



## Data Article

# Dataset-chemokines, cytokines, and biomarkers in the saliva of children with Sjögren's syndrome



Miyuraj Harishchandra Hikkaduwa Withanage<sup>a</sup>,  
M. Paula Gomez Hernandez<sup>b</sup>, Emily E. Starman<sup>c</sup>, Andrew B. Davis<sup>d</sup>,  
Erliang Zeng<sup>a</sup>, Scott M. Lieberman<sup>e</sup>, Kim A. Brogden<sup>b</sup>,  
Emily A. Lanzel<sup>f,\*</sup>

<sup>a</sup> Division of Biostatistics and Computational Biology, College of Dentistry, University of Iowa, Iowa City, IA United States

<sup>b</sup> Pediatric Dentistry, College of Dentistry, University of Iowa, Iowa City, IA, United States

<sup>c</sup> College of Dentistry, Iowa Institute for Oral Health Research, University of Iowa, Iowa City, IA, United States

<sup>d</sup> Department of Otolaryngology, College of Medicine, University of Iowa, Iowa City, IA, United States

<sup>e</sup> Stead Family Department of Pediatrics-Division of Rheumatology, Allergy and Immunology, Carver College of Medicine, University of Iowa, Iowa City, IA, United States

<sup>f</sup> Department of Oral Pathology, Radiology and Medicine, College of Dentistry, University of Iowa, Iowa City, IA, United States

## ARTICLE INFO

## Article history:

Received 9 April 2021

Revised 30 April 2021

Accepted 5 May 2021

Available online 14 May 2021

## Keywords:

Sjögren's syndrome

Children

Saliva

Chemokines

Cytokines

Biomarkers

## ABSTRACT

Sjögren's syndrome is an autoimmune disease that can also occur in children. The disease is not well defined and there is limited information on the presence of chemokines, cytokines, and biomarkers (CCBMs) in the saliva of children that could improve their disease diagnosis. In a recent study [1], we reported a large dataset of 105 CCBMs that were associated with both lymphocyte and mononuclear cell functions [2] in the saliva of 11 children formally diagnosed with Sjögren's syndrome and 16 normal healthy children. Here, we extend those findings and use the Mendeley dataset [2] to identify CCBMs that have predictive power for Sjögren's syndrome in female children. Datasets of CCBMs from all saliva samples and female children saliva samples were standardized. We used machine learning methods to select Sjögren's syndrome associated CCBMs and assessed the

\* Corresponding author.

E-mail address: [emily-lanzel@uiowa.edu](mailto:emily-lanzel@uiowa.edu) (E.A. Lanzel).

predictive power of selected CCBMs in these two datasets using receiver operating characteristic (ROC) curves and associated areas under curve (AUC) as metrics. We used eight classifiers to identify 16 datasets that contained from 2 to 34 CCBMs with AUC values ranging from 0.91 to 0.94.

© 2021 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Specifications Table

Subject	Immunology
Specific subject area	Sjögren's syndrome in children
Type of data	Figure and Table
How data were acquired	<ol style="list-style-type: none"> <li>1. Fluorescent microparticle-based immunoassays (Luminex Human Magnetic Assay, R&amp;D Systems, Minneapolis, MN)</li> <li>2. Luminex model 100 IS (Luminex, Austin, TX USA)</li> <li>3. xPonent v3.1 software (Luminex, Austin, TX)</li> <li>4. Milliplex Analyst v5.1 software (EMD Millipore, Billerica, MA)</li> </ol>
Data format	Raw and analyzed
Parameters for data collection	CCBM concentrations in saliva samples from children with Sjögren's syndrome and from healthy children of the same gender and age
Description of data collection	Saliva samples were collected from August 30, 2016 to May 23, 2017. CCBM data was collected from July 12, 2019 to August 16, 2019
Data source location	Institution: University of Iowa College of Dentistry City/Town: Iowa City Country: USA Location: 41.6628 (41°39'46"N), -91.5511 (91°33'4"W)
Data accessibility	Repository name: Mendeley Data Identification number: <a href="http://dx.doi.org/10.17632/yphm77tg24.1">http://dx.doi.org/10.17632/yphm77tg24.1</a> Direct URL to data: <a href="https://data.mendeley.com/datasets/yphm77tg24/1">https://data.mendeley.com/datasets/yphm77tg24/1</a>
Related research article	M.P. Gomez Hernandez, E.E. Starman, A.B. Davis, M.H. Hikkaduwa Withanage, E. Zeng, S.M. Lieberman, K.A. Brogden, E.A. Lanzel. A unique profile of chemokines, cytokines, and biomarkers in the saliva of children with Sjögren syndrome. <i>Rheumatology</i> (2021) 1–13. <a href="https://doi.org/10.1093/rheumatology/keab098">https://doi.org/10.1093/rheumatology/keab098</a>

## Value of the Data

- The Mendeley dataset is among the first to report 105 CCBMs that were associated with both lymphocyte and mononuclear cell functions in the saliva of children with Sjögren's syndrome.
- The dataset can be used to identify smaller groups of CCBMs that can serve as predictor biomarkers for Sjögren syndrome diagnosis in children.
- Receiver operating characteristic (ROC) curves and associated areas under curve (AUC) are the metrics of machine learning methods that were used to evaluate the predictive power of these CCBMs for Sjögren's syndrome.
- Datasets with high AUC values in saliva of female children samples indicated that those CCBMs can serve as predictor biomarkers for Sjögren syndrome diagnosis in children.

## 1. Data Description

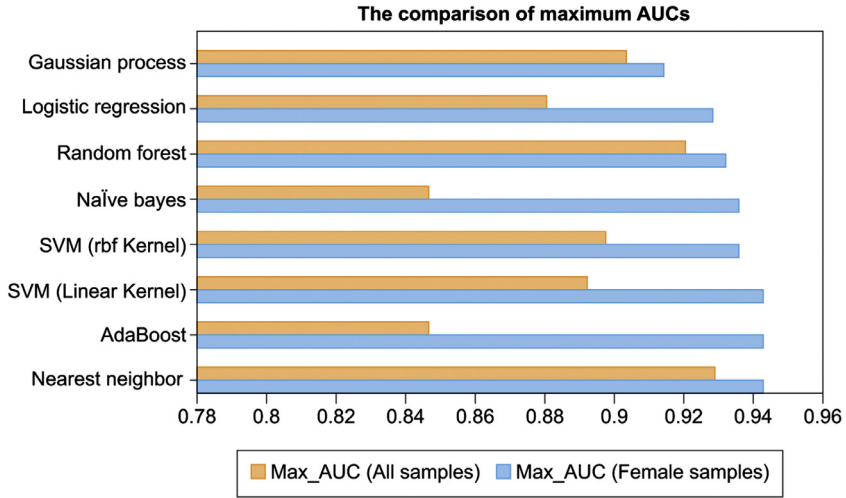
Sjögren's syndrome in children is not well defined. Xerostomia and xerophthalmia are not often the primary manifestations in children. Half of the children with Sjögren's syndrome present with parotitis and the other half present with less-specific clinical features. Therefore, there is a need to identify specific CCBMs in saliva of children with Sjögren's syndrome to increase child-specific diagnostic criteria. In a recent study, we used multiplex fluorescent microparticle-based immunoassays to determine the concentrations of 105 CCBMs in the saliva of children with Sjögren's syndrome and in the saliva of 16 normal healthy children [1]. The CCBMs we selected for that study were related to leukocyte activities and functions [1]. The dataset is accessible in the Mendeley Data Repository [2]. Specifically, 43/105 CCBMs were different ( $p < 0.05$ ) in children with Sjögren's syndrome compared to the healthy study controls [1]. Elevated CCBMs in IPA annotations were associated with autoimmune diseases and specific leukocyte functions including those associated with cellular movement, immune cell trafficking, and cell signaling. ROC curves and AUC values identified smaller datasets of CCBMs (e.g., IL27 and CCL4, k-Nearest Neighbor; AUC=0.93) that could be used as predictors for Sjögren's syndrome in children.

In adults, the ratio of females to males with Sjögren's syndrome is 9:1 but in children, the ratio of females to males is 5:1. Previously, we used the Mendeley dataset [2] to identify the CCBMs that can have predictive power for Sjögren's syndrome in children [1]. In this report, we used the Mendeley dataset [2] to identify CCBMs that can have predictive power for Sjögren's syndrome in female children.

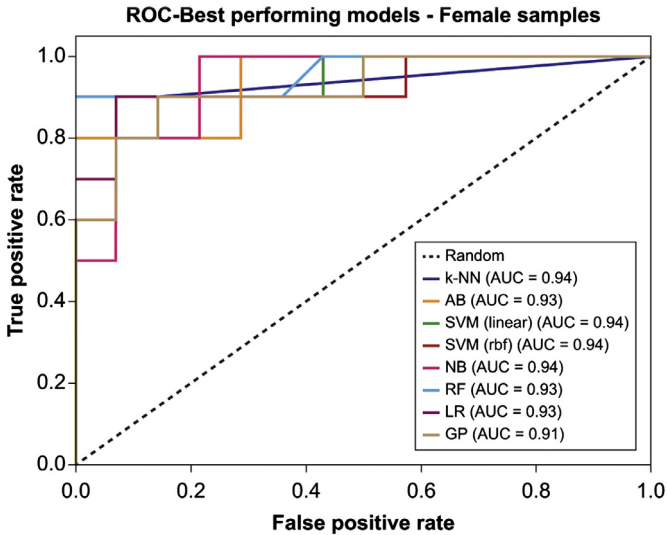
The data for this study is in an excel file called "Dataset-CCBMs, saliva, children, SS (01-02-21).xlsx". posted in the Mendeley database: <http://dx.doi.org/10.17632/yphm77tg24.1>. This file contains the replications of 105 CCBM concentrations (pg/ml) for the 27 subjects. The excel file itself is a bundle of multiple spreadsheets identified by different tabs (starting-Data, step2\_replicateGroups, stp3\_groupMeans, stp4\_meanDataRowsAdded, stp5\_OnlyMeanRowsRemain, input\_allData, input\_FemaleData\_only, Description). Each spreadsheet stores aforementioned data at different preprocessing stages including the starting raw data and final pre-processed data. Tab names of spreadsheets are enumerated and briefly described below.

1. StartingData The spreadsheet contains the CCBM measurements before any preprocessing.
2. Stp2\_replicateGroups: The spreadsheet contains the grouped replicate information of the three replicates. We annotate grouped replicates as 1,2,3,...,26, 27. Annotation are in the second column of the spreadsheet.
3. Stp3\_groupMeans: The spreadsheet contains group means. The rows containing group means are annotated as "new" in the second column of the spreadsheet.
4. Stp4\_meanDataRowsAdded: The spreadsheet in which a labeled row (i.e., labeled to represent each group) was added and defined for mean values. These rows have been annotated as "new" in the second column of the spreadsheet.
5. Stp5\_OnlyMeanRowsRemain: The spreadsheet contains only the mean data rows after the raw data rows are discarded.
6. Input\_alldata The spreadsheet was created based on stp5\_OnlyMeanRowsRemain spreadsheet. It contains data from both female and male samples. All the CCBM data columns were standardized using <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html#sklearn.preprocessing.MinMaxScaler>.

In Fig. 1, we compared the maximum AUC values resulting from different classifiers (i.e., best performing models) constructed based on all samples (orange) and constructed based only on female samples (blue). The maximum AUC values reported from classifier models constructed based only on female samples were higher than the AUC values reported by models constructed based on all samples at all times. Differences appeared greater using the classifiers Logistic Regression, Naïve Bayes, Support Vector Machine with radial basis function (rbf) Kernel, Support Vector Machine with Linear Kernel, and AdaBoost than with the classifiers Gaussian Process, Random Forest, and Nearest Neighbor.



**Fig. 1.** Comparison of the highest area under curve (AUC) values resulting from classifiers using two datasets: maximum AUC of all samples (red) and maximum AUC of female samples only (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 2.** Comparison of receiver operating characteristic (ROC) curves of classifiers on different feature sets. The area under curve (AUC) values indicate CCBMs can be served as predictor biomarkers for Sjögren syndrome diagnosis in children. Eight classifiers were used, including k-NN: k-Nearest Neighbor, RF: Random Forest, GP: Gaussian Process, SVM (rbf): Support Vector Machine with rbf Kernel, SVM (Linear): Support Vector Machine with Linear Kernel, LR: Logistic Regression, AB: AdaBoost, and NB: Naïve Bayes.

In Fig. 2, we compared the ROC curves of classifiers on different feature sets. The higher AUC values indicated that CCBMs can serve as predictor biomarkers for Sjögren syndrome diagnosis in children. Classifiers included the k-Nearest Neighbor, Random Forest, Gaussian Process, Support Vector Machine with rbf Kernel, Sup-

**Table 1**

The area under curve (AUC) values of best performing models.

Classifier	Model Name	Feature set (a)	AUC <sup>a</sup>
Random Forest (RR)	RR_fe_1	IL27, MIA, CCL4, CXCL11	0.93
	RR_fe_2	IL27, MIA, CCL4, CXCL11, IL23A	
	RR_fe_3	IL27, MIA, CCL4, CXCL11, TNFRSF18	
Logistic Regression (LR)	LR_fe_1	IL27, CCL4, CXCL11	0.93
Gaussian Process (GP)	GP_fe_1	IL27, CCL4, CXCL11, IL23A	0.91
	GP_fe_2	IL27, CCL4, CXCL11, MIA	
	GP_fe_3	IL27, CCL4, CXCL11, MIA, TNFRSF18, CCL19, IL12B, TNFRSF9, LIF, CCM2, GZMB	
	GP_fe_4	IL27, CCL4, CXCL11, IL23A, TNFRSF18, CCL19, IL12B, TNFRSF9, LIF	
k-Nearest Neighbor (NN)	NN_fe_1	IL27, CCL4	0.94
SVM (rbf)	SVMr_fe_1	CCL4, IL27, CXCL11, TNFRSF18	0.94
	SVMr_fe_2	CCL4, IL27, CXCL11	
SVM (Linear)	SVML_fe_1	IL27, CCL4, CXCL11, TNFRSF18	0.94
	SVML_fe_2	IL27, CCL4, CXCL11, TNFRSF18, MIA	
AdaBoost (AB)	AB_fe_1	CCL4, IL27, CXCL11, TNFRSF18, IL12B, ALCAM, CCL19, IL23A, TSLP, IL16, LIF, TNFSF5, TNFRSF8, CCL20, IRX1, CCL15, IL15, TNFRSF9, CXCL13, CCM2, CD40, CCL21, IL1B, MIA, XCL1, MMP9, CCL11, S100A8, GZMB, ULBP2, TNFSF11, CCL5, CXCL10, IFNB1	0.94
Naïve Bayes (NB)	NB_fe_1	IL27, CCL4, CXCL11, MIA, IL23A, CCL21, CCL19, ACAN, CCM2, TNFRSF9	0.94
	NB_fe_2	IL27, CCL4, CXCL11, MIA, IL23A, CCL21, CCL19, ACAN, CCM2, TNFRSF9, IL12B	

<sup>a</sup> Two different feature sets could have the same prediction power, that is, same AUC values. For example, models "SVML\_fe\_1" and "SVML\_fe\_1" have the same AUC value (0.94).

port Vector Machine with Linear Kernel, Logistic Regression, AdaBoost, and Naïve Bayes.

Eight classifiers identified 16 datasets containing 2–34 CCBMs with AUC values of 0.91 to 0.94 (Table 1). It is interesting to note that different feature sets could have the same prediction power, that is, the same AUC values, yet contain different datasets of CCBMs. This ranged from the Gaussian Process classifier, which contained 4 datasets of 4–11 CCBMs with an AUC of 0.91 to the Nearest Neighbor, SVM (rbf), SVM (Linear), AdaBoost, and Naïve Bayes classifiers, which contained 8 datasets of 2–34 CCBMs with an AUC of 0.94.

## 2. Experimental Design, Materials and Methods

### 2.1. Samples

Unstimulated whole saliva was collected as approved by the Human Institutional Review Board of the University of Iowa (IRB ID #: 200907702) (1). Consent was obtained, background material was recorded, and 1.0 to 7.0 ml saliva was collected from 11 children formally diagnosed with Sjögren's syndrome and 16 normal healthy children, matched for gender and age. After collection, saliva was stored at  $-80^{\circ}\text{C}$  until analysis. The features of these children have been previously reported (1).

After all saliva samples were collected, they were thawed on ice and centrifuged at 16,100 RCF (13,200 RPM, Eppendorf, 5415D centrifuge, Brinkmann Instruments, Inc., Westbury, NY) for 5 min at  $24^{\circ}\text{C}$  to pellet particulates and debris. Sample supernatants were decanted to another tube and held on ice.

**Table 2**

Catalog and lot numbers for kits of fluorescent microparticle-based immunoassays used to determine the concentrations of chemokines, cytokines, and biomarkers (CCBMs) in the saliva of children formally diagnosed with Sjögren's syndrome and from normal healthy children, matched for gender and age, who served as study controls.

Kit/Lot <sup>a</sup>	No. of CCBMs in Plex	CCBMs in kit
LXSAHM-07 (lot 129,053)	7	CCL5, CCL17, IL12A, TIMP1, TNFRSF13B, TNFRSF17, TNFRSF1A
LXSAHM-36 (lot 129,052)	36	B2M, CALCA, CCL1, CCL11, CCL20, CCL22, CCL26, CCL28, CCL3, CCL7, CCL8, CD274, CD40, CXCL13, CXCL14, CXCL4, CXCL9, FCER1G, GZMA, IFNA1, IFNB1, IFNG, IFNGR1, IL16, IL1A, IL1B, IL1R2, IL21, IL6, LGALS3, LGALS9, SLPI, TNFRSF7, TNFRSF8, TNFSF11, TNFSF5
LXSAHM-22 (lot 128,990)	22	IL11RA, C9, CCL21, CST3, FSTL1, IFNL3, IGFBP3, IL12B, IL15, IL27, LGALS3BP, LIF, LTF, MIA, NAGLU, S100A8, S100A9, TNFRSF18, TNFRSF1B, TSLP, ULBP2, XCL1
LXSAHM-40 (lot 129,245)	40	A2M, ACAN, ALCAM, AMBP, C5, CA9, CCL13, CCL14, CCL15, CCL18, CCL19, CCL2, CCL23, CCL24, CCL25, CCL27, CCL4, CCM2, CD276, CTSS, CXCL10, CXCL11, FASLG, FSTL3, GAS6, GDF2, GZMB, IFNL2, IL10, IL23A, IL2RA, IL7, IRX1, MMP9, PECAM1, SELL, TNFA, TNFRSF9, TNFSF13, TNFSF13B

<sup>a</sup> Luminex Human Magnetic Assay, R&D Systems, Minneapolis, MN USA.

## 2.2. Determination of CCBMs

The concentrations (pg/ml) of 105 CCBMs were determined in each supernatant, in triplicate, using multiplex fluorescent microparticle-based immunoassays (Luminex Human Magnetic Assay, R&D Systems, Minneapolis, MN). Detection of 105 CCBM analytes was performed in 4 assay runs and the product catalog numbers, lot numbers, and CCBM composition in each kits is listed in Table 2. Briefly, three 50  $\mu$ l aliquots of saliva supernatant were added to immunoassay plates. Magnetic microparticles with attached anti-human analyte-specific antibodies were added and the immunoassay plates were incubated on an orbital shaker (Titer Plate Shaker, Lab-Line Instruments, Inc., Melrose Park, IL USA) at 4 °C. After 18.0 h, the microparticles were washed (ELx405TS magnetic plate washer, BioTek, Winooski, VT USA) and incubated with biotinylated anti-human analyte-specific antibodies at room temperature for 1.0 h in the dark. These antibodies were diluted as indicated in the instructions on the inserts in each kit. The microparticles were washed, incubated with streptavidin-phycoerythrin conjugate at room temperature for 0.5 h, washed, and suspended in buffer. Median fluorescent intensity (MFI) values of the bound phycoerythrin were determined in the Luminex model 100 IS (Luminex, Austin, TX USA).

CCBM concentrations were interpolated from their MFI values using five parameter logistic curves created from the standard concentrations and their respective MFI readings on the Luminex 100 IS using xPonent v3.1 software (Luminex, Austin, TX) or on the readout files using Milliplex Analyst v5.1 software (EMD Millipore, Billerica, MA) as previously described (1). CCL27 concentrations were below the standard curve and these concentrations were extrapolated from their MFI values using curves created from zero concentration to the lowest standard concentration and their respective MFI readings. B2M concentrations were above the standard curve and these concentrations were extrapolated from their MFI values using curves extended beyond the highest concentration and their respective MFI readings.

## 2.3. Analysis

For analysis, the data went through a series of transformations as described in the Mendeley dataset. The replicates were first grouped, a mean was calculated, and the means were used for determining statistical analysis among groups, hierarchical clustering, principal component analysis, and ingenuity pathway analysis (1). The dataset of mean concentrations of CCBMs from

both female and male children saliva samples were then standardized and the dataset of mean concentrations of CCBMs from female children saliva samples were also standardized. Machine learning methods were then used to assess the predictive power of the CCBMs in these two datasets using ROC and AUC metrics [1]. Correlation [3], Information Gain [4], Information Gain Ratio [5], Symmetrical Uncertainty [6], and ReliefF [7] feature selection methods were used to rank CCBMs. An ensemble method of the aggregated ranks of five feature selection methods was also used to rank CCBMs. Discrete sets of CCBMs (i.e., Union, AtLeast2, AtLeast3, AtLeast4, SelectedByFive) were obtained using Venn diagram, representing overlapped features of five feature selection methods. The “Union” feature set at a given rank  $k$  is the union of all top  $k$  features selected by five methods. The “SelectedByFive” feature set at a given rank  $k$  is the common features of all top  $k$  feature sets each selected by five methods. This definition is analogous for feature sets of “AtLeast2”, “AtLeast3”, and “AtLeast4”. For example, “AtLeast2” feature set at a given rank  $k$  includes top  $k$  features that are selected by at least 2 feature selection methods.

Each top  $k$  feature set was evaluated using classification methods: K-Nearest Neighbor ( $k = 3$ ) [8], AdaBoost (trees=100) [9], Support Vector Machine (Linear Kernel) [10], Support Vector Machine (radial basis function (rbf) Kernel) [11], Naïve Bayes [12], Random Forest (trees=100) [13], Logistic Regression, and Gaussian Process [14]. The classifier used selected features to train a model that then was used to predict Sjögren’s Syndrome. The performance of a classifier on a specific feature set was evaluated using leave-one-out cross-validation and ROC curve. The average AUC was used to measure the performance.

## Ethics Statement

Twenty-seven children were consented and 1.0 to 7.0 ml of unstimulated whole saliva was collected as approved by the Human Institutional Review Board of the University of Iowa (IRB ID #: 200907702).

## CRediT Author Statement

**Miyuraj Harishchandra Hikkaduwa Withanage:** Methodology, Formal analysis, Writing – original draft, Writing – review & editing; **M. Paula Gomez Hernandez:** Methodology, Conceptualization, Writing – original draft, Writing – review & editing; **Emily E. Starman:** Methodology, Writing – original draft, Writing – review & editing; **Andrew B. Davis:** Methodology; **Erliang Zeng:** Formal analysis, Writing – original draft, Writing – review & editing; **Scott M. Lieberman:** Conceptualization, Writing – original draft, Writing – review & editing, Funding acquisition; **Kim A. Brogden:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition; **Emily A. Lanzel:** Conceptualization, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

## Acknowledgments

We acknowledge the support and help of C. Michael Knudson MD, PhD, Kristen Coleman MS, Rita D. Sigmund MA and Joe Galbraith, Tissue Procurement Core, University of Iowa Carver College of Medicine, The University of Iowa, Iowa City, IA 52242. We acknowledge the support

and help of Kevin Knudtson, PhD, Director, Genomics Division, Iowa Institute of Human Genetics, University of Iowa Carver College of Medicine, The University of Iowa, Iowa City, IA USA.

## References

- [1] M.P. Gomez Hernandez, E.E. Starman, A.B. Davis, M.H. Hikkaduwa Withanage, E. Zeng, S.M. Lieberman, K.A. Brogden, E.A. Lanzel, A unique profile of chemokines, cytokines, and biomarkers in the saliva of children with Sjögren syndrome, *Rheumatology* (2021) 1–13 <https://doi.org/>, doi:10.1093/rheumatology/keab098.
- [2] M.H. Hikkaduwa Withanage, M.P. Gomez Hernandez, E.E. Starman, A.B. Davis, E. Zeng, S.M. Lieberman, K.A. Brogden, E.A. Lanzel, Dataset-chemokines, cytokines, and biomarkers in the saliva of children with Sjögren's syndrome, V1, 2021 <http://dx.doi.org/>, doi:10.17632/yphm77tg24.1.
- [3] M. Hall, Correlation-based feature selection for discrete and numeric class machine learning, in: *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000)*, Stanford University, Stanford, 2000, pp. 359–366.
- [4] X. Jiang, J. Jao, R. Neapolitan, Learning predictive interactions using information gain and bayesian network scoring, *PLoS ONE* 10 (2015) e0143247 <https://doi.org/>, doi:10.1371/journal.pone.0143247.
- [5] R. Praveena Priyadarsini, M.L. Valarmathi, S. Sivakumari, Gain ratio based feature selection method for privacy preservation, *ICTACT J. Soft Comput.* 1 (2011) 201–205 [http://ictactjournals.in/paper/IJSC\\_V1\\_I4\\_6\\_201-205.pdf](http://ictactjournals.in/paper/IJSC_V1_I4_6_201-205.pdf).
- [6] B. Singh, N. Kushwaha, O.P. Vyas, A feature subset selection technique for high dimensional data using symmetric uncertainty, *J. Data Anal. Inform. Process.* 2 (2014) 95–105 <http://dx.doi.org/>, doi:10.4236/jdaip.2014.24012.
- [7] I. Kononenko, Estimating attributes: analysis and extensions of RELIEF, in: F. Bergadano, L. De Raedt (Eds.), *Proceeding of the Machine Learning: ECML-94: European Conference on Machine Learning*, Springer, Berlin Heidelberg, Catania, 1994, pp. 171–182.
- [8] T.M. Cover, P.E. Hart, Nearest neighbor pattern classification, *IEEE T. Inform. Theory* 13 (1967) 21–27 <https://dx.doi.org/>, doi:10.1109/IT.1967.1053964.
- [9] T.H. Kim, D.C. Park, D.M. Wou, T. Jeong, S.Y. Min, Multi-class classifier-based adaboost algorithm, in: Y. Zhang, Z.H. Zhou, C. Zhang, L. Ying (Eds.), *Intelligent Science and Intelligent Data Engineering*, Springer Berlin Heidelberg, Berlin, 2012, pp. 122–127.
- [10] T. Hastie, J. Friedman, R. Tibshirani, Support vector machines and flexible discriminants, in: T. Hastie, J. Friedman, R. Tibshirani (Eds.), *The Elements of Statistical learning: Data Mining, Inference, and Prediction*, Springer, New York, New York, 2001, pp. 371–409.
- [11] W.J. Wang, Z.B. Xu, W.Z. Lu, X.Y. Zhang, Determination of the spread parameter in the Gaussian kernel for classification and regression, *Neurocomputing* 55 (2003) 643–663 <https://doi.org/>, doi:10.1016/S0925-2312(02)00632-X.
- [12] J.D.M. Rennie, L. Shih, J. Teevan, D.R. Karger, Tackling the poor assumptions of naive bayes text classifiers, in: *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, 20, AAAI Press, Washington, 2003, p. 616.
- [13] L. Breiman, Random Forests, *Mach. Learn* 45 (2001) 5–32 <https://doi.org/>, doi:10.1023/A:1010933404324.
- [14] C.E. Rasmussen, C.K.I. Williams, *Gaussian Processes For Machine Learning*, The MIT Press, Cambridge, 2005.