



OPEN

Using machine learning model explanations to identify proteins related to severity of meibomian gland dysfunction

Andrea M. Storås^{1,2✉}, Fredrik Fineide^{2,3}, Morten Magnø^{4,5}, Bernd Thiede⁶, Xiangjun Chen^{4,7,8}, Inga Strümke⁹, Pål Halvorsen^{1,2}, Hilde Galtung¹⁰, Janicke L. Jensen¹¹, Tor P. Utheim^{2,3,4,7,12} & Michael A. Riegler^{1,2,13}

Meibomian gland dysfunction is the most common cause of dry eye disease and leads to significantly reduced quality of life and social burdens. Because meibomian gland dysfunction results in impaired function of the tear film lipid layer, studying the expression of tear proteins might increase the understanding of the etiology of the condition. Machine learning is able to detect patterns in complex data. This study applied machine learning to classify levels of meibomian gland dysfunction from tear proteins. The aim was to investigate proteomic changes between groups with different severity levels of meibomian gland dysfunction, as opposed to only separating patients with and without this condition. An established feature importance method was used to identify the most important proteins for the resulting models. Moreover, a new method that can take the uncertainty of the models into account when creating explanations was proposed. By examining the identified proteins, potential biomarkers for meibomian gland dysfunction were discovered. The overall findings are largely confirmatory, indicating that the presented machine learning approaches are promising for detecting clinically relevant proteins. While this study provides valuable insights into proteomic changes associated with varying severity levels of meibomian gland dysfunction, it should be noted that it was conducted without a healthy control group. Future research could benefit from including such a comparison to further validate and extend the findings presented here.

Meibomian gland dysfunction (MGD) is a condition characterized by an impairment of the physiologic function of the meibomian glands (MGs) and a reduced quantity or quality of the secreted meibum. Healthy meibum lowers the surface tension of the tear film and reduces ocular surface evaporation¹. In patients with MGD, this loss of function increases tear evaporation, cause hyperosmolarity in the tears, and ocular irritations. MGD is one of the leading causes of dry eye disease (DED), affecting several hundred million people globally². Both MGD and DED are tied to significantly impaired quality of life and substantial economic and social burdens^{3,4}.

Despite the primary presentation of MGD being diminished function of the ocular surface lipid layer, it has long been thought that understanding the role of the tear film proteins may be key to understanding the etiology of MGD and related symptoms⁵. Despite the top 20 most abundant proteins accounting for around 90% of the total tear film proteins⁶, more than 1500 unique proteins have been identified in the tear film⁷, many of which likely have important bioactive and cellular signalling roles⁸.

¹Department of Holistic Systems, Simula Metropolitan Center for Digital Engineering, Oslo, Norway. ²Department of Computer Science, OsloMet - Oslo Metropolitan University, Oslo, Norway. ³The Norwegian Dry Eye Clinic, Oslo, Bergen, Norway. ⁴Department of Ophthalmology, Sørlandet Hospital Arendal, Arendal, Norway. ⁵Department of Plastic and Reconstructive Surgery, Oslo University Hospital, Oslo, Norway. ⁶Department of Biosciences, University of Oslo, Oslo, Norway. ⁷Department of Medical Biochemistry, Oslo University Hospital, Oslo, Norway. ⁸Department of Ophthalmology, Vestre Viken Hospital Trust, Drammen, Norway. ⁹Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway. ¹⁰Institute of Oral Biology, University of Oslo, Oslo, Norway. ¹¹Department of Oral Surgery and Oral Medicine, University of Oslo, Oslo, Norway. ¹²Department of Ophthalmology, Oslo University Hospital, Oslo, Norway. ¹³Department of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway. ✉email: andrea@simula.no

Artificial intelligence, and in particular the subfield called machine learning (ML), is technology that can successfully solve a large variety of tasks in the field of ophthalmology⁹ and DED¹⁰. ML models learn from data to reach a predefined goal, such as estimating the tear film break-up time from clinical dry eye test results¹¹. By exploring the inner workings of the resulting ML models, insights can be gained about which variables, also called *features*, in the dataset are important for the models when, e.g., making a diagnosis. This opens up for discovering new medical factors that are associated with medical conditions and diseases.

The aim of the current study was to explore potential relationships between the degree of MGD and protein expressions in tears and at the same time to build an ML pipeline. Emphasis was put on studying differences in tear protein expression across different severity levels of MGD, rather than simply comparing the proteomic profiles of healthy individuals to that of MGD patients. Techniques from ML were used to train predictive models and investigate what proteins the models regarded as important for predicting levels of MGD. Several of the proteins regarded as important by the models have previously been confirmed to be altered in the presence of DED by earlier studies. From a medical perspective, other insights might be gained by carefully investigating the proteins considered as important by the models. Some of the identified proteins might for example play a role in the pathogenesis of MGD and/or serve as potential targets for new treatments. They could also provide information about the severeness and prognosis of MGD.

Methods

This section describes the dataset, analysis methods, ML models and feature importance techniques applied in this work. A flowchart of the process is given in Fig. 1.

Protein analysis and clinical tests

The dataset that the current work is based on includes detailed protein measurements from the tears of 320 patients with DED visiting the Norwegian Dry Eye Clinic during the years 2017 to 2019. No exclusion criteria were applied. The Regional Medical Ethics Committee of South-East Norway (REK) had approved the study (reference id 6892), and all medical procedures were performed in compliance with the Declaration of Helsinki. Written informed consent was given by all the patients. Since the present work focused on MGD, only the subset of patients with MGD were applied. As a result, the dataset in the current work included observations from 234 patients. Further descriptions about the tear protein collection, analysis and additional clinical tests for the original dataset are provided below.

Protein data is available from samples collected from Schirmer strips from the left eye of each patient. The quantity of tears equalled the full amount collected by the strips. The strips were stored in 500 μ l phosphate buffered saline (PBS) and kept frozen until analyzed. Proteins were extracted from the strips using precipitation with cold acetone before trypsin digestion was performed. The tear proteins were then analyzed using untargeted liquid chromatography coupled to mass spectrometry (LC-MS) as previously described¹². The PEAKS X Pro software was applied for relative protein quantification and protein identification as described earlier¹³: for label-free quantification (LFQ) using PEAKS X Pro, the following parameters were applied on protein level: false discovery rate (FDR) \leq 1%, fold change \geq 2, significance method analysis of variance (ANOVA) with at least 2 peptides. For normalization, the total ion current (TIC) was used. The LC-MS data was searched against the human Uniprot database (20384 entries) with PEAKS X Pro software version 10.5 (Bioinformatics Solutions, Waterloo, ON, Canada). The following parameters were used: digestion enzyme: trypsin, maximum missed cleavage: 1, fragment ion mass error tolerance: 0.05 Da and parent ion error tolerance: 10 ppm. Oxidation of methionine and acetylation of the N-terminus were specified as variable modifications. The maximum number of posttranslational modifications (PTMs) per peptide was set to 2. An FDR of 1% was applied to the datasets.

Additional results from clinical tests are available in the dataset applied in the current work, including fluorescein break-up time (FBUT), results from Schirmer test and the level of MGD. To measure FBUT, 5 μ l 2% fluorescein was instilled with a micropipette and the seconds after a blink and before the tear film broke up was counted digitally. Unanesthetized Schirmer test was performed with sterile strips (0–30 mm/5 min) following standard protocols as described by Bron et al.¹⁴. Meibum quality and MG expressibility together with interpretation of meibography images were used to determine the severity level of MGD. To assess meibum quality and MG expressibility, a cotton swab was used to put a light pressure on the lower lid margin. Meibum quality was assessed according to a four-point scale (0: clear, 1: cloudy, 2: granular and 3: toothpaste), and a sum score

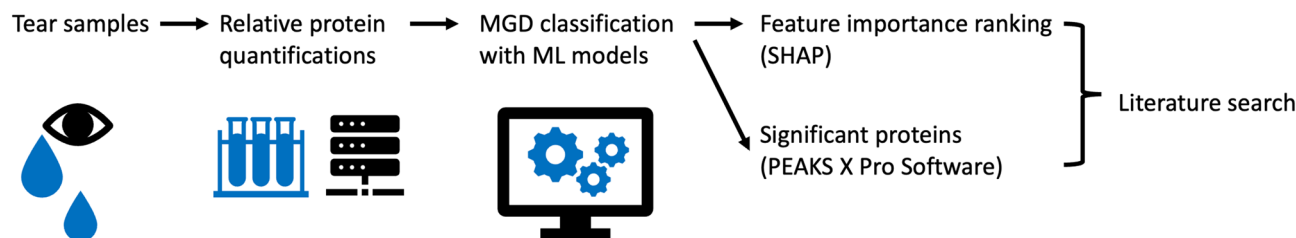


Figure 1. Overview of the workflow. Tear samples are extracted and the proteins quantified. Further on, ML models classify degree of MGD from the protein quantifications. The proteins are ranked by their importance in the models using Shapley Additive Explanations (SHAP), while PEAKS Pro X Software identify significant proteins. Identified proteins are investigated for potential relationship to MGD and DED through literature search.

for the central eight glands was calculated (range: 0–24). MG expressibility was graded on a four-point scale based on the number of expressible MGs among the central five glands (0: all glands expressible; 1: 3–4 glands expressible; 2: 1–2 glands expressible; 3: no glands expressible). Infrared (IR) meibography images were obtained using the Keratograph 5M (Oculus, Wetzlar, Germany). The amount of MG dropout was graded according to the Meiboscale by Pult^{15,16}. The scale ranges from 0 to 4, where level 0 represents no glandular atrophy and the increasing grades represent increasing glandular loss. The MGD grading was performed as described in The International Workshop on Meibomian Gland Dysfunction¹.

Data preprocessing

Because the relative protein quantification dataset applied in the present work only includes one patient with MGD level 1, this patient was excluded. Consequently, the final dataset contained MGD grade data from 233 patients. Because tears were sampled from the left eye, the MGD level for the left eye was considered. This ensured that the tear proteins correspond to the severity of MGD in the same eye. The level of MGD ranged from level 2 to 4 with the majority of observations belonging to level 3. Patient characteristics are included in Table 1. From the table, it is observed that the majority of patients were female for all levels of MGD. The mean age tended to increase slightly with increasing severity of MGD even though the differences were not statistically significant. Further on, higher levels of MGD were associated with shorter FBUTs and higher values of meibum expressibility, meibum quality and MG dropout, showing statistically significant differences between the three MGD levels (p values < 0.05). The significant differences are not unexpected because these clinical parameters are affected during MGD, and meibum expressibility and quality and MG dropout are all parameters included in MGD staging. The Schirmer test results were not statistically significantly different between the different levels of MGD.

From an ML perspective, input features that hold the identical value for all observations in the dataset do not provide any useful information to the model. Consequently, features representing proteins that were not identified in any of the patients, i.e., the feature values were 'NULL' for all observations, were removed from the dataset. Moreover, proteins with quantifications registered as 0 were replaced with 1. These proteins represent protein inference, meaning that the measurements do not necessarily belong to the protein of interest. Missing protein peak areas were replaced with 0, since missing values for the proteins mean that the proteins were not detected in the tear sample. The peak area should therefore be 0. Finally, proteins arising from contamination of the tear samples were removed prior to analysis. This includes keratins, except keratins 18 and 19, all dermcidins and dermacolines, as well as trypsin and putative trypsin-6. Most keratins, dermcidins and dermacolines are contaminations from the skin and should therefore not be included. Trypsin and putative trypsin-6 are proteins added to digest the proteins into smaller peptides. The final number of proteins included in the analysis was 2188. Logarithmic transformation followed by standard scaling was performed for the proteins prior to analysis due to wide value ranges and skewness in the value distributions of the relative protein quantifications.

Machine learning analysis

A boosting tree-based ML algorithm called LGBMClassifier was trained to predict the level of MGD from the relative protein quantifications. The unbalanced numbers of patients in each MGD level was taken into account during training by setting the 'is_unbalanced' hyperparameter to *True*. To avoid overfitting, the 'num_leaves' hyperparameter was set to 2. For the remaining hyperparameters, default values were applied¹⁷.

First, an LGBMClassifier was trained to predict the MGD levels 2 to 4 from the relative protein quantifications. Next, the three MGD levels were predicted independently as three different binary classification problems. This was done to investigate if the different MGD levels relate to different proteins as features. Specifically, one model was trained to predict whether the patient had MGD level 2 or not, one model predicted whether the MGD level was 3 or not, and one model predicted whether the MGD level was 4 or not.

The full dataset was used to train the models, and model performance was calculated on the same dataset. The reason the data was not split into training and test sets was because the aim was to investigate relationships between protein expression and level of MGD rather than developing models for predictive purposes. By splitting the dataset, less information would be available for the models to learn the task. Due to class imbalance, model performance was evaluated using balanced accuracy, F1 score and Matthews correlation coefficient (MCC). For prediction of all MGD levels by the same model, the weighted average was used to calculate the F1 score.

| MGD level | N (female/male) | Age (years) | OSDI | FBUT (s) | Schirmer (mm) | ME | MQ | MG dropout |
|-----------|-----------------|-------------|-------------|-----------|---------------|-----------|------------|------------|
| 2 | 31 (23/8) | 51.9 (17.0) | 32.0 (24.7) | 5.1 (4.1) | 15.5 (10.2) | 1.5 (0.7) | 6.0 (2.4) | 2.1 (1.0) |
| 3 | 126 (88/38) | 54.3 (17.8) | 27.4 (16.4) | 4.5 (3.6) | 14.7 (10.5) | 2.0 (0.5) | 8.3 (3.6) | 2.4 (0.9) |
| 4 | 76 (59/17) | 55.8 (17.9) | 34.3 (22.0) | 3.3 (2.7) | 14.4 (9.3) | 2.7 (0.5) | 11.1 (6.7) | 2.8 (0.9) |
| Total | 233 (170/63) | 54.5 (17.7) | 30.3 (19.8) | 4.2 (3.5) | 14.7 (10.1) | 2.2 (0.6) | 8.9 (5.0) | 2.5 (1.0) |

Table 1. Characteristics of the patients included in the analyses. Values are reported as mean with standard deviations in parentheses. Dry eye test results are for the left eye. ME, meibum expressibility; MQ, meibum quality; N, number of observations; OSDI, ocular surface disease index score.

Feature importance

To investigate feature importances, Shapley Additive Explanations (SHAP)¹⁸ was applied. SHAP approximates Shapley values, which arrive from game theory and measure how much each feature contributes to the final model prediction¹⁹. Shapley values represent the fair share of the total payout (or model prediction) each player (or input feature) participating in a collaborative game should receive. Shapley values exhibit several attractive properties that make the methods based on these values popular for explaining ML models²⁰. For instance, features that contribute equally to the model prediction receive the same Shapley value, while features that do not contribute receive a Shapley value of 0. The widely used library SHAP is based on the Shapley value and therefore shares the above-mentioned attractive properties¹⁸. It has established itself as a commonly used method for attributing importance of input features to ML models, including for tree-based models²¹, and this explanation method was chosen to explain the ML models in this study. The 15 most important features according to SHAP were examined for each of the four ML models. Because a low SHAP value represents a low feature importance for the model, the lowest-ranking features were not considered. The cutoff at 15 was chosen in order to focus on a limited number of proteins to investigate: the proteins represented by the 15 most important features according to SHAP were inspected in more detail by medical experts that are highly experienced in biomedical analyses, DED and MGD. Literature searches were performed by the experts for all the detected proteins to identify prior studies investigating the relevance of the proteins with respect to DED and MGD.

In addition to the original importance values provided by SHAP, a novel technique that weights the importance values by the model uncertainty was also proposed. This approach emphasizes feature importances from predictions the model is more certain about. The proposed weighting technique is outlined below.

For each observation in the dataset, the classification models output a probability score for each possible class. The probability score ranges from 0 to 1, where higher scores indicate that the models are more certain that the corresponding class is correct. Scores close to 0, on the other hand, mean that the models are certain that the corresponding class is not correct. If the probability scores are close to 0.5, the models are highly uncertain about which class the observation belongs to. The final model prediction is the class with the highest probability score. Based on the predicted probabilities, the predictions are divided into three groups: 'Positive', 'Negative' and 'Borderline'. If the predicted class is correct and the corresponding predicted probability is > 0.6 , the predicted sample is 'Positive'. However, if the predicted class is wrong, but the corresponding predicted probability is still > 0.6 , the predicted sample is 'Negative'. All other predictions (correct and incorrect with predicted probabilities of 0.6 or below) are classified as 'Borderline'. This is because the model is less certain about these predictions.

Next, the SHAP values for the corresponding predicted class are retrieved. If the sample is 'Positive', the SHAP value is increased according to the predicted probability as follows:

$$SHAP\ value_{weighted} = \frac{SHAP\ value_{original}}{1 - predicted\ probability} \quad (1)$$

For negative samples, however, the SHAP values are decreased as follows:

$$SHAP\ value_{weighted} = SHAP\ value_{original} \cdot (1 - predicted\ probability) \quad (2)$$

Because the predicted probabilities never exceed 1, this will result in reduced SHAP values for the wrong predictions and increased values for the correct predictions. In order to increase the SHAP value more as the predicted probability increases for the correct predictions, the value is divided by $(1 - predicted\ probability)$ rather than by the predicted probability directly. The same logic applies for the negative group. The SHAP values are not changed for the borderline group.

In the case of multiclass classification and correct class predictions, the SHAP values for the other classes are downweighted. Similarly, for incorrect predictions, SHAP values for the correct class are upweighted, while the wrong classes get their SHAP values downweighted. The weighting is performed as described above. Next, the sum of the SHAP values is calculated and the feature ranking is compared with the ranking of the original SHAP values. The weighted SHAP values are summarized in two different ways. First, samples from all three groups are included, i.e., the 'Positive', 'Negative' and 'Borderline' groups. As an alternative when no ground truth annotations are available, the SHAP values for borderline samples are simply excluded. This is because the model is less certain about these samples, and the feature importances for these samples can consequently be less reliable. The SHAP values for the rest of the samples are not changed. As for the model explanations using unweighted SHAP values, the proteins representing the 15 most important features following uncertainty weighting were examined by medical experts.

Finally, to compare the explainable artificial intelligence (XAI) methods with a more traditional approach for detecting proteins, an additional analysis using the PEAKS X Pro software was performed. This analysis identified the proteins that were significantly different between MGD levels 2, 3 and 4 and was described above in "Protein analysis and clinical tests" section.

Technical details

Python version 3.9.2 was used for all experiments. Scikit-learn²² was used for data preparation and model evaluation, and lightgbm for training the LGBMClassifier¹⁷. The SHAP library version 0.40.0 was used in the current study¹⁸. For reproducibility, the source code is made available online as described in Code availability.

Results

Machine learning models

When the classifier trained to predict MGD levels 2 to 4 was evaluated on the same data as used for training, the balanced accuracy was 72%, the F1 score was 0.82, and the MCC was 0.72. For the model predicting MGD level 2, the balanced accuracy, F1 score and MCC were 95%, 0.80 and 0.78, respectively. The corresponding model performance metrics for the model predicting MGD level 3 were 85%, 0.86 and 0.69, while for the model predicting MGD level 4, balanced accuracy, F1 score and MCC were 87%, 0.82 and 0.73, respectively.

Unweighted feature importance

To determine which proteins were regarded as most important for predicting MGD severity, feature importance values were estimated for the final models. A comparison of the relative quantifications in MGD levels 2 to 4 for the proteins regarded as most important by the ML models are provided in Supplementary Table A.1. The importance plot for the model predicting all levels of MGD is shown in Fig. 2a. Higher values mean higher importance. When searching the literature, the features representing the proteins S100-A8 and proline-rich protein 4 (PRP4) are upregulated and downregulated in DED⁸. Several of the other proteins listed in Fig. 2a are also reported to be related to DED and MGD: The polymeric immunoglobulin receptor^{23–26}, prostaglandin reductase 1²⁷ and immunoglobulin kappa variable 2-24 (IGKV2-24)²⁸.

Corresponding feature importance plots for the most relevant features for the three binary classification models are provided in Fig. 2b–d. For the model predicting MGD level 2 (Fig. 2b), the most important model feature represented tissue alpha-L fucosidase. Moreover, three of the other important features represented proteins that are upregulated (protein S100-A8) or downregulated (lysozyme C and PRP4) in DED⁸. Further on, features representing the immune system associated proteins IGKV2-24 and the polymeric immunoglobulin receptor were regarded as important features. These two proteins were also identified by the multiclass ML model. Serotransferrin and immunoglobulin alpha-2 heavy chain (IGA2) are additional proteins represented among the top features that are reported to be related to DED^{6,29}.

For predictions of MGD level 3 (Fig. 2c), features representing ubiquitin-conjugating enzyme E2 variant and cytoplasmic malate dehydrogenase were regarded as most important. Prostaglandin reductase 1 and IGKV2-24 were also considered as important according to the SHAP values. Other highly ranked proteins that are reported as being related to DED are immunoglobulin lambda variable 8-61 (IGLV8-61)³⁰, immunoglobulin lambda variable 6-57 (IGLV6-57)³⁰, immunoglobulin gamma-1 heavy chain (IGG1)³¹, all-trans retinol dehydrogenase-7 (ADH7)^{32,33} and S100 calcium-binding protein A7, also known as psoriasis^{34–36}.

Finally, for predictions of MGD level 4 (Fig. 2d), the feature representing heterogeneous nuclear ribonucleoprotein M was regarded as the most important feature. PRP4, which is downregulated in DED⁸, is also present in the figure. In addition, the features representing the polymeric immunoglobulin receptor and prostaglandin reductase 1 were considered important by the model according to SHAP.

Weighted feature importance

The ranking of the proteins according to the weighted SHAP values are provided in Tables 2, 3, 4 and 5. Proteins that were not included in the corresponding feature importance plots using the original SHAP values are

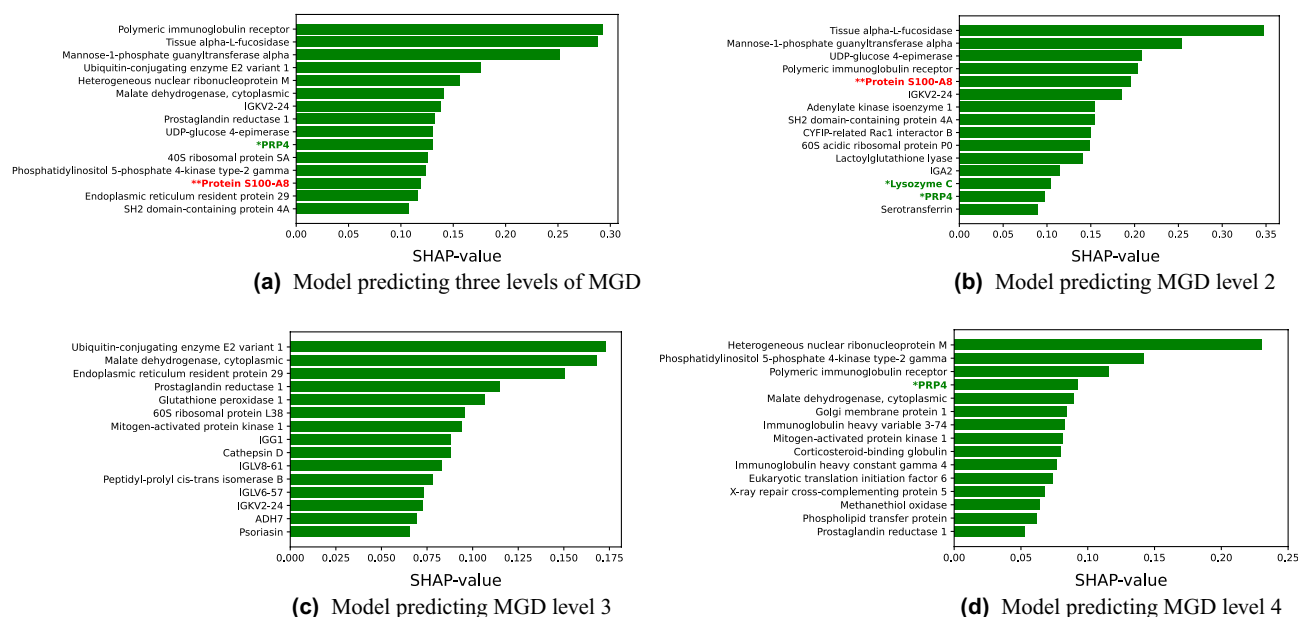


Figure 2. Unweighted feature importances for the four classifiers predicting MGD levels from relative protein quantifications using SHAP. * (green font): Protein identified as potential biomarker; downregulated in DED, ** (red font): Protein identified as potential biomarker; upregulated in DED⁸. Figures are best viewed with zoom.

| Rank | Protein (accession number), incl. BL | Protein (accession number), excl. BL |
|------|---|---|
| 1 | Ubiquitin-conjugating enzyme E2 variant 1 (Q13404) | Polymeric immunoglobulin receptor (P01833) |
| 2 | Endoplasmic reticulum resident protein 29 (P30040) | Mannose-1-phosphate guanyltransferase alpha (Q96IJ6) |
| 3 | Polymeric immunoglobulin receptor (P01833) | Tissue alpha-L-fucosidase (P04066) |
| 4 | Tissue alpha-L-fucosidase (P04066) | Ubiquitin-conjugating enzyme E2 variant 1 (Q13404) |
| 5 | Malate dehydrogenase, cytoplasmic (P40925) | Heterogeneous nuclear ribonucleoprotein M (P52272) |
| 6 | Prostaglandin reductase 1 (Q14914) | Endoplasmic reticulum resident protein 29 (P30040) |
| 7 | Mannose-1-phosphate guanyltransferase alpha (Q96IJ6) | IGKV2-24 (A0A0C4DH68) |
| 8 | IGKV2-24 (A0A0C4DH68) | 40S ribosomal protein SA (P08865) |
| 9 | Glutathione peroxidase 1 (P07203) | Prostaglandin reductase 1 (Q14914) |
| 10 | IGG1 (P0DOX5) | UDP-glucose 4-epimerase (Q14376) |
| 11 | Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma (Q8TBX8) | Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma (Q8TBX8) |
| 12 | Heterogeneous nuclear ribonucleoprotein M (P52272) | PRP4 (Q16378) |
| 13 | Psoriasin (P31151) | Protein S100-A8 (P05109) |
| 14 | ADH7 (P40394) | SH2 domain-containing protein 4A (Q9H788) |
| 15 | PRP4 (Q16378) | Malate dehydrogenase, cytoplasmic (P40925) |

Table 2. Feature importances for the classifier predicting three levels of MGD from relative protein quantifications using SHAP weighted by model uncertainty. **Bold font:** Not among the original top ranked features. BL, borderline predictions.

| Rank | Protein (accession number), incl. BL | Protein (accession number), excl. BL |
|------|--|--|
| 1 | Tissue alpha-L-fucosidase (P04066) | Tissue alpha-L-fucosidase (P04066) |
| 2 | Mannose-1-phosphate guanyltransferase alpha (Q96IJ6) | Mannose-1-phosphate guanyltransferase alpha (Q96IJ6) |
| 3 | UDP-glucose 4-epimerase (Q14376) | UDP-glucose 4-epimerase (Q14376) |
| 4 | IGKV2-24 (A0A0C4DH68) | Protein S100-A8 (P05109) |
| 5 | Polymeric immunoglobulin receptor (P01833) | Polymeric immunoglobulin receptor (P01833) |
| 6 | Protein S100-A8 (P05109) | IGKV2-24 (A0A0C4DH68) |
| 7 | CYFIP-related Rac1 interactor B (Q9NUQ9) | SH2 domain-containing protein 4A (Q9H788) |
| 8 | 60S acidic ribosomal protein P0 (P05388) | Adenylate kinase isoenzyme 1 (P00568) |
| 9 | Adenylate kinase isoenzyme 1 (P00568) | 60S acidic ribosomal protein P0 (P05388) |
| 10 | SH2 domain-containing protein 4A (Q9H788) | Lactoylglutathione lyase (Q04760) |
| 11 | Lactoylglutathione lyase (Q04760) | CYFIP-related Rac1 interactor B (Q9NUQ9) |
| 12 | IGA2 (P0DOX2) | IGA2 (P0DOX2) |
| 13 | Dynactin subunit 2 (Q13561) | Lysozyme C (P61626) |
| 14 | Lysozyme C (P61626) | PRP4 (Q16378) |
| 15 | Alpha-centractin (P61163) | Serotransferrin (P02787) |

Table 3. Feature importances for the classifier predicting MGD level 2 from relative protein quantifications using SHAP weighted by model uncertainty. **Bold font:** Not among the original top ranked features. BL, borderline predictions.

highlighted with bold font in the tables. For all the models, the feature rankings changed when the SHAP values were weighted. The relative quantifications in MGD levels 2 to 4 for the proteins are included in Supplementary Table A.1. Considering the multiclass classifier predicting all three levels of MGD, four of the top 15 most important features represented proteins that differ from the ranking with the original SHAP values. Three of these four proteins stood out as particularly interesting with respect to DED: glutathione peroxidase 1^{37–39}, psoriasin and ADH7. For the classifier predicting MGD level 2, dynactin subunit 2, being a part of a complex that is involved in the secretion of proteins from cells in the lacrimal glands⁴⁰, was one of the new highly ranked proteins. Regarding the MGD level 3 model, no new proteins were identified using weighted SHAP values, while for the MGD level 4 classifier, none of the new highest ranked proteins were found to be related to MGD or DED.

Label-free quantification using PEAKS

According to the PEAKS analysis of the protein quantifications for the patients with MGD levels 2–4, 13 proteins were found to be significantly different between the three levels of MGD. The proteins are listed in Supplementary Table A.2. Among these, thymidine phosphorylase has been associated with DED^{41,42}. Supplementary Table A.2 also reports the changes in relative protein quantifications between the three different levels of MGD for the significant proteins.

| Rank | Protein (accession number), incl. BL | Protein (accession number), excl. BL |
|------|--|--|
| 1 | Ubiquitin-conjugating enzyme E2 variant 1 (Q13404) | Ubiquitin-conjugating enzyme E2 variant 1 (Q13404) |
| 2 | Endoplasmic reticulum resident protein 29 (P30040) | Endoplasmic reticulum resident protein 29 (P30040) |
| 3 | Malate dehydrogenase, cytoplasmic (P40925) | Malate dehydrogenase, cytoplasmic (P40925) |
| 4 | Glutathione peroxidase 1 (P07203) | Prostaglandin reductase 1 (Q14914) |
| 5 | Cathepsin D (P07339) | Glutathione peroxidase 1 (P07203) |
| 6 | 60S ribosomal protein L38 (P63173) | Mitogen-activated protein kinase 1 (P28482) |
| 7 | Prostaglandin reductase 1 (Q14914) | 60S ribosomal protein L38 (P63173) |
| 8 | IGG1 (P0DOX5) | Cathepsin D (P07339) |
| 9 | Mitogen-activated protein kinase 1 (P28482) | IGLV8-61 (A0A075B610) |
| 10 | Peptidyl-prolyl cis-trans isomerase B (P23284) | IGG1 (P0DOX5) |
| 11 | IGLV6-57 (P01721) | Peptidyl-prolyl cis-trans isomerase B (P23284) |
| 12 | IGLV8-61 (A0A075B610) | IGLV6-57 (P01721) |
| 13 | IGKV2-24 (A0A0C4DH68) | ADH7 (P40394) |
| 14 | ADH7 (P40394) | IGKV2-24 (A0A0C4DH68) |
| 15 | Psoriasis (P31151) | Psoriasis (P31151) |

Table 4. Feature importances for the classifier predicting MGD level 3 from relative protein quantifications using SHAP weighted by model uncertainty. BL, borderline predictions.

| Rank | Protein (accession number), incl. BL | Protein (accession number), excl. BL |
|------|---|---|
| 1 | Heterogeneous nuclear ribonucleoprotein M (P52272) | Heterogeneous nuclear ribonucleoprotein M (P52272) |
| 2 | Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma (Q8TBX8) | Phosphatidylinositol 5-phosphate 4-kinase type-2 gamma (Q8TBX8) |
| 3 | Polymeric immunoglobulin receptor (P01833) | Polymeric immunoglobulin receptor (P01833) |
| 4 | Golgi membrane protein 1 (Q8NB4) | Golgi membrane protein 1 (Q8NB4) |
| 5 | Corticosteroid-binding globulin (P08185) | PRP4 (Q16378) |
| 6 | PRP4 (Q16378) | Immunoglobulin heavy variable 3-74 (A0A0B4J1X5) |
| 7 | Immunoglobulin heavy variable 3-74 (A0A0B4J1X5) | Corticosteroid-binding globulin (P08185) |
| 8 | Immunoglobulin heavy constant gamma 4 (P01861) | Mitogen-activated protein kinase 1 (P28482) |
| 9 | Malate dehydrogenase, cytoplasmic (P40925) | Malate dehydrogenase, cytoplasmic (P40925) |
| 10 | X-ray repair cross-complementing protein 5 (P13010) | Immunoglobulin heavy constant gamma 4 (P01861) |
| 11 | Mitogen-activated protein kinase 1 (P28482) | Eukaryotic translation initiation factor 6 (P56537) |
| 12 | Phospholipid transfer protein (P55058) | Methanethiol oxidase (Q13228) |
| 13 | Eukaryotic translation initiation factor 6 (P56537) | X-ray repair cross-complementing protein 5 (P13010) |
| 14 | Methanethiol oxidase (Q13228) | Phospholipid transfer protein (P55058) |
| 15 | Protein NDRG1 (Q92597) | Prostaglandin reductase 1 (Q14914) |

Table 5. Feature importances for the classifier predicting MGD level 4 from relative protein quantifications using SHAP weighted by model uncertainty. **Bold font:** Not among the original top ranked features. BL, borderline predictions.

Summary of the findings

To summarize, several of the most important features in each of the four ML models represented proteins that are known to be upregulated or downregulated in DED. One of these was protein S100-A8, which is upregulated in DED⁸. A boxplot of the relative quantifications of protein S100-A8 for the different MGD levels in our dataset is provided in Fig. 3a. It is observed from the plot that the protein levels were mostly stable, but for the higher MGD levels, the levels of protein S100-A8 were skewed towards higher values. The differences in protein quantifications were statistically significant (p value < 0.05) between MGD levels 2 and 3, while the differences were borderline significant between MGD levels 2 and 4 (p value = 0.052). PRP4 is downregulated in DED⁸. Figure 3b plots PRP4 quantifications for the MGD levels. Even though the plot shows a trend toward higher values for MGD levels 2 and 3 compared to level 4, the differences were not statistically significant. Since PRP4 has been found not to be significantly downregulated in lipid deficient dry eye, which includes MGD, this might explain why the levels were more stable in this plot⁴³. Finally, relative quantifications of lysozyme C, which is known to be downregulated in DED⁸, are plotted for different levels of MGD in Fig. 3c. The protein amounts tended to be higher in tears from patients with MGD levels 2 compared to level 3 and 4. However, no statistically significant changes were found. The significant protein differences for protein S100-A8 support that this proteins might play a central role in MGD. Still, because several of the MGD level groups expressed similar protein levels, it suggests that the ML models also identified patterns that go beyond alterations in single protein quantifications.

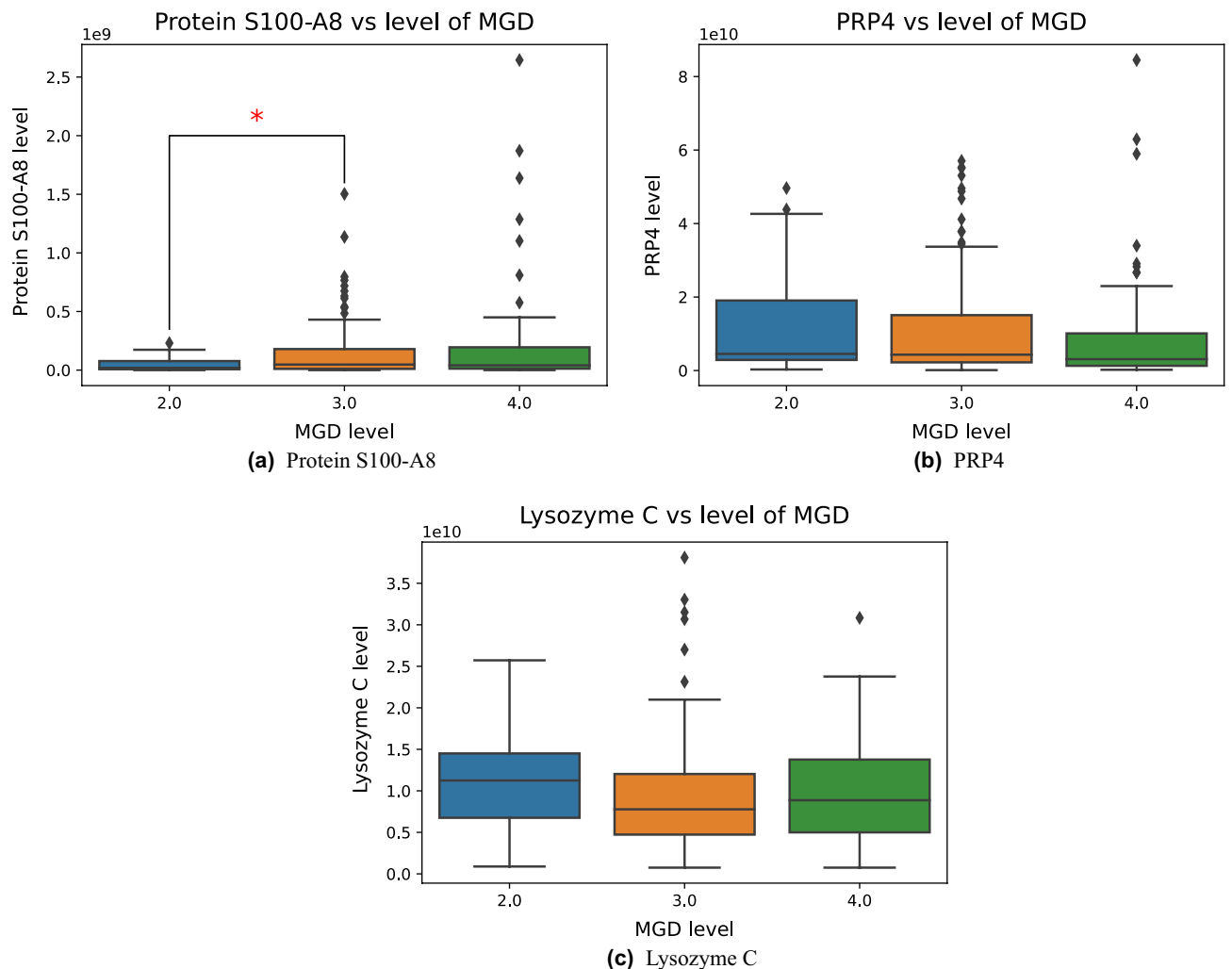


Figure 3. Relative quantifications of protein S100-A8, PRP4 and lysozyme C for different levels of MGD. Statistically significant changes in protein levels (p values < 0.05) are marked with *.

Corresponding fold changes in the relative quantifications of protein S100-A8, PRP4 and lysozyme C between MGD levels 2, 3 and 4 are included in Supplementary Table A.1.

Discussion

The work presented in this paper shows that ML models combined with the feature importance method SHAP are able to detect proteins that are relevant for DED and MGD. Because one aim of the study was to explore proteomic changes in subjects with increasing severity of disease, the models were trained on patients with MGD only. Among the top 15 ranked features in each of the four ML models, there were several proteins that are interesting with respect to DED and MGD. These proteins are PRP4, S100 calgranulin A8, lysozyme C, prostaglandine reductase 1, psoriasis, serotranferrin, ADH7, the polymeric immunoglobulin receptor, and the immunoglobulins IGKV2-24, IGLV8-61, IGLV6-57, IGA2 and IGG1. Their relationships to DED and MGD are discussed below.

Jackson et al.⁸ performed a literature review about proteins in DED and identified some proteins that are promising biomarkers for DED. Three of the proteins detected with the method proposed in the current study were found to be potential biomarkers that are upregulated or downregulated in DED according to the review: PRP4, S100 calgranulin A8 and lysozyme C. These three proteins are marked in Fig. 2a–d. Compared to healthy controls, the expression of PRP4 and S100 calgranulin A8 were significantly different in both aqueous-deficient and combined aqueous-deficient and evaporative DED, while lysozyme C was only significantly differentially expressed in combined aqueous-deficient and evaporative DED⁶.

Further on, prostaglandin reductase 1 is involved in the breakdown of prostaglandins⁴⁴. Prostaglandins trigger inflammatory responses on the ocular surface and were considered as a contributor to the pathogenesis of DED²⁷. Prostaglandin reductase 1 might therefore be a key protein with respect to DED and MGD.

Psoriasis, being associated with the immune system and inflammation, has been found to be present at high levels in MGs from individuals without any diseases affecting or involving the lacrimal glands³⁴. Moreover, the gene for psoriasis was upregulated in diseased MGs³⁵ as well as in saliva from patients with primary Sjögren

syndrome (pSS)³⁶. It should, however, be noted that all included subjects in the latter study were females. Taken together, psoriasin could be of value for the grading of MGD and serve as a potential biomarker for DED.

Serotransferrin binds and transports iron and exhibits antimicrobial effects by keeping iron unavailable to pathogens⁴⁵. MGD is associated with microbial infections. Serotransferrin was increased in aqueous deficient and combined aqueous deficient and evaporative DED⁶. Moreover, levels of serotransferrin increased with increasing age⁴⁶. Since the risk of experiencing DED also increases with increasing age, this indicates a potential relationship between serotransferrin and DED. Also considering its antimicrobial effects and that the protein levels were altered in several types of DED, serotransferrin might play a role in the pathogenesis of MGD.

ADH7 is part of a class IV alcohol dehydrogenase primarily involved in the oxidation of retinol to retinaldehyde and possibly retinoic acid synthesis^{47,48}. Retinoids are associated with the proliferation, differentiation, keratinization and apoptosis of corneal epithelial cells and deficiency of vitamin A can cause both DED and keratopathy³². A metabolite of vitamin A, 13-cis retinoic acid, is known to cause MGD and has been shown to alter gene expression related to inflammation and differentiation as well as inducing apoptosis of MG epithelial cells *in vitro*³³.

The polymeric immunoglobulin receptor was significantly downregulated in patients with aqueous-deficient DED compared to individuals without DED²³. The protein was also decreased in tears from rabbits with Sjögren syndrome-associated dry eye²⁴. Moreover, the protein was suppressed in reflex tears, that are produced in response to irritant stimulations²⁵. Huang et al. found, on the contrary, that the polymeric immunoglobulin receptor was upregulated in tears from patients with DED²⁶. Despite contradicting findings, changes in the protein expression have been detected in different types of DED, suggesting that it could play a role in the pathogenesis of the disease. Further on, IGA2 is important for the adaptive immune response and binds to the polymeric immunoglobulin receptor²⁹. Amorim et al.²⁸ found the levels of IGKV2-24 to be significantly higher in tears from patients with proliferative diabetic retinopathy compared to non-diabetic controls. In this study, 74% of the proliferative diabetic retinopathy patients exhibited Schirmer test results < 10 mm/5 min, and tear film breakup-times were significantly reduced compared to the controls. In our dataset, FBUT was significantly lower for patients with MGD level 4 compared to MGD levels 2 and 3. Consequently, the importance of IGKV2-24 in our models could also be affected by differences in FBUT between the patient groups. Still, considering their roles in the immune response and the fact that MGD facilitates microbial infections in the eye⁴⁹, IGKV2-24's, IGA2's and the polymeric immunoglobulin receptor's roles in MGD could be investigated further.

Both IGLV8-61 and IGLV6-57 are immunoglobulins involved in the immune response. In a study investigating tear proteomics following laser-assisted in-situ keratomileusis (LASIK) and small incision lenticule extraction (SMILE) surgery, IGLV8-61 and IGLV6-57 were both upregulated one week after LASIK surgery³⁰. The patients had a mean tear film break-up time of 6.1 and Schirmer test of approximately 9 mm/5 min, which are clinical signs of DED. Even though the observed alterations in IGLV8-61 and IGLV6-57 probably arose from the surgery, they could also be connected to dry eye-related pathology.

IGG1, also involved in the adaptive immune response, could play a role in the development of dry eyes. Mackie et al. reported increased tear levels of immunoglobulin gamma in patients with DED compared to healthy controls³¹. A study comparing allergic conjunctivitis in mice with and without DED observed significantly increased levels of IGG1 regardless of DED, suggesting that DED did not affect the protein expression⁵⁰. However, since allergic conjunctivitis activate the adaptive immune response, this might mask possible alterations during DED. A reduction of IGG1 in Descemet's membrane, which is a part of the cornea, was observed for patients with Fuchs endothelial corneal dystrophy⁵¹. Taken together, alterations of IGG1 expression are observed for several diseases on the ocular surface, including DED.

When weighting the SHAP importance values by the model uncertainty, two additional proteins that are potentially important for MGD were identified. These proteins were glutathione peroxidase 1 and dynactin subunit 2. Moreover, psoriasin, IGG1 and ADH7 were ranked among the top 15 features for the multiclass model, which was not the case without uncertainty weighting. The relevance of glutathione peroxidase 1 and dynactin subunit 2 with respect to MGD and DED is discussed below.

Studies have shown that oxidative stress plays a role in the mechanism of DED⁵². Glutathione peroxidase 1 controls the levels of reactive oxygen species (ROS) and is important for the normal function of the MGs³⁷. Hyperosmolarity, which is often observed with DED and pSS are both associated with reduced levels of this protein^{38,39}. Consequently, glutathione peroxidase 1 stands out as a potentially promising biomarker for MGD.

Dynactin subunit 2 is a part of the dynactin complex, which binds to and activates dynein. Further on, this facilitates exocytosis from acinar cells, which reside in the lacrimal glands and secrete proteins onto the ocular surface⁴⁰. Studies of mouse models for pSS indicate that the protein secretion from acinar cells are altered independently of inflammatory responses^{53,54}. Future research should therefore look into whether human lacrimal gland secretion is affected by changes in levels of the dynactin subunit 2 and whether such changes are associated with MGD and DED.

Regarding the significant proteins detected by the PEAKS X Pro software, thymidine phosphorylase is potentially relevant for DED and MGD. Thymidine phosphorylase is associated with endothelial cell survival and function⁴¹. Loss-of-function has been reported to give dry eyes in a case report, although as part of a complex systemic clinical picture⁴². Further research is required to investigate the potential role of thymidine phosphorylase in MGD.

Remarkably, only one of the significant proteins detected by PEAKS Pro X, mannose-1-phosphate guanylyltransferase alpha, was also found by the feature importance approach. Consequently, it seems like these two approaches complement each other, and applying both of them can result in a higher number of identified potential biomarkers for diseases.

When using the feature importance approaches, most of the proteins identified as promising biomarkers for MGD are involved in the immune response and inflammation. Other proteins, such as PRP4, ADH7 and

dynactin subunit 2, contribute to the normal function of the eye. These findings confirm that MGD is a complex condition associated with disturbance of processes in the healthy eye and activation of several inflammatory and immunologic pathways. There is most likely not one single biomarker, but rather a panel of biomarkers, that characterize MGD.

Even though several of the important features identified in this study represented proteins that were previously found to be related to DED, the majority of the features did not. Highly ranked proteins such as mannose-1-phosphate guanyltransferase alpha, 60S ribosomal proteins L6 and L13 and corticosteroid-binding globulin have not been extensively studied with respect to DED and MGD. However, mannose-1-phosphate guanyltransferase alpha is associated with disease in general, and the other proteins are involved in the regulation of inflammation. Investigating the relevance of these and other highly ranked proteins with respect to DED and MGD might give rise to novel medical discoveries.

Some of the detected proteins that are potential biomarkers for MGD, for example psoriasin and the polymeric immunoglobulin receptor, are altered in pSS. It is not unreasonable that some proteins can play key roles in both MGD and pSS, especially because pSS has dry eyes as a primary symptom and has been associated with MGD^{55,56}. Still, the most upregulated tear protein in pSS, neutrophil gelatinase-associated lipocalin⁵⁷, was not among the detected proteins in the current study. One possible explanation is that patients with MGD make up a more diverse group, also including patients without pSS. Even though MGD and pSS can occur simultaneously, these are two separate diagnoses, and it is expected that there are differences in how they affect the ocular surface and tear composition.

The feature importance plots differed between the MGD models. This is not surprising, as we expect different protein levels to be of higher or lower importance according to the severity of the MGD. Because MGD is associated with inflammation, one would for example expect proteins related to inflammation to be ranked differently for the different levels of MGD. Still, many of the same proteins were highly ranked for all models. This means that there were also similarities between the different levels of MGD.

The reliability of the estimated feature importance values will be affected if the ML models are not performing well. In this work, the balanced accuracy for the multiclass ML model was 72%, while the balanced accuracies for the three binary ML models were 85% or higher. The high model performances indicate that the predicted important features should be robust. At the moment, there exists no method that takes the model uncertainty into account when determining the feature importance. In the current work, the estimated importance values were weighted based on the predicted probabilities as a mean of including uncertainty into the feature ranking. When the ground truth annotations are available, the importance values for correct predictions with high certainty from the model can be assigned more weights than incorrect and/or uncertain predictions. By following this, a change in the feature ranking was observed, and new potentially MGD-related proteins were identified. However, when the ground truth is not available, the correct model predictions cannot be identified. In this case, only the original importance values for the most certain model predictions were included. Using this technique, no new proteins were present among the 15 most important features. This is probably due to the low number of uncertain predictions. The results indicate that the weighting is most effective when the ground truth is known.

The dataset applied in this study has several strengths. First, it includes a high number of patients. Moreover, the comprehensive proteomic analyses provide detailed information about the tear composition for the patients and were combined with several clinical parameters including level of MGD. Finally, because all patients were examined at the same eye clinic, variations between the performance of the various procedures are expected to be minimal. Taken together, these strengths are adding reliability to the reported results.

This study has some potential limitations. First, the ML models were not evaluated using an external test set. As a result, the models' abilities to generalize to new patients are not known. The main reason for not dividing the dataset into training and test sets was because the aim was to describe patterns in the data rather than developing ML models for predictive purposes. This approach is motivated by a method called 'microscope artificial intelligence (AI)', where models are developed and explained in order to gain a deeper understanding of the data used to train the models^{58,59}. Similar to applying a microscope for studying our surroundings in more detail, explaining the ML model can enable us to view the data from a different angle, potentially leading to new discoveries. According to microscope AI, the goal is to extract knowledge out of the data rather than creating models for automatic decision-making⁵⁸. Still, the ML models from the current work should not be used for diagnostic purposes since they have not been externally evaluated.

Another potential limitation is the chosen tear collection method. The tears in the present study were collected using Schirmer strips, adhering to previous protocols presented by our group^{57,60}. However, the optimal methodology for collecting tear samples remains an area of debate. Collection methods such as microcapillary tubes with or without saline flush, Schirmer strips as well as the intrinsic individual variability may contribute to both composition and concentration of collected proteins. Some studies found similar results regarding protein expression following collection with microcapillary tubes and Schirmer strips^{61,62}, while another study indicates that the collection method might impact the proteins detected in the sample⁶³. In the present project, unanesthetized Schirmer tests were conducted, which can stimulate reflex tearing. Indeed, at least 15 proteins have been noted to differ between reflex and basal tears²⁵, where basal tears might be regarded as more relevant when studying MGD. However, it can be argued that the previous studies mentioned above indicate that the applied collection method using Schirmer strips ensures a representative collection of biological material as well as a low degree of contamination. Moreover, although an anesthetized Schirmer test might be more representative concerning basal tears, the topical anesthesia might serve as a contaminant and alter tear composition.

DED is broadly divided into aqueous deficient DED and evaporative DED⁶⁴. The latter is the more common of the two and MGD is the most common cause of evaporative DED. These subdivisions of DED are not mutually exclusive. Instead, they often overlap along a spectrum referred to as mixed DED, which is commonly seen in the clinic. In the MGD dataset used in the current work, some of the patients with MGD also exhibit signs of

aqueous deficient DED. This might be a potential limitation because some proteins have been noted to be altered as a result of lacrimal tear production⁶⁵, which further on can give rise to differences in protein measurements between individuals with and without aqueous deficient DED. On the other hand, the fraction of patients with Schirmer test values < 5 mm was relatively low and stable for each of the MGD levels 2 to 4, ranging between 16 and 22%. Consequently, if the protein measurements were affected due to reduced tear production, it would most likely have similar effects across all these MGD levels. The inclusion of patients with mixed DED probably had little effect on the reported results. Still, it would be interesting to conduct studies specifically targeting tear proteomic patterns in patients with mixed aqueous deficient DED and evaporative DED. While the dataset in the present work only includes 46 out of 233 patients with mixed aqueous deficient DED and MGD, data collection from a larger cohort is necessary to get reliable results for this subgroup of DED patients. Future work should look into this topic.

This study explored which proteins are regarded as important for ML models predicting different levels of MGD. Several of the detected features represented proteins that from earlier research are known to be altered in DED. What might seem unusual with the presented work is that a control group without MGD was not included for training the ML models. However, the aim was not to distinguish healthy eyes from eyes diagnosed with MGD, but rather to explore the expression of proteins in tears for different levels of MGD. Individuals without MGD were not regarded as relevant to include in the training dataset because they might obscure the proteomic differences within the MGD patient group. However, future work should look into how the current results compare to a control group without MGD. Investigating the protein rankings in healthy controls can strengthen the findings in the present study.

In conclusion, this study successfully combined ML and proteomics to explore relationships between MGD severity and protein expression in tears. Examination of which proteins the ML models regarded as important for predicting levels of MGD showed that several of the highest ranked proteins are known to be upregulated or downregulated in DED. Other proteins that might be relevant in MGD were also detected. Future work should explore these proteins further with the aim to increase the understanding of MGD and develop improved treatment options. Moreover, the proposed method for detecting potentially relevant proteins could be applied to other types of medical conditions and diseases with the aim to discover new medical knowledge.

Data availability

The data is currently not publicly available since it contains sensitive patient data. The interested reader is encouraged to contact the corresponding author if access to the data is needed.

Code availability

The code is made publicly available and can be accessed through the following link: <https://github.com/Andre aStoraas/MGD-proteins-XAI>.

Received: 14 July 2023; Accepted: 19 December 2023

Published online: 22 December 2023

References

- Nichols, K. K. *et al.* The international workshop on meibomian gland dysfunction: Executive summary. *Investig. Ophthalmol. Vis. Sci.* **52**, 1922–1929. <https://doi.org/10.1167/iovs.10-6997a> (2011).
- Stapleton, F. *et al.* TFOS DEWS II epidemiology report. *Ocul. Surf.* **15**, 334–365. <https://doi.org/10.1016/j.jtos.2017.05.003> (2017).
- Morthen, M. K. *et al.* The physical and mental burden of dry eye disease: A large population-based study investigating the relationship with health-related quality of life and its determinants. *Ocul. Surf.* **21**, 107–117. <https://doi.org/10.1016/j.jtos.2021.05.006> (2021).
- Yu, J., Asche, C. V. & Fairchild, C. J. The economic burden of dry eye disease in the United States: A decision tree analysis. *Cornea* **30**, 379–387. <https://doi.org/10.1097/ICO.0b013e3181f7f363> (2011).
- Tong, L., Zhou, L., Beuerman, R. W., Zhao, S. Z. & Li, X. R. Association of tear proteins with Meibomian gland disease and dry eye symptoms. *Br. J. Ophthalmol.* **95**, 848–852. <https://doi.org/10.1136/bjo.2010.185256> (2011).
- Perumal, N., Funke, S., Pfeiffer, N. & Grus, F. H. Proteomics analysis of human tears from aqueous-deficient and evaporative dry eye patients. *Sci. Rep.* **6**, 1–12. <https://doi.org/10.1038/srep29629> (2016).
- Zhou, L. *et al.* In-depth analysis of the human tear proteome. *J. Proteom.* **75**, 3877–3885. <https://doi.org/10.1016/j.jprot.2012.04.053> (2012).
- Jackson, C. J., Gundersen, K. G., Tong, L. & Utheim, T. P. Dry eye disease and proteomics. *Ocul. Surf.* **24**, 119–128. <https://doi.org/10.1016/j.jtos.2022.03.001> (2022).
- Ting, D. S. W. *et al.* Artificial intelligence and deep learning in ophthalmology. *Br. J. Ophthalmol.* **103**, 167–175. <https://doi.org/10.1136/bjophthalmol-2018-313173> (2019).
- Storås, A. M. *et al.* Artificial intelligence in dry eye disease. *Ocul. Surf.* **23**, 74–86. <https://doi.org/10.1016/j.jtos.2021.11.004> (2022).
- Fineide, F. *et al.* Predicting an unstable tear film through artificial intelligence. *Sci. Rep.* **12**, 21416. <https://doi.org/10.1038/s41598-022-25821-y> (2022).
- Aqrawi, L. A. *et al.* Proteomic and histopathological characterisation of sicca subjects and primary Sjögren's syndrome patients reveals promising tear, saliva and extracellular vesicle disease biomarkers. *Arthritis Res. Ther.* <https://doi.org/10.1186/s13075-019-1961-4> (2019).
- Hynne, H. *et al.* Proteomic profiling of saliva and tears in radiated head and neck cancer patients as compared to primary Sjögren's syndrome patients. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms23073714> (2022).
- Bron, A. J. *et al.* Methodologies to diagnose and monitor dry eye disease: Report of the diagnostic methodology subcommittee of the International Dry Eye Workshop (2007). *Ocul. Surf.* **5**, 108–152. [https://doi.org/10.1016/S1542-0124\(12\)70083-6](https://doi.org/10.1016/S1542-0124(12)70083-6) (2007).
- Pult, H. & Riede-Pult, B. Non-contact meibography: Keep it simple but effective. *Contact Lens Anterior Eye* **35**, 77–80. <https://doi.org/10.1016/j.clae.2011.08.003> (2012).
- Pult, H. & Riede-Pult, B. H. An assessment of subjective and objective grading of meibography image. *Investig. Ophthalmol. Vis. Sci.* **53**, 588 (2012).

17. Ke, G. *et al.* LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems* Vol. 30 (eds Guyon, I. *et al.*) (Curran Associates, Inc., 2017).
18. Lundberg, S. M. *et al.* Explainable ai for trees: From local explanations to global understanding. <https://doi.org/10.48550/arXiv.1905.04610> (2019). arXiv:1905.04610.
19. Shapley, L. S. A value for n-person games. In *Contributions to the Theory of Games (AM-28), Volume II* (1953).
20. Huettner, F. & Sunder, M. Axiomatic arguments for decomposing goodness of fit according to Shapley and Owen values. *Electron. J. Stat.* **6**, 1239–1250. <https://doi.org/10.1214/12-EJS710> (2012).
21. Lundberg, S. M., Erion, G. G. & Lee, S.-I. Consistent individualized feature attribution for tree ensembles. <https://doi.org/10.48550/ARXIV.1802.03888> (2018).
22. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
23. Srinivasan, S., Thangavelu, M., Zhang, L., Green, K. B. & Nichols, K. K. iTRAQ quantitative proteomics in the analysis of tears in dry eye patients. *Investig. Ophthalmol. Vis. Sci.* **53**, 5052–5059. <https://doi.org/10.1167/iovs.11-9022> (2012).
24. Zhou, L. *et al.* Proteomic analysis revealed the altered tear protein profile in a rabbit model of Sjögren's syndrome-associated dry eye. *Proteomics* **13**, 2469–2481. <https://doi.org/10.1002/pmic.201200230> (2013).
25. Perumal, N., Funke, S., Wolters, D., Pfeiffer, N. & Grus, F. H. Characterization of human reflex tear proteome reveals high expression of lacrimal proline-rich protein 4 (prp4). *Proteomics* **15**, 3370–3381. <https://doi.org/10.1002/pmic.201400239> (2015).
26. Huang, Z., Du, C.-X. & Pan, X.-D. The use of in-strip digestion for fast proteomic analysis on tear fluid from dry eye patients. *PLoS ONE* **13**, 1–12. <https://doi.org/10.1371/journal.pone.0200702> (2018).
27. Nebbioso, M. *et al.* Analysis of the pathogenic factors and management of dry eye in ocular surface disorders. *Int. J. Mol. Sci.* **18**, 1764. <https://doi.org/10.3390/ijms18081764> (2017).
28. Amorim, M. *et al.* Putative biomarkers in tears for diabetic retinopathy diagnosis. *Front. Med.* <https://doi.org/10.3389/fmed.2022.873483> (2022).
29. Pilette, C., Ouadrhiri, Y., Godding, V., Vaerman, J.-P. & Sibille, Y. Lung mucosal immunity: Immunoglobulin-A revisited. *Eur. Respir. J.* **18**, 571–588. <https://doi.org/10.1183/09031936.01.00228801> (2001).
30. Liu, Y.-C. *et al.* Comparison of tear proteomic and neuromediator profiles changes between small incision lenticule extraction (SMILE) and femtosecond laser-assisted in-situ keratomileusis (LASIK). *J. Adv. Res.* **29**, 67–81. <https://doi.org/10.1016/j.jare.2020.11.001> (2021).
31. Mackie, I. A. & Seal, D. V. Diagnostic implications of tear protein profiles. *Br. J. Ophthalmol.* **68**, 321–324. <https://doi.org/10.1136/bjo.68.5.321> (1984).
32. Latta, L. *et al.* Similarities in DSG1 and KRT3 downregulation through retinoic acid treatment and PAX6 knockdown related expression profiles: Does PAX6 affect RA signaling in limbal epithelial cells?. *Biomolecules* **11**, 1651. <https://doi.org/10.3390/biom11111651> (2021).
33. Ding, J., Kam, W. R., Dieckow, J. & Sullivan, D. A. The influence of 13-cis retinoic acid on human meibomian gland epithelial cells. *Investig. Ophthalmol. Vis. Sci.* **54**, 4341–4350. <https://doi.org/10.1167/iovs.13-11863> (2013).
34. Garreis, F. *et al.* Expression and regulation of antimicrobial peptide Psoriasin (S100A7) at the ocular surface and in the lacrimal apparatus. *Investig. Ophthalmol. Vis. Sci.* **52**, 4914–4922. <https://doi.org/10.1167/iovs.10-6598> (2011).
35. Liu, S. *et al.* Changes in gene expression in human meibomian gland dysfunction. *Investig. Ophthalmol. Vis. Sci.* **52**, 2727–2740. <https://doi.org/10.1167/iovs.10-6482> (2011).
36. Baldini, C. *et al.* Proteomic analysis of saliva: A unique tool to distinguish primary Sjögren's syndrome from secondary Sjögren's syndrome and other sicca syndromes. *Arthritis Res. Ther.* **13**, 1–16. <https://doi.org/10.1186/ar3523> (2011).
37. Nezzar, H. *et al.* Investigation of antioxidant systems in human meibomian gland and conjunctival tissues. *Exp. Eye Res.* **165**, 99–104. <https://doi.org/10.1016/j.exer.2017.09.005> (2017).
38. Deng, R. *et al.* Oxidative stress markers induced by hyperosmolarity in primary human corneal epithelial cells. *PLoS ONE* **10**, 1–16. <https://doi.org/10.1371/journal.pone.0126561> (2015).
39. Čejková, J. *et al.* Decreased expression of antioxidant enzymes in the conjunctival epithelium of dry eye (Sjögren's syndrome) and its possible contribution to the development of ocular surface oxidative injuries. *Histol. Histopathol.* **23**, 1477–1483 (2008).
40. Wu, K. *et al.* Molecular mechanisms of lacrimal acinar secretory vesicle exocytosis. *Exp. Eye Res.* **83**, 84–96. <https://doi.org/10.1016/j.exer.2005.11.009> (2006).
41. Baris, A., Fraile-Bethencourt, E., Eubanks, J., Khou, S. & Anand, S. Thymidine phosphorylase facilitates retinoic acid inducible gene-1 induced endothelial dysfunction. *Cell Death Dis.* **14**, 294. <https://doi.org/10.1038/s41419-023-05821-0> (2023).
42. Blázquez, A. *et al.* Increased muscle nucleoside levels associated with a novel frameshift mutation in the thymidine phosphorylase gene in a Spanish patient with MNGIE. *Neuromuscul. Disord.* **15**, 775–778. <https://doi.org/10.1016/j.nmd.2005.07.008> (2005).
43. Perumal, N., Funke, S., Wolters, D., Pfeiffer, N. & Grus, F. H. In-depth protein profiling and identification of tear fluid biomarkers in different subgroups of dry eye disease: Proline-rich protein 4 (prp4) as a potential biomarker for aqueous-deficient dry eye syndrome. *Investig. Ophthalmol. Vis. Sci.* **55**, 2002 (2014).
44. Wang, X. *et al.* Prostaglandin reductase 1 as a potential therapeutic target for cancer therapy. *Front. Pharmacol.* <https://doi.org/10.3389/fphar.2021.717730> (2021).
45. La, O. A. A. & Brock, J. H. Iron and transferrin. Research and therapeutic applications. *Biotechnol. Appl.* **18**, 1–9 (2001).
46. Näntinen, J. *et al.* Age-associated changes in human tear proteome. *Clin. Proteom.* <https://doi.org/10.1186/s12014-019-9233-5> (2019).
47. National Library of Medicine. *ADH7 alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide [Homo sapiens (human)]*, accessed 02 June 2023; <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=131#gene-expression> (2023).
48. Allali-Hassani, A., Peralba, J. M., Martras, S., Farrés, J. & Parés, X. Retinoids, ω -hydroxyfatty acids and cytotoxic aldehydes as physiological substrates, and H2-receptor antagonists as pharmacological inhibitors, of human class IV alcohol dehydrogenase. *FEBS Lett.* **426**, 362–366. [https://doi.org/10.1016/S0014-5793\(98\)00374-3](https://doi.org/10.1016/S0014-5793(98)00374-3) (1998).
49. Baudouin, C. *et al.* Revisiting the vicious circle of dry eye disease: A focus on the pathophysiology of meibomian gland dysfunction. *Br. J. Ophthalmol.* **100**, 300–306. <https://doi.org/10.1136/bjophthalmol-2015-307415> (2016).
50. Kishimoto, T., Ishida, W., Nakajima, I., Fukuda, K. & Yamashiro, K. Aqueous-deficient dry eye exacerbates signs and symptoms of allergic conjunctivitis in mice. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms23094918> (2022).
51. Kuot, A. *et al.* Reduced expression of apolipoprotein E and immunoglobulin heavy constant gamma 1 proteins in Fuchs endothelial corneal dystrophy. *Clin. Exp. Ophthalmol.* **47**, 1028–1042. <https://doi.org/10.1111/ceo.13569> (2019).
52. Dogru, M., Kojima, T., Simsek, C. & Tsubota, K. Potential role of oxidative stress in ocular surface inflammation and dry eye disease. *Investig. Ophthalmol. Vis. Sci.* **59**, DES163–DES168. <https://doi.org/10.1167/iovs.17-23402> (2018).
53. Robinson, C. P., Yamamoto, H., Peck, A. B. & Humphreys-Beher, M. G. Genetically programmed development of salivary gland abnormalities in the nod (nonobese diabetic)-scidmouse in the absence of detectable lymphocytic infiltration: A potential trigger for sialoadenitis of nod mice. *Clin. Immunol. Immunopathol.* **79**, 50–59. <https://doi.org/10.1006/clin.1996.0050> (1996).
54. da Costa, S. R. *et al.* Male NOD mouse external lacrimal glands exhibit profound changes in the exocytotic pathway early in postnatal development. *Exp. Eye Res.* **82**, 33–45. <https://doi.org/10.1016/j.exer.2005.04.019> (2006).
55. Ramos-Casals, M., Brito-Zeron, P., Siso-Almirall, A., Bosch, X. & Tzioufas, A. G. Topical and systemic medications for the treatment of primary Sjögren's syndrome. *Nat. Rev. Rheumatol.* **8**, 399–411. <https://doi.org/10.1038/nrrheum.2012.53> (2012).

56. Sullivan, D. A. *et al.* Meibomian gland dysfunction in primary and secondary Sjögren syndrome. *Ophthalmic Res.* **59**, 193–205. <https://doi.org/10.1159/000487487> (2018).
57. Aqrabi, L. A. *et al.* Identification of potential saliva and tear biomarkers in primary Sjögren's syndrome, utilising the extraction of extracellular vesicles and proteomics analysis. *Arthritis Res. Ther.* **19**, 1–15. <https://doi.org/10.1186/s13075-017-1228-x> (2017).
58. Hubinger, E. An overview of 11 proposals for building safe advanced AI. <https://doi.org/10.48550/arXiv.2012.07532> (2020). arXiv: 2012.07532.
59. Hubinger, E. *Chris Olah's views on AGI safety*, accessed 18 October 2023; <https://www.alignmentforum.org/posts/X2i9dQK3gETCYqh2/chris-olah-s-views-on-agi-safety> (2019).
60. Fineide, F. *et al.* Characterization of lipids in saliva, tears and minor salivary glands of Sjögren's syndrome patients using an HPLC/MS-based approach. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms22168997> (2021).
61. Grus, F. H. *et al.* SELDI-TOF-MS ProteinChip array profiling of tears from patients with dry eye. *Investig. Ophthalmol. Vis. Sci.* **46**, 863–876. <https://doi.org/10.1167/iovs.04-0448> (2005).
62. Posa, A. *et al.* Schirmer strip vs. capillary tube method: Non-invasive methods of obtaining proteins from tear fluid. *Ann. Anat. Anat. Anz.* **195**, 137–142. <https://doi.org/10.1016/j.aanat.2012.10.001> (2013).
63. Green-Church, K. B., Nichols, K. K., Kleinholz, N. M., Zhang, L. & Nichols, J. J. Investigation of the human tear film proteome using multiple proteomic approaches. *Mol. Vis.* **14**, 456–470 (2008).
64. Craig, J. P. *et al.* TFOS DEWS II definition and classification report. *Ocul. Surf.* **15**, 276–283. <https://doi.org/10.1016/j.jtos.2017.05.008> (2017) (TFOS International Dry Eye WorkShop (DEWS II)).
65. Willcox, M. D. *et al.* TFOS DEWS II tear film report. *Ocul. Surf.* **15**, 366–403. <https://doi.org/10.1016/j.jtos.2017.03.006> (2017) (TFOS International Dry Eye WorkShop (DEWS II)).

Author contributions

A.M.S., F.F. and M.A.R. conceptualized and designed the study. A.M.S., M.M., B.T. and X.C. analyzed and processed the data. A.M.S. implemented and performed the machine learning and model explanation experiments. A.M.S., F.F., M.M., X.C., H.G., J.L.J. and T.P.U. interpreted and discussed the results. A.M.S. drafted the manuscript, which was critically revised by F.F., M.M., B.T., X.C., I.S., M.A.R., P.H., H.G., J.L.J. and T.P.U.. All authors read and approved the final manuscript and had final responsibility for the decision to submit for publication.

Competing interests

TPU is co-founder and co-owner of The Norwegian dry eye clinic and the Clinic of eye health, Oslo, Norway, which delivers talks for and/or receives financial support from the following: ABIGO, Alcon, Allergan, AMWO, Bausch & Lomb, Bayer, European school for advanced studies in ophthalmology, InnZ Medical, Medilens Nordic, Medistim, Novartis, Santen, Specsavers, Shire Pharmaceuticals and Thea Laboratories. He has served on the global scientific advisory board for Novartis and Alcon as well as the European advisory board for Shire Pharmaceuticals. Utheim is the Norwegian Global Ambassador for Tear Film and Ocular Surface Society (TFOS), a Board Member of the International Ocular Surface Society, an International Member of the Japanese Lid and Meibomian gland working group (LIME), a Consultant at the Norwegian Association for the Blind and Partially Sighted, the President of the Oslo Society of ophthalmology, and the Editor-in-Chief of *Oftalmolog*, an eye journal distributed to all eye doctors in the Nordic region since 1980. Besides publishing articles of presumed interest to our readers, *Oftalmolog* publishes advertisements from pharmaceutical companies, companies selling ophthalmological equipment, and associations organizing conferences and events in ophthalmology. For more information, visit: oftalmolog.com. All other authors declare no financial or non-financial competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-50342-7>.

Correspondence and requests for materials should be addressed to A.M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023