameter values, and random sampling is performed to elicit different combinations. As the SVC model performed relatively the best, the tuned parameters are *kernel*: linear, *gamma*: 1.18737 and *C*: 0.05635.

While all classifiers achieved relatively similar acceptable average performances, SVC had marginally superior scores as can be surmised from Table I across accuracy, specificity and F1-score.

TABLE I: Quantitative model performance metrics

Model	Accuracy	Sensitivity	Specificity	F1-Score
SVC	75.7	83.7	74.9	40.7
LR	75.6	83.8	74.6	40.5
KNN	72.9	80.9	72.0	37.3
LGB	73.3	84.3	72.1	38.6
XGB	75.2	83.9	72.0	40.2
CB	74.5	83.6	73.5	39.5
RF	72.5	84.3	71.2	38.0

For explaining which feature contributed to a single patient's outcomes, and by *how much*, the Shapley value analysis was performed on the best (SVC) model and aggregated as in Figure 2.

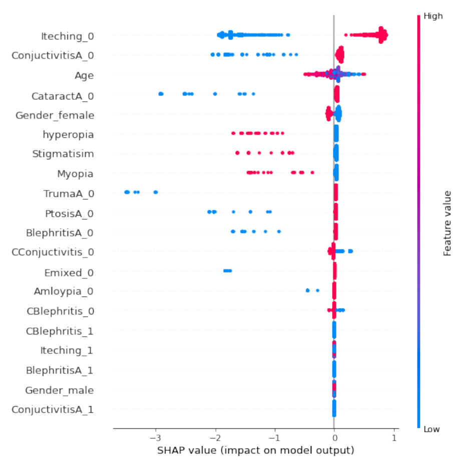

Fig. 2: Global explanations with Shapley Values

IV. Discussion

Traditional models and ensemble models obtaining scores near to each other implies a certain level of linearity assumption is valid for the dataset. As such, it appears that the complex nature of ensemble models to capture higher dimensional relationships is not necessary. Noticeably, only a few features are numerical and largely categorical features are treated differently by each algorithm internally. The balance between precision and recall is relatively low as denoted by F1-Score, owing to a likely saturation of borderline cases, where patients have symptoms very similar to that of a confirmed Vitamin A deficit patient, yet belongs to the healthy class. It is also possible that the stochastic behavior of random undersampling applied for correcting imbalances led to unfavorable instances. These "harder" to classify instances are where separation boundaries are not clearly defined within the scope of the available features. Synthetic augmentation was not utilized in this work for two reasons. First was to observe the baseline performance, and second was to keep the instance distributions as close as to the original patient data without major external intervention.

Aposteriori analysis was preferred as almost all features are positively correlated with the target variable of Vitamin A with reasonable strength. Permutation feature importance methods (for tree-based models and apriori feature selection is based on diminishing model performance, whereas Shapley values is based on magnitude of feature attributions. As mentioned previously, Figure 2 confers a notion of interpretability on a granular level to aid in the reduction of false positives and false negatives. Upon observation of Figure 2, it appears empirically that absence of conjunctivitis, cataracts, and itching but the presence of a singular refractive error among older children increases potential of deficiency. Although itching and conjunctivitis can share a similar pathology with other ocular diseases, their emergence as xerophthalmia in accompaniment with refractive errors and cataracts can be indicative of severe Vitamin A deficiency [15]. Approaches like this can help doctors prioritize certain predictor features on a patient-by-patient basis to provide precision medicine, provided that XAI techniques' results are viewed in purview of domain expertise, as posited by [3].

V. Conclusion

This study is one of the first works to explore the clinical utility of electronic records containing symptoms and prior diagnoses only for the detection of Vitamin A deficiency with an XAI perspective. A potential limitation in medical record acquisition is process uniformity, measurement units and nature of variables collected. This dataset is obtained from a single region, constraining any inferences to the local population demographics. Recent research dictates site-agnostic data harmonization to be a valid mitigation technique. Future work can consider them in addition to employing various synthetic oversampling and undersampling methods for better utilization of available data, as well as incorporation of visual acuity prescription values for improved model performance measures.

References

[1] *Vitamin A Deficiency*. URL: https://www.who.int/data/maternal-newborn-child-adolescent-ageing/advisory-groups/gama/gama-related-resources/nlis (visited on 02/15/2022).

[2] Felix Y. Chau, Dana Wallace, Thasarat Vajaranant, Leon Herndon, Paul Lee, Pratap Challa, Rand Allingham, and Irene Maumenee. "Chapter 31 - Osteogenesis Imperfecta and the Eye". In: *Osteogenesis Imperfecta*. Ed. by Jay R. Shapiro, Peter H. Byers, Francis H. Glorieux, and Paul D. Sponseller. San Diego: Academic Press, Jan. 1, 2014, pp. 289–303. URL: https://www.sciencedirect.com/science/article/pii/B9780123971654000319 (visited on 11/24/2021).

[3] Jamie Duell, Xiuyi Fan, Bruce Burnett, Gert Aarts, and Shang-Ming Zhou. "A Comparison of Explanations Given by Explainable Artificial Intelligence Methods on Analysing Electronic Health Records". In: *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*. July 2021, pp. 1–4.

[4] Ada E. Aghaji, Roseline Duke, and Ugochukwu C. W. Aghaji. "Inequitable Coverage of Vitamin A Supplementation in Nigeria and Implications for Childhood Blindness". In: *BMC Public Health* 19.1 (Mar. 8, 2019), p. 282. URL: https://doi.org/10.1186/s12889-019-6413-1 (visited on 02/15/2022).

[5] Kwadwo Owusu Akuffo, Ronel Sewpaul, Natisha Dukhi, Akosua Kesewah Asare, David Ben Kumah, Emmanuel Kofi Addo, Eldad Agyei-Manu, and Priscilla Reddy. "Eye Care Utilization Pattern in South Africa: Results from SANHANES-1". In: *BMC Health Services Research* 20.1 (Aug. 17, 2020), p. 756. URL: https://doi.org/10.1186/s12913-020-05621-8 (visited on 02/15/2022).

[6] Cara S Kosack, Anne-Laure Page, Greet Beelaert, Tumwesigye Benson, Aboubacar Savane, Anne Ng'ang'a, Bita Andre, Jean-Paul BN Zahinda, Leslie Shanks, and Katrien Fransen. "Towards More Accurate HIV Testing in Sub-Saharan Africa: A Multi-Site Evaluation of HIV RDTs and Risk Factors for False Positives". In: *Journal of the International AIDS Society* 20.1 (Mar. 24, 2017), p. 21345. pmid: 28364560. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5467586/ (visited on 02/15/2022).

[7] G. Shobana and S. Nikkath Bushra. "Classification of Myopia in Children Using Machine Learning Models with Tree Based Feature Selection". In: *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). Nov. 2020, pp. 1599–1605.

[8] Xu Yang, Guo Chen, Yunchong Qian, Yuhan Wang, Yisong Zhai, Debao Fan, and Yang Xu. "Prediction of Myopia in Adolescents through Machine Learning Methods". In: *International Journal of Environmental Research and Public Health* 17.2 (Jan. 2020), p. 463. pmid: 31936770. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7013571/ (visited on 11/24/2021).

[9] Haotian Lin, Erping Long, Xiaohu Ding, Hongxing Diao, Zicong Chen, Runzhong Liu, Jialing Huang, Jingheng Cai, Shuangjuan Xu, Xiayin Zhang, Dongni Wang, Kexin Chen, Tongyong Yu, Dongxuan Wu, Xutu Zhao, Zhenzhen Liu, Xiaohang Wu, Yuzhen Jiang, Xiao Yang, Dongmei Cui, Wenyan Liu, Yingfeng Zheng, Lixia Luo, Haibo Wang, Chi-Chao Chan, Ian G. Morgan, Mingguang He, and Yizhi Liu. "Prediction of Myopia Development among Chinese School-Aged Children Using Refraction Data from Electronic Medical Records: A Retrospective, Multicentre Machine Learning Study". In: *PLOS Medicine* 15.11 (Nov. 6, 2018), e1002674. URL: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002674 (visited on 11/24/2021).

[10] Tao Tang, Zekuan Yu, Qiong Xu, Zisu Peng, Yuzhuo Fan, Kai Wang, Qiushi Ren, Jia Qu, and Mingwei Zhao. "A Machine Learning-Based Algorithm Used to Estimate the Physiological Elongation of Ocular Axial Length in Myopic Children". In: *Eye and Vision* 7.1 (Oct. 22, 2020), p. 50. URL: https://doi.org/10.1186/s40662-020-00214-2 (visited on 11/24/2021).

[11] Tae Keun Yoo, Ik Hee Ryu, Geunyoung Lee, Youngnam Kim, Jin Kuk Kim, In Sik Lee, Jung Sub Kim, and Tyler Hyungtaek Rim. "Adopting Machine Learning to Automatically Identify Candidate Patients for Corneal Refractive Surgery". In: *npj Digital Medicine* 2.1 (Dec. 2019), p. 59. URL: http://www.nature.com/articles/s41746-019-0135-8 (visited on 11/24/2021).

[12] José Jiménez-Luna, Francesca Grisoni, and Gisbert Schneider. "Drug Discovery with Explainable Artificial Intelligence". In: *Nature Machine Intelligence* 2.10 (10 Oct. 2020), pp. 573–584. URL: https://www.nature.com/articles/s42256-020-00236-4 (visited on 04/04/2022).

[13] Clare Gilbert. "The Eye Signs of Vitamin A Deficiency". In: *Community Eye Health* 26.84 (2013), pp. 66–67. pmid: 24782581. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3936686/ (visited on 04/19/2022).

[14] Scott M. Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., Dec. 4, 2017, pp. 4768–4777.

[15] Matthew Hao Lee, Marc G. Sarossy, and Ehud Zamir. "Vitamin A Deficiency Presenting with 'Itchy Eyes'". In: *Case Reports in Ophthalmology* 6.3 (Dec. 2015), pp. 427–434.