Original Article

Prediction of T2DM Using Conjunctival Sac Microbiota, a Machine Learning Approach

Soroush Setareh [1], MS; Azadeh Kavianfar [2], MS; Hamidreza Taherkhani [2,*], MS

1. Department of Biotechnology, College of Science, University of Tehran, Tehran, Iran.
2. Laboratory of Systems Biology and Bioinformatics, Department of Bioinformatics, Kish Intl. Campus, University of Tehran, Tehran, Iran.

***Corresponding Author:** Hamidreza Taherkhani
E-mail: htaherkhani@ut.ac.ir

Abstract

Background: Association of T2DM and OS disorders addresses a human eye metagenome drift. Despite the clarity of diabetic retinopathy, process of involvement of conjunctival sac microbiota is still ambiguous. We seek predictive value of OS microbiota using ML-based methods, since the early diagnosis of diabetic retinopathy and conjunctival sac microbiome diabetic dysbiosis could be beneficial both for clinical and research purposes.

Material and Methods: 16S rRNA characterization of human eye metagenome for samples of 192 patients (with mean age of 66 years and 56 % females) with different onsets of T2DM is analyzed using various metrics including abundance and diversity indices and LDA at phyla, families, and genera levels. We took advantage of variance threshold, Chi-squared significance, and LDA Effect Size (LEfSe) feature selection strategies for inclusion of predictive families and genera in the T2DM prediction model. ML models with different algorithms including RF, GB, SVM, and ANN are implemented. Generalizability and robust performance of the models are also ensured using a 5-fold cross-validation process. DeLong's test is also used to investigate different performance of the methods.

Results: Microbiome analyses revealed that eye metagenome profiles of the patients with <15 years of T2DM history show significantly higher richness and diversity. ML model performance shows ROC-AUC of ~0.8. ML model with the superior performance exhibit sensitivity and accuracy of 0.86 and 0.68, respectively, in the prediction of T2DM occurrence.

Conclusion: significant correlation and co-occurrence of T2DM and eye microbiome dysbiosis is trackable and well-optimized ML-strategies can predict T2DM onsets based on the microbiome of conjunctival sac.

Keyword: Type 2 Diabetes; Mellitus; Conjunctival Sac Microbiome; Diabetic Life Style; Machine Learning.

Article Notes: Received: Sep. 26, 2020; Received in revised form: Nov. 06, 2020; Accepted: Nov. 22, 2020; Available Online: Apr. 4, 2021.

How to cite this article: Setareh S, Kavianfar A, Taherkhani H. Prediction of T2DM Using Conjunctival Sac Microbiota, a Machine Learning Approach. Journal of Ophthalmic and Optometric Sciences . 2021;5(2): 31-46.

Introduction

Diabetes is one of the most prevalent diseases worldwide and numerous studies has shown that Type 2 Diabetes Mellitus (T2DM), which forms up near 90 percent of the total diabetes mellitus incidences [1], is strongly correlated with other disorders and diseases, as the T2DM plays the role of common soil for many disorders including but not limited to cardiovascular [2], hepatic [3], oral [4], renal [5], ocular and retinal [6–8], pulmonary [9], and neurological disorders [10].

Retinal and Ocular Surface (OS) disorders in many cases have shown correlation with the onset of T2DM. Dry eye syndrome [11], tear-film dysfunction [12], diabetic papillopathy, glaucoma, and cataract are merely some of diabetes mellitus associated ocular complications [13].

However, it should be considered that the T2DM is not solely the disorder itself, but includes a complete lifestyle of its own. Diabetic lifestyle may alter one's diet, use of medicine, consumption of other narcotics, and generally its life, in many aspects. Hence, speaking of T2DM highlights the diabetic lifestyle as general, not the disease itself, since the patients suffering the T2DM might have lived entirely different or even might have not been aware of their condition [14].

Development and progression of diabetic retinopathy and ocular disorders is affected by patients' demographic features, onset of T2DM and many other factors. As a result, diagnosis of T2DM and monitoring the situation of OS during the T2DM, specifically in elders, is of great importance and concern. Non-invasive and inexpensive laboratory tests or In-Vitro-Diagnostic (IVD) devices are required to screen OS regularly and detect diabetic retinopathy as initial signs of OS dysbiosis emerge.

Diabetic retinopathy itself and necessity of required cares including consuming antibiotics in order to protect ocular surface from microbial infections, impacts the microbial community of the ocular surface [7,15,16] and a hypothetical dysbiosis is assumed, either as a result from T2DM retinopathy and lifestyle or as a cause for T2DM retinopathy. Higher frequency of positive cultures [17], overgrowth and colonization by potential pathogenic bacteria [18], and higher alpha diversity of microbiota associated with the conjunctival surface of diabetic patients [16] are mentioned previously. Hence, investigating conjunctival microbiome is of great importance and potential correlations are hypothesized.

In this study, we have shown that T2DM is in correlation with higher richness and diversity in microbial composition of Conjunctival Sac (CS) and we have developed a Machine-Learning (ML) model which is capable of classifying healthy and T2DM records based on their eye metagenome with outstanding performance.

Material and Methods

Data acquisition

16S rRNA sequencing profile of eye conjunctival sac samples from 192 patients is considered for the following analysis. Data includes records for 29 patients with ≥ 15 years of Type 2 Diabetes Mellitus (T2DM) history, 50 patients with < 15 years of T2DM history, and 113 samples with no T2DM history, considered as the 'non-diabetic' control group. Diabetic patients are divided into categories of diabetes duration.

It must be kept in mind that patients with T2DM history might vary fundamentally in terms of their physiological condition and well-being. Long history of T2DM has affected their lifestyles for many years and

as a result, deviations from normal lifestyle could be considered.

Prior to the profiling of conjunctival sac microbiome profiling, inclusion criteria of having a stable blood-sugar level (fasting blood-glucose < 8.3 mmol/L or HbA1c level ≤7 %) and not using eye drops from at least 3 months prior to the study has been applied. Samples with histories of lacrimal or blepharal diseases, corneal or conjunctival disorders, glaucoma, contact lens wearing, history of eye surgeries or with other systematic diseases other than T2DM were then omitted from the dataset with respect to the exclusion criteria [19]. All remaining 192 samples do satisfy the inclusion and exclusion criteria.

Reads along with samples metadata is publicly available in the European Bioinformatics Institute (EBI) database under the accession number of PRJNA629667. Records with the sample aliases corresponding to the history of cataract surgery or post-surgical consumption of antibiotics including Levofloxacin are removed from the original dataset, leaving merely 192 samples.

Model Input-Output correlation

Chi-squared and Fisher's test for patients' sex and age have been conducted to investigate whether the parameters show correlation with the T2DM disease stage or not. Baseline characteristics of the data in addition to P values for the statistical tests are represented in the table 1.

Microbiome data preprocessing

Reads for all samples are demultiplexed, denoised, and representative sequences are extracted. Then samples with frequency less than 50 are filtered out, leaving 187 samples. Representative sequences are then clustered to Operational Taxonomic Units (OTUs) with respect to a similarity threshold of 99 %, by utilizing a Naïve-Bayes (NB) classifier and the Greengenes 13.8 database. Classification confidence threshold of 0.7 has also been applied to the clustered OTUs, leaving all 187 samples and ensuring adequate confidence in the OUT calling process. All preprocessing steps are carried out using Qiime 2.0 package [20] and Qiime python scripts, in the Google Colab servers.

Microbiome data analysis

From a total 6653 unique representative sequences, 6070 OTUs are emerged and the rest are unidentified. Further analyses are conducted in the R environment (R-Studio 4.1.3). Qiime2 objects are converted to phyloseq class objects, an R-recognizable data structure, using phyloseq library [21]. Composition of the conjunctival

Table 1: Baseline characteristics of the variables present in the metadata

Parameter	Non-diabetic	< 15 years T2DM	≥ 15 years T2DM	P value
Continuous variable	Mean (± Standard deviation)			Fisher exact test P value
Age (in years)	65.37 (± 9.51)	66.26 (± 8.78)	69.17 (± 7.54)	0.1334
Categorical variable	Count (Percent)			Chi-squared test P value
Sex (female)	61 (53.98 %)	29 (58.00 %)	18 (62.07 %)	0.8219

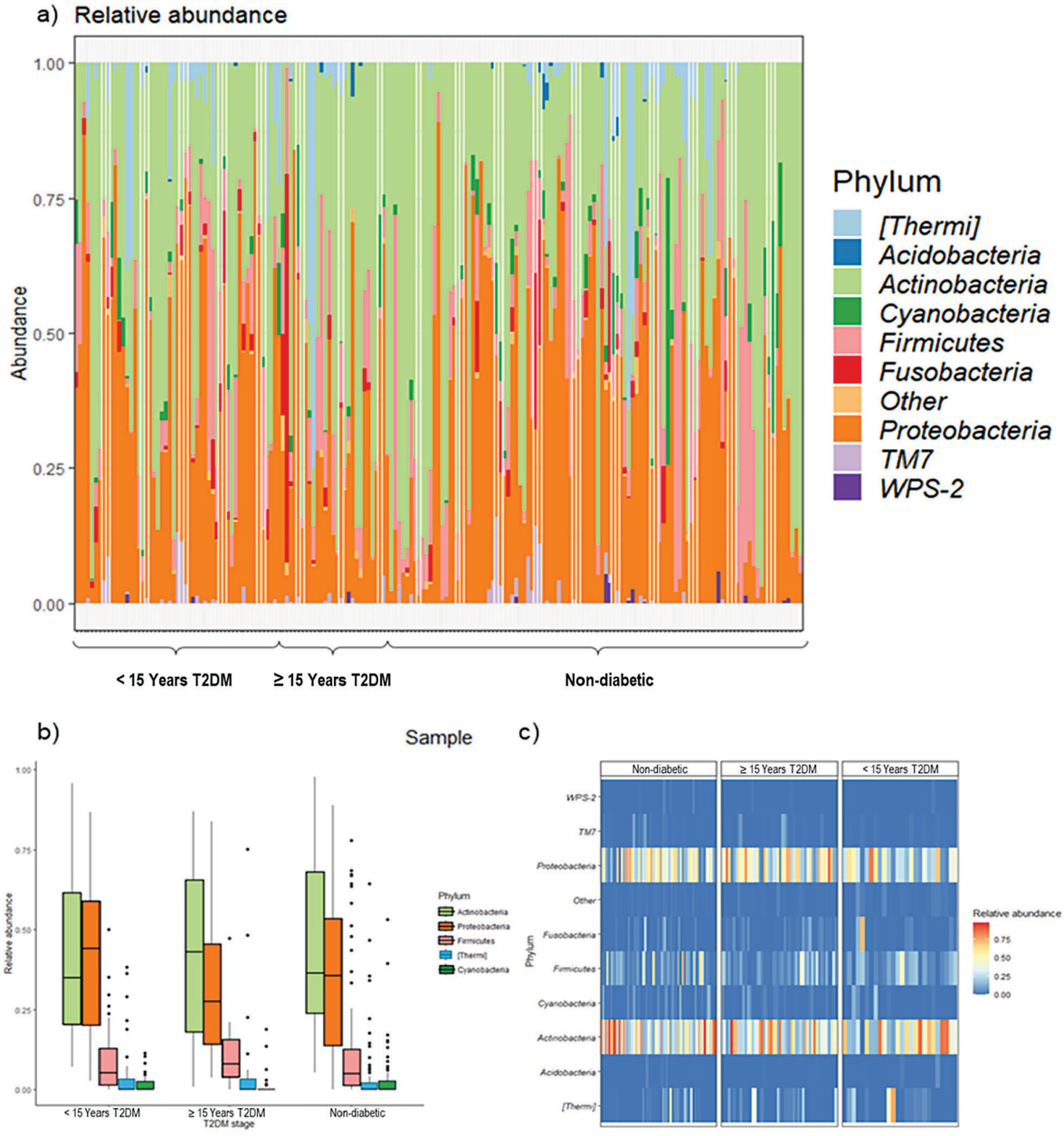

Figure 1: Composition analysis of eye microbiome at the phyla level. a) Relative abundances composition plot for phyla present in the eye conjunctival sac microbiome and b) Phyla relative abundance boxplot for the five most abundant phyla in different T2DM groups, reveal the Actinobacteria and Proteobacteria phyla are most abundant taxa where their distributions show drifts among different groups c) Phyla abundance correlation coefficient heatmap in different T2DM groups also reveals that some samples contain relatively higher counts of certain phyla

sac microbial community in the patients is compared using relative abundance of the OTUs at the Phylum level. α-diversity of the microbial communities in non-diabetic, < 15 years T2DM, and ≥ 15 years T2DM groups are investigated using various α-diversity metrics, including Inverse Simpson [22], Gini-Simpson [23], Shannon [24], Fisher [25], and Coverage [26] statistics. ß-diversities of the microbial communities in different groups are compared using

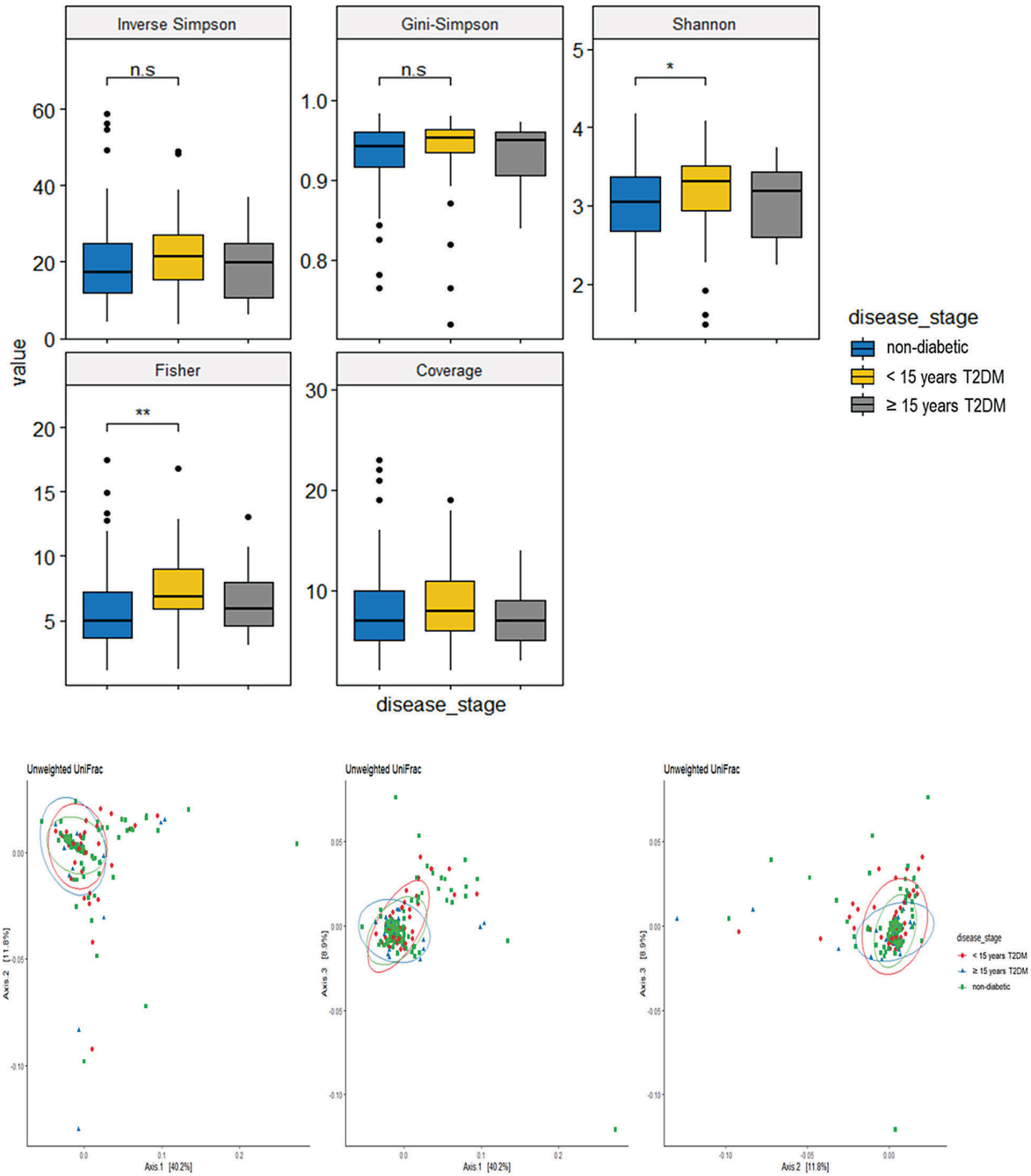

Figure 2: Alpha and beta diversity statistics for taxa at different T2DM groups. < 15 years history of T2DM is associated with the significant increment of alpha-diversity, as confirmed by paired t-tests. t-test is implemented and P values significant at with 90, 95, and 99 percent confidence levels are marked with Non-significant (n.s), asterisk (*), and double-asterisk (**) signs, respectively (Top). Unweighted UniFrac as a Beta diversity metric in the PCoA planes for different T2DM groups. All combinations of three principle component axes are considered and dispersion of samples in the PCoA planes shows no separation between samples from different T2DM groups (Bottom)

unweighted UniFrac [27], and the corresponding Principle Component Analysis (PCoA) [28] has been done. Linear Discriminant Analysis (LDA) of the Effect size (LEfSe) [29] is then used to identify taxa who contribute mostly to the groups difference. Significance of the LEfSe

Figure 3: LDA Scores and direction for families (left) and genera (right) with significant LEfSe. Significance of LDA scores are determined using a Kruskal-Wallis rank sum test and merely features with P value of < 0.05 are represented here, along with their direction of discrimination in a radar-plot

analysis and LDA score is then tested using Kruskal-Wallis rank sum test [30] and a threshold of 0.05 for the test's P value is considered, leaving merely taxa with a significant LEfSe. Pearson Correlation Coefficient (PCC) is also calculated between taxa at different genera or families to find taxa with same abundance.

T2DM prediction model

Machine learning strategies are implemented to solve a binary classification problem and T2DM prediction model is an effort to predict diabetes mellitus based on the abundance of microbial composition of the patients' conjunctival sac. Overall workflow of the T2DM prediction model is illustrated in figure 3.

Preprocessing

From the initial abundance matrices, OTUs of family and genus are aggregated by summation and of all abundances and abundance matrices are then formed for each rank. Counts of unique OTUs identified at each rank are illustrated in figure 4. Abundance matrices are normalized using Min-max normalization as described in [1]. Not normalized copies of the abundance matrices are also kept for further model trainings.

$$\hat{\theta}_{i,j} = \frac{\theta_{i.j} - \min(\theta_j)}{\max(\theta_j) - \min(\theta_j)} \quad (1)$$

Where $\hat{\theta}_{i,j}$ is the normalized abundance for $\theta_{i.j}$ which is abundance of the feature j in sample i. All unidentified phylum, family, and genus entities are merged as unidentified taxa and due to the absence of the missing values in the abundance tables, no missing value imputation strategy is used. Since a binary classification problem is considered, < 15 years T2DM and ≥ 15 years T2DM groups are merged, leaving diabetic and non-diabetic labels as only possible groups for each sample.

Feature selection

Due to relevantly high number of unique family and genus features, feature selection is crucial to reduce the dimensionality and prevent model's overfitting [31]. On the other hand, it is immensely hard to track the effect of

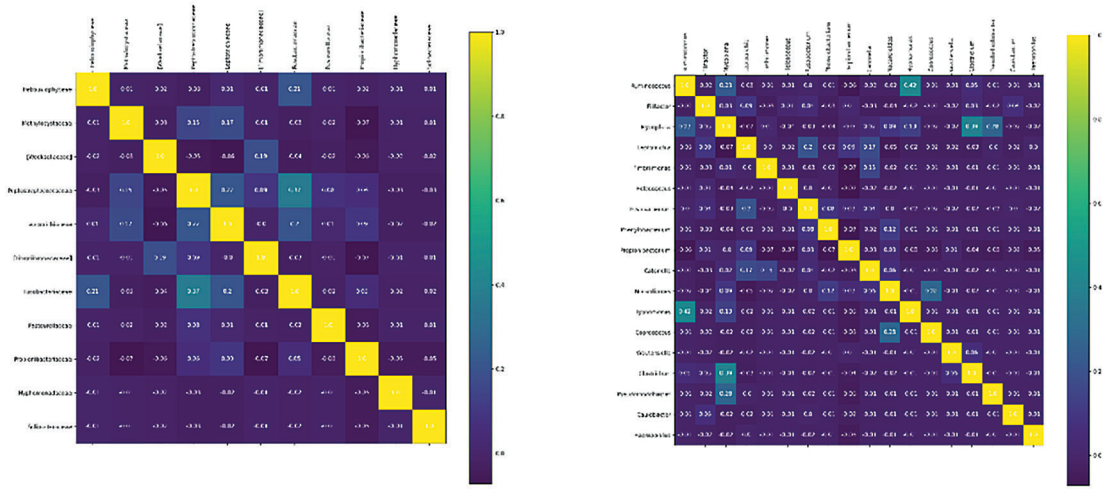

Figure 4. PCC Heatmap for families (left) and genera (right) with significant LEfSe. Pearson's Correlation Coefficient indicates taxa correlation. Taxa with PCC higher than > 0.7 are omitted to reduce dimensionality and here we depicted taxa with no significant correlation.

a large number of taxa at once and inferring the results in clinical and therapeutic applications becomes arduous as the number of parameters increase [32], making feature selection even more important.

Univariate and multi-variate feature selection methods are utilized. A variance threshold of 0.016 is used for all normalized features to omit low-variance features, selecting merely high-variance features and as a result, reducing the dimensionality. In another feature selection method, normalized features are subjected to the Chi-squared test and 15 features with the most significant chi2 values, with respect to the samples' classes, are selected [33]. Features whom were identified using LDA, as described before, are also selected as outcome of another feature selection procedure, since they have shown high contribution in inter-group difference and are probably meaningful features. In addition to Chi2, High-variance, and LDA univariate feature selection methods, between-taxa correlation in family level and genus level features are also investigated using Pearson's Correlation Coefficient (PCC). Features selected with the best univariate feature selection method, leading to the best performance of the model, are then subjected to a bi-variate feature selection using PCC, removing one of the features (typically the least significant one) from any pair of features who show high correlation coefficient ($r > 0.7$). Chi2 and High-variance feature selection procedures are implemented using the Scikit-learn library [34] functions in the python environment.

Machine Learning models

Several Machine Learning (ML) algorithms, including Random Forest (RF), Gradient Boosting (GB), Multi-Layer Perceptron (MLP) as a simple Neural Network (NN), and Support Vector Machine (SVM), are used for supervised binary classification of the records. All records are split among train and test groups, randomly assigning one-five of the records to the test group and the rest to the train group [35]. To ensure model generalizability, this process is repeated five times in a 5-fold cross validation fashion, assigning each record to test set exactly once. List of ML methods along with their hyper parameters is present in table 2.

Table 2: Machine Learning methods and utilized hyper parameters

Method	Parameter	Value
Random forest	Trees count	1000
	Fraction of samples at each leaf	0.1 % of all samples
	Used criterion	Gini
Gradient bossting	Total iterations of boosting	1000
	Training fraction	0.8
	Fraction of samples at each leaf	0.1 % of all samples
	Early steps of constant gradient criterion	3
	Maximum features considered during split	3
Support vector machine	C	1.0
	Type of the kernel	RBF
	Coefficient of the kernel	1/number of features
Artificial neural network	Hidden layers count	3
	Hidden layers' dimensionality	16, 8, 4
	Each hidden layer activation function	Tanh, Tanh, Tanh, Sigmoid

Model Performance

Performance of the models, are evaluated in terms of Receiver Operator Characteristics Area Under the Curve (ROC-AUC). For a total of 5 feature preparation methods and 4 ML models, mean and standard deviation of 5-fold cross validation ROC-AUC on the test data is calculated. To ensure that the models are performing uniquely, correlation of the predictions is investigated using the DeLong's test [36].

Results

Microbiome analysis

Composition Plot

Relative abundance composition plot for top 10 phylum, with respect to phyla abundance, has been illustrated in figure 1. A minor dysbiosis in the microbiome's composition is addressable through the phyla relative abundances. Box and whisker plot is also illustrated to depict the relative abundance of top five phyla. In addition, phyla abundance heat map is also representing each phylum propensity for presence in each group. Overall, relative abundance and composition analysis is highlighting Actinobacteria and Proteobacteria phyla as the most abundant phyla and a minor dysbiosis in the microbiome's composition is addressable through the phyla relative abundances.

Alpha diversity

Box and whisker plot of alpha diversity statistics are represented in figure 2a. Significant difference in alpha diversities between subgroups has been tested using t-test and the resultant P values are marked as **, *, and n.s for 99, 95, and 90 percent confidence levels, respectively. As the statistical inference of alpha diversities suggests, early emergence of the T2DM exhibits a significantly more diverse conjunctival microbiota in terms of alpha diversity, as confirmed by almost all alpha diversity statistics.

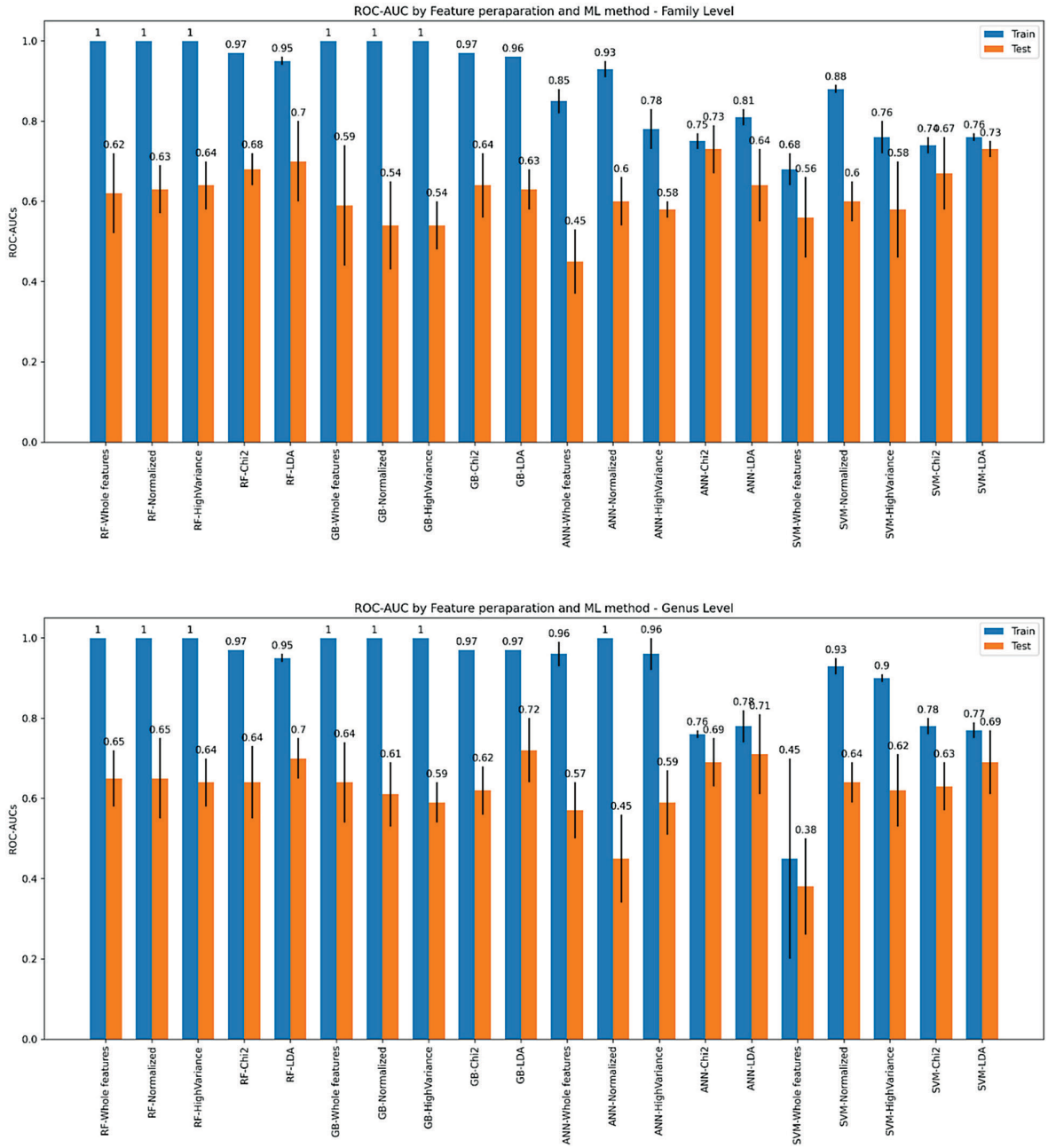

Figure 5: ROC-AUC Barplots for different feature preparation methods and different ML techniques. Mean ROC-AUC among 5-fold cross-validation iterations are represented on top of each barplot and error bars are demonstrating one standard deviation interval. Results for classification at both families (top) and genera (bottom) are depicted

Beta diversity

Despite the higher alpha diversities, beta diversities show no significant difference between the groups. Three principle components are selected with respect to their highest explained variances (PC1: 40.2 %, PC2: 11.8 %, PC3: 8.9 %, of the total variance explained), and PCoA plots for each combination of these axes are represented in figure 2b. Ovals are sketched to address the central tendency and dispersion of samples for each group. High dispersion of the samples in

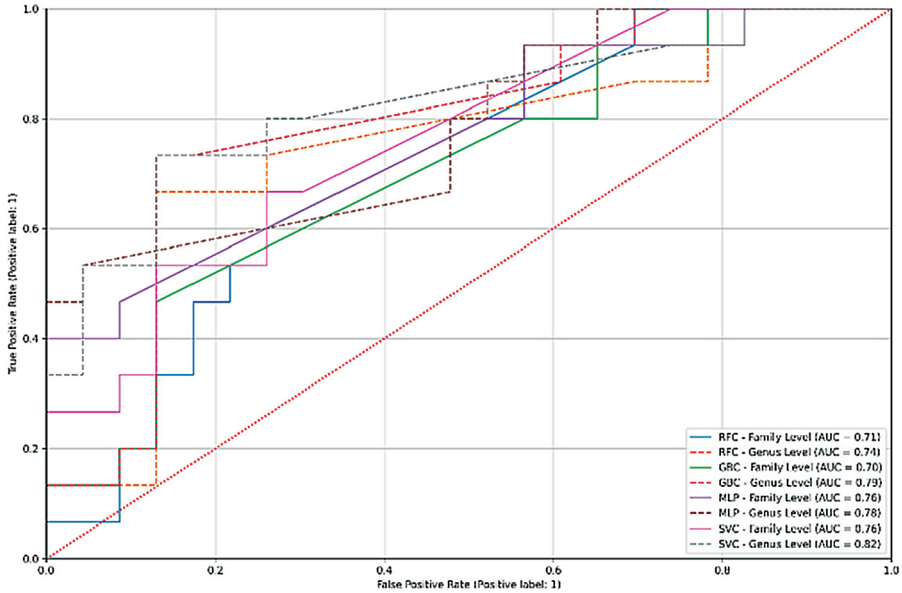

Figure 7: ROC-AUC plot for different ML methods performed on LDA-selected families and genera features. ROC-AUC of models are represented versus the random state prediction, which is shown as red bisector with AUC of 0.50. SVC is out-performing other techniques with the highest ROC-AUC of 0.76 and 0.82 at family and genus levels, respectively

the principle component planes and confusion of the samples is noticeable. Significance of the inter-group diversity has been assessed using Adonis test, where P value of 0.37 highlights no significant difference between groups.

LefSe analysis results assigns a LDA score to each genus or family, indicating how strong and in which direction a taxon has contributed to the intergroup difference. Higher values of LDA score in each direction implies that the prevalence of the corresponding taxon is more enriched in that direction and thus, exhibits a higher effect size. Radar plot of distinct families and genera, with significant Kruskal-Wallis difference (P value < 0.05), is depicted in figure 3. As the radar plots show, Propionibacteriaceae family and propionibacterium genus are enriched in the non-diabetic patients, where Methylocystaceae and Leptotrichiaceae families and Leptotrichia, Helococcus, and Catonella genera are enriched in the eye microbiota of the patients with < 15 years history of T2DM.

Taxa correlation

Investigation of taxa correlation for LEfSe significant taxa using PCC revealed that some families and genera show high degrees of correlation and exact same abundance. Correlated taxa are removed and heatmaps in the figure 4, show that there are no correlated taxa left, which ensures no redundancy from duplicated taxa.

Using Machine Learning to predict diabetes mellitus based on conjunctival sac's microbiome

Performance of the ML models in the prediction of T2DM occurrence based on the conjunctival sac's microbiome, in terms of ROC-AUC is reported in Supplementary table 1. Values indicate average and standard deviation of ROC-AUC in overall 5-fold cross-validation process. Higher means imply better performance where lower standard deviation

indicate higher robustness.

As it is also clearly depicted in figure 5, feature selection using LDA features improves the performance on one hand, and on the other hand, prevents model overfitting.

Model Performance

DeLong P values for the difference of the models are represented in a heatmap in Supplementary figure 1. ROC-AUC diagram in the figure 7 shows how different classifiers perform at the families and genera features selected by LDA. Once the best models are identified with respect to their highest ROC-AUC, additional performance metrics, including model accuracy, sensitivity, specificity, Negative Predictive Value (NPV), and Positive Predictive Value (PPV), are also calculated and reported for each top model, in Supplementary table 1.

Discussion

From 10 phyla with highest relative abundances, each may contribute to diabetic retinopathy

Actinobacteria, Proteobacteria, Firmicutes, Thermi, and Cyanobacteria, phyla which are the most abundant phyla in the OS microbiome, shows no significant difference between different disease stages. Alterations in Actinobacteria, Proteobacteria, and Firmicutes phyla between T2DM and healthy patients is negligible also in the gut microbiome [37]. However, reduced Actinobacteria abundance in the patients with diabetic retinopathy is reported, which is differs with higher Actinobacteria abundance in patients with ≥ 15 years history of T2DM in this study, which might show diabetic retinopathy.

Ocular surface microbial community in the onset of T2DM, exhibits higher level of diversity as significantly higher alpha diversity metrics suggest, which affirms previous findings [15, 19, 38]. Increased microbial diversity with the occurrence of diabetes mellitus is also addressed in other microbiotas, including fecal [39], vaginal [40], and oral [41] microbial communities. Presence of higher concentrations of blood sugar favors growth of the microbes, which results in the diversity and richness of ocular surface microbiome. Besides, metabolic drift in the diabetic patients and altered concentrations of Interleukins [41], in response to inflammation, also show correlations with a more diverse flora. However, the causality is studies merely [42], and still seems unclear.

Predictive values of features selected through the LDA with highest LEfSe could be immensely important in studying diabetic retinopathy and the progression of diabetes-associated ocular diseases including dry eye symptoms and tear-film dysfunction [43], tear analysis for non-invasive detection of T2DM [44], tracking the dysbiosis in the ocular surface in following of T2DM occurrence, and a plethora of other vital considerations against diabetes mellitus.

Significant increase in the abundance of Fusobacteriaceae family and in the T2DM patients and positive correlation of T2DM with the bundance of Fusobacterium genus is reported [45, 46] previously, and is also denoted in this study, as Fusobacteriaceae family and Fusobacterium genus features are both remarked with LEfSe in the direction of T2DM.

Propionibacteriaceae family on the other hand, is contributing to non-diabetic direction. Bacteria from the Propionibacterium genus are commensal bacteria which could turn into opportunistic pathogens in occasions [47] and act as main agents for the Propionibacterium acnes [48] and occurrence of endophthalmitis [49].

Since the Propionibacterium genus is directing to non-diabetic group, a change in the behavior of these bacteria in response to the diabetic retinopathy and associated cares [50] is possible, which makes the commensal-to-pathogenic drift in the Propionibacterium behavior.

Hence, not merely the count and the abundance of the bacteria affects or is affected by the T2DM physiology, but this correlation could also make the behavior of them to undergo some alterations. As a result, to make this correlation clearer and the results more interpretable, additional information about microbial 'omics are required.

Effect of a plethora of parameters are tried to be minimal to limit the results to the T2DM state. Chi-squared and Fisher tests are utilized to investigate the model input-output correlation. Furthermore, effect of ethnicity, geographical location, and sex are assumed to make no significant impact on the OS microbiome [51]. But even though the correlation between input parameters including age and sex and the output parameter of T2DM state is not significant to affect the results and produce an input-output correlation bias, and inclusion and exclusion criteria are selected in a way to minimize life-style bias, there could still be factors which can mislead the interpretations. Well-organized clinical trials and establishment of Design-of-Experiment (DoE) setups are suggested in order to minimize the effect of unseen parameters and improve model robustness and capability.

A common problem emerges during the classification of the representative sequences and further aggregation of features at taxonomical ranks. Utilized classifier and the corresponding database can strongly affect the outcome and produce a classification bias error. During an error-prone classification of features as different taxa, some features might get classified as a same taxon. As a result, distinct features with different abundance profiles, might correspond to same taxon. These features would be identified as duplicate records, despite their different patterns among different samples. Here an aggregation error is possible. Consider A and B features with 1000,1 and 1,1000 abundance profiles in diabetic and non-diabetic samples, respectively [52]. After classification with a benchmark classifier, both could be classified as the genus C with unidentified species. After aggregation of features as the genus level with two popular aggregation strategies, aggregation by summation and aggregation by dropping duplicates, an information loss occurs, which might cause further errors. As genera are aggregated, differential abundance of A and B features are vanished, leaving no trails of their significance.

Sequencing profile of the V3-V4 hypervariable region of the 16s rRNA gene, provides satisfactory information on the composition of a microbial community and ensures adequate resolution due to specificity of this region. However, misclassification of the duplicate taxa, might be reduced by changing the sequencing region or considering more than one target region, which may increase specificity [53].

Conclusion

Regardless the unspecified causality direction, T2DM shows strong association with conjunctival sac microbiome dysbiosis and further ocular disorders. Here by comparing several feature selection strategies and algorithms, we proposed a ML method for prediction of T2DM based on the conjunctival sac microbial composition, which is performing precisely and accurately.

However, it should be kept in mind that not

solely the T2DM itself is in correlation with the eye microbiome drift, but the diabetic lifestyle, embedding diabetic diet, drug habits, abstinences and the T2DM pathogenicity itself, is the main reason of eye metagenome drift. Diabetic lifestyle components are the main debate and future researches may clear this debate.

Abbreviations

T2DM: Type-2 Diabetes Mellitus. OS: Ocular Surface.
ML: Machine Learning.
LDA: Linear Discriminant Analysis.
RF: Random Forest.
GB: Gradient Boosting.
SVM: Support Vector Machine.
ANN: Artificial Neural Network.
ROC-AUC: Receiver Operator Characteristic Area Under the Curve

Authors ORCIDs

Soroush Setareh:
https://orcid.org/0000-0002-1095-9499
Hamidreza Taherkhani:
https://orcid.org/0000-0002-5853-7260

References

1. Tao Z, Shi A, Zhao J. Epidemiological Perspectives of Diabetes. Cell Biochem Biophys. 2015;73(1):181–5.
2. Stern MP. Perspectives in Diabetes Diabetes and Cardiovascular Disease The "Common Soil" Hypothesis. (4):369–74.
3. Leite NC, Villela-Nogueira CA, Pannain VLN, Bottino AC, Rezende GFM, Cardoso CRL, et al. Histopathological stages of nonalcoholic fatty liver disease in type 2 diabetes: Prevalences and correlated factors. Liver Int. 2011;31(5):700–6.
4. Ponte E, Tabaj D, Maglione M, Melato M. Diabetes mellitus and oral disease. Acta Diabetol. 2001;38(2):57–62.
5. Mogensen CE, Andersen MJF. Increased kidney size and glomerular filtration rate in untreated juvenile diabetes: Normalization by insulin-treatment. Diabetologia. 1975;11(3):221–4.
6. Misra SL, Braatvedt GD, Patel D V. Impact of diabetes mellitus on the ocular surface: a review. Clin Exp Ophthalmol. 2016;44(4):278–88.
7. Markoulli M, Flanagan J, Tummanapalli SS, Wu J, Willcox M. The impact of diabetes on corneal nerve morphology and ocular surface integrity. Ocul Surf. 2018;16(1):45–57.
8. Luu CD, Szental JA, Lee SY, Lavanya R, Wong TY. Correlation between retinal oscillatory potentials and retinal vascular caliber in type 2 diabetes. Investig Ophthalmol Vis Sci. 2010;51(1):482–6.
9. Irfan M, Jabbar A, Haque AS, Awan S, Hussain SF. Pulmonary functions in patients with diabetes mellitus. Lung India. 2011;28(2):89–92.
10. Fiore V, De Rosa A, Falasca P, Marci M, Guastamacchia E, Licchelli B, et al. Focus on the Correlations between Alzheimer's Disease and Type 2 Diabetes. Endocrine, Metab Immune Disord - Drug Targets. 2019;19(5):571–9.
11. Odinaeva KN, Al Khateeb NS, Frolov MA. Dry eye syndrome in patients with gout. Point

View East – West. 2021;2016(4):6–11.
12. Manaviat MR, Rashidi M, Afkhami-Ardekani M, Shoja MR. Prevalence of dry eye syndrome and diabetic retinopathy in type 2 diabetic patients. BMC Ophthalmol. 2008;8:1–5.
13. Howells LH. Ocular complications of diabetes mellitus. Br J Ophthalmol. 1953;37(12):716–24.
14. Kavianfar A, Salimi M, Taherkhani H. A Review of the Management of Eye Diseases Using Artificial Intelligence , Machine Learning , and Deep Learning in Conjunction with Recent Research on Eye Health Problems. J Ophthalmic Optom Sci. 2021;5(2).
15. Li S, Yi G, Peng H, Li Z, Chen S, Zhong H, et al. How ocular surface microbiota debuts in type 2 diabetes mellitus. Front Cell Infect Microbiol. 2019;9(JUN):1–9.
16. Aragona P, Baudouin C, Benitez del Castillo JM, Messmer E, Barabino S, Merayo-Lloves J, et al. The ocular microbiome and microbiota and their effects on ocular surface pathophysiology and disorders. Surv Ophthalmol. 2021;66(6):907–25.
17. Martins EN, Alvarenga LS, Höfling-Lima AL, Freitas D, Zorat-Yu MC, Farah ME, et al. Aerobic Bacterial Conjunctival Flora in Diabetic Patients. Cornea. 2004;23(2):136–42.
18. Aragona P, Giuffrida S, Di Stefano G et al. Ocular surface changes in type 1 diabetic patients. Adv Exp Med Biol. 2005;(506):667–72.
19. Zhu X, Wei L, Rong X, Zhang Y, Zhang Q, Wen X, et al. Conjunctival Microbiota in Patients With Type 2 Diabetes Mellitus and Influences of Perioperative Use of Topical Levofloxacin in Ocular Surgery. Front Med. 2021;8(April):1–10.
20. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852–7.
21. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One. 2013;8(4).
22. Hill MO. Diversity and Evenness: A Unifying Notation and Its Consequences. Ecology. 1973;54(2):427–32.
23. Jost L. Entropy and diversity. Oikos. 2006;113(2):363–75.
24. Spellerberg IANF, Fedor PJ. Tribute To Claude Shannon (1916-2001). Glob Ecol Biogeogr. 2003;177–9.
25. Fisher RA, Corbet AS, Williams CB. The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. J Anim Ecol. 1943;12(1):42.
26. Willis AD. Rarefaction, alpha diversity, and statistics. Front Microbiol. 2019;10(OCT).
27. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol. 2007;73(5):1576–85.
28. Gower JC. Principal Coordinates Analysis. Encycl Biostat. 2005;
29. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6).
30. Kruskal WH, Wallis WA. Use of Ranks in One-Criterion Variance Analysis. J Am Stat Assoc. 1952;47(260):583–621.
31. Lever J, Krzywinski M, Altman N. Points of Significance: Model selection and overfitting. Nat Methods. 2016;13(9):703–4.
32. Brüssow H. Problems with the concept of gut microbiota dysbiosis. Microb Biotechnol. 2020;13(2):423–34.
33. Maghsoudloo M, Azimzadeh Jamalkandi

S, Najafi A, Masoudi-Nejad A. An efficient hybrid feature selection method to identify potential biomarkers in common chronic lung inflammatory diseases. Genomics. 2020 Sep;112(5):3284–93.
34. Fabian Pedregosa. Scikit-learn: Machine Learning in Python. Environ Health Perspect. 2019;127(9):2825–30.
35. Hooshmand SA, Zarei Ghobadi M, Hooshmand SE, Azimzadeh Jamalkandi S, Alavi SM, Masoudi-Nejad A. A multimodal deep learning-based drug repurposing approach for treatment of COVID-19. Mol Divers. 2021;25(3):1717–30.
36. DeLong ER, DeLong DM, Clarke-Pearso DL. Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics. 1998;44(3):837–45.
37. Das T, Jayasudha R, Chakravarthy SK, Prashanthi GS, Bhargava A, Tyagi M, et al. Alterations in the gut bacterial microbiome in people with type 2 diabetes mellitus and diabetic retinopathy. Sci Rep. 2021;11(1):1–15.
38. Zhang Z, Zou X, Xue W, Zhang P, Wang S, Zou H. Ocular surface microbiota in diabetic patients with dry eye disease. Investig Ophthalmol Vis Sci. 2021;62(12):21–3.
39. Hu J, Nomura Y, Bashir A, Fernandez-Hernandez H, Itzkowitz S, Pei Z, et al. Diversified microbiota of meconium is affected by maternal diabetes status. PLoS One. 2013;8(11).
40. Tejesvi M V., Nissi R, Saravesi K, Pirttilä AM, Markkola A, Talvensaari-Mattila A, et al. Association of prevalent vaginal microbiome of mother with occurrence of type I diabetes in child. Sci Rep. 2019;9(1):1–6.
41. Benjamin Chun-Kit Tong. 乳鼠心肌提取 HHS Public Access. Physiol Behav. 2017;176(5):139–48.
42. Zhao L. The gut microbiota and obesity: From correlation to causality. Nat Rev Microbiol. 2013;11(9):639–47.
43. Zou X, Lu L, Xu Y, Zhu J, He J, Zhang B, et al. Prevalence and clinical characteristics of dry eye disease in community-based type 2 diabetic patients: The Beixinjing eye study. BMC Ophthalmol. 2018;18(1):1–7.
44. Baca JT, Finegold DN, Asher SA. Tear glucose analysis for the noninvasive detection and monitoring of diabetes mellitus. Ocul Surf. 2007;5(4):280–93.
45. Gurung M, Li Z, You H, Rodrigues R, Jump DB, Morgun A, et al. Role of gut microbiota in type 2 diabetes pathophysiology. EBioMedicine. 2020;51:1–9.
46. Matsha TE, Prince Y, Davids S, Chikte U, Erasmus RT, Kengne AP, et al. Oral Microbiome Signatures in Diabetes Mellitus and Periodontal Disease. J Dent Res. 2020;99(6):658–65.
47. Achermann Y, Goldstein EJC, Coenye T, Shirtliffa ME. Propionibacterium acnes: From Commensal to opportunistic biofilm-associated implant pathogen. Clin Microbiol Rev. 2014;27(3):419–40.
48. Deramo VA, Ting TD. Treatment of Propionibacterium acnes endophthalmitis. Curr Opin Ophthalmol. 2001;12(3):225–9.
49. Chan WM, Lam DSC, Liu DTL, Wong TH, Yuen KSC. Photodynamic therapy for recurrent myopic choroidal neovascularisation after limited macular translocation surgery [5]. Br J Ophthalmol. 2003;87(9):1188–9.
50. Lohmann CP, Linde HJ, Reischl U. Improved detection of microorganisms by polymerase chain reaction in delayed endophthalmitis after cataract surgery. Ophthalmology. 2000;107(6):1047–51.
51. Gomes JÁP, Frizon L, Demeda VF. Ocular surface microbiome in health and disease. Asia-Pacific J Ophthalmol. 2020;9(6):505–11.

52. Abbasi K, Razzaghi P, Poso A, Ghanbari-Ara S, Masoudi-Nejad A. Deep Learning in Drug Target Interaction Prediction: Current and Future Perspectives. Curr Med Chem. 2021;28(11):2100–13.
53. Kavianfar A, Taherkhani H, Ghorbani F. Utilizing Microbiome Approaches for Antibiotic Resistance Analysis ; an Ocular Case Evaluation. J Ophthalmic Optom Sci. 2021;5(1).

Footnotes and Financial Disclosures

Conflict of interest:

The authors have no conflict of interest with the subject matter of the present manuscript.