

Supplementary Information

Attenuated total reflection Fourier-transform infrared spectral discrimination in human bodily fluids of oesophageal transformation to adenocarcinoma

Ishaan Maitra,^{a*} Camilo L. M. Morais,^a Kássio M. G. Lima,^{a,b*} Katherine M. Ashton,^c
Ravindra S. Date^c and Francis L. Martin^{a*}

*^aSchool of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston
PR1 2HE, UK*

*^bInstitute of Chemistry, Biological Chemistry and Chemometrics, Federal University of Rio
Grande do Norte, Natal, Brazil*

*^cLancashire Teaching Hospitals NHS Foundation Trust, Royal Preston Hospital, Preston
PR2 9HT, UK*

*ishaan.maitra@googlemail.com; kassiolima@gmail.com; flmartin@uclan.ac.uk

Table S1: Quality metrics (accuracy, sensitivity, specificity and F-scores) and multivariate classification methods (PCA-QDA and SPA-QDA) results for Normal vs. Inflammatory vs. Barrett's vs. LGD vs. HGD vs. OAC using **plasma** samples.

FOM	PCA-QDA model for Oesophageal stages					
	<i>Normal</i>	<i>Inflammatory</i>	<i>Barrett's</i>	<i>LGD</i>	<i>HGD</i>	<i>OAC</i>
Accuracy (%)	100	100	100	92.3	92.3	100
Sensitivity (%)	100	100	100	100	91.3	100
Specificity (%)	100	100	100	0	100	100
F-Scores (%)	100	100	100	0	95.4	100
FOM	SPA – QDA model for Oesophageal stages					
	<i>Normal</i>	<i>Inflammatory</i>	<i>Barrett's</i>	<i>LGD</i>	<i>HGD</i>	<i>OAC</i>
Accuracy (%)	84.6	84.6	96.1	100	96.7	92.3
Sensitivity (%)	94.7	86.3	95.2	100	100	95.2
Specificity (%)	57.1	75.0	100	100	66.6	80
F-Scores (%)	71.2	80.2	97.5	100	80	86.9

Table S2: Category-distinguishing wavenumbers for Normal vs. Inflammatory vs. Barrett's vs. LGD vs. HGD vs. OAC using **plasma** samples obtained for SPA-QDA and GA-QDA models.

Wavenumbers (cm ⁻¹)	Tentative Assignments
929	Left-handed helix DNA (Z form)
952	Symmetric stretching vibration of $\nu_1\text{PO}_4^{3-}$ (phosphate of HA)
987	OCH ₃ (polysaccharides-cellulose)
999	Ring stretching vibrations mixed strongly with CH in-plane bending
1018	$\nu(\text{CO})$, $\nu(\text{CC})$, $\delta(\text{OCH})$, ring (polysaccharides, pectin)
1053	$\nu\text{C-O}$ & $\delta\text{C-O}$ of carbohydrates
1072	Phosphate I band for two different C-O vibrations of Deoxyribose in DNA in disordering structure
1134	Oligosaccharide C-OH stretching band 2-Methylmannoside
1180	Amide III band region
1354	Stretching C-O, deformation C-H, deformation N-H
1381	δCH_3 Stretching C-O, deformation C-H, deformation N-H
1388	Carbon particle
1392	Less characteristic, due to aliphatic side groups of the amino acid residues
1431	$\delta(\text{CH}_2)$ (polysaccharides, cellulose)
1485	Deformation C-H
1539	Protein Amide II absorption- predominately β -sheet of Amide II
1585	Ring C-C stretch of phenyl
1624	Amide I region
1643	Amide I band (arises from C=O stretching vibrations)
1681	C=O Guanine deformation N-H in plane
1712	C=O
1724	C=O stretching band mode of the fatty acid ester

Table S3: Quality metrics (accuracy, sensitivity, specificity and F-scores) and multivariate classification methods (PCA-QDA, SPA-QDA and GA-QDA) results for Normal vs. Inflammatory vs. Barrett’s vs. LGD vs. HGD vs. OAC using **saliva** samples.

FOM – PCA-QDA	Oesophageal stages					
	<i>Normal</i>	<i>Inflammatory</i>	<i>Barrett’s</i>	<i>LGD</i>	<i>HGD</i>	<i>OAC</i>
Accuracy (%)	92.6	100	96.3	88.8	92.6	92.6
Sensitivity (%)	89.4	100	100	96	91.6	100
Specificity (%)	100	100	80	0	100	60
F-Scores (%)	94.4	100	88.9	0	95.6	75
FOM – SPA-QDA	Oesophageal stages					
	<i>Normal</i>	<i>Inflammatory</i>	<i>Barrett’s</i>	<i>LGD</i>	<i>HGD</i>	<i>OAC</i>
Accuracy (%)	88.8	85.2	96.2	100	96.3	96.3
Sensitivity (%)	100	86.9	95.4	100	100	95.4
Specificity (%)	62.5	75	100	100	66.6	100
F-Scores (%)	76.9	80.5	97.6	100	80	97.7
FOM – GA-QDA	Oesophageal stages					
	<i>Normal</i>	<i>Inflammatory</i>	<i>Barrett’s</i>	<i>LGD</i>	<i>HGD</i>	<i>OAC</i>
Accuracy (%)	92.5	92.6	100	92.6	88.9	96.2
Sensitivity (%)	100	91.3	100	100	91.6	95.4
Specificity (%)	75	100	100	0	66.6	100
F-Scores (%)	85.7	95.4	100	0	77.1	97.7

Table S4: Category-distinguishing wavenumbers for Normal vs. Inflammatory vs. Barrett's vs. LGD vs. HGD vs. OAC using **saliva** samples obtained for SPA-QDA and GA-QDA models.

Wavenumbers (cm ⁻¹)	Tentative Assignments
902	Phosphodiester region
991	C-O ribose, C-C
1003	Ring stretching vibrations mixed strongly with CH in-plane bending
1014	$\nu(\text{CO})$, $\nu(\text{CC})$, $\delta(\text{OCH})$, ring (polysaccharides, pectin)
1068	Stretching C-O ribose
1099	Stretching PO_2^- symmetric (phosphate II)
1107	$\nu(\text{CO})$, $\nu(\text{CC})$, $\delta(\text{OCH})$, ring (polysaccharides, pectin)
1431	$\delta(\text{CH}_2)$ (polysaccharides, cellulose)
1558	Ring base
1585	Ring C-C stretch of phenyl
1589	Ring C-C stretch of phenyl
1604	Adenine vibration in DNA
1624	Peak of nucleic acids due to the base carbonyl stretching and ring breathing mode
1643	Amide I band (arises from C=O stretching vibrations)
1689	Peak of nucleic acids due to the base carbonyl stretching and ring breathing mode
1697	C ₂ =O guanine
1701	C5=O guanine
1716	C=O thymine
1743	C=O stretching mode of lipids
1778	$\nu(\text{C}=\text{C})$ lipids, fatty acids
1786	$\nu(\text{C}=\text{C})$ lipids, fatty acids

Table S5: Quality metrics (accuracy, sensitivity, specificity and F-scores) and multivariate classification methods (PCA-QDA, SPA-QDA and GA-QDA) results for Normal vs. Inflammatory vs. Barrett’s vs. LGD vs. HGD vs. OAC using **serum** samples.

FOM – PCA-QDA	Oesophageal stages					
	<i>Normal</i>	<i>Inflammatory</i>	<i>Barrett’s</i>	<i>LGD</i>	<i>HGD</i>	<i>OAC</i>
Accuracy (%)	92.0	92.0	96.0	92.0	92.0	96.0
Sensitivity (%)	100	90.4	100	95.6	95.4	95.2
Specificity (%)	71.4	100	80	50.0	66.6	100
F-Scores (%)	83.3	95.0	88.8	65.6	78.5	97.5
FOM – SPA-QDA	Oesophageal stages					
	<i>Normal</i>	<i>Inflammatory</i>	<i>Barrett’s</i>	<i>LGD</i>	<i>HGD</i>	<i>OAC</i>
Accuracy (%)	92.0	88.0	92.0	92.0	92.0	96.0
Sensitivity (%)	100	90.4	95.0	95.6	95.4	95.2
Specificity (%)	71.4	75.0	80.0	50.0	66.6	100
F-Scores (%)	93.3	82.0	86.8	65.6	78.5	97.5
FOM – GA-QDA	Oesophageal stages					
	<i>Normal</i>	<i>Inflammatory</i>	<i>Barrett’s</i>	<i>LGD</i>	<i>HGD</i>	<i>OAC</i>
Accuracy (%)	96.0	96.0	96.0	92.0	96.0	100
Sensitivity (%)	100	95.2	100	95.6	95.4	100
Specificity (%)	85.7	100	80.0	50.0	100	100
F-Scores (%)	92.3	97.5	88.8	65.6	97.6	100

Table S6: Category-distinguishing wavenumbers for Normal vs. Inflammatory vs. Barrett's vs. LGD vs. HGD vs. OAC using **serum** samples obtained for SPA-QDA and GA-QDA models.

Wavenumbers (cm ⁻¹)	Tentative Assignments
1041	Symmetric PO ₂ ⁻ stretching in RNA and DNA
1315	Amide III band components of proteins
1319	Amide III band components of proteins Collagen
1330	CH ₂ wagging
1338	CH ₂ wagging
1435	δ (CH ₂) (polysaccharides, cellulose)
1442	δ(CH ₂), lipids, fatty acids
1446	δ(CH ₂), lipids, fatty acids
1477	CH ₂ bending of the methylene chains in lipids
1492	C=C, deformation C-H
1539	Amide II
1573	C=N adenine
1593	Ring C-C stretch of phenyl
1600	Amide I band of proteins
1631	Ring C-C stretch of phenyl
1654	Amide I (of proteins in α-helix conformation)
1662	Amide I band
1743	C=O stretching mode of lipids

Table S7: Quality metrics (accuracy, sensitivity, specificity and F-scores) and multivariate classification methods (PCA-QDA, SPA-QDA and GA-QDA) results for Normal vs. Inflammatory vs. Barrett’s vs. LGD vs. HGD vs. OAC using urine samples.

FOM – PCA-QDA	Oesophageal stages					
	<i>Normal</i>	<i>Inflammatory</i>	<i>Barrett’s</i>	<i>LGD</i>	<i>HGD</i>	<i>OAC</i>
Accuracy (%)	100	100	100	100	100	100
Sensitivity (%)	100	100	100	100	100	100
Specificity (%)	100	100	100	100	100	100
F-Scores (%)	100	100	100	100	100	100
FOM – SPA-QDA	Oesophageal stages					
	<i>Normal</i>	<i>Inflammatory</i>	<i>Barrett’s</i>	<i>LGD</i>	<i>HGD</i>	<i>OAC</i>
Accuracy (%)	66.6	77.7	92.6	96.2	92.6	77.7
Sensitivity (%)	78.9	78.2	95.4	96.0	100	90.9
Specificity (%)	35.5	75.0	80.0	100	33.3	20.0
F-Scores (%)	50.8	76.6	87.1	97.9	50.0	32.7
FOM – GA-QDA	Oesophageal stages					
	<i>Normal</i>	<i>Inflammatory</i>	<i>Barrett’s</i>	<i>LGD</i>	<i>HGD</i>	<i>OAC</i>
Accuracy (%)	74.1	85.2	92.6	92.6	96.3	85.2
Sensitivity (%)	78.9	86.9	95.4	96.0	95.8	100
Specificity (%)	62.5	75.0	80.0	50.0	100	20.0
F-Scores (%)	69.7	80.5	87.0	65.7	97.8	33.3

Table S8: Category-distinguishing wavenumbers for Normal vs. Inflammatory vs. Barrett's vs. LGD vs. HGD vs. OAC using **urine** samples obtained for SPA-QDA and GA-QDA models.

Wavenumbers (cm ⁻¹)	Tentative Assignments
906	Phosphodiester region
956	C-O deoxyribose, C-C
995	Ring breathing
1018	Glycogen
1030	Glycogen
1095	$\nu_{as}PO_2^-$
1118	$\nu(CO)$, $\nu(CC)$, $\delta(OCH)$, ring (polysaccharides, pectin)
1141	Symmetric PO_2^- stretching in RNA and DNA
1242	Amide III collagen
1253	Amide III
1334	$\delta(CH)$, ring (polysaccharides, pectin)
1381	δCH_3
1431	Lipids
1446	$\delta(CH_2)$, lipids, fatty acids
1500	In-plane CH bending vibration from the phenyl rings
1550	Amide II of proteins
1562	Amide II of proteins (<i>e.g.</i> , side-chain carboxyl groups)
1577	Glutamate carboxylate stretching
1600	C=N cytosine, N-H adenine
1651	Amide I
1681	C=O Guanine deformation N-H in plane
1712	C=O thymine
1724	C=O stretching band mode of the fatty acid ester
1735	C=O stretching (lipids)
1777	$\nu(C=C)$ lipids, fatty acids