Chapter - 4

Results & Discussion

Chapter 4

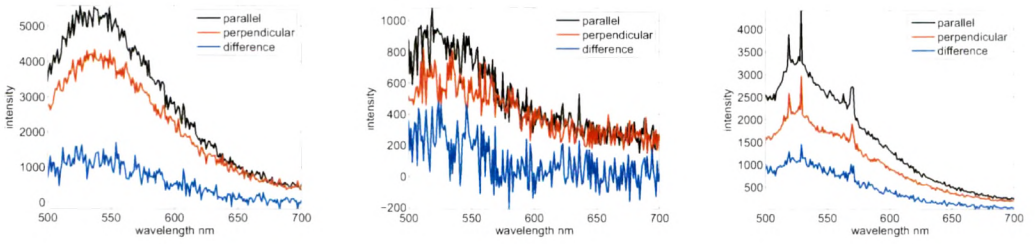
Results & Discussion

Analysis of Human Breast Tissue
Application of Wavelet Transform

In recent time wavelet transform which is also known as a mathematical microscope, has emerged as a powerful tool for the analysis of transient data. It is useful in disentangling characteristic variations at different scales (1), due to its multi resolution ability. Hence one can simultaneously have both time and frequency information, which is not possible through standard approaches like Fourier transform. This linear transform isolates local features and leads to a convenient dimensional reduction of the data in the form of average coefficients, resembling the data itself. The wavelet coefficients, at various levels, encapsulate the variations at corresponding scales.

4.1 Discrete Wavelet Transform

For our application, we make use of the simplest Haar wavelet, since the interpretation of the wavelet coefficients is quite transparent here; it is also free from artifacts arising due to the finite size of the data. Haar basis is special, since it is symmetric and compactly supported. If the data set contains 2^N elements, with N an integer then one can have a N-level decomposition. The pictorial demonstration of discrete wavelet transform is presented in fig. 3.8. $N = 0$ level represents



(a) Cancer (b) Benign (c) Normal

Figure 4.1: Typical autofluorescence intensity plots for parallel & perpendicular component and their differences for (a)Cancer, (b) Benign and (c) Normal tissues.

the original signal. In level one i.e., $N = 1$ Haar wavelet decomposition, the nearest neighbor averages and differences are calculated with the normalization factor of $1/\sqrt{2}$, which leaves half of the data in the form of low-pass coefficients and the other half in terms of level-1 high-pass coefficients. Subsequently, the same procedure can be applied once more to the low-pass coefficients to decompose them into level-2 high-pass coefficients and level-2 low-pass coefficients. In total N level decomposition can be carried out. It is worth mentioning that in case of Haar wavelet finding low pass coefficients at different levels amounts to binning of data in appropriate window sizes.

The samples were excited by 488-nm wavelength plane polarized light from an Ar-ion laser and the components of fluorescence light that are parallel and perpendicular to the incident polarized light were measured in the 500- to 700-nm wavelength region. Typical plots of the parallel, perpendicular and difference of parallel and perpendicular components of fluorescence spectra of cancer, benign and normal tissues are shown in fig. 4.1. Emission peaks of flavins and flavoproteins are identifiable in normal tissues at lower wavelengths.

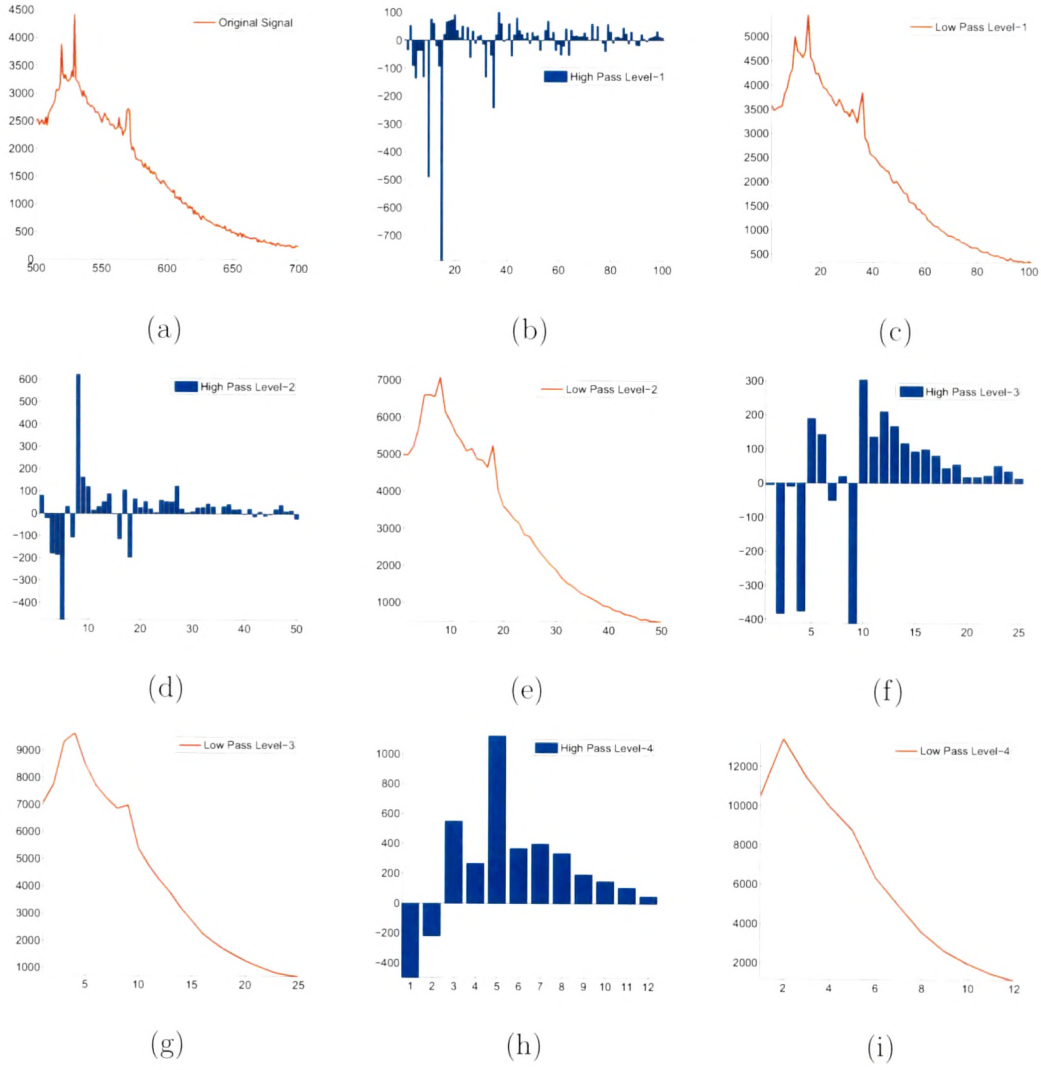


Figure 4.2: Represents, (a) the fluorescence spectrum (parallel component) of a normal tissue, whose high-pass wavelet coefficients are displayed in (b) level-1, (d) level-2, (f) level-3, (h) level-4 and (c), (e), (g), (i) showing the corresponding low pass coefficients.

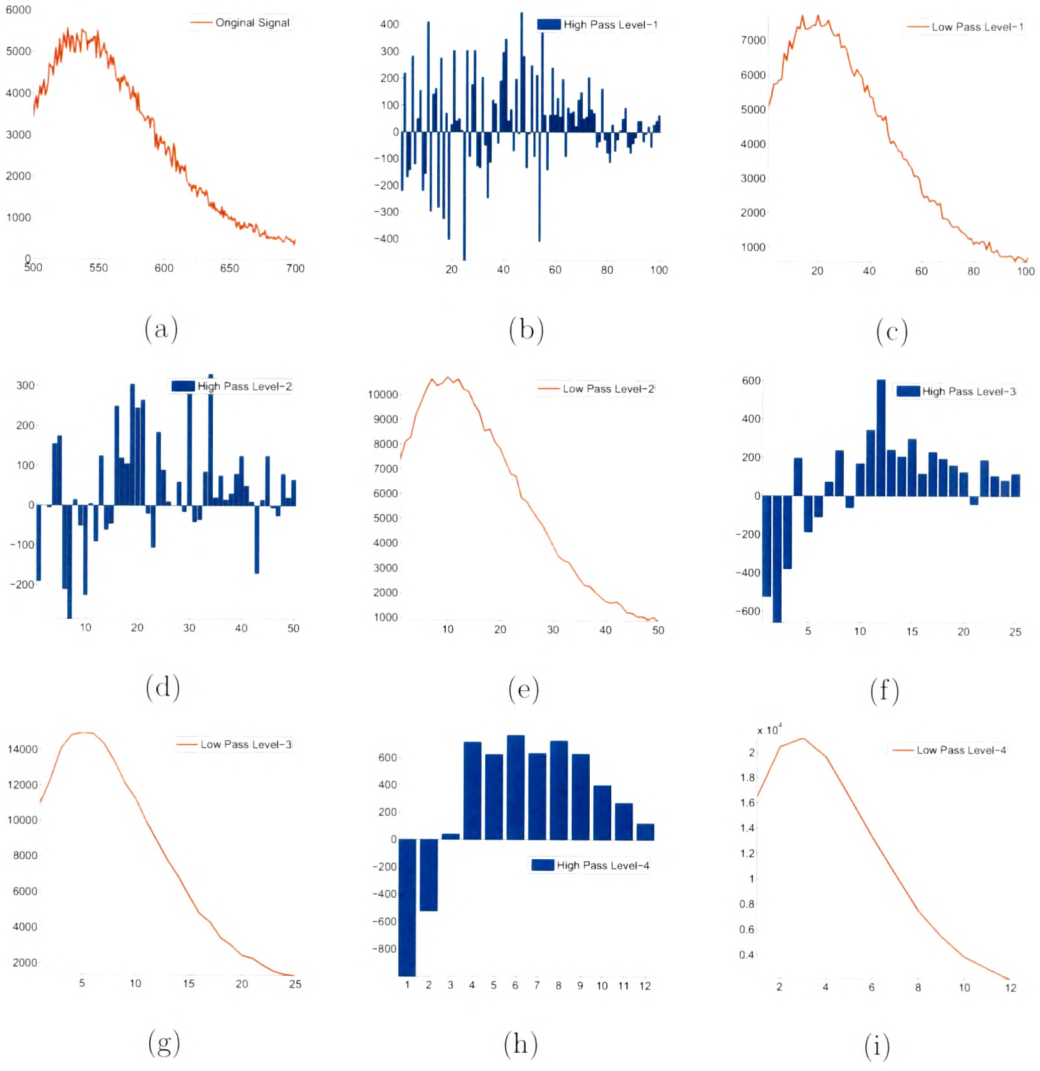


Figure 4.3: Represents, (a) the fluorescence spectrum (parallel component) of a cancer tissue, whose high-pass wavelet coefficients are displayed in (b) level-1, (d) level-2, (f) level-3, (h) level-4 and (c), (e), (g), (i) showing the corresponding low pass coefficients.

We have considered 192 intensity values, from 500nm to 691nm for analysis. The level-1 low pass coefficients are the averages of nearest neighbor data points, level-2 low pass coefficients represent averages of 4-data points etc. It is clear that, there are 96 level-1 coefficients, 48 level-2 coefficients etc. For 192 data points, we can have a six level decomposition, with 3 sixth level low pass coefficients. It is observed that, progressive averaging of data in discrete wavelet transform reduces statistical and experimental uncertainties. The variations at lower level represent the high frequency components, corresponding to variations in smaller window sizes. The original spectral profiles (for parallel component) and their four level decomposition for normal, cancerous and benign tissues are given in fig. 4.2, 4.3 & 4.4 respectively. Several features like the resemblance of the low-pass coefficients to corresponding spectral profiles, prominent high-pass coefficients at various levels, showing significant variations are clearly identifiable. While computing wavelet power one take the sum of the square of wavelet coefficients at particular level e.g., low pass power at level-4 corresponds to sum of square of low pass coefficients of 4th level. Usually for the sake of normalization, we normalize it i.e., divides it by total power P which is the sum of the squares of all the data points. Similarly one can compute high pass power. In wavelet transform total power is conserved i.e., sum of low pass and high pass power yield the total power. It is worth mentioning that wavelet decomposition obeys Parseval's theorem. Physically it means that,

$$P \equiv \sum {I_i}^2 = \sum {c_k}^2 + \sum_{j,k} d_{j,k}^2 \tag{4.1}$$

Fig. 4.5 depicts the histogram plots of the high pass coefficients of level-1 of the different tissue types, showing variations between the tissues type.

4.1.1 Results and Discussion of Difference of Parallel and Perpendicular Component

In this section the results of systematic analysis of the average coefficients of the wavelet transform, of the difference between parallel and perpendicular components of the fluorescence spectra, from normal and malignant human breast

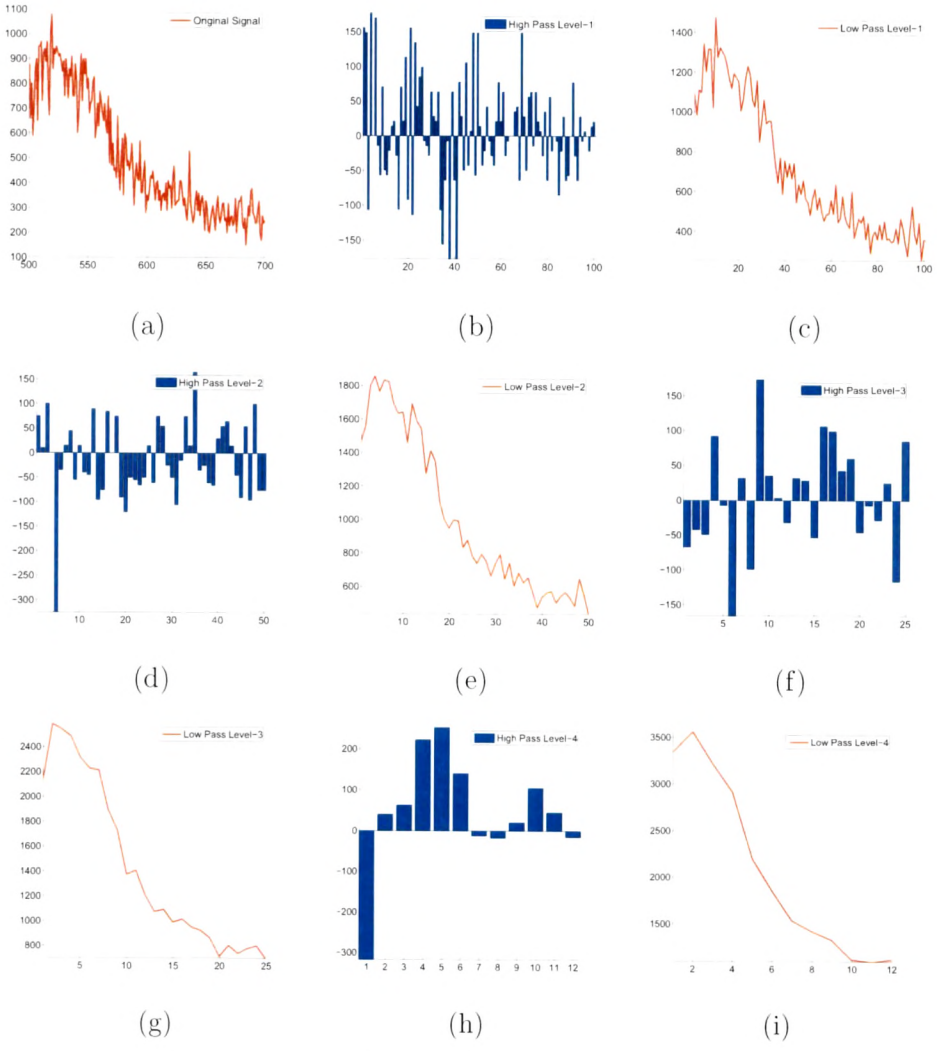
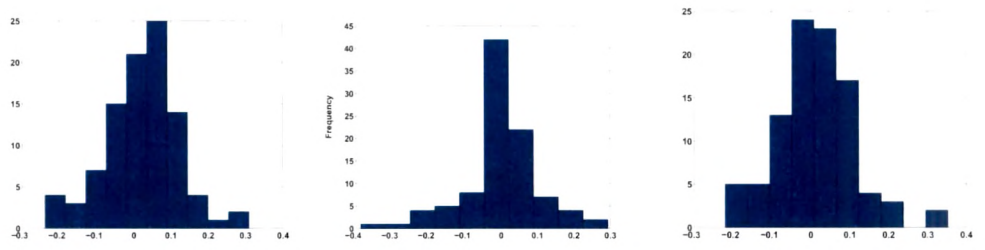


Figure 4.4: Represents, (a) the fluorescence spectrum (parallel component) of a benign tissue, whose high-pass wavelet coefficients are displayed in (b) level-1, (d) level-2, (f) level-3, (h) level-4 and (c), (e), (g), (i) showing the corresponding low pass coefficients.



(a) (b) (c)

Figure 4.5: Histogram of the normalized wavelet coefficient for (a) cancer, (b) benign and (c) normal tissues of level-1.

tissues are presented. The average coefficients in the wavelet domain are less sensitive to experimental and statistical uncertainties. The parallel component of the fluorescence suffers fewer scattering events. In comparison, the intensity of the perpendicular component is not only affected more by scatterers, but is also quite sensitive to absorption, since the path traversed by the same in the tissue medium is more. Hence, the difference of parallel and perpendicular intensities, apart from being relatively free from the diffusive component (2; 3), can be quite sensitive to microscopic biochemical changes including the effects of absorption in different tissue types. A number of parameters, capturing spectral features and subtle changes in the intensity profile of the diseased tissues, as compared to their normal counterparts, are identified in the wavelet domain. The physical origin of one of the distinguishing parameters may be ascribed to the changes in the concentration of porphyrin and the density of cellular organelles present in tumors (4; 5).

It is found that a clear difference between cancer and normal tissues emerge at fourth level. In particular, relatively higher value for a low pass coefficient at the third quarter of the 4^{th} level decomposition is seen in cancer tissues as shown in fig. 4.6(b). It should be noted that at the 4^{th} level, each coefficient corresponds to average of 16 original intensity values. This structure is absent at the fifth level, as seen in fig. 4.6(c), because of averaging over a bigger window size and is not transparent at the third level as it is clear from fig. 4.6(a). So a suitable block averaging highlights the presence of a weak fluorophore (porphyrin) in cancerous spectra which is not present in the normal case. The 5^{th} level low pass coefficients

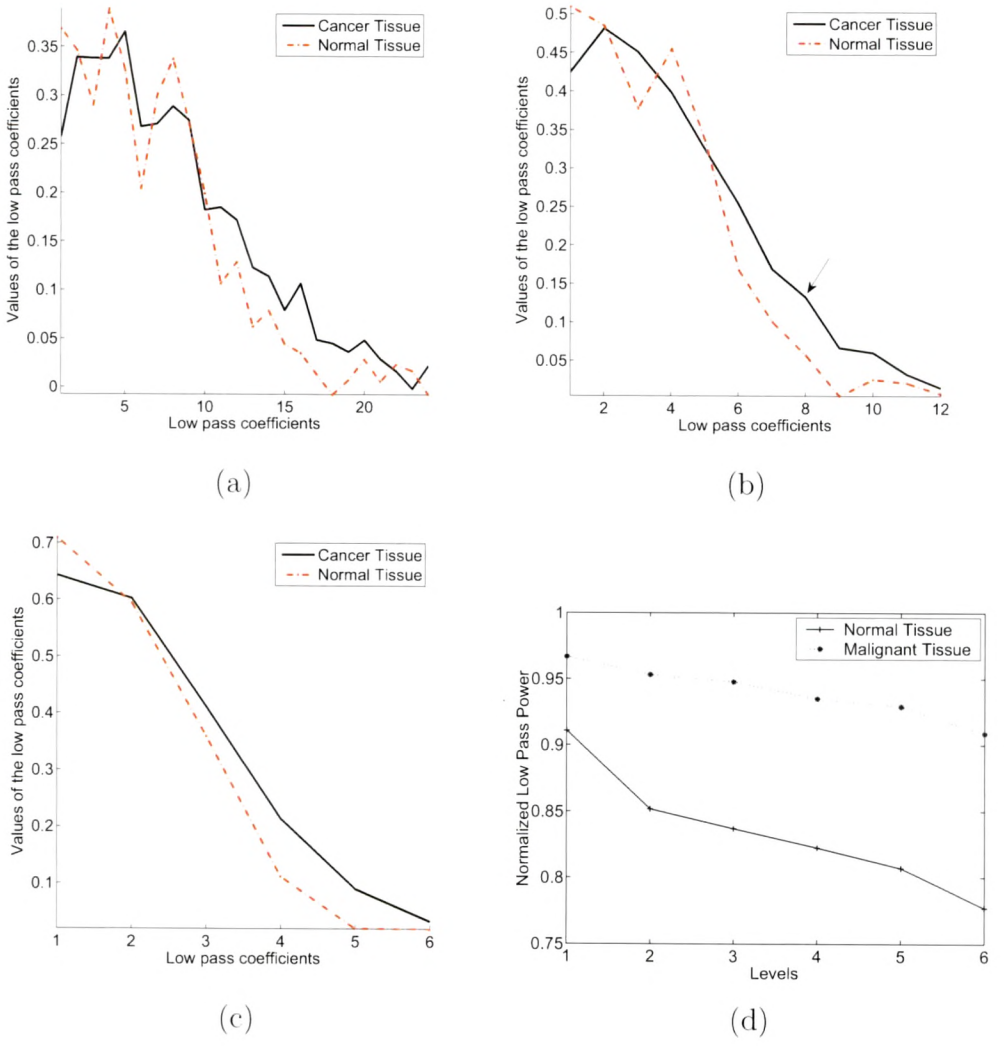
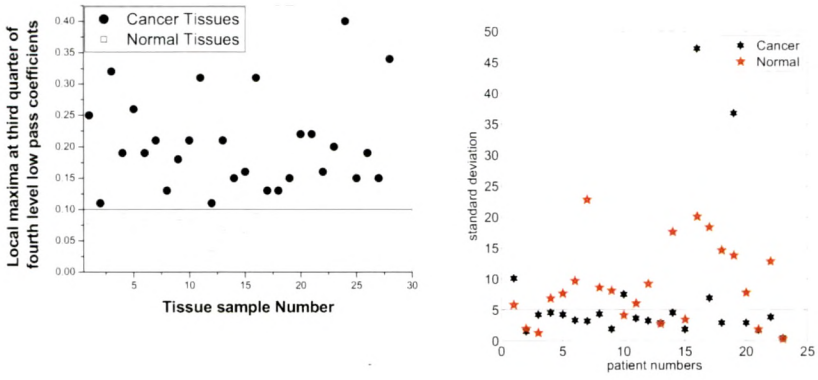


Figure 4.6: Plots of low pass coefficients of cancer and normal tissues for (a) level-3 (b) level-4 and (c) level-5 average coefficients. The fourth-level coefficients highlight the 630 nm weak emission peak in cancerous tissue. 6(d) shows a typical low pass power plot of the difference of parallel and perpendicular components of the fluorescence spectra of cancer and normal tissues. The rate of decrease of low pass power, as a function of levels, is slower in cancer tissues.



(a) (b)

Figure 4.7: Typical (a) Local maxima at third quarter of fourth level low-pass coefficients of cancer and normal breast tissues and (b) Standard deviations of the spectral fluctuations of the cancer and normal tissues, captured through the high pass coefficients at level-1.

do not show significant variation between cancer and normal tissues, except at the tail end, where the cancer values fall off slowly as compared to their normal counter parts. For the purpose of comparison between normal and cancer tissue fluorescence the low pass coefficients at 3^{rd}, 4^{th} and 5^{th} levels are shown in fig. 4.6. As pointed out earlier, at the 4^{th} level the low pass coefficients of the cancer tissue fluorescence show a minor peak at the 3^{rd} quadrant. The same is averaged out at 5^{th} and at 3^{rd} level, one sees many minor peaks apart from one at the 3^{rd} quadrant. Significantly the normalized wavelet powers representing the strength of the variation at different levels also differ between normal and cancer tissues at the same level. A typical plot revealing this aspect is given in fig. 4.6(d). It may be noted from equation(4.1) that the low pass power is the sum of square of low pass coefficients at that level and hence would highlight the differences in low pass coefficients between normal and cancerous tissues. The behavior of the low pass powers at different levels also reveal substantial differences between tissue types, as is clearly indicated in Table-1.

In case of Haar wavelet fourth level corresponds to 16 intensities values of original data set in third level it corresponds to 8 points whereas in fifth level it corresponds to 32 points. The fact that porphyrin peak manifests at the fourth

level indicates that the spectral width of this fluorophores is of the order of above window size. The absence of this fluorophores in case of normal samples makes the intensities fall faster in 620-670 nm regime. When a block averaging with window size ranging from 10, 12, 14, 16 and 18 nm is done, it is found that the peak at the third quadrant begins to manifest from the window size 14 nm and starts decreasing around 18 nm.

These normalized low-pass coefficients in the third quarter of the 4th level have been considered for illustrating the minor peak. The local maxima at third quarter of fourth level low-pass coefficients of cancer samples are more than 0.1 while those of normal tissues are less than 0.1, with a sensitivity of 100% and specificity of 83%. The scatter plot is shown in fig. 4.7(a). It should be noted here that the values for normal tissues which are more than 0.1 still show lower values than the corresponding tumors, consistent with all the other samples. Thus intra-patient diagnosis gives a clear distinction between cancer and normal tissues. Variations in inter-patient diagnosis may be due to the fact that, the growth of tumor depends on genetic (major genes, modifier genes) and non-genetic factors (birth, age, weight/diet, exercise, environmental exposures, etc) (6). The above fourth level low-pass coefficient, originate from the fluorescence signals around 630nm of the original data, corresponding to the porphyrin emission peak.

FAD and porphyrin are the major fluorophores that fluoresce in the visible wavelength regime, with peak intensities at 530 and 630 nm respectively. These fluorophores are considered as contrast agents for cancer detection (7; 8). It has been suggested that deficiency of ferrochelatase, the enzyme required for conversion of protoporphyrin IX (PpIX) to heme, in tumors results in accumulation of PpIX in these tissues relative to the normal ones (8). Such accumulation changes the relative concentration of these fluorophores thus altering the fluorescence spectra significantly, which in turn changes the peak heights of the emission bands of the two fluorophores. The scattering centers are known to enhance the fluorescence intensity (9). Thus the large size of cell suspensions, higher density of cells and accumulation of more porphyrin in tumors may possibly contribute to the small peak at 630 nm wavelength region. Since all tissues studied in this work were non-necrotic and non-fungating, as corroborated histopathologically, protoporphyrin emission may not be due to bacterial action (10).

A systematic analysis yields that, the low pass powers at 4^{th} level consistently differ between tissue types. Table-1 depicts the ratios of normalized low pass powers of cancerous and corresponding normal tissues at the fourth level. The third column shows the ratios and the second indicates the tissue type. It is observed from the table that, the ratio of the low pass powers is greater than one in ductal carcinoma (grade I, II, III) with a sensitivity of 80%. Interestingly it is found that, this ratio is less than one in tissues showing metastasis or with fatty tissues attached. The power at different levels captures the variance at that level.

It is found that, normalized low pass powers of cancerous tissues decrease more slowly as a function of levels, as compared to their normal counter parts. Table-1 also gives the corresponding slope values, capturing these differences in the rate of fall; the fourth column shows the differences between the slope values. The slopes have been calculated by the linear fit of low pass power profile. The values of normal tissues are more negative than the cancerous ones, indicating a sharper fall in the intensity profile. Very interestingly, like the low pass power, the cancerous samples with attached fatty tissues and the tissue types which are positive for metastasis show a higher steepness than the corresponding normal ones.

It may be noted that the above-mentioned three parameters also distinguish tumors of different grades. It is found that for grades I and II cancerous tissues, the values of the local low-pass maxima at the third quadrant are less than 0.2, but more than 0.2 in the grade III cancers, with a sensitivity of 75%.

It is worth emphasizing that in the discrete wavelet transform, the average low pass coefficients capture the broad band features in the spectra at a suitable scale and the high pass coefficients extract the high frequency behavior. The orthogonal property ensures that the extracted features at different scales are independent of each other. It is clear that the high pass coefficients can capture characteristic spectral variations, as also spectral fluctuations at different scales, whose statistical property can be studied for tissue discrimination.

In order to show the advantage of wavelet transform over the straightforward binning procedure (low pass), the high- pass features (since the high pass coefficients are differences of intensities) have been carefully analyzed. It is found

TABLE-1

S. No.	Description of tissue type	Ratio of 4^{th} level low pass power of cancer and normal	Slope values of the normalized six level low pass power		
			Cancer	Normal	Difference
1	Ductal carcinoma grade-III	1.2259	-0.011474	-0.037925	-0.0265
2	Cystosarcoma	1.4979	-0.102240	-0.108020	-0.0058
3	Ductal carcinoma grade-III	0.5897	-0.012917	-0.012797	0.0001
4	Intraductal carcinoma (poorly defined)	1.0675	-0.011890	-0.016320	-0.0044
5	Ductal carcinoma grade-III	1.1369	-0.102240	-0.108020	-0.0058
6	Ductal carcinoma	1.0230	-0.008624	-0.016169	-0.0075
7	Cystosarcoma	0.8361	-0.037675	-0.020604	0.0171
8	Ductal carcinoma grade-II positive for metastasis	0.9059	-0.023613	-0.015374	0.0082
9		1.0135	-0.015041	-0.017573	-0.0025
10	Ductal carcinoma grade-III positive for metastasis	0.9926	-0.010214	-0.011873	-0.0017
11	Ductal carcinoma grade-II	1.0386	-0.031467	-0.026810	0.0047
12	Ductal carcinoma	0.5150	-0.015834	-0.025814	-0.0100
13	Ductal carcinoma	1.0015	-0.005429	-0.006476	-0.0010
14	Ductal carcinoma grade-II	1.0187	-0.006719	-0.015287	-0.0086
15	Ductal carcinoma grade-II	1.1886	-0.005074	-0.059195	-0.0541
16	Ductal carcinoma grade-III	1.0205	-0.009140	-0.019149	-0.0100
17	Ductal carcinoma grade-II	1.1026	-0.007558	-0.021904	-0.0143
18	Ductal carcinoma grade-III	1.0083	-0.006816	-0.092203	-0.0854
19	Ductal carcinoma grade-III positive for metastasis	0.9964	-0.012288	-0.011054	0.0012
20	Ductal carcinoma grade-I	1.0259	-0.006358	-0.014993	-0.0086
21	Ductal carcinoma grade-II	1.0814	-0.005952	-0.024105	-0.0182
22	Cystosarcoma	1.0820	-0.007244	-0.016744	-0.0072
23	Lobular carcinoma	1.3962	-0.064706	-0.076012	-0.0113

that high level high pass coefficients do not give significant discrimination between tissue types. However, the percentage spectral fluctuations (the high-pass coefficients divided by their corresponding low-pass coefficients which are much more reliable for a statistical analysis) at level-1 show clear distinction between diseased and normal tissues. In particular the standard deviation of the normalized high pass level-1 high pass coefficients show considerable discrimination. For cancer tissues, as seen in fig. 4.7(b) standard deviations are much lower than their normal counterparts.

4.2 Continuous Wavelet Transform

Here, we investigate the subtle differences in the florescence spectra of normal and cancer breast tissues through continuous Morlet wavelets. The fact that, continuous wavelets provide an over complete basis, as compared to the complete orthonormal basis set in the discrete wavelets, makes them ideal to extract subtle changes in the spectral features (1). Morlet wavelet has been found to be quite effective in many areas of data analysis and hence the same is employed here. It is found that Morlet wavelet coefficients clearly identify distinguishing features between the tissue types. The intensity differences of parallel and perpendicularly polarized fluorescence spectra is subjected to investigation, since the same is relatively free of the diffusive background. After pin-pointing these robust features in the wavelet scalogram, we systematically study the autocorrelation property of the wavelet coefficients of the fluorescence spectra, since these coefficients show different periodic modulations with decreasing amplitude for cancer and normal cases. It is found the correlation properties strongly differ in the two tissue types, which can differentiate normal and malignant tissues with high sensitivity.

The wavelet scalograms of the intensity differences, as seen in fig. 4.8e and fig. 4.8f, show a clear difference between cancer and normal samples, in the middle scale range. Interestingly this difference is not captured by the parallel and perpendicular components alone (fig. 4.8a, b, c, d); this may be due to the fact that differences of parallel and perpendicular intensities, are relatively free from the diffusive component (2; 3). Hence, it can be quite sensitive to microscopic biochemical changes including the effects of absorption in different tissue types.

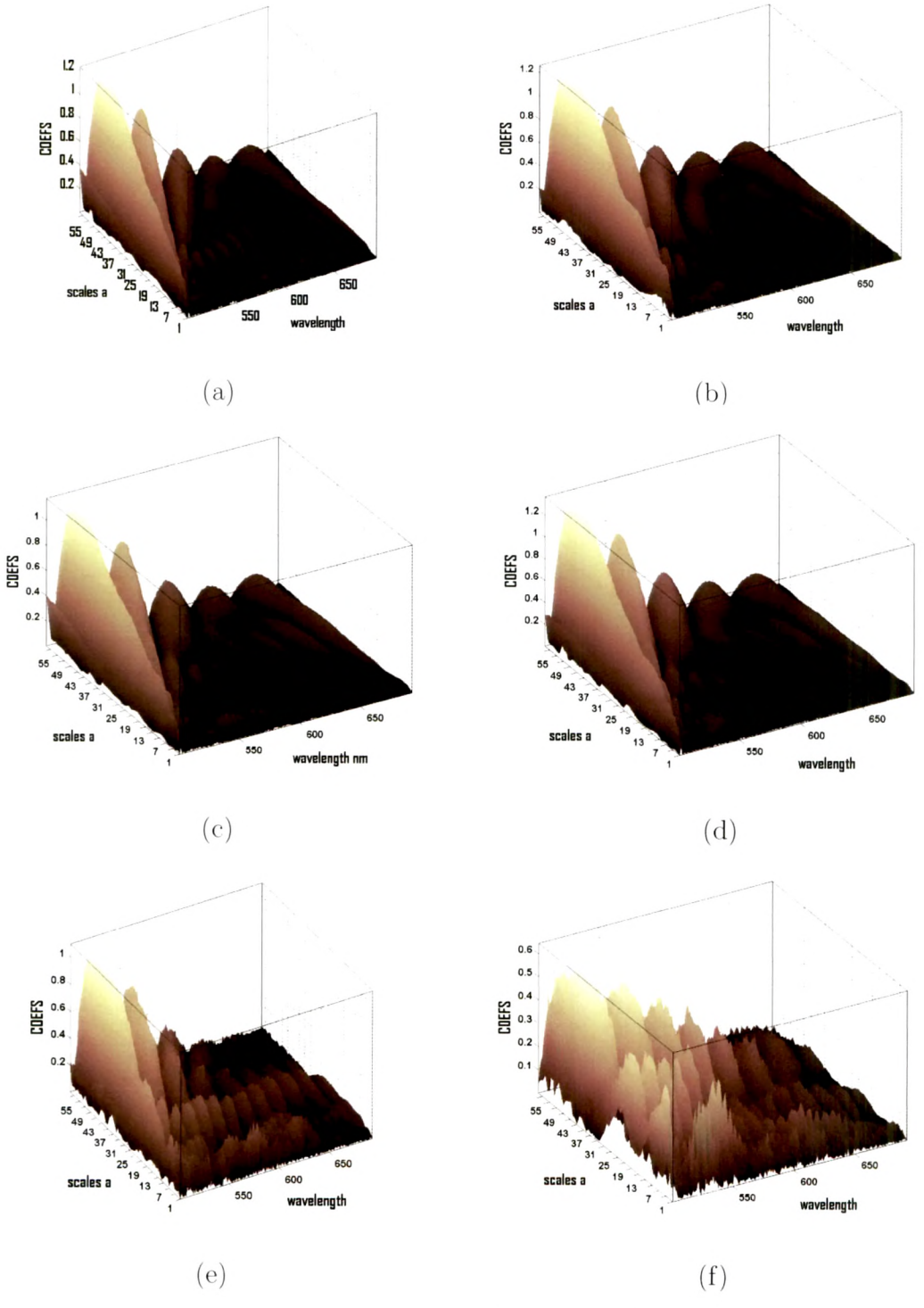
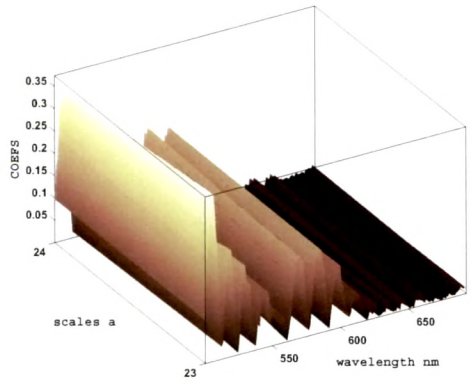


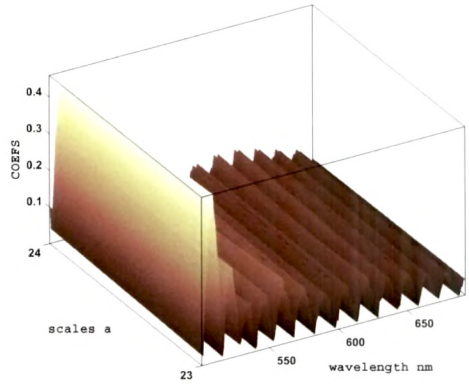
Figure 4.8: Continuous wavelet transform scalograms of (a), (b) parallel, (c), (d) perpendicular components and (e), (f) of difference of parallel and perpendicular components of the fluorescence spectra, respectively for cancer and normal tissue samples using Morlet wavelet.

This periodic variation in the scalogram in normal tissues are captured around 24th scale, whereas this type of feature is either absent in cancerous tissues or it is present with reduced amplitude. Fig. 4.9a, b depicts the periodic variation of the dominant wavelet coefficient. The fact that Morlet wavelet is a Gaussian function multiplied by a cosine modulation implies that smooth local variations can be detected through this wavelet. The Gaussian function provides a window in which the cosine function captures the modulations. Akin to the Fourier transform, if the modulations are similar to the cosine function, the wavelet coefficients are large. The window size can be varied by changing width of the Gaussian. The scale refers to this choice of the window size. Small scale corresponds to narrow Gaussian in which high frequency variations can be identified. Large scale values correspond to wider Gaussian windows, in which small frequency broader variations can be captured. Specifically, scale is inversely proportional to frequency. That data being considered here is fluorescence intensity versus wavelength from 500 to 700 nm, indicates the presence of fluorophores, like flavin and porphyrin in this domain. This fluorophore activity starts at different wavelength and hence can given rise to modulations in the data. This modulations can be of different sizes depending on the line shape of the fluorophore spectra. Morlet wavelet at different scales can ideally capture this smooth variations. It also needs to be mentioned that blood absorption can also create modulations around 540 nm. The scale 50 approximately corresponds to half the data size, since at that scale the Gaussian window is of that size. Scale 24 approximately corresponds to spectral width of 50 nm. The fact that these variations appear in the first half of wavelength axis pin points this fluoroscence activity to flavins and possibly blood absorption around 540 nm.

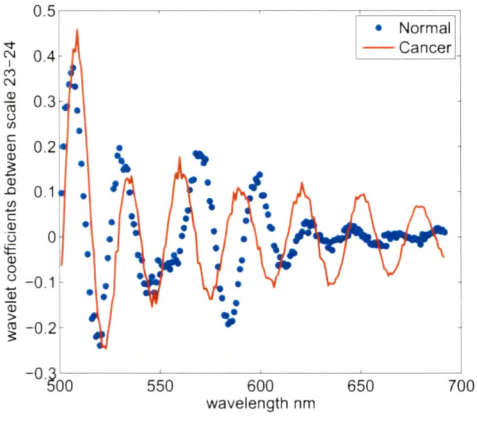
In order to properly study the scalogram behavior at different scales, we have computed the cumulative sum of the log of the magnitude of wavelet coefficients over wavelength. It is very interesting to note that, in all the tissue samples, cumulative power shows out of phase behavior between normal and cancer tissues at certain scale range. A typical plot of cumulative power as a function of scale values between 1-50 is shown in fig. 4.10a. This behavior is observed in almost all samples, albeit at different scale values, as summarized in the table-1. It is found that the cancer power is less in ductal carcinoma at the scale where a



(a)

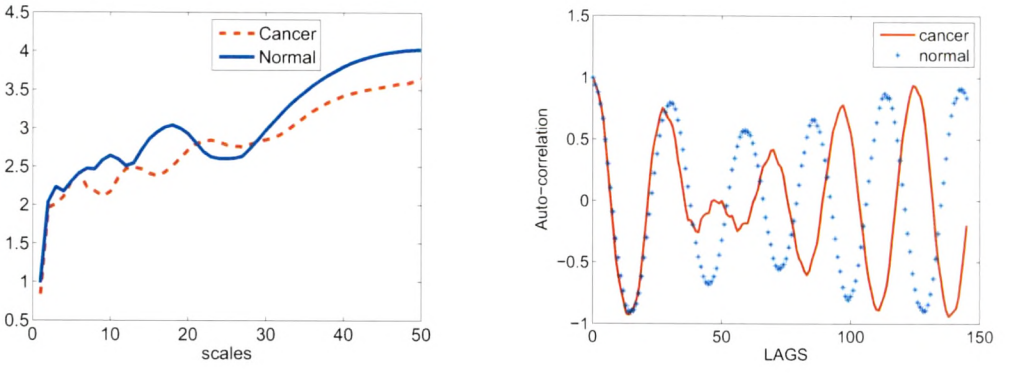


(b)



(c)

Figure 4.9: Plots of the prominent periodic modulations of the Morlet wavelet coefficients of (a) a normal and its (b) cancer counterpart, (c) the plot of the dominant wavelet coefficient showing significantly different behavior for cancer and normal tissues.



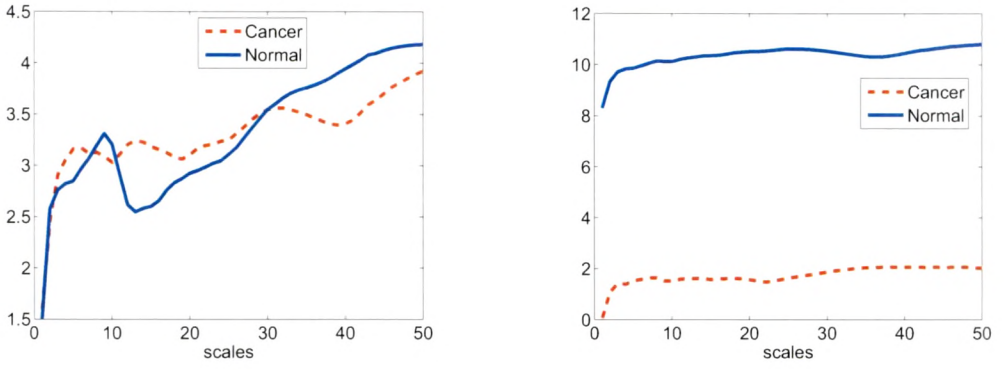
(a) (b)

Figure 4.10: Plots of (a) the cumulative sum of the CWT coefficients over all the wavelengths, over 1-50 scales, for the differences of the intensities of parallel and perpendicular components of fluorescence spectra of a typical Ductal sample and (b) autocorrelation of wavelet coefficients, corresponding to the scale of local maxima.

strong periodic variation is observed. It may be attributed to different flourophore activity in different wavelength domain. Similar studies of the power in the wavelength domain did not yield substantial difference between the tissue types.

Subsequently we have calculated the autocorrelation of these wavelet coefficients corresponding to the scale of the local maxima. A typical plot of autocorrelation is shown in fig. 4.10b, which exhibits a periodic behavior. It is noteworthy that the recurrence pattern for normal data type is rather consistent for all values of the lags, whereas for the cancer case, the periodicity seems to be absent for lags 40-60. It is found in all samples with the sensitivity of 83%. Since the autocorrelation is computed for a given scale over all wavelengths, the autocorrelation plot suggests that the period of recurrence present in a middle scale for cancer is 25, while for normal it is 30. All normal samples showed prominent wavelet coefficients around scale 24 unlike their cancer counterparts, thereby making specificity close to 100%.

It was found that Cystosarcoma and Lobular tumors showed different behavior in wavelet domain as compared to Ductal carcinoma. Their respective



(a) (b)

Figure 4.11: Plots of the cumulative sum of the CWT coefficients over all the wavelengths, over 1-50 scales, for the differences of the intensities of parallel and perpendicular components of fluorescence spectra of a typical (a) Cystosarcoma sample and (b) Lobular sample.

cumulative sums of wavelet power are shown in fig. 4.11a. and fig. 4.11b, which clearly differ from Ductal carcinoma in fig. 4.10a.

Application of SVD & PCA

4.3 Correlation Behavior of Tissue Auto fluorescence (Human Breast Tissue)

In this section the results of a systematic analysis of the correlation properties of the parallel and perpendicular components of the fluorescence spectra from human breast tissues for malignant and normal cases are presented.

It is physically clear that normal and cancer tissues have different morphological and biochemical compositions, which can give rise to different correlations in the tissue polarized fluorescence spectra. Hence the parallel and perpendicular

4.3 Correlation Behavior of Tissue Auto fluorescence (Human Breast Tissue)

TABLE-2

Sr. No.	Description of tissue	Range of Scale Values Displaying out of phase Behaviour of significant Wavelet Coefficients.
1	Ductal carcinoma grade-III	5 to 18
2	Cystosarcoma	
3	Ductal carcinoma grade-III	5 to 24
4	Intraductal carcinoma (poorly defined)	5 to 30
5	Ductal carcinoma grade-III	4 to 35
6	Ductal carcinoma	4 to 30
7	Cystosarcoma	5 to 30
8	Ductal carcinoma grade-II positive for metastasis	5 to 30
9		4 to 28
10	Ductal carcinoma grade-III positive for metastasis	5 to 38
11	Ductal carcinoma grade-II	5 to 28
12	Ductal carcinoma	4 to 24
13	Ductal carcinoma	5 to 24
14	Ductal carcinoma grade-II	4 to 30
15	Ductal carcinoma grade-II	5 to 35
16	Ductal carcinoma grade-III	5 to 28
17	Ductal carcinoma grade-II	5 to 18
18	Ductal carcinoma grade-III	4 to 30
19	Ductal carcinoma grade-III positive for metastasis	4 to 24
20	Ductal carcinoma grade-I	4 to 36
21	Ductal carcinoma grade-II	5 to 22
22	Cystosarcoma	5 to 18
23	Lobular carcinoma	

components of the tissue autofluorescence are expected to reveal different aspects of tissue characteristics. Figures 4.12(a) and 4.12(b) depict the correlation matrices for cancer and normal fluorescence spectra in the parallel polarization component. Figures 4.12(c) and 4.12(d) depict the same for the perpendicularly polarized light. The color code has been chosen such that red denotes the highest and blue the lowest value. One clearly sees domains of high correlation differing significantly in the two tissue types. Difference of intensities are also analyzed, since the same is relatively free of random components of the spectra. It is clearly seen in Figs. 4.12(e) and 4.12(f) that the correlation domain is much bigger here compared to parallel and perpendicular ones.

All the correlation matrices show block diagonal forms. There are overlapping regions of high correlation, possibly indicating different activities, e.g., emission, absorption, etc. Three distinct domains in both cancerous and normal correlation matrices are seen for the parallel case. In cases of normal samples, the last two domains do not overlap, whereas in cancer samples there is strong overlap in the middle wavelength band, i.e., 610 to 660 nm. This is the regime where porphyrin emission takes place. We also note that the absorption activity of blood takes place in the range of 540 to 580 nm. The first domain extends from 500 to 550 nm in both tissue types. This may be due to flavin emission. We observe that in the case of normal fluorescence, this domain has partial overlap with the second domain, which may be due to the presence of intracellular riboflavin, flavin coenzymes, and flavoproteins that fluoresce in the 540- to 560-nm domain. It is worth mentioning that in the case of cancer, one observes that there is no overlap between the first two domains. The second domain in both cases corresponds to blood absorption around 580 nm. The overlap in second and third domains in cancer may be due to the porphyrin activity in this wavelength regime.

In the perpendicular channel two distinct domains in normal tissue types with a small overlap, whereas in cancer three domains where the first two overlap strongly are found. Since perpendicular components travel more in the medium, it is more sensitive to absorption. This strong overlap is reflective of this fact and the porphyrin emission at 630 nm.

Next, eigenvectors corresponding to the dominant eigenvalues are analyzed. These are relatively free from statistical and experimental uncertainties and also

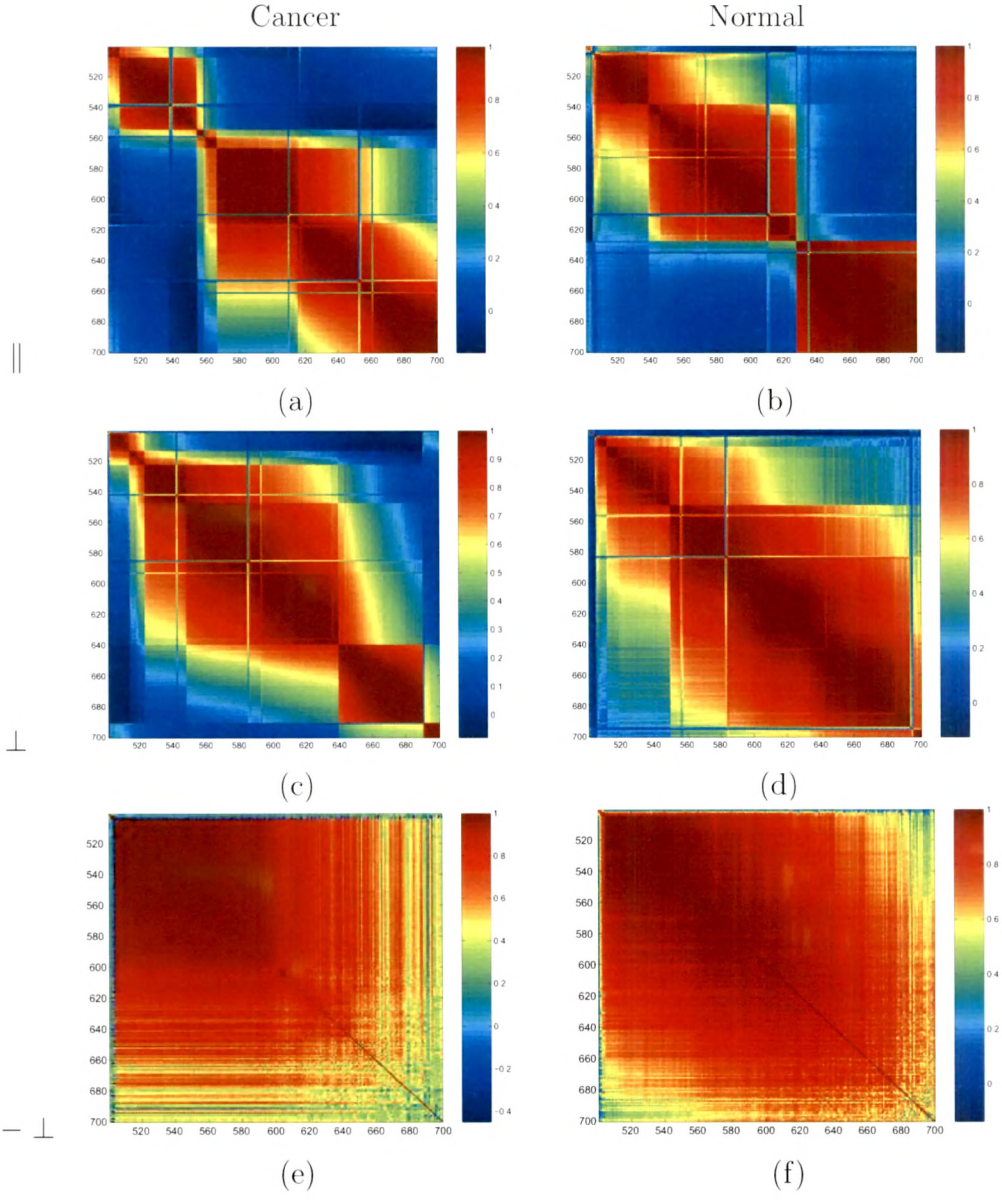
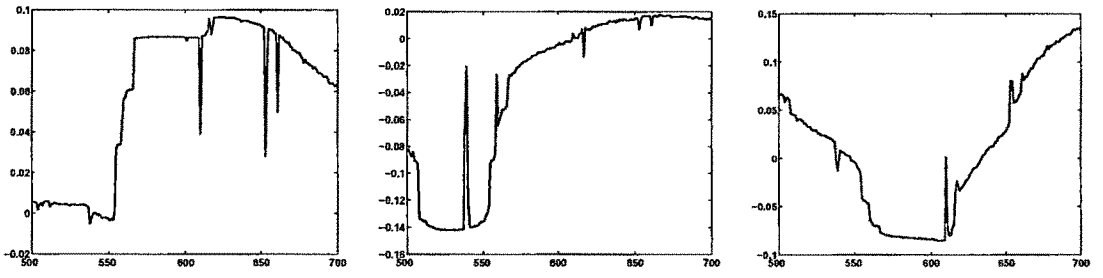
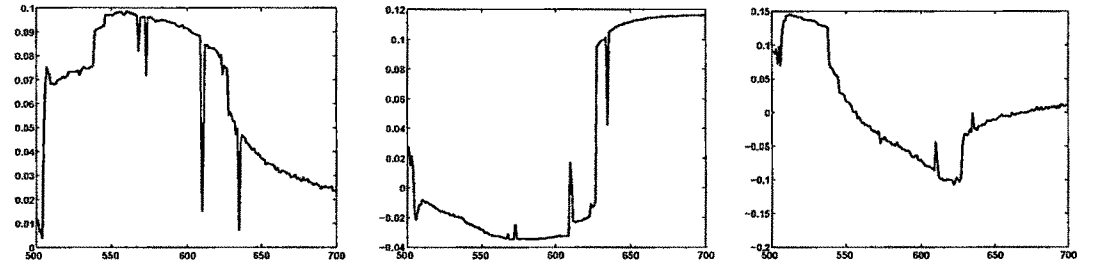


Figure 4.12: Correlation matrices of cancer and normal tissue samples: (a) and (b) parallel, (c) and (d) perpendicular components, and (e) and (f) difference of parallel and perpendicular components of polarized fluorescence data. The nature of correlations, as seen through different sized domains, are clearly different for cancer and normal fluorescence intensities. The difference of intensities yields much stronger correlations in the 500- to 700-nm range for normal breast samples.



(a) Cancer ||



(b) Normal ||

Figure 4.13: Plots of the entries in the eigenvectors corresponding to the three highest eigenvalues (from left to right) for (a) cancer and (b) normal parallel components. The x axis corresponds to the wavelength range of 500 to 700 nm. The observed domain type structures in the correlation matrices give rise to similar structures in the entries of the eigenvectors.

reflect the correlation properties observed earlier. The elements of the eigenvectors show different characteristics. As seen in Figs. 4.13 and 4.14, eigenvectors of the correlation matrix show significant differences in normal and cancerous tissues in both parallel and perpendicular components. Large eigenvalues carry information about dominant fluorophores. The block diagonal forms seen in the correlation matrix manifests in similar structures as seen in the highest eigenvectors. It is particularly transparent for the cancerous case.

In parallel components, there is lower contribution in the 500- to 550-nm range in the eigenvector corresponding to the highest eigenvalue, whereas in normal, there is more contribution. In this case also one sees rather small contributions from some wavelengths. These differ in cancer and normal cases. In the case

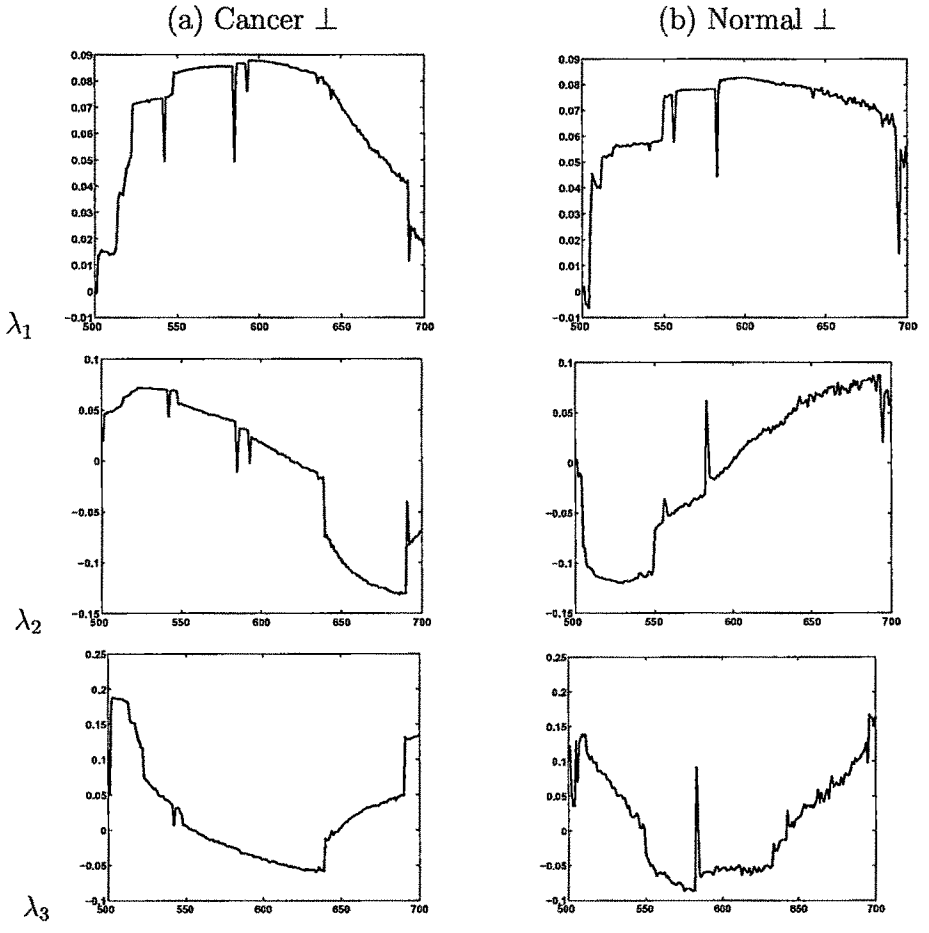


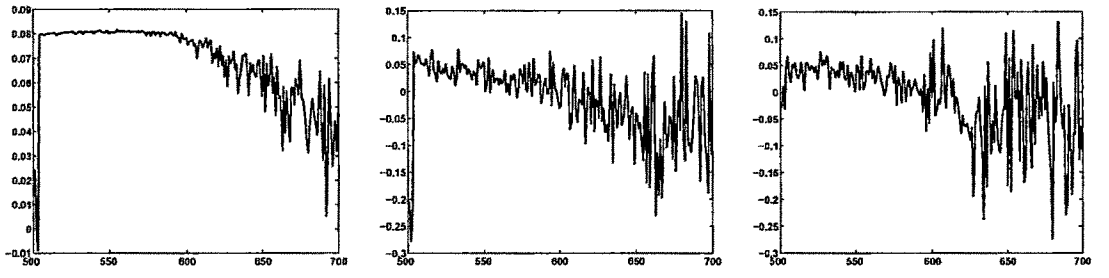
Figure 4.14: Plots of the entries in the eigenvectors corresponding to the three highest eigenvalues (from left to right) for (a) cancer and (b) normal perpendicular components. The x axis corresponds to the wavelength range of 500 to 700 nm. In the larger eigenvalues, one sees a domain-type structure. For the largest one, the normal sample shows much more correlation over the entire wavelength domain, since all the entries have almost identical values. For the normal component, one sees dominant contribution from certain wavelength sectors in the eigenvalues λ_2 and λ_3, as seen through the presence of peaks.

of the eigenvector corresponding to the second highest eigenvalue, one sees large contributions at certain wavelengths in both tissue types. In the third eigenvector, the middle wavelength domain contributes less for cancer, as compared to normal.

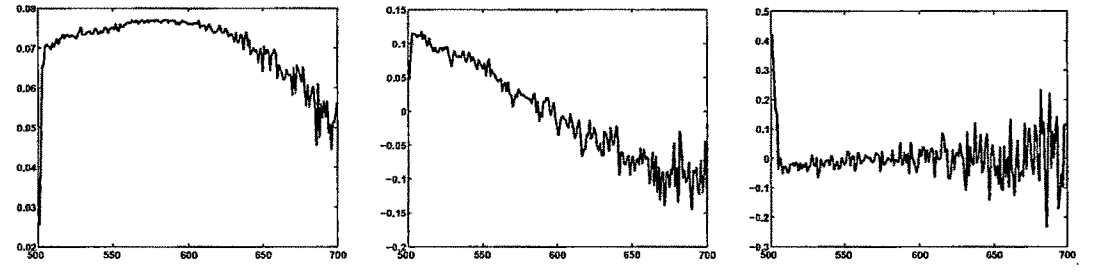
The dominant eigenvector for the perpendicular case does not show a significant difference between the two tissue types. Since the dominant eigenvector relates to the largest domain in the correlation matrix, physically this implies that the large domain structures in the correlation matrix are not very different for the two tissue types. The second one shows some characteristically high contributions at certain wavelengths, implying isolated smaller domains in that spectral regime. In the third case, the differences are not significant. Hence the parallel component retains more information about the tissue types.

The difference of intensities of parallel and perpendicular components are relatively free of the diffusive component. Hence, correlation in the intrinsic fluorescence is expected to better manifest in the corresponding correlation matrix. In Figs. 4.12(e) and 4.12(f), the correlation matrices of cancer and normal diffuse fluorescence differences indeed show very strong correlation over a much bigger wavelength regime than the cancer case. The correlation is stronger in the lower wavelength regime, with a small band around the 660-nm area. As seen in Fig. 4.15, for the $I_{\parallel} - I_{\perp}$, there is uniform values for all the entries, showing very high degrees of correlation of all the elements with each other. It is also found, particularly for the lower eigenvalues, higher contributions from certain sections than others. There are only a few dominant eigenvalues.

The eigenvalue distributions, as seen in Fig. 4.16, are also quite different for parallel and perpendicular components of cancer and normal tissue correlation matrices. The eigenvalues of the perpendicular component clearly show the difference in the nature of the spectral fluctuations of the perpendicular and parallel components. The solid line fits in Fig. 4.16 depict the fitting of probability distribution given in eq. 3.20, with the eigenvalues of the correlation matrix in the parallel and perpendicular channels. As is clear, the fit is extremely good in the perpendicular channel for cancerous tissues. The corresponding fit for the parallel case is poor, as is for the normal samples. This distribution has been derived under the assumption that the correlation matrix has been derived from Gaussian random numbers.(13) Hence, a good fit in the perpendicular channel



(a) Cancer $\| - \perp$



(b) Normal $\| - \perp$

Figure 4.15: Plots of the entries in the eigenvectors corresponding to the three highest eigenvalues (from left to right) for (a) cancer and (b) normal tissue intensity differences of parallel and perpendicular components. The x axis corresponds to the wavelength range of 500 to 700 nm. For cancer and normal cases, we find only a few large eigenvalues. The correlation domains are much bigger, since the entries in the eigenvector are almost identical over the entire wavelength domain.

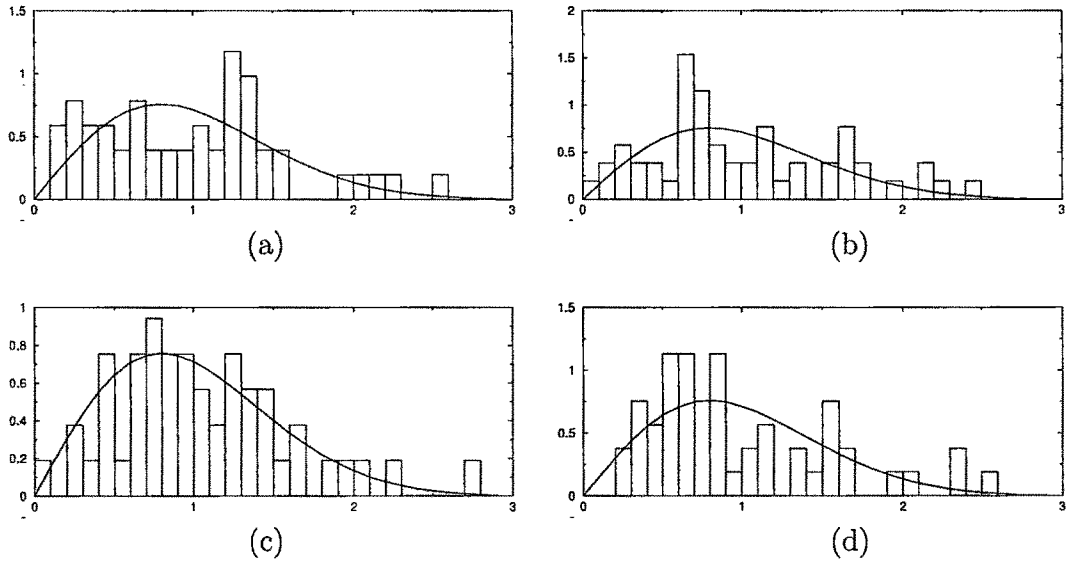


Figure 4.16: Histogram plots of the level spacing distribution for the unfolded eigenvalues. Histograms (a) and (b) correspond to the eigenvalues of cancer and normal tissues, respectively, in the parallel channel. (c) and (d) give the histogram plots for the eigenvalues of cancer and normal tissues in the perpendicular channel. Solid lines depict the fit of the $\rho_c(\lambda)$ on the eigenvalues. One sees an extremely good fit for the perpendicular case.

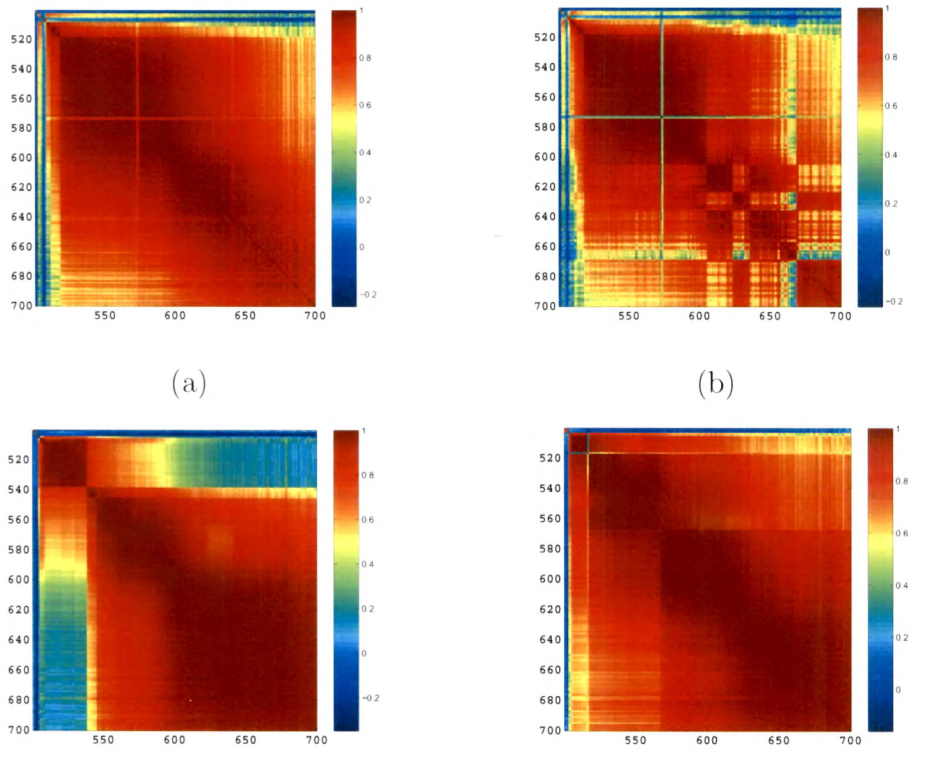
for the cancerous case indicates that the corresponding spectral fluctuations are more randomized. This is not the case for the parallel component for both normal and cancerous tissues.

It is to be emphasize once more that the choice of the lowest and highest eigenvalues is dictated by the random matrix theory. For randomized entries in the correlation matrix, the lowest and highest eigenvalues are given in eq. 3.19 following the random matrix approach. These are determined from the number of rows and columns in the matrix (Q=200/N, in the present case). Hence, they were the same in all the figures. The solid lines are fits pertaining to the distribution function given in eq. 3.20, which was derived for random entries. For smaller eigenvalues, one also sees a Gaussian distribution of the entries of the eigenvector [Fig. 4.14(a)] showing random characteristics. This clearly corroborates the wavelet based analysis, which indicated that the spectral

fluctuations in the perpendicular channel in cancerous tissues are much more randomized, as compared to their normal counterparts.(11; 12)

It is worth mentioning that the parameters of autofluorescence from tissue used for diagnosis include differences in the static spectra or decay kinetics (fluorescence lifetime). Among these, the difference in the static spectra of cancerous and normal tissues has been the more widely explored for diagnosis of cancer. The observed differences in the static fluorescence spectra between cancerous and normal tissues may arise due to several factors, like a change in concentrations, quantum efficiency of the tissue fluorophores, or a change in the modulation of the wavelength-dependent scattering and absorption properties of tissue. The usual strategy has been to develop a suitable diagnostic algorithm that can best extract the diagnostic features from the recorded bulk tissue fluorescence spectra for discriminating between cancerous and normal tissues. The present study exploits the spectral correlation properties of polarized fluorescence spectra (both the parallel and perpendicular components of fluorescence) of normal and cancerous human breast tissues for diagnosis. The results reveal interesting spectral correlation behavior of the parallel and perpendicular components of fluorescence. While the spectral fluctuations in the perpendicular component of fluorescence showed randomization effects, the parallel component revealed intrinsic information on tissue fluorophores like flavin and its derivatives, as well as porphyrin. Further, the perpendicular component also yielded complementary information on the wavelength domain where absorption takes place. The correlation properties of the polarized fluorescence spectra may thus form a useful biological metric for the diagnosis of cancer.

Benign tissues have different morphological and bio-chemical compositions, which can give rise to different correlations in the tissue polarized fluorescence spectra. Benign tissues of pericanalicular type behaved closer to cancerous tissues and the intracanalicular type like that of normal samples. Even though more overlapping is observed in pericanalicular as compared to intracanalicular benign tissues, no distinct domain is seen in benign tissues for different activities as it was clear in cancer and normal tissues [Fig. 4.17].



(a) (b)

(c) (d)

Figure 4.17: Correlation matrices of benign intracanalicular and pericanalicular tissue samples, (a), (c) parallel, and (b), (d) perpendicular components of polarized fluorescence data. The nature of correlations, as seen through different sized domains, are clearly different for benign intracanalicular and pericanalicular fluorescence intensities.

4.4 Study of Kernel-Smoother density, employed to Principal components

Here the results of a systematic analysis of the Kernel-Smoother density, employed to Principal components is presented. It is observed that the nature of the Kernel-Smoother (ks) density of the eigenvector corresponding to the highest eigenvalue entries i.e., PC1 (Fig.4.18) show remarkable difference between benign, normal and cancer tissues. A secondary peak is seen to emerge, along with the primary peak in all the three type of tissues. With the smoothening of data, by considering the low-pass coefficients, through Haar wavelets, the secondary peak becomes more prominent (Fig.4.19 & 4.20), in case of cancer crossed component as the intensity of the perpendicular component is not only affected more by scatterers, but is also quite sensitive to absorption, since the path traversed by the same in the tissue medium is more. It is found that, the ks density shows significantly better discrimination. The scatter plot of PC1 versus PC2 (Fig.4.21) reveals considerable flattening in benign and normal tissue samples as compared to the cancer ones. It is significant to observe that the perpendicular component shows much better differentiation between the tissue types(12).

Significantly, the probability density of second principal component i.e., PC2 of low pass coefficients clearly separate out cancer tissue from benign and normal (Fig.4.22). This motivates one to identify the correlation domains, corresponding to these two components, in order to identify the flourophores responsible for the spectral differences between the tissue types. For this purpose, the correlation matrix using these two highest eigenvalues independently have reconstructed. Two distinct domains, associated with the two principal component i.e., PC1 and PC2 are observed capturing different spectral features (Fig.4.23 & 4.24).

In parallel component, the domain associated with PC1 shifts towards higher wavelength in case of cancer, as compared to the normal and benign tissues. At lower wavelength flavin and its derivatives are the active flourophores, whereas porphyrin emits at higher wavelength; hence this shifting of domain may be attributed to more porphyrin emission in case of cancer case, as the parallel component is more sensitive to the emission of fluorophores present in the tissue. Large eigenvalues carry the information about dominant fluorophores. In case

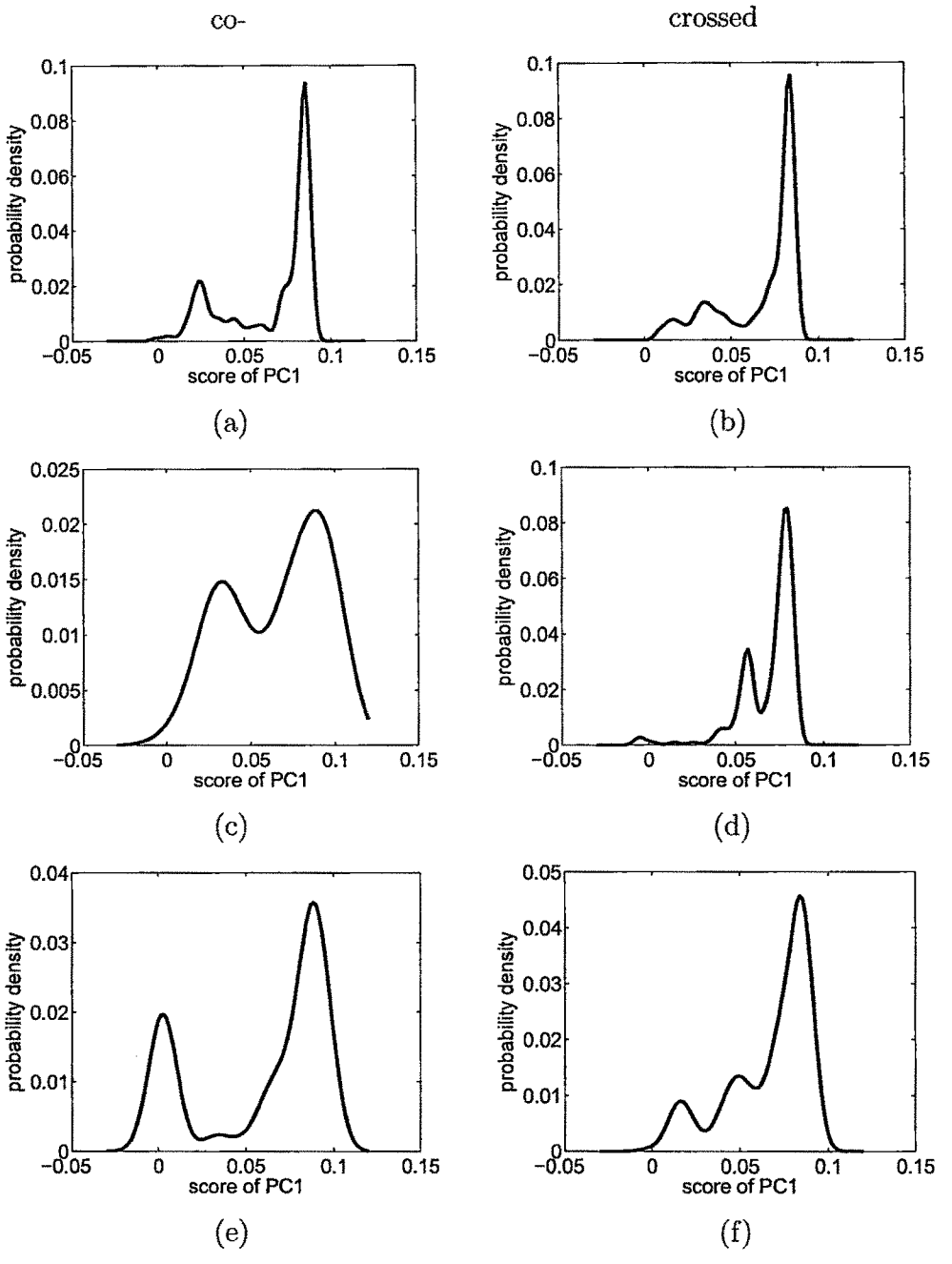


Figure 4.18: Probability density of PC1 for (a) & (b) benign, (c) & (d) normal and (e) & (f) cancer tissue types. The probability density is estimated on 100 equally spaced points that span the range of values observed in the principal component using a kernel smoothing technique, whereby a 'normal' kernel type was chosen for estimation.

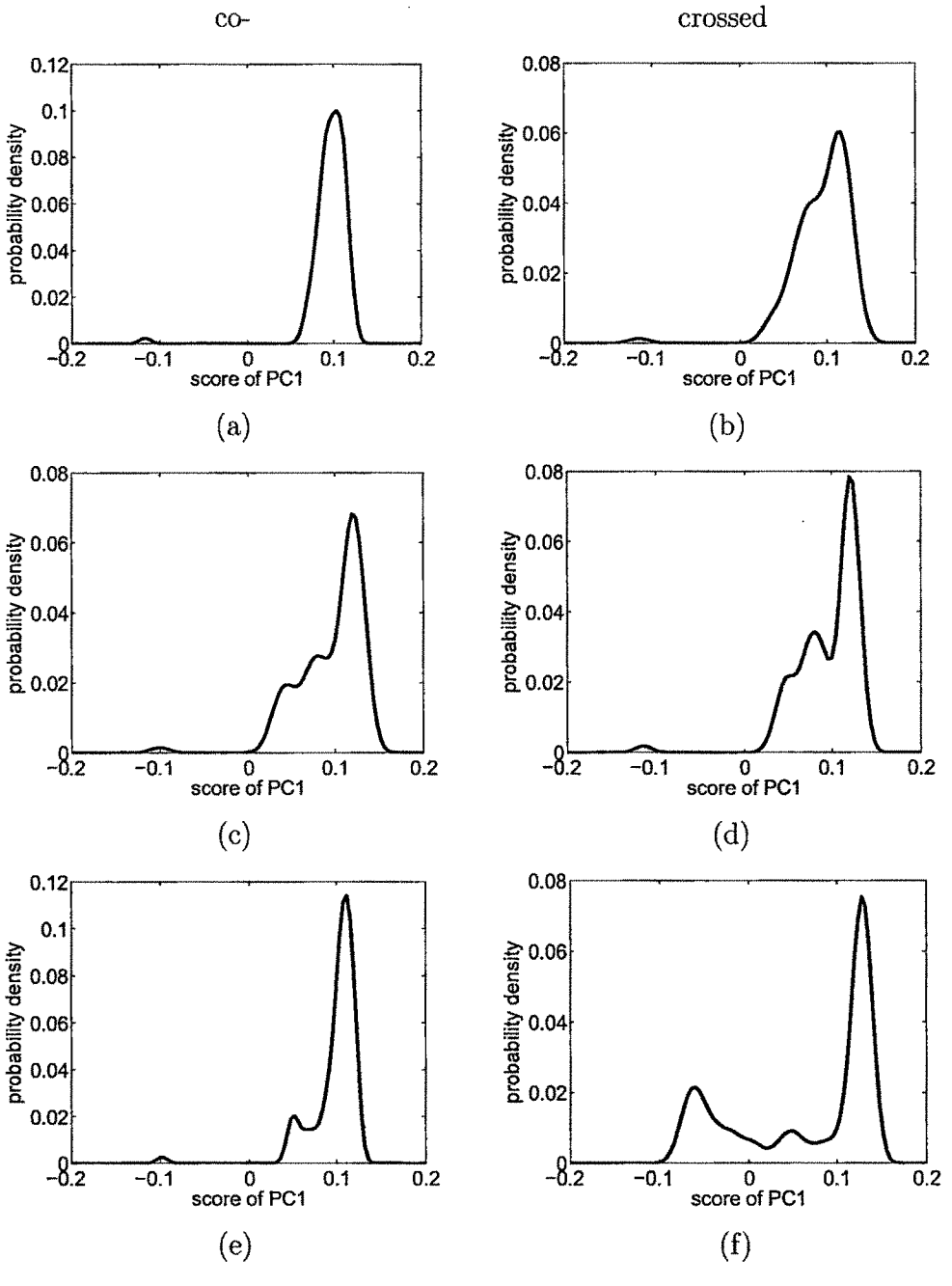


Figure 4.19: Probability density of PC1 of level-1 low pass coefficients for (a) & (b) benign, (c) & (d) normal and (e) & (f) cancer tissue types.

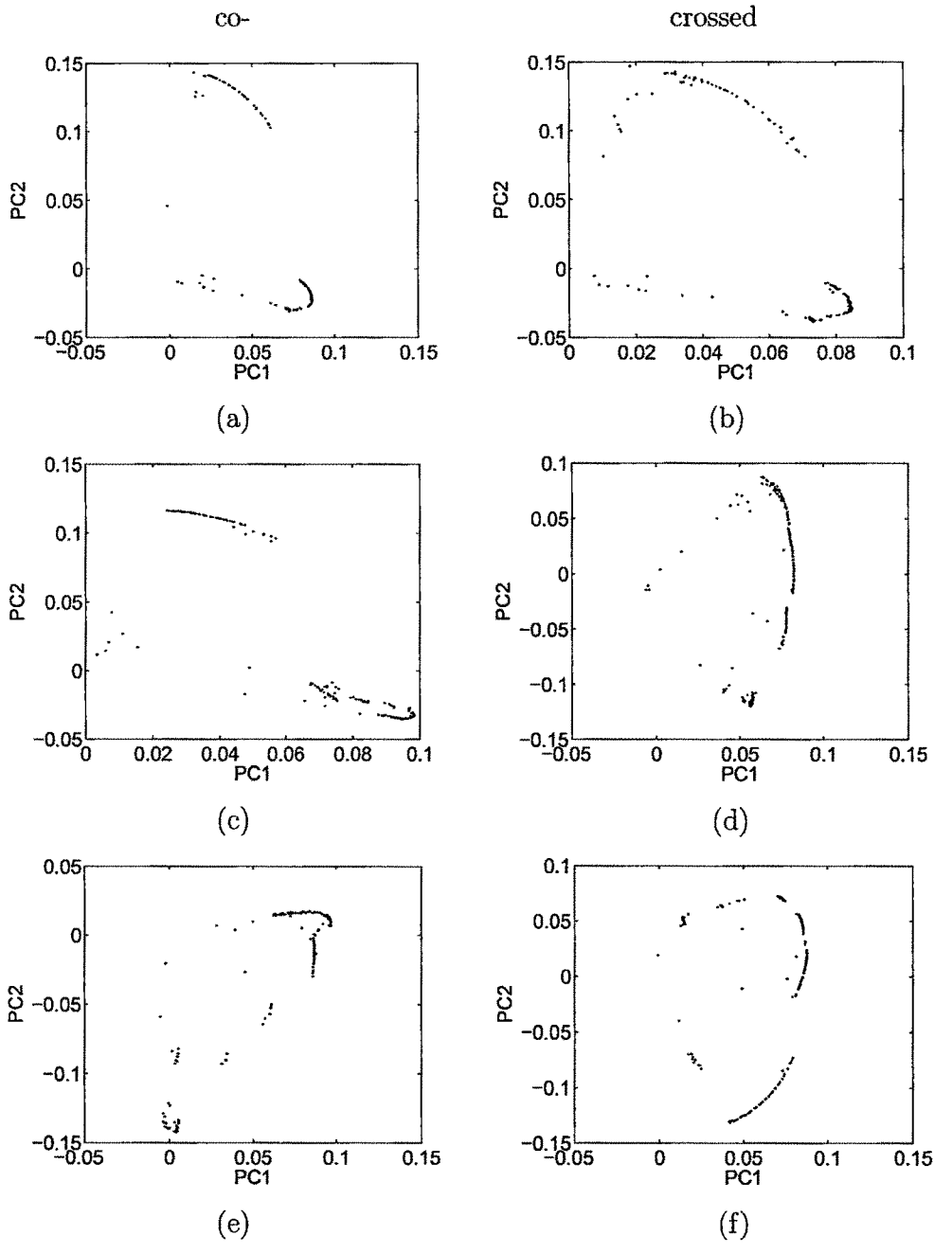
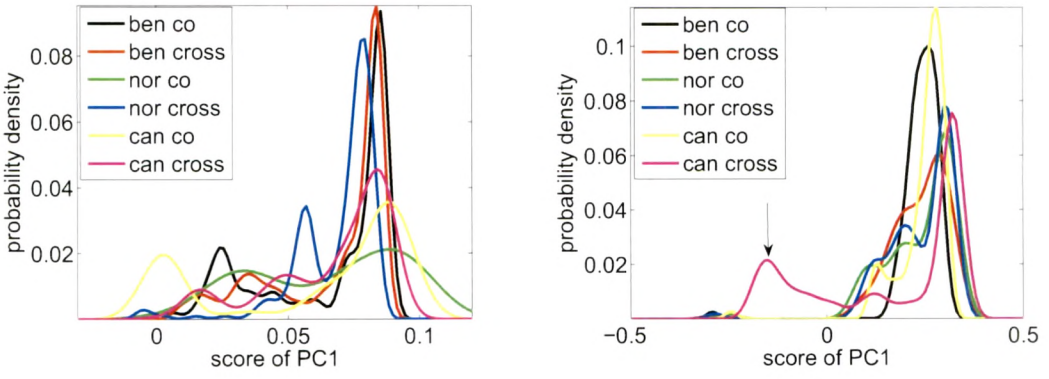
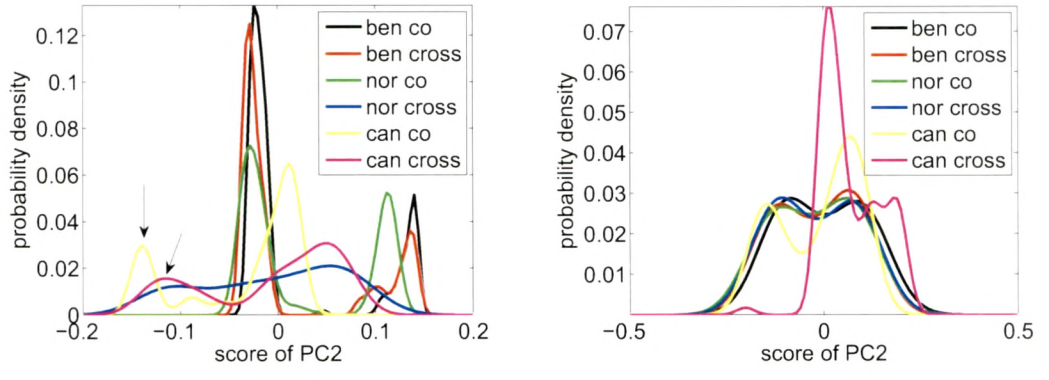


Figure 4.20: Scatter plot of PC1 and PC2 for (a) & (b) benign, (c) & (d) normal and (e) & (f) cancer tissue types. PC1 (eigenvector corresponding to highest eigenvalue) captures the highest proportion of variance in the time series, while PC2 captures the second highest proportion of variance.



(a) (b)

Figure 4.21: Probability density for PC1 of (a) intensity and (b) low pass coefficients (level-1).



(a) (b)

Figure 4.22: Probability density for PC2 of (a) intensity and (b) low pass coefficients (level-1).

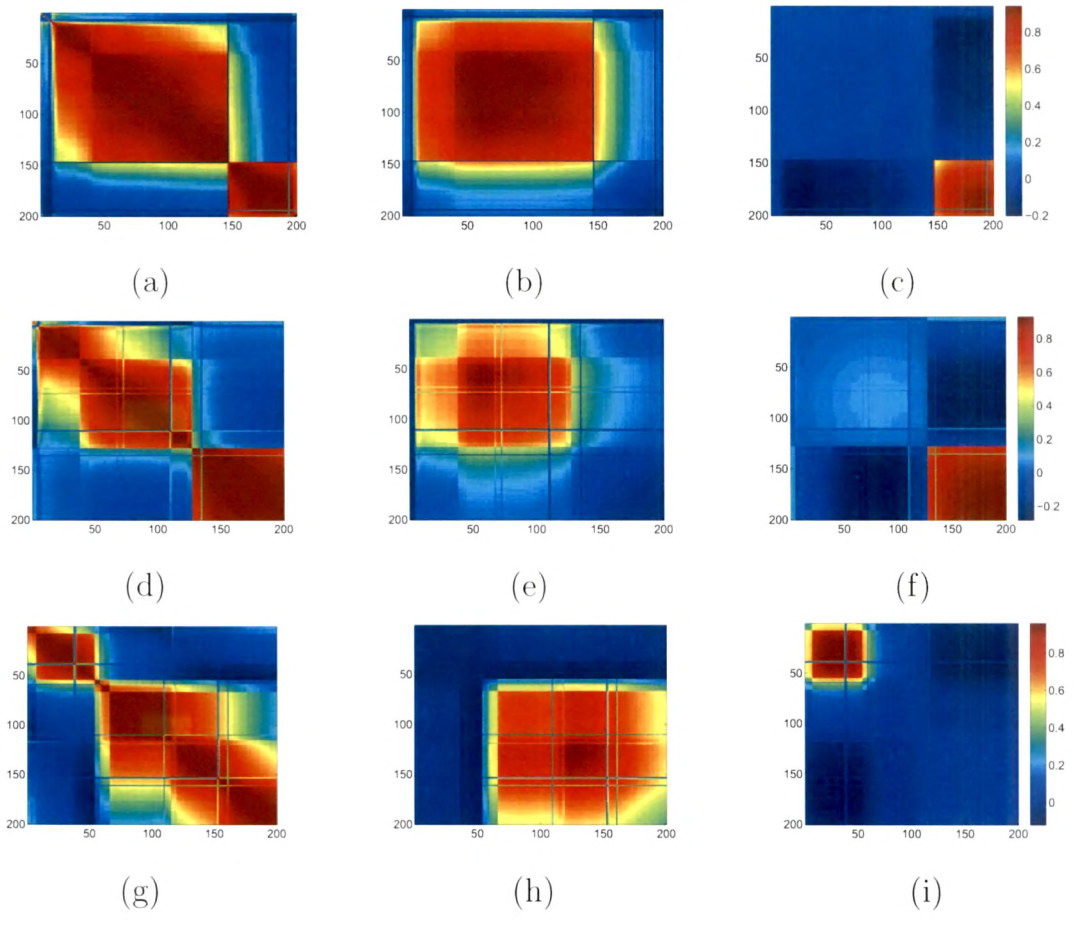


Figure 4.23: Correlation matrices of benign, normal and cancer tissue samples of (a), (d), (g) original intensity, (b), (e), (h) using PC1 and (c), (f), (i) using PC2 of parallel component of polarized fluorescence data.

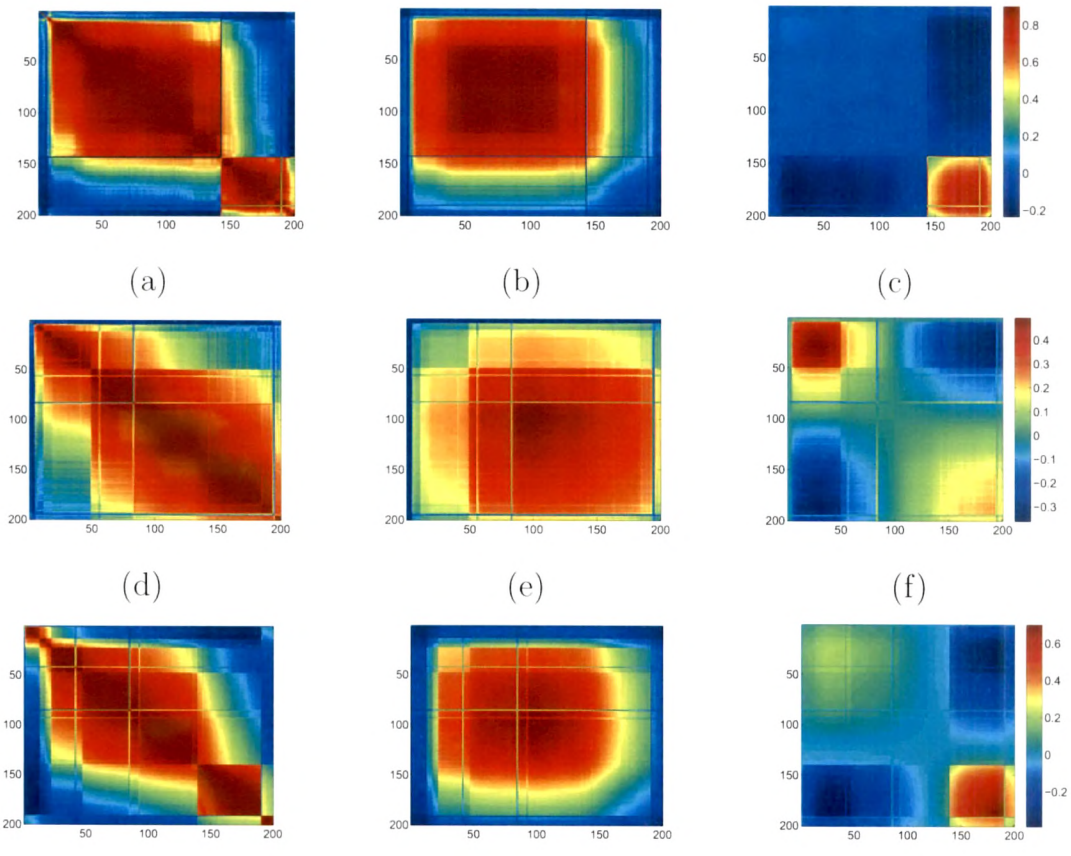


Figure 4.24: Correlation matrices of benign, normal and cancer tissue samples of (a), (d), (g) original intensity, (b), (e), (h) using PC1 and (c), (f), (i) using PC2 of perpendicular component of polarized fluorescence data.

of the eigenvector corresponding to the second highest eigenvalue, one sees large contributions at higher wavelengths, in cancer tissue. For analysis of the same use of a kernal smoother have made. Use of the same allow one to isolate the distributions responsible for the features in the correlation matrix.

The parallel component reveals intrinsic fluorescence more clearly whereas perpendicular component shows the medium effects, since this component travels the longer route. We studied both of them to isolate spectral features, differentiating tissue types.

4.4.1 Conclusion

In conclusion the use of wavelet transform in conjunction with the singular value decomposition led to a transparent distinction between the tissue types. Wavelets allowed a scale dependent separation of average behavior, which is less prone to statistical and experimental uncertainties. Singular value decomposition then enabled one to achieve dimensional reduction and pin-point the dominant distinguishing correlated features between the tissue types. One clearly observes two distinct domains in the correlation matrix of the SVD, highlighting two dominant spectral features. The corresponding eigen vectors capture these information, which gave significant differences between the cancer, normal and benign tissues. Use of a Kernel smoother, efficiently extracted the distribution of the entries of the dominant eigenvectors, which contained information about the correlation aspects of the fluorescence data. Significantly, clear distinction between the tissue types emerged in the second principal component, which corresponds to the smaller correlated domain in the correlation matrix. It is also interesting to note that better tissue differentiation is achieved in the perpendicular component, which shows that the present method extracts the characteristic medium effect, since the perpendicular component is more sensitive to the same, due its relatively larger travel path.

4.5 Correlation Behavior of Tissue Auto fluorescence (Human Cervical Tissue) with SVD & PCA

In case of cervical tissues the excitation source in the fluorimeter was a 450W Xenon lamp with 350 nm excitation wavelength and the spectra were recorded for the wavelength range of 380-650 nm. Typical spectra recorded from both malignant and normal tissues of same patient are shown in Fig.4.25. It is observed here that the overall emissions are possibly more in cancer as compared to that of the normal counterpart. The reason for the same can be attributed to the fact that the amount of the major fluorophores. The averaging effect of the lamp due to large beam size will give rise to decrease in the random fluctuations, thereby accentuating weak spectral features.

In total 46 samples were collected from various patients belonging to different age groups and also socio-economic backgrounds. In order to compare their spectral properties data sets are first normalized and dominant principal components i.e. the eigenvectors corresponding to the largest eigenvalues of the covariance matrix, extracted using singular value decomposition. In this study, over 95% of the variance in the data was explained by the first two components of the PCA. The first principal component contained a strong steady-state signal. Since the first and second principal components (PC1 & PC2) capture the highest proportion of variance present in the data, the aim of this study is to analyze the two highest principal components using scatter plots and probability density functions. The scatter plot of PC1 versus PC2 (Fig.4.26) shows considerable differences between different tissue types. The difference of the component shows remarkable difference as this is free from diffusive component, so the intrinsic fluorescence is expected to better manifest.

Eigenvectors are relatively free from statistical and experimental uncertainties. Large eigenvalues carry information about dominant fluorophores. Here the result of a systematic analysis of the Kernel-Smoother density, employed to

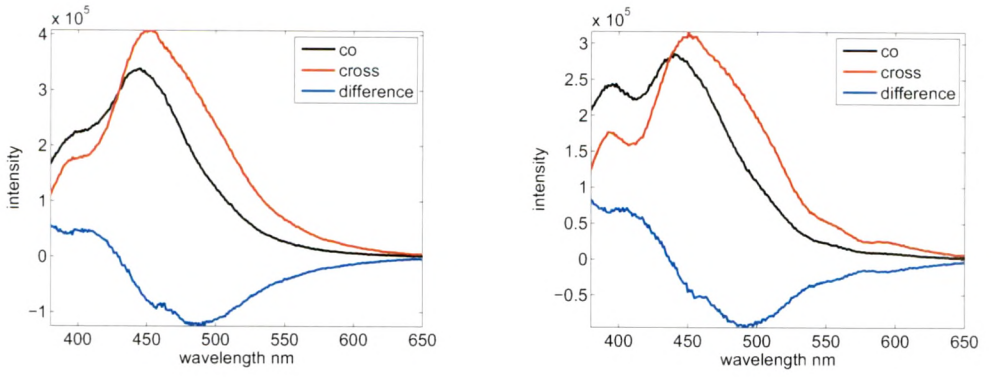


Figure 4.25: Typical autofluorescence intensity plots for co- & crossed- component and their differences for (a) Cancer and (b) Normal cervix tissues.

Principal components is presented. It is observed that the nature of the Kernel-Smoother (ks) density of the eigenvector corresponding to the highest eigenvalue entries (Fig.4.27) show remarkable difference between cancer and normal tissues. A secondary peak emerges along with primary peak in case PC1 of cancer tissues in both the channels as well as in the difference also. In both PC1 and PC2 distinct peaks are seen in case of cancer using difference of intensity. It may be attributed to the contribution of collagen (present in the stromal layer of cervical tissue) fluorescence decreases and the contributions of reduced nicotinamide adenine dinucleotide, NADH (present in the epithelial layer of cervical tissue) fluorescence and the increase of haemoglobin absorption as the tissue progresses from normal to dysplastic condition. The correlation property becomes more random in malignancy state.

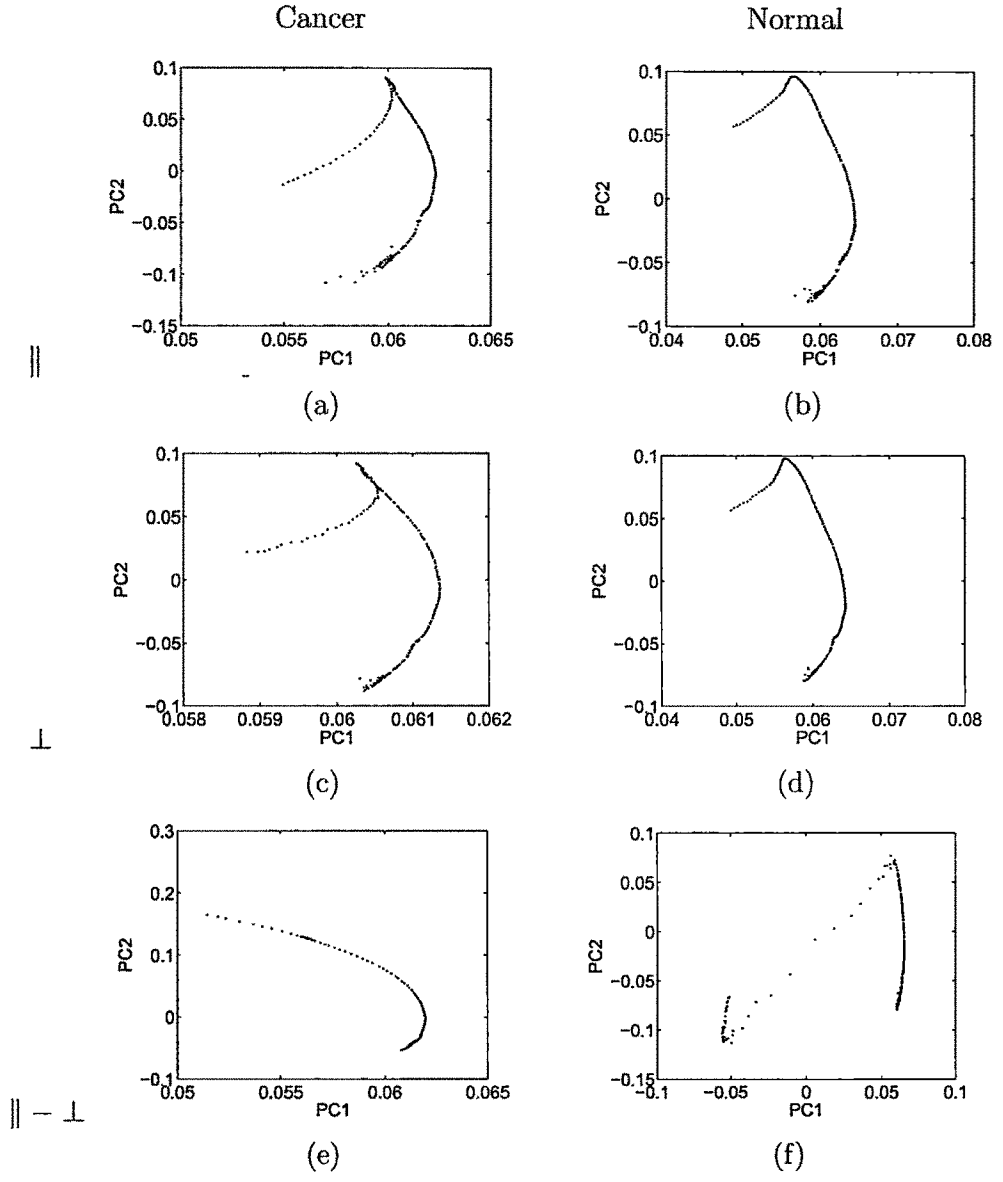
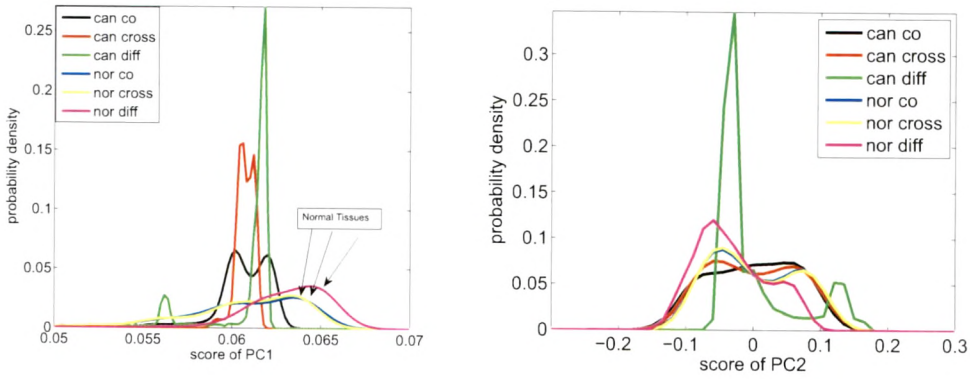


Figure 4.26: Scatter plot of PC1 and PC2 for (a) & (b) co-, (c) & (d) crossed- and (e) & (f) difference of co- and crossed- component of cancer and normal tissue types. PC1 (eigenvector corresponding to highest eigenvalue) captures the highest proportion of variance in the time series, while PC2 captures the second highest proportion of variance.



(a) (b)

Figure 4.27: Probability density for (a) PC1 and (b) PC2. The probability density is estimated on 100 equally spaced points that span the range of values observed in the principal component using a kernel smoothing technique, whereby a 'normal' kernel type was chosen for estimation.

References

[1] I. Daubechies, *Ten Lectures on Wavelets* (CBMS-NSF regional conference series in applied mathematics, Philadelphia, PA, 1992, Vol. 64). 61, 73

[2] V. Backman, R. Gurjar, K. Badizadegan, I. Itzkan, R. R. Dasari, L. T. Perelman and M. S. Field, "Polarized light scattering spectroscopy for quantitative measurement of epithelial cellular structures *in situ*," *IEEE Journal of Selected Topics in Quantum Electronics.* **5**, 1019 (1999). 67, 73

[3] S. G. Demos, W. B. Wang and R. R. Alfano, Applied Opt. **37**, 792 (1998). 67, 73

[4] L. T. Perelman, V. Backman, M. Wallace, G. Zonios, R. Manoharan, A. Nusrat, S. Shields, M. Seiler, C. Lima, T. Hamano, I. Itzkan, J. Van Dam, J. M. Crawford, and M. S. Feld, "Observation of Periodic Fine Structure in Reflectance from Biological Tissue: A New Technique for Measuring Nuclear Size Distribution," *Phys. Rev. Lett.* **80**, 627 (1998). 67

[5] M. S. Nair, N. Ghosh, N. S. Raju and A. Pradhan, Appl Opt. **41**, 4024 (2002). 67

[6] E. Levy-Lahad and S.E. Plon, Science **302**, 574 (2003). 70

[7] R. R. Alfano, G. C. Tang, A. Pradham, W. Lam, D. S. J. Choy, and E. Opher, "Fluorescence spectra from cancerous and normal human breast and lung tissues," *IEEE Journal of Quantum Electronics* **23**, 1806 (1987). 70

[8] G.A. Wagniers, W.M. Star and B.C. Wilson, Photochem. Photobiol. **68**, 603 (1998). 70

[9] N. C. Biswal, S. Gupta, N. Ghosh and A. Pradhan, Opt. Exp. **11**, 3320 (2003). 70

[10] F.N. Ghadially and W.J.P. Neish, Nature **88**, 1124-1124 (1960). 70

[11] N. Agarwal, S. Gupta, Bhawna, A. Pradhan, K. Vishwanathan, and Prasanta K. Panigrahi, "Wavelet-based characterization of spectral fluctuations in normal, benign, and cancerous human breast tissues," *IEEE J. Sel. Top. Quantum Electron.* **9**, 154 (2003). 87

[12] Sharad Gupta, Maya S. Nair, Asima Pradhan, Nrusingh C. Biswal, Nidhi Agarwal, Asha Agarwal, and Prasanta K. Panigrahi, "Wavelet transform of breast tissue fluorescence spectra: a technique for diagnosis of tumors," *Journal of Biomedical Optics* **10(5)**, 054012 (September/October 2005). 87, 89

[13] A. M. Sengupta and P. P. Mitra, "Distributions of singular values for some random matrices," *Phys. Rev. E* **60**, 3389-3392 (1999). 84