Supplementary Information (SI) for Analyst.

Supplementary Materials

Non-invasive detection of Oncostatin M and TNF-α biomarkers for inflammatory bowel disease in a microphysiological chip with embedded pH tuning capabilities

Daniele Bellisario[a], Alessandra Calogiuri[a*], Elisa Sciurti[a], Laura Blasi[a], Vanessa Esposito[a], Flavio Casino[a], Pietro Siciliano[a], Antonio Della Torre[a] and Luca Francioso[a]

[a]Institute for Microelectronics and Microsystems CNR-IMM, Via per Monteroni "Campus Ecotekne", Lecce, Italy

* Author to whom correspondence should be addressed: *alessandracalogiuri@cnr.it*

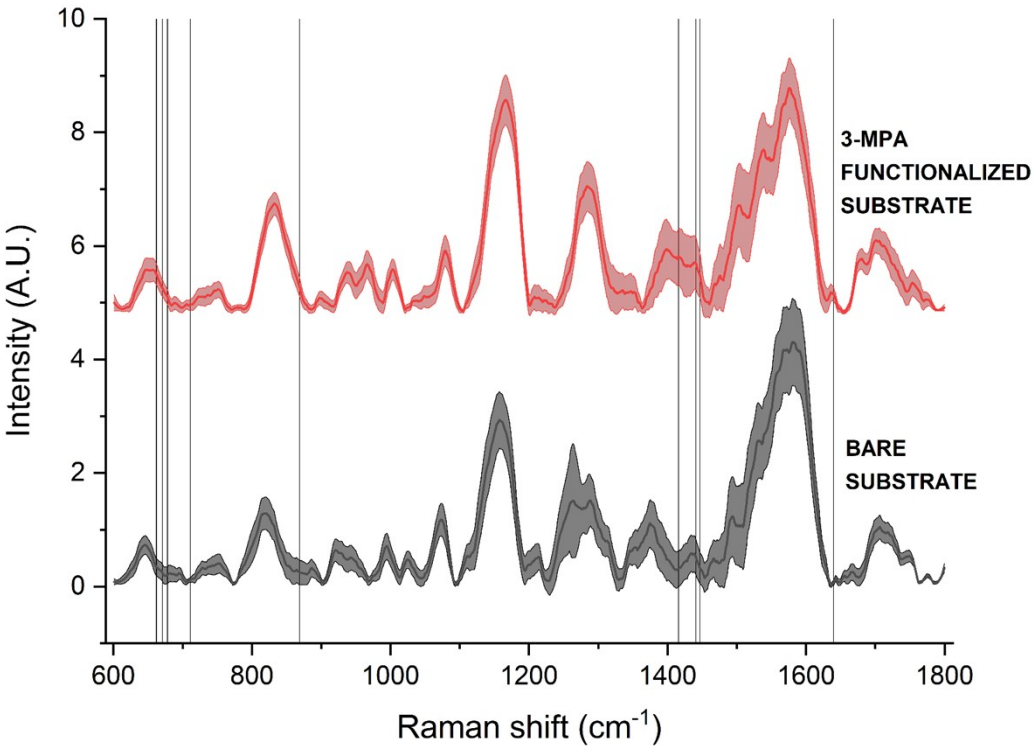

Figure S1: Off-chip reference SERS spectra of the commercial substrate before and after 3-MPA functionalization. Shaded areas represent the standard deviation (n = 64). Compared to the bare substrate, the functionalized substrate presents additional peaks consistent with a 3-MPA layer, In particular, the main peaks observed after 3-MPA functionalization are: 662, 671, 678, 711 cm^{-1} (ν(C–S)), 869 cm^{-1} (thiol-related mode), 1416 cm^{-1} (carboxylate symmetric stretch, νs(COO^-)), 1441 and 1447 cm^{-1} (δ(CH_2) with contributions from C–O/OH modes), and 1640 cm^{-1} (carbonyl/carboxylate region) [1-3].

Table S.1 The assignments of Raman peaks for characterization of OSM.

Peak no	Peak Center (cm^{-1})	Assignment
1	688	C-S stretch
2	722	C-S stretch
3	733	C-S stretch
4	763	Tryptophan
5	805	Proline, hydroxyproline, tyrosine
6	814	Proline, hydroxyproline, tyrosine
7	823	Out-of-plane ring breathing, tyrosine
8	852	Proline, hydroxyproline, tyrosine, glutamine
9	885	Proteins, including collagen
10	895	C-C skeletal
11	919	Proline, hydroxyproline
12	969	Phosphate monoester groups of phosphorylated protein
13	1004	Phenylalanine
14	1032	CH_2CH_3 bending modes
15	1060	C-C in-plane bending
16	1063	C-C skeletal stretch random conformation
17	1202	Aromatic C-O and C-N
18	1450	CH_2 bending, glutamine
19	1573	C-C stretching
20	1579	C-C stretching
21	1593	C=N and C=C stretching
22	1678	C=O stretching
23	1683	Amide I (disordered structure; non-hydrogen bonded)
24	1692	Amide I of proteins

Table S.2 The assignments of Raman peaks for characterization of TNF-α.

Peak no	Peak Center (cm^{-1})	Assignment
1	661	C-S stretching mode of cystine
2	722	C-S stretch
3	738	C-S stretch
4	768	Tryptophan
5	771	Tryptophan
6	846	Tyrosine
7	861	Tyrosine
8	877	C-C stretching, hydroxyproline
9	901	C-C stretching
10	907	C-C stretching
11	919	Proline, hydroxyproline
12	925	n(C-C), proline & valine (protein band)
13	1101	Phenylalanine (proteins)
14	1105	Phenylalanine (proteins)
15	1182	C-H bending mode of structural proteins
16	1198	Amide III-Due to C-N stretching and N-H bending

17	1202	Amide III & CH_2 wagging vibrations from glycine backbone & proline side chains
18	1450	CH_2 bending, glutamine
19	1453	C-H bending mode of structural proteins, glutamine
20	1457	CH_2/CH_3 deformation
21	1573	C-C stretching
22	1579	C-C stretching
23	1583	C=C bending mode of phenylalanine
24	1653	Amide I of proteins
25	1665	Amide I of proteins (disordered structure-solvated)
26	1667	Amide I (antiparallel b-sheet)
27	1678	Amide I (b-sheet)
28	1692	Amide I of proteins
29	1697	Amide I (turns and bands)

Enzyme-linked Immunosorbent Assay (ELISA) Results

In order to verify whether the washing step with acetone effectively detached the selected proteins, we performed an Enzyme-linked Immunosorbent Assay (ELISA) to quantitatively measure OSM and TNF-α before and after the acetone treatment.

In the case of substrates exposed to OSM, the mean absorbance values decreased from 1.681 ± 0.357 prior to acetone washing to 0.531 ± 0.125 following the treatment, thus approaching the blank control value of 0.590 ± 0.106. A reproducible reduction in ELISA signal was also observed for TNF-α substrates, with mean absorbance values decreasing from 0.166 ± 0.004 to 0.144± 0.008 following acetone washing. In the case of TNF-α, the blank control values were 0.089 ± 0.023).

The different extent of ELISA signal reduction observed for OSM and TNF-α following acetone washing may be attributed to differences in protein size, protein concentration, conformation, and interaction strength with the SERS surface. Although acetone treatment significantly suppressed the specific ELISA signal for both analytes, the post-washing absorbance values remained slightly higher than the respective blank controls, particularly in the case of TNF-α, thus indicating that complete removal of the bound proteins or background contributions cannot be entirely excluded.

Finally, to test the surface reusability, subsequent to acetone washing, the regenerated SERS substrates were exposed to OSM and analysed once more by ELISA. A measurable ELISA signal was obtained on all samples, although with reduced intensity in comparison to pristine substrates (0,49±0,06). This result indicates a partial loss of surface activity following repeated use and regeneration. In line with these results, the trained ML algorithm also shows reduced efficiency after washing compared to pristine substrates.

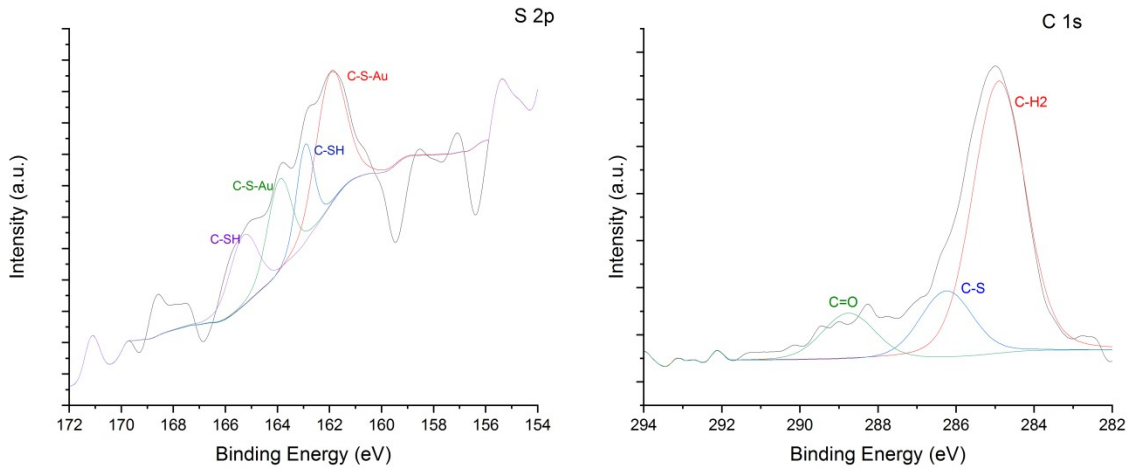

Fig. S2: (a-b) High-resolution spectra for S 2p and C 1s regions are reported with the curve-fitting results and peak component attributions.

In addition to ELISA measurements, XPS analysis was performed on the SERS substrates following acetone washing in order to provide a more comprehensive assessment of surface regeneration at the chemical level. As demonstrated in Fig S1, the S 2p and C 1s spectra obtained following the washing process demonstrate consistency with the presence of the 3-MPA monolayer, with characteristic contributions attributed to C–S/Au–S and C–H/C–S bonds. These results provide further evidence for the effective protein removal while preserving the underlying surface functionalization.

<u>OSM classification</u> - Linear Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) data processing

The Raman shifts intensities utilized to train the ML models were extracted using the Peak Analyzer tool in OriginPro. Positive peaks were identified by determining local maxima in the spectra using the second derivative method, coupled with a quadratic Savitzky-Golay smoothing (window size of 5), to enhance peak detection while minimizing noise [4], [5]. Each peak's intensity at its corresponding Raman shift was extracted, and only peaks occurring consistently in at least 50% of the measurements for each condition (32 out of 64 spectra for each condition) were retained. After filtering, the final datasets contained 378 peaks for pH 8.5, 327 for pH 7.8, 295 for pH 7.4, 244 peaks for pH 6.8, 301 for pH 6.4 and 297 for pH 4. The dataset was first divided into training (80%) and testing (20%) sets [6]. Feature ranking was then performed exclusively on the training set, using two supervised learning models. The extracted peaks were ranked through an embedded method; therefore, two different linear supervised learning models were trained [7]. A linear Support Vector Machine (SVM) with a box constraint of 10, ranks features based on the weights derived from its linear kernel [8], [9]. Similarly, Linear Discriminant Analysis (LDA) ranked features based on LDA weights [10], [11]. To assess stability across varying data distributions the classification was performed using 10-fold cross-validation [12]. The classification performance of these models was compared using two different feature sets: the top 40 and the top 100 ranked features [13]. The confusion matrix obtained from the test phase of the models with the highest test

accuracy for each pH were illustrated in FigureS1 and FigureS2. In binary classification, the confusion matrix is a 2×2 table in which the rows represent the true classes and the columns represent the predicted classes. The diagonal cells (True Negative + True Positive) represent correct predictions, while the cells outside the diagonal (False Positive + False Negative) represent errors.
TableS1 and TableS2 summarize the resulting performance metrics of the test, moreover, include accuracy derived from the training and test confusion matrix. The performance metrics reported in Tables S1–S4 were derived from the confusion matrices obtained during the test phase with the following formulas:

$$Sensitivity\ (or\ Recall) = \frac{TP}{(TP\ +\ FN)}$$

$$Specificity\ = \frac{TN}{(TN\ +\ FP)}$$

$$Precision\ = \frac{TP}{(TP\ +\ FP)}$$

$$F1\_score = 2\ \times \frac{(Precision\ \times\ Sensitivity)}{(Precision\ +\ Sensitivity)}$$

$$TP = True\ Positive\ TN = \ True\ Negative\ FP = False\ Positive\ FN = False\ Negative$$

<u>TNF-alpha classification</u> - Linear Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) data processing

Peak centers and Raman intensity were identified for each spectrum for all pH values considered, using the peak finding procedure described for Oncostatin M. After filtering, the final datasets contained 337 peaks for pH 8, 264 for pH 7.7, 316 for pH 7.4, 276 peaks for pH 7, 247 for pH 6.4, 249 for pH 5.8, 304 peaks for pH 5.2 and 259 for pH 4. For a more in-depth examination of pH-dependent spectral differences, a supervised classification approach was employed. First, the dataset was randomly divided into training and test sets with an 80/20 ratio, and then the Raman peak centers were ranked based on SVM weights and LDA coefficients, according to the previous analysis. Also, this time, the performance of 3 models was evaluated using 2 sets of ranked features, the top 40 and top 100 features, using 10-fold cross-validation. The classification results and the performance metrics of the test were reported in TableS3 and TableS4. Moreover, the confusion matrix of the models with the highest accuracy in the test phase for each pH was illustrated in FigureS4 and FigureS5. As illustrated in TableS3 and TableS4, 100 features often result in the best way to achieve optimal classification performance when models using features of both LDA and SVM ranking.

Table S.3 Performance metrics of qSVM, cSVM, and LDA models for Oncostatin M detection using LDA-ranked features with the top 40 and top 100 features at different pH values. Bold values indicate the model with the highest test accuracy for each pH condition.

	Model	n° features	Specificity	Precision	Sensitivity	F1_score	Train Accuracy	Test Accuracy
pH 4	q SVM	40	0,577	0,593	0,640	0,616	66.3%	60.8%
		100	0,577	0,607	0,680	0,641	64.7%	62.7%
	c SVM	40	0,462	0,517	0,600	0,555	65.3%	52.9%
		100	0,500	0,519	0,560	0,539	64.1%	52.9%

	LDA	40	0,731	0,696	0,640	0,667	74.5%	68.6%
		100	**0,692**	**0,704**	**0,760**	**0,731**	**80.9%**	**72.5%**
pH 6.4	**q SVM**	40	1,000	1,000	0,920	0,958	90.7%	96.1%
		100	1,000	1,000	0,960	0,980	87.3%	98%
	c SVM	40	1,000	1,000	0,920	0,958	87.8%	96.1%
		100	**1,000**	**1,000**	**0,960**	**0,980**	**87.9%**	**98%**
	LDA	40	0,962	0,958	0,920	0,939	88.3%	94.1%
		100	0,962	0,958	0,920	0,939	88.3%	94.1%
pH 6.8	**q SVM**	40	0,962	0,962	0,962	0,962	95.7%	96.2%
		100	0,962	0,963	1,000	0,981	97.1%	98.1%
	c SVM	40	1,000	1,000	0,962	0,981	94.7%	98.1%
		100	0,962	0,963	1,000	0,981	96.2%	98.1%
	LDA	40	0,962	0,963	1,000	0,981	97.1%	98.1%
		100	**1,000**	**1,000**	**1,000**	**1,000**	**99%**	**100%**
pH 7.4	**q SVM**	40	0,680	0,704	0,679	0,691	70.7%	67.9%
		100	0,760	0,793	0,821	0,807	77.7%	77.4%
	c SVM	40	0,720	0,731	0,679	0,704	69.8%	69.8%
		100	0,680	0,742	0,821	0,780	76.8%	75.5%
	LDA	40	0,800	0,821	0,821	0,821	82.3%	81.1%
		100	**0,920**	**0,920**	**0,821**	**0,868**	**87%**	**86.8%**
pH 7.8	**q SVM**	40	0,808	0,844	0,900	0,871	82.5%	85.7%
		100	0,731	0,794	0,900	0,844	87%	82.1%
	c SVM	40	0,769	0,812	0,867	0,839	83.8%	82.1%
		100	0,769	0,818	0,900	0,857	86.1%	83.9%
	LDA	40	0,769	0,824	0,933	0,875	83.9%	85.7%
		100	**0,808**	**0,853**	**0,967**	**0,906**	**95.5%**	**89.3%**
pH 8.5	**q SVM**	40	0,808	0,815	0,880	0,846	89.8%	84.3%
		100	0,846	0,857	0,960	0,906	93.7%	90.2%
	c SVM	40	0,846	0,852	0,920	0,885	91.2%	88.2%
		100	0,846	0,862	1,000	0,926	93.2%	92.2%
	LDA	40	0,808	0,828	0,960	0,889	95.1%	88.2%
		100	**1,000**	**1,000**	**1,000**	**1,000**	**97.5%**	**100%**

Table S.4 Performance metrics of qSVM, cSVM, and LDA models for Oncostatin M detection using SVM-ranked features with the top 40 and top 100 features at different pH values. Bold values indicate the model with the highest test accuracy for each pH condition.

	Model	**n° features**	**Specificity**	**Precision**	**Sensitivity**	**F1_score**	**Train Accuracy**	**Test Accuracy**
pH 4	**q SVM**	40	0,654	0,640	0,640	0,640	56%	64.7%
		100	0,654	0,640	0,640	0,640	64.2%	64.7%
	c SVM	40	0,615	0,565	0,520	0,542	60.3%	56.9%
		100	0,615	0,615	0,640	0,627	60.7%	62.7%
	LDA	**40**	**0,769**	**0,769**	**0,800**	**0,784**	**74%**	**78.4%**

		100	0,731	0,741	0,800	0,769	76.4%	76.5%
pH 6.4	**q SVM**	40	1,000	1,000	0,920	0,958	93.2%	96.1%
		100	1,000	1,000	0,960	0,980	91.8%	98%
	c SVM	40	0,962	0,960	0,960	0,960	90.3%	96.1%
		100	**1,000**	**1,000**	**0,960**	**0,980**	**92.2%**	**98%**
	LDA	40	0,962	0,958	0,920	0,939	91.7%	94.1%
		100	1,000	1,000	0,960	0,980	88.8%	98%
pH 6.8	**q SVM**	**40**	**1,000**	**1,000**	**1,000**	**1,000**	**100%**	**100%**
		100	1,000	1,000	1,000	1,000	99.5%	100%
	c SVM	40	1,000	1,000	1,000	1,000	99.5%	100%
		100	1,000	1,000	1,000	1,000	99.5%	100%
	LDA	40	1,000	1,000	1,000	1,000	99.5%	100%
		100	1,000	1,000	1,000	1,000	97.5%	100%
pH 7.4	**q SVM**	40	0,760	0,760	0,679	0,717	81%	71.7%
		100	0,720	0,759	0,786	0,772	84.2%	75.5%
	c SVM	40	0,720	0,708	0,607	0,654	78.2%	66%
		100	0,720	0,750	0,750	0,750	73.6%	79.7%
	LDA	40	0,760	0,769	0,714	0,740	77.7%	73.6%
		100	**0,760**	**0,806**	**0,893**	**0,847**	**87.4%**	**83%**
pH 7.8	**q SVM**	40	0,923	0,933	0,933	0,933	88.8%	92.9%
		100	0,885	0,909	1,000	0,952	92.8%	94.6%
	c SVM	40	0,962	0,964	0,900	0,931	88.4%	92.9%
		100	**0,962**	**0,967**	**0,967**	**0,967**	**92.8%**	**96.4%**
	LDA	40	0,846	0,857	0,800	0,828	89.3%	82.1%
		100	1,000	1,000	0,867	0,929	92.8%	92.9%
pH 8.5	**q SVM**	40	0,769	0,793	0,920	0,852	92.2%	84.3%
		100	0,808	0,828	0,960	0,889	93.2%	88.2%
	c SVM	40	0,846	0,862	1,000	0,926	92.7%	92.2%
		100	0,885	0,889	0,960	0,923	94.1%	92.2%
	LDA	40	0,846	0,862	1,000	0,926	95.6%	92.2%
		100	**0,923**	**0,926**	**1,000**	**0,962**	**97%**	**96.1%**

Table S.5 Performance metrics of qSVM, cSVM, and LDA models for TNF-α detection using LDA-ranked features with the top 40 and top 100 features at different pH values. Bold values indicate the model with the highest test accuracy for each pH condition.

	Model	**n° features**	**Specificity**	**Precision**	**Sensitivity**	**F1_score**	**Train Accuracy**	**Test Accuracy**
PH 4	**q SVM**	40	0,923	0,920	0,920	0,920	82,90%	92,20%
		100	0,962	0,955	0,840	0,894	86,30%	90,20%
	c SVM	40	0,885	0,889	0,960	0,923	81%	92.2%
		100	0,962	0,957	0,880	0,917	87.3%	92.2%
	LDA	40	0,846	0,852	0,920	0,885	84.9%	88.2%

		100	**0,885**	**0,889**	**0,960**	**0,923**	**96.6%**	**92.2%**
PH 5.2	**q SVM**	40	0,778	0,778	0,840	0,808	74.6%	80.8%
		100	0,926	0,917	0,880	0,898	84.7%	90.4%
	c SVM	40	0,778	0,769	0,800	0,784	75.1%	78.8%
		100	0,926	0,917	0,880	0,898	88%	90.4%
	LDA	40	0,889	0,875	0,840	0,857	81.8%	86.5%
		100	**0,852**	**0,857**	**0,960**	**0,906**	**94.2%**	**90.4%**
PH 5.8	**q SVM**	40	0,857	0,833	0,741	0,784	63.6%	80%
		100	**0,821**	**0,821**	**0,852**	**0,836**	**73.2%**	**83.6%**
	c SVM	40	0,714	0,714	0,741	0,727	64.1%	72.7%
		100	0,821	0,808	0,778	0,793	73.2%	80%
	LDA	40	0,786	0,727	0,593	0,653	68.6%	69.1%
		100	0,786	0,769	0,741	0,755	78.2%	76.4%
PH 6.4	**q SVM**	40	0,760	0,769	0,769	0,769	83%	76.5%
		100	0,800	0,815	0,846	0,830	89.3%	82.4%
	c SVM	40	0,840	0,833	0,769	0,800	81%	80.4%
		100	0,840	0,840	0,808	0,824	89.7%	82.4%
	LDA	40	0,640	0,679	0,731	0,704	82.5%	68.6%
		100	**0,680**	**0,765**	**1,000**	**0,867**	**91.2%**	**84.3%**
PH 7	**q SVM**	40	0,800	0,815	0,846	0,830	84%	82.4%
		100	0,760	0,786	0,846	0,815	86.5%	80.4%
	c SVM	40	0,800	0,808	0,808	0,808	80.5%	80.4%
		100	0,760	0,786	0,846	0,815	86.5%	80.4%
	LDA	**40**	**0,840**	**0,852**	**0,885**	**0,868**	**77.8%**	**86.3%**
		100	0,800	0,821	0,885	0,852	80.5%	84.3%
PH 7.4	**q SVM**	40	0,962	0,955	0,840	0,894	80.3%	90.2%
		100	0,962	0,952	0,800	0,869	83.8%	88.2%
	c SVM	40	0,846	0,833	0,800	0,816	78.8%	82.4%
		100	0,962	0,947	0,720	0,818	82.2%	84.3%
	LDA	40	0,846	0,826	0,760	0,792	79.9%	80.4%
		100	**1,000**	**1,000**	**0,840**	**0,913**	**90.6%**	**92.2%**
PH 7.7	**q SVM**	40	0,654	0,679	0,760	0,717	72.7%	70.6%
		100	0,731	0,750	0,840	0,792	76.6%	78.4%
	c SVM	40	0,654	0,679	0,760	0,717	73.2%	70.6%
		100	0,769	0,778	0,840	0,808	80.5%	80.4%
	LDA	**40**	**0,769**	**0,786**	**0,880**	**0,830**	**78.5%**	**82.4%**
		100	0,808	0,808	0,840	0,824	76.6%	82.4%
PH 8	**q SVM**	40	0,560	0,633	0,731	0,678	62.4%	64.7%
		100	0,760	0,778	0,808	0,793	74.6%	78.4%
	c SVM	40	0,640	0,667	0,692	0,679	65.8%	66.7%
		100	0,720	0,741	0,769	0,755	73.7%	74.5%
	LDA	40	0,680	0,714	0,769	0,740	73.1%	72.5%
		100	**0,840**	**0,833**	**0,769**	**0,800**	**92.2%**	**80.4%**

Table S.6 Performance metrics of qSVM, cSVM, and LDA models for TNF-α detection using SVM-ranked features with the top 40 and top 100 features at different pH values. Bold values indicate the model with the highest test accuracy for each pH condition.

	Model	n° features	Specificity	Precision	Sensitivity	F1_score	Train Accuracy	Test Accuracy
pH 4	q SVM	40	0,923	0,923	0,960	0,941	86,9%	94,1%
		100	0,962	0,962	1,000	0,981	91,7%	98,0%
	c SVM	40	0,923	0,926	1,000	0,962	89,3%	96,1%
		100	**1,000**	**1,000**	**1,000**	**1,000**	**93,6%**	**100,0%**
	LDA	40	0,962	0,962	1,000	0,981	86,9%	98,0%
		100	0,846	0,846	0,880	0,863	89,2%	86,3%
pH 5.2	q SVM	40	0,852	0,852	0,920	0,885	85,7%	88,5%
		100	0,889	0,875	0,840	0,857	91,9%	86,5%
	c SVM	40	0,815	0,828	0,960	0,889	84,7%	88,5%
		100	0,852	0,852	0,920	0,885	91,4%	88,5%
	LDA	40	0,815	0,821	0,920	0,868	86,6%	86,5%
		100	**0,889**	**0,893**	**1,000**	**0,943**	**89,5%**	**94,2%**
pH 5.8	q SVM	40	0,964	0,952	0,741	0,833	73,6%	85,5%
		100	0,893	0,870	0,741	0,800	79,1%	81,8%
	c SVM	**40**	**0,929**	**0,913**	**0,778**	**0,840**	**75,5%**	**85,5%**
		100	0,893	0,870	0,741	0,800	81,8%	81,8%
	LDA	40	0,929	0,895	0,630	0,739	74,5%	78,2%
		100	0,893	0,880	0,815	0,846	79,5%	85,5%
pH 6.4	q SVM	40	0,800	0,800	0,769	0,784	90,7%	78,4%
		100	0,800	0,815	0,846	0,830	95,1%	82,4%
	c SVM	40	0,880	0,870	0,769	0,816	92,7%	82,4%
		100	**0,880**	**0,880**	**0,846**	**0,863**	**93,2%**	**86,3%**
	LDA	40	0,760	0,769	0,769	0,769	91,7%	76,5%
		100	0,840	0,852	0,885	0,868	90,3%	86,3%
pH 7	q SVM	40	0,800	0,839	1,000	0,912	84,0%	90,2%
		100	0,800	0,821	0,885	0,852	86,9%	84,3%
	c SVM	40	0,800	0,839	1,000	0,912	86,0%	90,2%
		100	0,800	0,821	0,885	0,852	86,4%	84,3%
	LDA	**40**	**0,880**	**0,897**	**1,000**	**0,946**	**84,5%**	**94,1%**
		100	0,920	0,923	0,923	0,923	82,5%	92,2%
pH 7.4	q SVM	40	0,769	0,778	0,840	0,808	87,1%	80,4%
		100	0,885	0,875	0,840	0,857	89,2%	86,3%
	c SVM	40	0,692	0,724	0,840	0,778	86,2%	76,5%
		100	0,808	0,808	0,840	0,824	86,8%	82,4%
	LDA	**40**	**0,885**	**0,880**	**0,880**	**0,880**	**88,7%**	**88,2%**
		100	0,923	0,905	0,760	0,826	92,6%	84,3%
pH 7.7	q	40	0,846	0,840	0,840	0,840	87,8%	84,3%

	SVM	100	0,846	0,846	0,880	0,863	83,9%	86,3%
	c SVM	40	0,846	0,840	0,840	0,840	86,3%	84,3%
		100	0,846	0,840	0,840	0,840	84,3%	84,3%
	LDA	40	0,923	0,920	0,920	0,920	95,1%	92,2%
		100	**0,962**	**0,962**	**1,000**	**0,981**	**92,2%**	**98,0%**
pH 8	**q SVM**	40	0,640	0,700	0,808	0,750	78,0%	72,5%
		100	0,680	0,714	0,769	0,740	81,0%	72,5%
	c SVM	40	0,640	0,700	0,808	0,750	78,1%	72,5%
		100	0,720	0,750	0,808	0,778	81,9%	76,5%
	LDA	**40**	**0,640**	**0,727**	**0,923**	**0,813**	**79,5%**	**78,4%**
		100	0,680	0,733	0,846	0,785	87,3%	76,5%

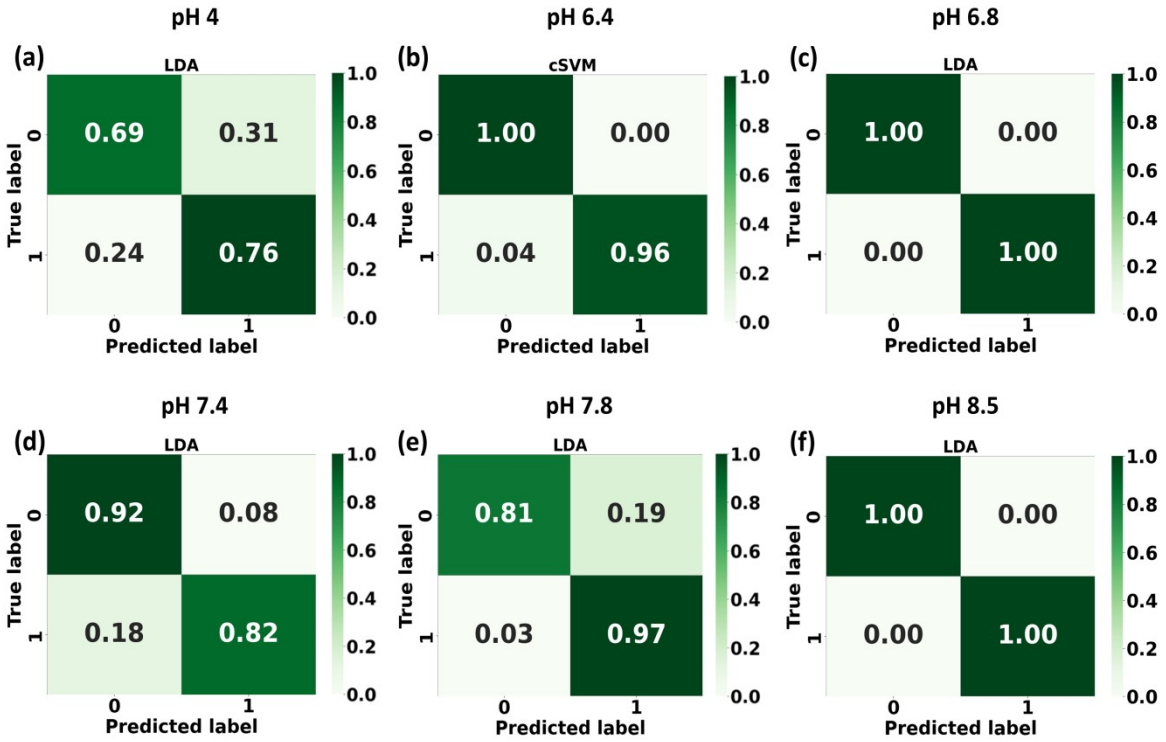

Figure S3: Confusion matrices for the test phase of OSM detection models with the highest accuracy obtained with LDA ranking for each pH. (a) pH 4, LDA model, 100 features; (b) pH 6.4, cSVM model, 100 features; (c) pH 6.8, LDA model, 100 features; (d) pH 7.4, LDA model, 100 features; (e) pH 7.8, LDA model, 100 features; (f) pH 8.5, LDA model, 100 features.

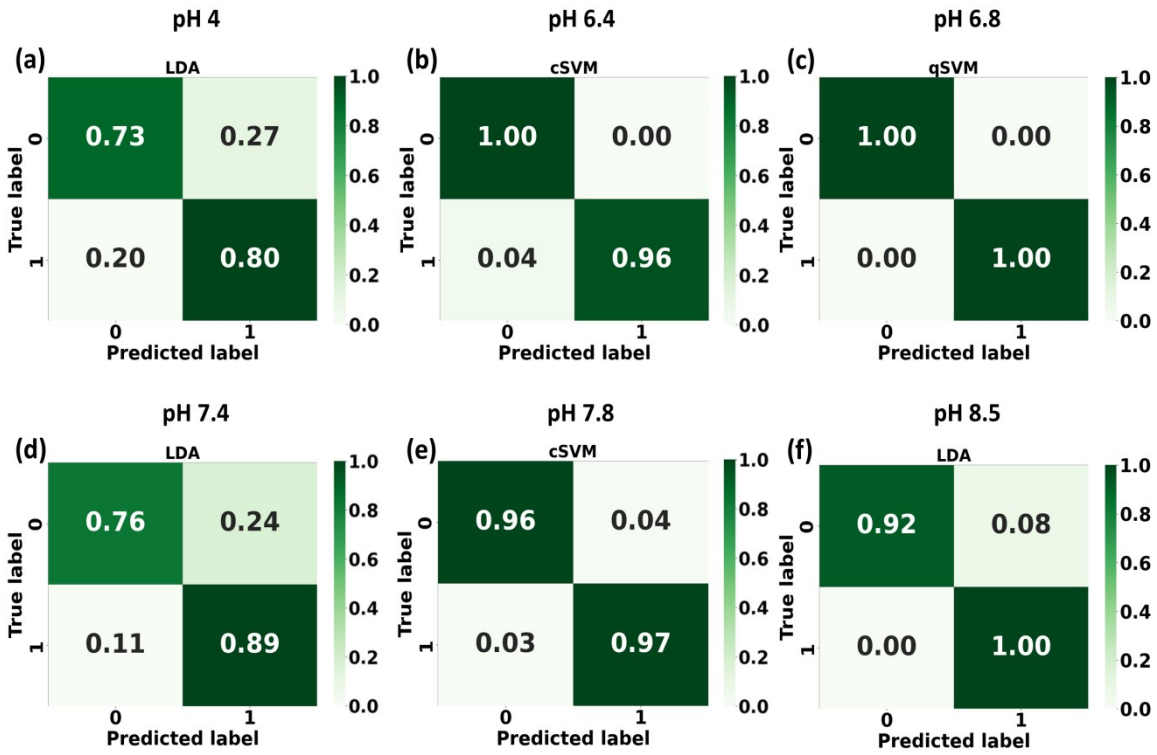

Figure S4: Confusion matrices for the test phase of OSM detection models with the highest accuracy with SVM ranking for each pH. (a) pH 4, LDA model, 40 features; (b) pH 6.4, cSVM, 100 features; (c) pH 6.8, qSVM, 40 features; (d) pH 7.4, LDA model,100 features; (e) pH7.8, cSVM,100 features; (f) pH 8.5, LDA model, 100 features

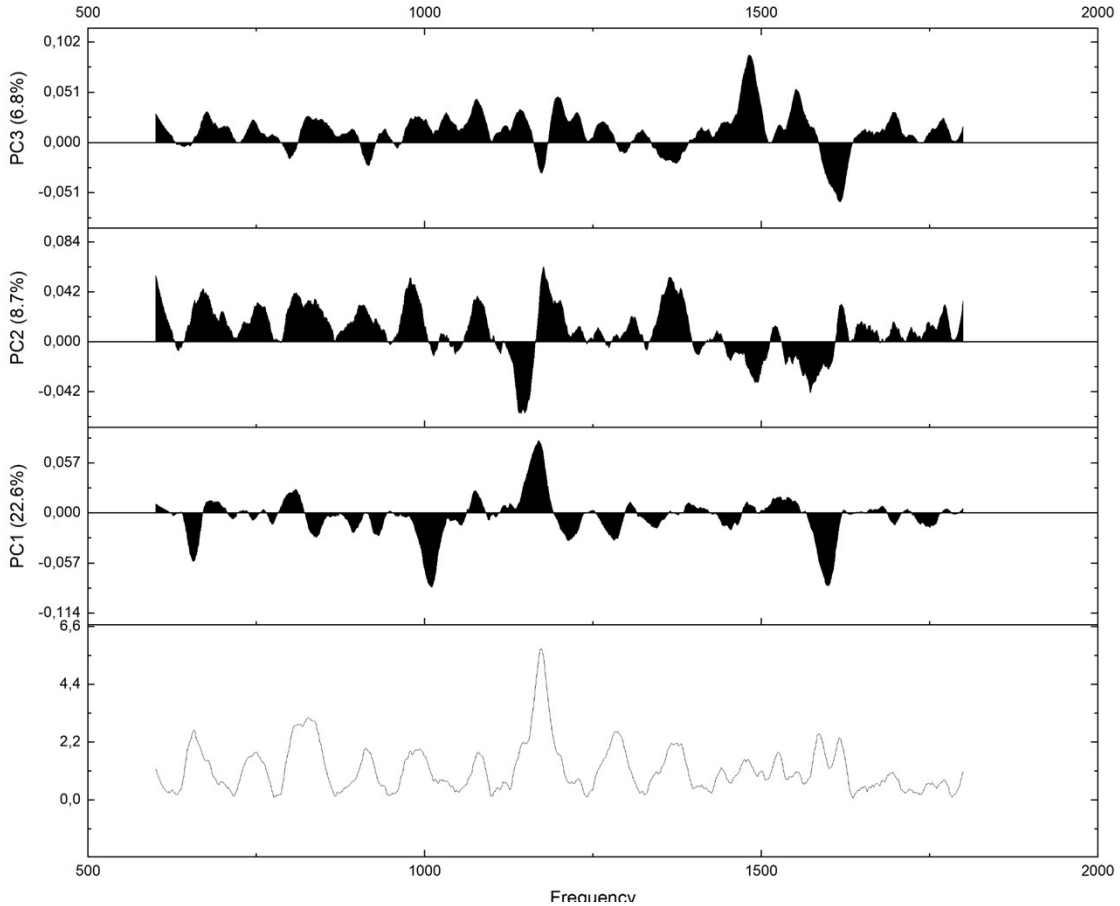

Figure S5: Loading plots of the first three principal components (PC1, PC2, and PC3) from PCA performed between F12K medium and F12K + OSM 25 ng/mL at optimal pH. The mean spectra for OSM acquisitions at the optimized pH are also shown.

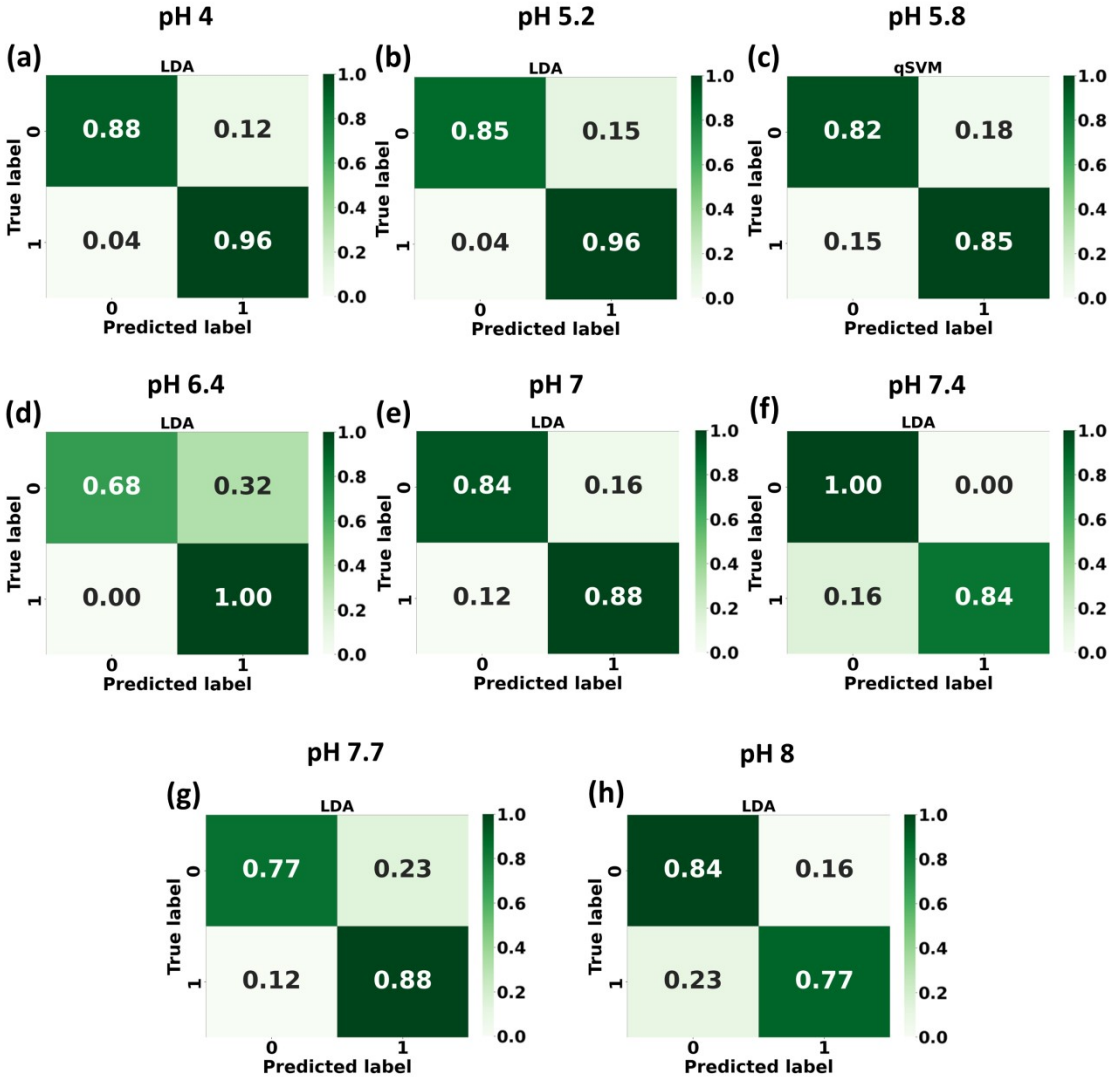

Figure S6: Confusion matrices for the test phase of TNF detection models with the highest accuracy with LDA ranking for each pH. (a) pH 4, LDA model,100 features; (b) pH 5.2, 100 features; LDA model (c) pH 5.8, qSVM model, 100 features; (d) pH 6.4, LDA model, 100 features; (e) pH 7, LDA model, 40 features; (f) pH 7.4, LDA model, 100 features; (g) pH 7.7, LDA model, 40 features; (h) pH 8, LDA model, 100 features

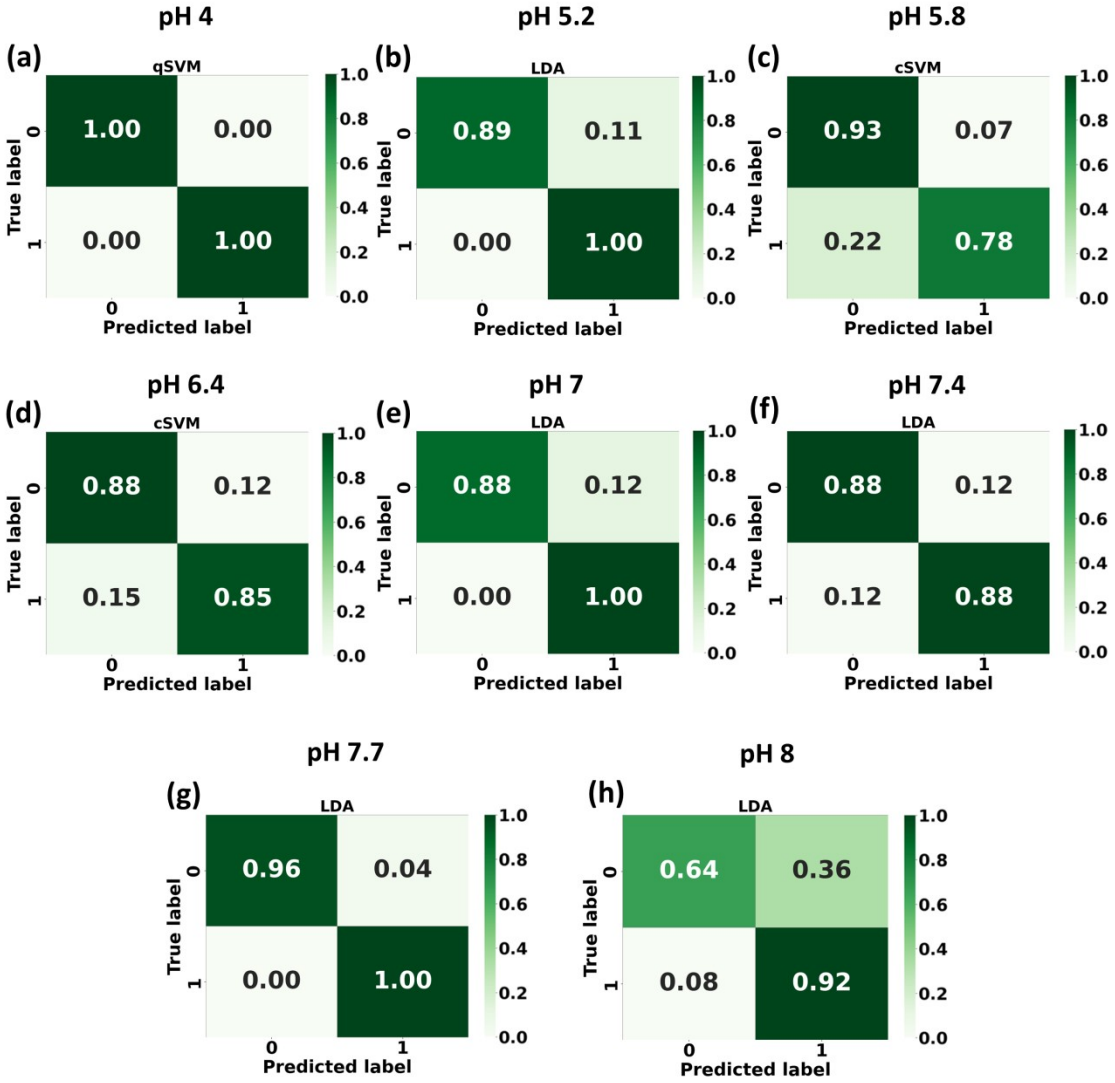

Figure S7: Confusion matrices for the test phase of TNF detection models with the highest accuracy with SVM ranking for each pH. (a) pH 4, qSVM model, 100 features; (b) pH 5.2, LDA model, 100 features; (c) pH 5.8, cSVM, 40 features; (d) pH 6.4, cSVM model, 100 features; (e) pH 7, LDA model, 40 features; (f) pH 7.4, LDA model, 40 features; (g) pH 7.7, LDA model, 100 features; (h) pH 8, LDA model, 40 features

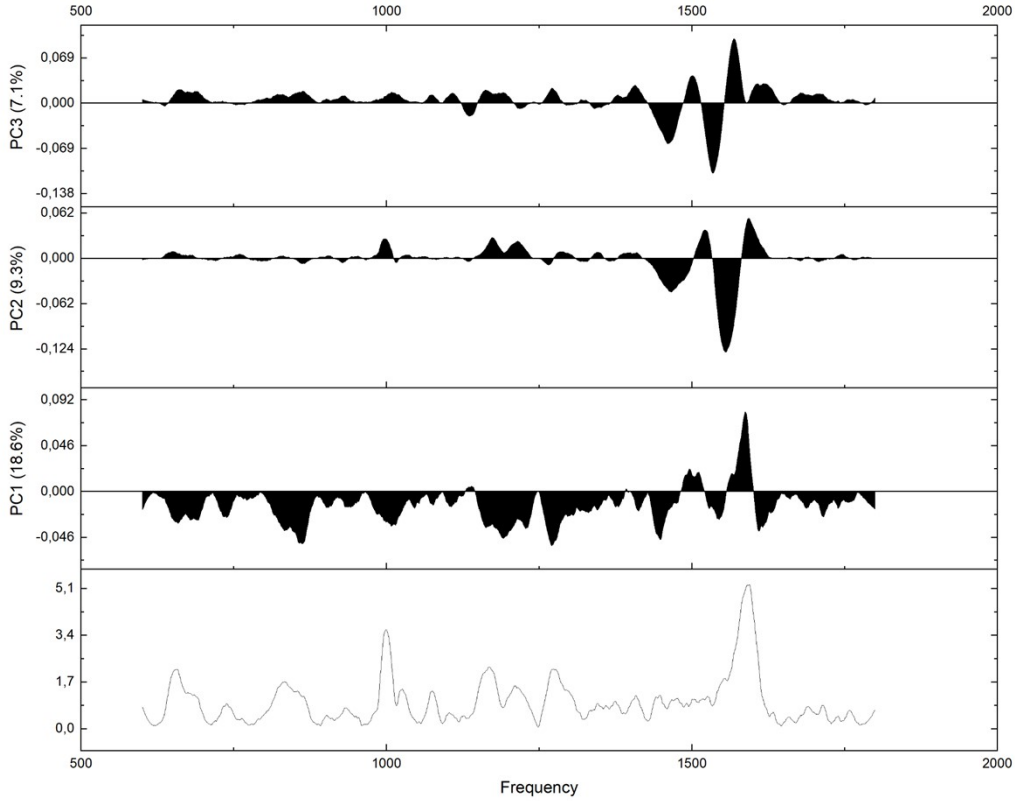

Figure S8: Loading plots of the first three principal components (PC1, PC2, and PC3) from PCA performed between F12K medium and F12K + TNF-α 100 pg/mL at optimal pH. The mean spectra for TNF-α acquisitions at the optimized pH are also reported.

Table S7: Optimized hyperparameters for Oncostatin M obtained model at optimal pH

Model	Hyperparameter	Value
Linear SVM	Kernel	Linear
	C	0.5801
	Coef0	1
	Gamma	Scale
	Tol	0.0077

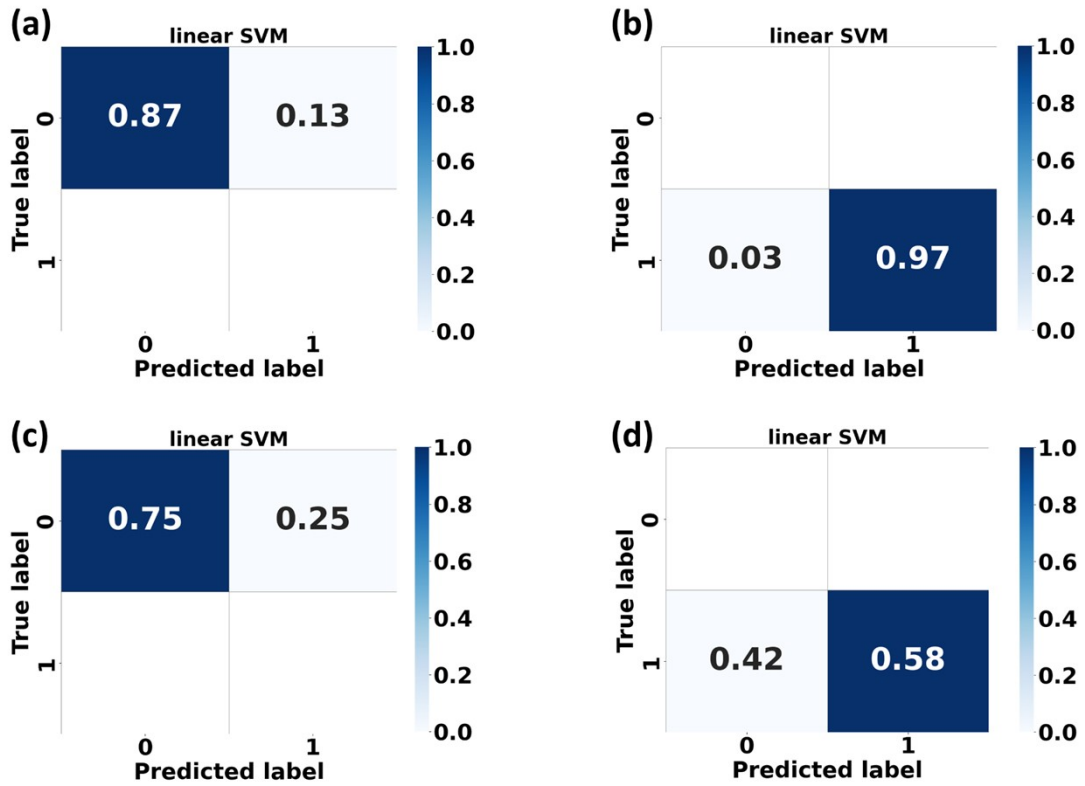

Figure S9: Confusion matrices for the test phase of the optimized Linear SVM models at optimal pH (6.8) for OSM detection, shown class by class. (a) Class 1: F12K medium (64 spectra), (b) Class 2: F12K + OSM (64 spectra), (c) Class 1: F12K + TNF (64 spectra), and (d) Class 2: F12K + TNF + OSM (64 spectra).

Table S8: Optimized hyperparameters for the TNF-α model obtained at optimal pH

Model	Hyperparameter	Value
Linear SVM	Kernel	Linear
	C	2.3369
	Coef0	1
	Gamma	Auto
	Tol	0.0482

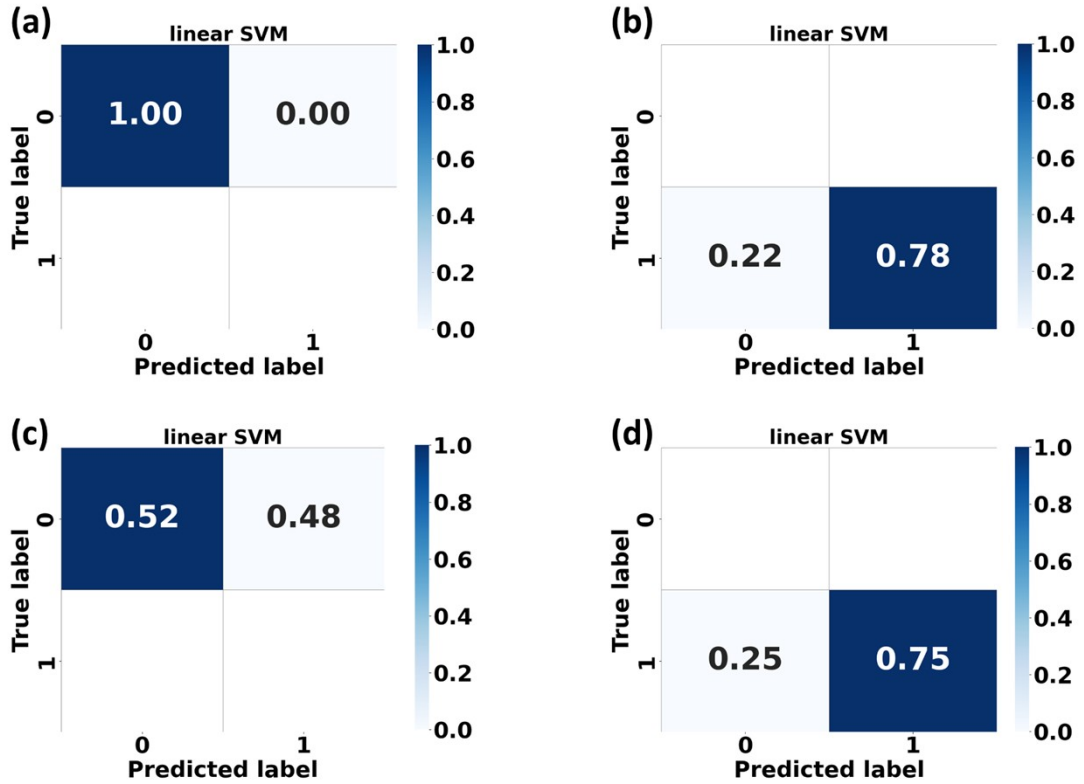

Figure S10: Confusion matrices for the test phase of the optimized Linear SVM models at optimal pH (4) for TNF detection, shown class by class. (a) Class 1: F12K medium (64 spectra), (b) Class 1: F12K + TNF (64 spectra), (c) Class 2: F12K + OSM (64 spectra), and (d) Class 2: F12K + TNF + OSM (64 spectra).

Supplementary Bibliography

[1] Azhar, U.; Ahmed, Q.; Ishaq, S.; Alwahabi, Z.T.; Dai, S. “Exploring Sensitive Label-Free Multiplex Analysis with Raman-Coded Microbeads and SERS-Coded Reporters”. Biosensors, 12, 121, Feb. 2022. doi: 10.3390/bios12020121

[2] Marques, F. C., Oliveira, G. P., Teixeira, R. A. R., Justo, R. M. S., Neves, T. B. V., & Andrade, G. F. S. . “Characterization of 11-mercaptoundecanoic and 3-mercaptopropionic acids adsorbed on silver by surface-enhanced Raman scattering”. *Vibrational Spectroscopy*, *98*, 139–144, (Aug. 2018). doi: 10.1016/j.vibspec.2018.07.015

[3] Kudelski, A. . Structures of monolayers formed from different HS—(CH $_2$) $_2$ —X thiols on gold, silver and copper: comparitive studies by surface-enhanced Raman scattering. *Journal of Raman Spectroscopy*, *34*(11), 853–862, (Oct. 2003). doi: 10.1002/jrs.1062

[4] I. N. Saraeva *et al.*, “Analysis of Skin Neoplasms’ Raman Spectra Using the Lorentz Approximation Method: Pilot Studies,” *JETP Lett*, vol. 119, no. 7, pp. 556–563, Apr. 2024, doi: 10.1134/S0021364023604153.

[5] M. A. Czarnecki, “Resolution Enhancement in Second-Derivative Spectra,” *Appl Spectrosc*, vol. 69, no. 1, pp. 67–74, Jan. 2015, doi: 10.1366/14-07568.

[6] S. Lee *et al.*, “Early-stage diagnosis of bladder cancer using surface-enhanced Raman spectroscopy combined with machine learning algorithms in a rat model,” *Biosens Bioelectron*, vol. 246, p. 115915, Feb. 2024, doi: 10.1016/j.bios.2023.115915.

[7] N. Pudjihartono, T. Fadason, A. W. Kempa-Liehr, and J. M. O’Sullivan, “A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction,” *Frontiers in Bioinformatics*, vol. 2, Jun. 2022, doi: 10.3389/fbinf.2022.927312.

[8] Z. Wang *et al.*, “Rapid Biomarker Screening of Alzheimer’s Disease by Interpretable Machine Learning and Graphene-Assisted Raman Spectroscopy,” *ACS Nano*, vol. 16, no. 4, pp. 6426–6436, Apr. 2022, doi: 10.1021/acsnano.2c00538.

[9] J. Brank, M. Grobelnik, N. Milic-Frayling, and D. Mladenic, “Feature Selection Using Linear Support Vector Machines,” Jun. 2002. [Online]. Available: https://www.microsoft.com/en-us/research/publication/feature-selection-using-linear-support-vector-machines/

[10] Y.-H. Lee, J. H. Won, Q.-S. Auh, and Y.-K. Noh, “Age group prediction with panoramic radiomorphometric parameters using machine learning algorithms,” *Sci Rep*, vol. 12, no. 1, p. 11703, Jul. 2022, doi: 10.1038/s41598-022-15691-9.

[11] T. Wang, M. Shoaran, and A. Emami, “Towards Adaptive Deep Brain Stimulation in Parkinson’S Disease: Lfp-Based Feature Analysis and Classification,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Apr. 2018, pp. 2536–2540. doi: 10.1109/ICASSP.2018.8462472.

[12] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, in IJCAI’95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995, pp. 1137–1143. doi: 10.5555/1643031.1643047.

[13] M. CHEMMAKHA, O. HABIBI, and M. LAZAAR, “Improving Machine Learning Models for Malware Detection Using Embedded Feature Selection Method,” *IFAC-PapersOnLine*, vol. 55, no. 12, pp. 771–776, 2022, doi: 10.1016/j.ifacol.2022.07.406.