Journal of Artificial Intelligence for Medical Sciences
Vol. 3 (1–2), December 2022, pp. 1–7
Journal home: http://oapublishing-jaims.com/jaims
DOI: https://doi.org/10.55578/joaims.220819.001; eISSN: 2666-1470

RESEARCH ARTICLE

Discriminative Machine Learning Analysis for Skin Microbiome: Observing Biomarkers in Patients with Seborrheic Dermatitis

H.E.C. van der Wall[1,2,*], R.J. Doll[1], G.J.P. van Westen[2], T. Niemeyer-van der Kolk[1], G. Feiss[6], H. Pinckaers[3], M.B.A. van Doorn[4], T. Nijsten[4], M.G.H. Sanders[4], A.F. Cohen[1,2], J. Burggraaf[1,2,5], R. Rissmann[1,2], L.M. Pardo[4]

[1]*Method Development, Centre for Human Drug Research, The Netherlands*
[2]*Drug Discovery and Safety, Leiden Academic Centre for Drug Research, The Netherlands*
[3]*Department of Pathology, Radboud University Medical Center, The Netherlands*
[4]*Department of Dermatology, Erasmus Medical Center, The Netherlands*
[5]*Department of Clinical Pharmacology, Leiden University Medical Center, The Netherlands*
[6]*GL Feiss Consulting, LLC, Telford, Pennsylvania, United States*

ARTICLE DATA

Article History

Received 28 April 2022
Revised 19 July 2022
Accepted 18 August 2022

Keywords
Microbiome
Artificial intelligence
Machine learning
Seborrheic dermatitis
Data science

ABSTRACT

In recent years the skin microbiome has taken center stage as drug target and as disease biomarker. Computational analyses of microbiome sequencing data from patients with skin diseases, for example seborrheic dermatitis, can be performed to identify discriminative biomarkers in the microbiome profile. The aim of the present study was twofold, namely to employ machine learning to predict disease from the microbiome dataset, and to identify discriminative biomarkers in the microbiome of patients with seborrheic dermatitis versus healthy controls using machine learning techniques. The population consisted of 97 patients with seborrheic dermatitis and 763 healthy controls. Skin swabs were taken from naso-labial fold (lesional skin: n = 22; non-lesional skin: n = 75, controls: n = 763). Using an extra trees machine learning model, differences between the skin microbiome of patients with seborrheic dermatitis versus healthy controls were characterized. Subsequently, the most important microorganisms for discrimination were determined by feature analysis and SHapley Additive exPlanations (SHAP) values. The accuracy of the prediction models to discriminate between skin affected by seborrheic dermatitis and facial skin from healthy subjects was 77% and the ROC-AUC was 83%. Next to *Cutibacterium* and *Staphylococcus*, the most important organisms for discrimination had a relatively low occurrence. Our study showed that machine learning can be utilized to identify discriminating biomarkers in the microbiome skin. This indicates that machine learning can be of major importance in basic skin research, and in the discovery and development of new individualized therapies, involving the microbiome.

INTRODUCTION

The skin is the largest organ of the human body and is colonized by a wide range of microorganisms [1]. Many of the microorganisms living on the skin (its microbiome) are harmless and, in some cases, provide vital functions.

Despite the great interest of the skin as an ecosystem during the past decade, the study of the skin microbiome was until recently restricted by the low host-commensal cell ratio and the high taxonomical divergence among skin sites [2]. This changed by the introduction of methodology to remove microbial DNA from low biomass skin samples such as described by Garcia-Garcera et al. in 2013. The authors utilized a combination of molecular techniques that involved standard, quantitative PCR and amplicon sequencing of 16S rRNA, which significantly improved the field of skin microbiome research. At present, the skin microbiome is known to be involved in several skin diseases [3]. This breakthrough has led to additional knowledge on specific microorganisms that play a role in some of these skin disorders, for instance the role of *Staphylococcus aureus* in atopic dermatitis and *Cutibacterium acnes* in acne vulgaris. However, the role of microorganisms that are less abundant is still largely unknown. It is plausible that the presence of a combination of several different organisms forming a specific microbial profile might also contribute to the development and subtype of skin disease. A challenge

*Corresponding author. Email: hvdwall@chdr.nl

to explore this hypothesis is however hampered by the magnitude of the data which analysis is frequently beyond conventional data analyses. Machine learning may offer a solution because the underlying computational analyses may facilitate the identification of specific patterns of microorganisms that are discriminative for a specific type of skin diseases [4].

Machine learning has been rapidly adopted in microbiome studies for diagnosing clinical diseases. Modelling of the human microbiome by machine learning offers the potential to identify specific microbial biomarkers and may aid in the diagnosis of many clinical diseases. For instance, machine learning has already shown its ability to identify key features (markers) and modelling predictive biomarker signature in a variety of fields, including oncology [5–7], neurology [5–7], immunology [8], gastroenterology [9], diabetes [10], and skin diseases [11,12]. The advantages of machine learning techniques over classical statistical models are to infer relationships between variables for automatic pattern discovery and handling with multi-dimensional data [13]. By training a highly accurate model it is easy to find out which features are most informative for classification. For a dataset with many different features, in this case more than 600, machine learning is therefore saving time and effort, compared to existing statistical methods. In addition, the benefits of machine learning comprise flexibility and scalability compared with conventional statistical approaches, which makes it deployable for several tasks, such as diagnosis and classification, and survival predictions [14]. As a result, machine learning may be highly informative for the development of therapeutic modalities to ameliorate the microbial imbalance and to counteract certain pathogens.

The aim of the present study was twofold; to employ machine learning to predict disease from the microbiome dataset, and to identify discriminative biomarkers in the microbiome of patients with seborrheic dermatitis versus healthy controls using machine learning techniques. We hypothesized that the microbiome-based biomarkers alone can be used to predict the diagnosis. These models can then be employed to identify discriminative biomarkers in the microbiome.

MATERIALS & METHODS

All data used in the present study were obtained from a previous study performed in participants from the Rotterdam Study [15]. This was a cross sectional study embedded in a population-based study. Skin swabs were taken from naso-labial fold from 97 participants with seborrheic dermatitis (lesional skin: $n = 22$; non-lesional skin $n = 75$) and controls without skin conditions on the face or scalp ($n = 763$). Participants with seborrheic dermatitis and involvement of the nasolabial fold were considered lesional cases, and those without involvement of the nasolabial fold non-lesional cases. The median age was 53 years in the control group, 56 years for non-lesional cases and 68 years for lesional cases (for further details see Sanders, Nijsten [15]). The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians.

Data Collection for the Model

In all participants included in the study, the skin microbiome was analyzed by amplifying the V1 to V3 variable regions of the 16S rRNA gene using the 27F-519R primer pair and dual indexing. The genes were annotated using the Silva database. In the current study, microbiome data from the three categories of skin from the face were analyzed; facial skin from controls, (facial) non-lesional and lesional skin from patients with seborrheic dermatitis. To show the clearly visible differences between the skin categories, average microbiome profiles were created for each category of skin by taking the average occurrence of each bacterium present in all pertaining subjects. Subsequently, by using machine learning the lesional skin from patients with seborrheic dermatitis was characterized in order to discriminate it from the skin of healthy subjects. The occurrence of 686 organisms at genus level present in either one of both datasets were used as features. In addition, three well known alpha-diversity indices, the Simpson's diversity index, the Shannon diversity index, and the Chao1 index, were used as features [16].

Data Pre-processing and Selection

As there were more observations available from healthy skin than from the skin affected by seborrheic dermatitis, there was an imbalance in the data set. A four-fold cross validation was used. Therefore, six microbiome profiles (close to 25 percent of the total number of affected profiles) were used for validation of the model each fold. The same number of healthy profiles were used to produce a balanced validation set. The remaining profiles were used to train the model (training set).

To create a balanced training set, the SMOTE (Synthetic Minority Oversampling Technique) algorithm was applied to produce 'synthetic' profiles of the skin affected by seborrheic dermatitis based on the values already present in these underrepresented microbiome profiles [17]. Next, the features in both the training and validation sets were standardized based on the mean and standard deviation of the features in the training set.

Feature Selection

As the last preprocessing step, feature selection is performed on all 689 features, including the diversity indices. When two features had a high correlation (> 0.9 or < -0.9), only the most important one – based upon the feature importance of fitting the model on the training set – was used [18]. The features were selected by the training set in unsupervised fashion.

Machine Learning

Several different machine learning algorithms which were obtained from the scikit-learn module version 1.0.1 in python 3.7.9 were employed on the data. A DecisionTree Classifier, a RandomForest Classifier, a GradientBoostingClassifier, a Support Vector Classifier, Logistic Regression, and a Extra Trees Classifier were used to make an attempt to distinguish healthy from affected skin. Prior to training, a nested cross-validation (within the training set) was used to optimize the model hyperparameters. This process of oversampling, feature selection,

optimization, training, and validation was repeated in each fold with different training and validation data. In each fold, the validation was performed with profiles that had not previously been in a validation set, so that all 22 different profiles from skin affected with seborrheic dermatitis were tested at least once. Two profiles were twice in the test set. The optimal models were evaluated on the validation fold with the accuracy, sensitivity, specificity, and Area Under the Receiver Operating Characteristic (ROC) Curves (AUC). An overview of the machine learning workflow is shown in Supplementary Figure 2. The best performing model was used for further anylyses. The performance of the optimized model using the selected features was compared with the performance of conventional logistic regression using all features.

To gain insight into the impact of the individual features on the predictions, SHAP (SHapley Additive exPlanations) values were calculated [19]. To validate the importance of the features, the feature values of the correctly and incorrectly predicted occasions were compared. The impact of the features was validated by means of the feature importance of the model.

RESULTS

The three skin categories showed many similarities in the microbiome. As expected, *Staphylococcus* and *Cutibacterium* showed the highest relative abundance (range 20–50%); *Cutibacterium* was highest on average in healthy profiles and in profiles of the non-lesional skin of patients with seborrheic dermatitis. *Staphylococcus* was highest in the skin affected by seborrheic dermatitis. Figure 1 shows the average skin microbiome profiles of lesional and non-lesional facial skin of patients with seborrheic dermatitis and of healthy skin. The non-lesional skin shows an average microbiome profile which is in between healthy and affected skin.

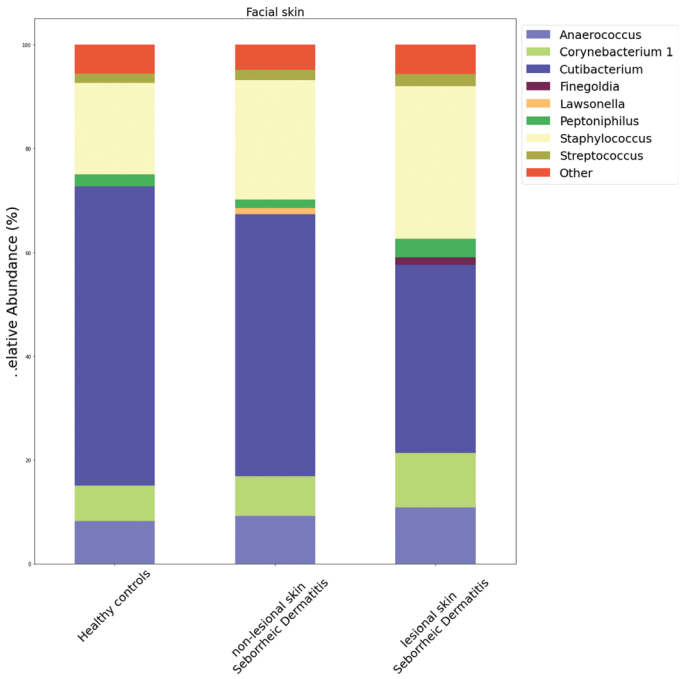


Figure 1 | Average microbiome profiles of the facial skin of healthy subjects and of the non-lesional and lesional (facial) skin of patients with seborrheic dermatitis. The *y*-axis shows the relative occurrences in percentages. The profiles show organisms at genus level that had on average an occurrence of more than 1%. Other organisms are combined as 'other'.

From all the models tested, the extra trees classifier performed best. The results of the employment of this model type are shown below.

Figure 2 shows the true labels (clinical diagnosis) versus the predicted labels for seborrheic dermatitis versus controls based on the skin microbiome. The models predicting seborrheic dermatitis had an overall accuracy of 77% (range 73–81%, compared to 48 ± 14% with conventional binary logistic regression). Out of the 24 profiles with seborrheic dermatitis, 18 were correctly predicted, indicating a sensitivity of 75% (range 65–85%). Of the 24 healthy profiles, 19 were correctly predicted, indicating a specificity of 79% (range 71–87%). The average AUC of the models was 83% (range 77–89%, Figure 3).

Figure 4 shows the impact of the occurrence of each organism on the predictions of seborrheic dermatitis versus healthy control. A low occurrence of *Cutibacterium* and a high occurrence of *Staphylococcus* was shown to be most predictive for the diagnosis of seborrheic dermatitis. It can be observed that the other

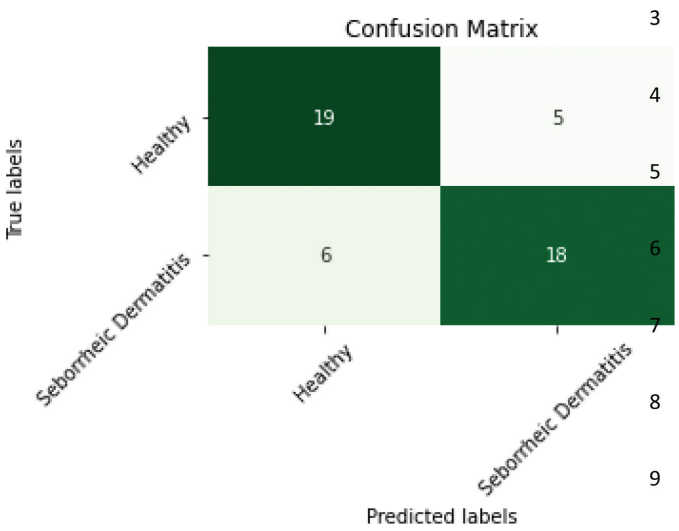


Figure 2 | Confusion matrix (predictive analysis tool) of the predicted diagnoses based on the skin microbiome profile of patients with seborrheic dermatitis by means of machine learning. The *x*-axis shows the predicted labels. The *y*-axis shows the true labels.

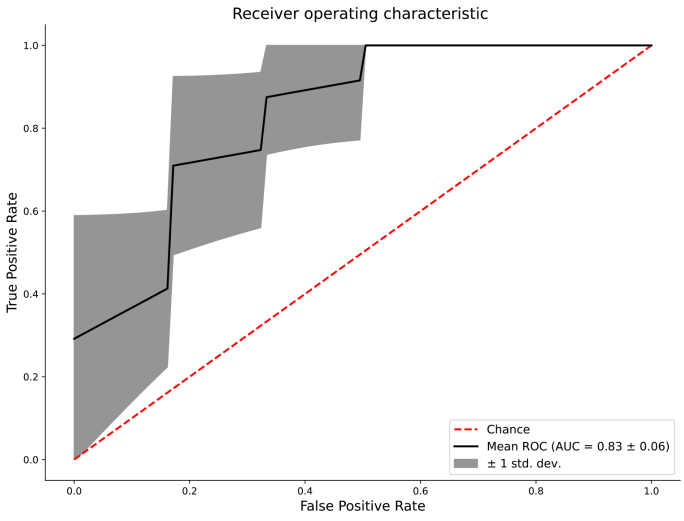


Figure 3 | Receiver Operating Characteristic (ROC) curve of the models predicting seborrheic dermatitis where the black line is the mean curve and the gray area is the standard deviation.

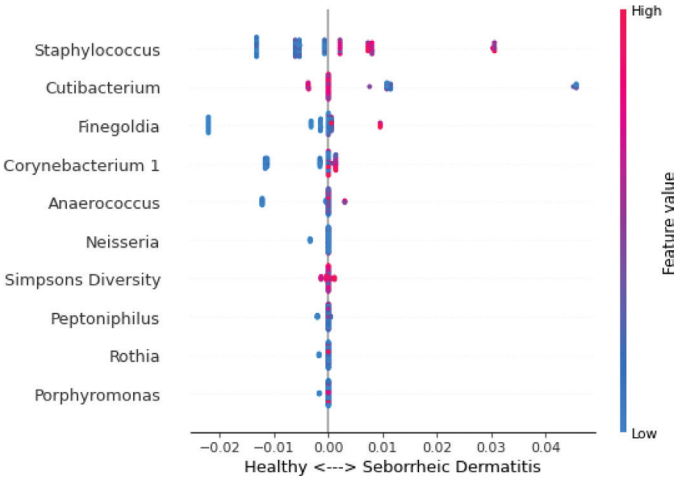


Figure 4 | SHAP values of the 10 features with the highest impact on the prediction of disease diagnosis for seborrheic dermatitis. A relatively high occurrence of a microorganism is shown in red, whereas a relatively low occurrence of a microorganism is shown in blue. The predicted diagnosis of control is on the left side of the *x*-axis, and of seborrheic dermatitis on the right side of the *x*-axis. Shorter bars mean less impact on diagnosis.

micro-organisms, which had any impact in the discrimination of seborrheic dermatitis, showed a relatively low occurrence.

The boxplots of the standardized values of the most important organisms for the correct and wrongly predicted profiles are shown in the Supplementary Material. These figures confirm the findings of the SHAP values.

DISCUSSION

In the present study we demonstrated that machine learning-based models may facilitate the identification of discriminative biomarkers in the microbiome of patients. These findings are particularly important for skin diseases, in which the microbiome has not been fully elucidated. Modulations of skin microbiome composition to restore host–microbiome homeostasis could become important future strategies to treat or prevent skin disease [3]. This highlights the potential important role of machine learning in the discovery of targets for new medical therapies.

Machine learning has been recently applied in microbiome studies for diagnosing clinical diseases in various fields, such as oncology, neurology, immunology and dermatology [4]. In the present study we aimed to employ a machine learning model to predict the diagnosis using skin microbiome profiles from patients with seborrheic dermatitis. The created models provided a unique insight into the types and complex patterns of micro-organisms involved this skin condition. Although many data are available on the skin microbiome in seborrheic dermatitis, our results show that bacteria with a low abundance are also valuable for disease discrimination. The factor of low *Cutibacterium* contributes to our models, being consistent with previous reports [20].

Many factors are known to significantly affect the skin microbiome, including weather conditions and washing behaviour, but also skin diseases and the use of therapeutic agents [21–24]. The current study showed that, based on the microbiome alone, machine learning models could predict a diagnosis of seborrheic dermatitis with an accuracy of 77%. The area under the curve from our models was 83%. There was a strong predictive correlation between the microbiome profile and the specific dermatologic disease. Given the large number of influencing factors that can affect the microbiome (although excluded as much as possible in the clinical trial) perfect predictive power using the microbiome cannot be expected. But, because of their high performances, these models could potentially be used to identify discriminative disease biomarkers in the microbiome.

Value of Machine Learning

Machine learning methods are being actively and widely used to elucidate the composition of microbiome and to investigate how they affect host phenotypes [23,25]. Various studies have already explored the power of machine learning to use microbiome patterns to predict host characteristics [23,26–28]. In addition, machine learning has earlier been applied on the skin microbiome to predict the postmortem interval [29]. In the current study we have shown that disease biomarkers can also be found using machine learning in the skin microbiome. Some of the parameters, such as a high abundance of *Staphylococcus*, and the diversity are already known to play a role in seborrheic dermatitis [30–32].

Moreover, machine learning identified distinctive organisms that are not initially considered important to investigate based on high occurrence rates. As shown in Figure 4, the occurrence of *Corynebacterium* 1, *Anaerococcus*, and *Finegoldia* also play a role in the distinction between facial skin from healthy individuals and subjects with seborrheic dermatitis. While the above-noted three organisms might have been identified through occurence alone (Figure 1), Gemella, Prevotella and Granullicatella would not have been identified as they occurred at less than one percent and are included in "other". Use of machine learning identified organisms which would otherwise have been overlooked. Future studies, including some association analysis or text analysis to describe the biological function mechanism of these biomarkers could be very interesting.

Using the same dataset, conventional binary logistic regression produced results with lower discrimination ability.

The results of this proof of concept study indicate that machine learning can be a valuable tool to find organisms that distinguish diseased skin from healthy skin. While the microbiology of seborrheic dermatitis has been thought to be fairly well elucidated, the importance of low occurrence organisms was shown. This suggest that in skin diseases with a less known and/or more complex microbiome profile machine learning could be a valuable investigative tool.

Limitations

Some limitations of the study should be noticed. Apart from a relatively low number of patients, DNA from microbial eukaryotes, such as yeast or fungi, could unfortunately not be classified meaning by this 16S gene screening method and limited primer set, only prokaryotic DNA, known in the database, could be recognized. This precludes recognition of fungi, such as yeast and bacteria not known in the database, while the fungal genus *Malassezia* is also known to be a potential biomarker for seborrheic dermatitis [33,34]. For future studies, sequencing of the

internal transcribed spacer (ITS) region [35], would enable the possibility of identifying fungi as potential biomarkers.

There is an age difference between the control group and lesional cases. This could be a factor of disease. Future studies, linking the organisms to an age group would be very interesting.

Genus level was the deepest screening level in this study, indicating a second limitation of screening by means of 16S DNA sequencing with a limited set of primers. Therefore, it has to be taken into account that the exact distribution of species within the genus is unknown. For example a large part of the 16S DNA recognized as *Staphylococcus* might not originate from the species *S. aureus* but from *S. epidermidis*, a species which is very common on healthy skin [36]. This should be investigated in future studies with novel techniques that have the ability to give deeper insight on species or even strain level.

CONCLUSION

In recent years, various elements in the skin microbiome have become of high interest for pharmaceutical companies as new drug targets. Despite the challenges and hurdles yet to overcome, it seems very likely that microbiome modulation will play a future role in the treatment of skin disease. From our study it has become clear that machine learning can be instrumental for the identification of biomarkers in the microbiome of skin. Consequently, machine learning can be of major importance in basic skin research, and in the discovery and development of new individualized therapies, involving the microbiome.

DECLARATIONS

Ethical Approval and Consent to Participate

All ethical approvals for the procedures have been obtained.

Consent for Publication

All authors approved the final article for publication.

Availability of Data and Materials

Not applicable.

Conflicts of Interest

The authors have no financial conflicts of interest to declare.

Authors' Contribution

van der WHEC contributed in the conception and design of the study, analysis and interpretation of data, drafting the article. DRJ, VWGJP, FG, PH, vDMBA, NT, SMGH, and BJ contributed in revising the study critically for important intellectual content and in the final approval of the version to be submitted. N-van der KT contributed in the analysis and interpretation of data. CAF gave his final approval of the version to be submitted. RR and PLM contributed in the analysis and interpretation of data, revising the study critically for important intellectual content, and in the final approval of the version to be submitted.

Funding

The authors declare no funding.

Abbreviations

SMOTE: Synthetic Minority Oversampling Technique
ROC: Receiver Operating Characteristic
AUC: Area Under the Curve
SHAP: SHapley Additive exPlanations

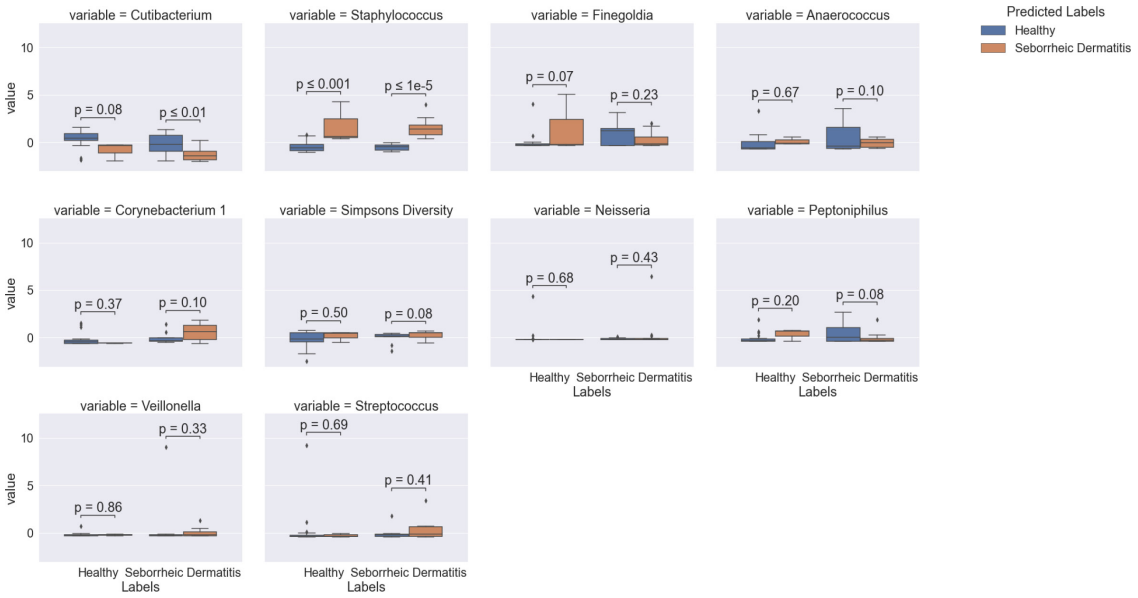
REFERENCES

[1] E.A. Grice, J.A. Segre, The skin microbiome, Nat. Rev. Microbiol. 9 (2011), 244–253.
[2] M. Garcia-Garcerà, K. Garcia-Etxebarria, M. Coscollà, A. Latorre, F. Calafell, A new method for extracting skin microbes allows metagenomic analysis of whole-deep skin, PLoS ONE. 8 (2013), e74914.
[3] P.L.J.M. Zeeuwen, M. Kleerebezem, H.M. Timmerman, J. Schalkwijk, Microbiome and skin diseases, Curr. Opin. Allergy Clin. Immunol. 13 (2013), 514–520.
[4] M. Leclercq, B. Vittrant, M.L. Martin-Magniette, M.P. Scott Boyer, O. Perin, A. Bergeron, et al., Large-scale automatic feature selection for biomarker discovery in high-dimensional OMICs data, Front. Genet. 10 (2019), 452.
[5] D. Zhang, D. Shen, The Alzheimer's Disease Neuroimaging Initiative, Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease, NeuroImage. 59 (2012), 895–907.
[6] G. Deshpande, L.E. Libero, K.R. Sreenivasan, H.D. Deshpande, R.K. Kana, Identification of neural connectivity signatures of autism using machine learning, Front. Hum. Neurosci. 7 (2013), 670.
[7] T. Fekete, N. Zach, L.R. Mujica-Parodi, M.R. Turner, Multiple kernel learning captures a systems-level functional connectivity biomarker signature in amyotrophic lateral sclerosis, PLoS ONE. 8 (2013), e85190.
[8] A. Sutherland, M. Thomas, R.A. Brandon, R.B. Brandon, J. Lipman, B. Tang, et al., Development and validation of a novel molecular biomarker diagnostic test for the early detection of sepsis, Crit. Care. 15 (2011), R149.
[9] A. Kohli, E.A. Holzwanger, A.N. Levy, Emerging use of artificial intelligence in inflammatory bowel disease, World J. Gastroenterol. 26 (2020), 6923–6928.
[10] Q.A. Hathaway, S.M. Roth, M.V. Pinti, D.C. Sprando, A. Kunovac, A.J. Durr, et al., Machine-learning to stratify diabetic patients using novel cardiac biomarkers and integrative genomics, Cardiovasc. Diabetol. 18 (2019), 78.
[11] V. Fortino, L. Wisgrill, P. Werner, S. Suomela, N. Linder, E. Jalonen, et al., Machine-learning–driven biomarker discovery for the discrimination between allergic and irritant contact dermatitis, Proc. Natl. Acad. Sci. U S A. 117 (2020), 33474–33485.
[12] H. Johansson, M. Lindstedt, A.-S. Albrekt, C.A.K. Borrebaeck, A genomic biomarker signature can predict skin sensitizers using a cell-based *in vitro* alternative to animal tests, BMC Genomics. 12 (2011), 399.
[13] L.J. Marcos-Zambrano, K. Karaduzovic-Hadziabdic, T.L. Turukalo, P. Przymus, V. Trajkovik, O. Aasmets, et al., Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment, Front. Microbiol. 12 (2021), 634511.

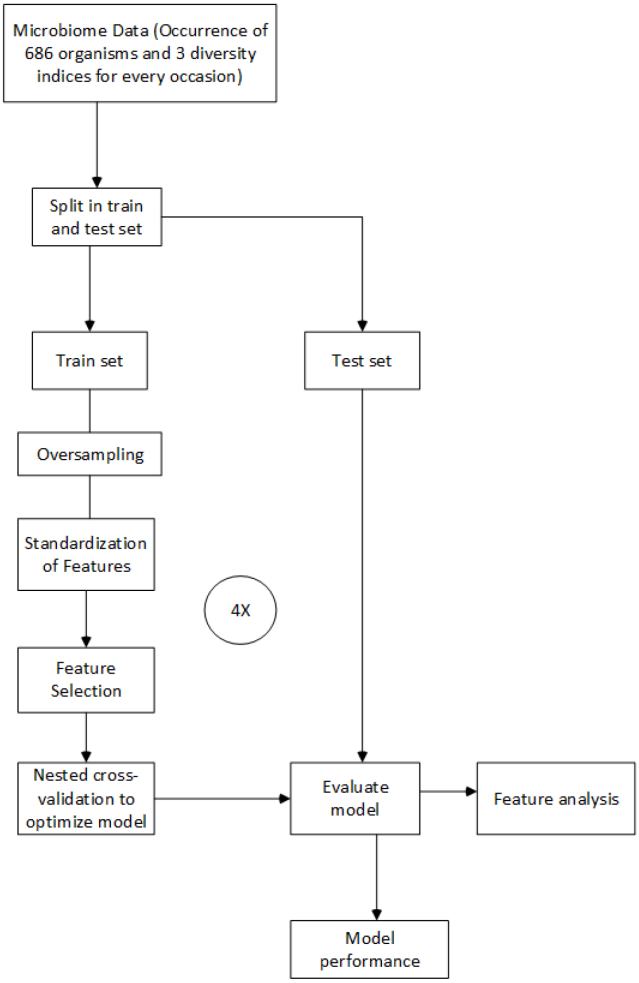
[14] H.S.R. Rajula, G. Verlato, M. Manchia, N. Antonucci, V. Fanos, Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment, Medicina. 56 (2020), 455.
[15] M.G.H. Sanders, T. Nijsten, J. Verlouw, R. Kraaij, L.M. Pardo, Composition of cutaneous bacterial microbiome in seborrheic dermatitis patients: a cross-sectional study, PLoS ONE. 16 (2021), e0251136.
[16] E.H. Simpson, Measurement of diversity, Nature. 163 (1949), 688.
[17] N.V. Chawla , K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artifi. Intellig. Res. 16 (2002), 321–357.
[18] H.E.C. van der Wall, R.J. Doll, G.J.P. van Westen, I. Koopmans, R.G. Zuiker, J. Burggraaf, et al., The use of machine learning improves the assessment of drug-induced driving behaviour, Accid. Anal. Prevention. 148 (2020), 105822.
[19] S. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, arXiv preprint, arXiv:1705.07874, 2017.
[20] Z. Xu, Z. Wang, C. Yuan, X. Liu, F. Yang, T. Wang, et al., Dandruff is associated with the conjoined interactions between host and microorganisms, Sci. Rep. 6 (2016), 24877.
[21] L.B. Nørreslet, T. Agner, M.-L. Clausen, The skin microbiome in inflammatory skin diseases, Curr. Derm. Rep. 9 (2020), 141–151.
[22] M. Picardo, M. Ottaviani, Skin microbiome and skin disease: the example of rosacea, J. Clin. Gastroenterol. 48 (2014), S85–S86.
[23] A.P. Carrieri, N. Haiminen, S. Maudsley-Barton, L.-J. Gardiner, B. Murphy, A.E. Mayes, et al., Explainable AI reveals changes in skin microbiome composition linked to phenotypic differences, Sci. Rep. 11 (2021), 4565.
[24] S.K. Shukla, N.S. Murali, M.H. Brilliant, Personalized medicine going precise: from genomics to microbiomics, Trends Mol. Med. 21 (2015), 461–462.
[25] J. Namkung, Machine learning methods for microbiome studies, J. Microbiol. 58 (2020), 206–216.
[26] D. Knights, E.K. Costello, R. Knight, Supervised classification of human microbiota, FEMS Microbio. Reviews. 35 (2011), 343–359.
[27] Y.-H. Zhou, P. Gallins, A review and tutorial of machine learning methods for microbiome host trait prediction, Front. Genet. 10 (2019), 579.
[28] L. Moitinho-Silva, G. Steinert, S. Nielsen, C.C.P. Hardoim, Y.-C. Wu, G.P. McCormack, et al., Predicting the HMA-LMA status in marine sponges by machine learning, Front. Microbiol. 8 (2017), 752.
[29] H.R. Johnson, D.D. Trinidad, S. Guzman, Z. Khan, J.V. Parziale, J.M. DeBruyn, et al., A machine learning approach for using the postmortem skin microbiome to estimate the postmortem interval, PLoS ONE. 11 (2016), e0167370.
[30] F. Tamer, M.E. Eren Yuksel, E. Sarifakioglu, Y. Karabag, *Staphylococcus aureus* is the most common bacterial agent of the skin flora of patients with seborrheic dermatitis, Dermatol. Pract. Concept. 8 (2018), 80–84.
[31] L. Vicent, M. Martínez-Sellés, Electrocardiogeriatrics: ECG in advanced age, J. Electrocardio. 50 (2017), 698–700.
[32] A. Tanaka, O. Cho, C. Saito, M. Saito, R. Tsuboi, T. Sugita, Comprehensive pyrosequencing analysis of the bacterial microbiota of the skin of patients with seborrheic dermatitis, Microbiol. Immunol. 60 (2016), 521–526.
[33] R.J. Hay, *Malassezia*, dandruff and seborrhoeic dermatitis: an overview, Br. J. Dermatol. 165 (2011), 2–8.
[34] A.K. Gupta, R. Batra, R. Bluhm, T. Boekhout, T.L. Dawson, Skin diseases associated with Malassezia species, J. Am. Acad. Dermatol. 51 (2004), 785–798.
[35] J.L. Metcalf, Z.Z. Xu, S. Weiss, S. Lax, W.V. Treuren, E.R. Hyde, et al., Microbial community assembly and metabolic function during mammalian corpse decomposition, Science. 351 (2016), 158–162.
[36] S. Seite, T. Bieber, Barrier function and microbiotic dysbiosis in atopic dermatitis, Clin. Cosmet. Investig. Dermatol. 8 (2015), 479–483.

SUPPLEMENTARY MATERIALS

Supplementary Figure 1 shows which of the 10 microorganisms caused skin profiles affected with seborrheic dermatitis to be wrongly predicted (a large difference between bars 1 and 2), and healthy profiles to be wrongly predicted as seborrheic dermatitis (a large difference between bars 3 and 4).



Supplementary Figure 1 | Boxplots of the standardized values of the 10 most important features for distinguishing seborrheic dermatitis from healthy skin based on the microbiome in order of importance of the models. On the left side are the healthy profiles and right are profiles affected by seborrheic dermatitis. Visualized in blue are the profiles predicted as healthy and the profiles predicted as affected are shown in orange. The *y*-axis shows the values of the standardized occurrences. The *p* values of an independent *t*-test between the right an wrong predicted profiles are shown.



Supplementary Figure 2 | Machine learning workflow (four-fold cross validation).