

The Potential utility of fourier transform infrared spectroscopy for the diagnosis and the prognosis of non-hodgkin lymphoma

Abstract

Background: Fourier Transform Infrared (FTIR) can detect slight change in the biochemical global composition in samples taken from patients with Non-Hodgkin's lymphomas (NHL) including changes in carbohydrates, proteins, nucleic acid, lipids or even water. These small changes might serve early detection before any morphological changes occur.

Patients and methods: The study included samples from (30) patients with diffuse large B-cell lymphomas (DLBCL) and samples from (33) healthy controls. The serum samples were analyzed using FTIR spectroscopy and the resulted data were analyzed using support vector machine (SVM) learning.

Results: Ranges 1240-1190 cm^{-1} and 1140-1000 cm^{-1} had the best true positive, false negative and accuracy results. True negative, false positive, specificity and precision were best seen in Range 1580-1480 cm^{-1} and sensitivity was best seen in Range 3500-2800 cm^{-1} . Ranges 1240-1190 cm^{-1} and 1140-1000 cm^{-1} represent frequencies that are assigned to bonds found in nucleic acids.

Conclusion: FTIR spectra ranges that could be recommended to be used as diagnostic and prognostic tool for DLBCL patients are 3500-2800, 1580-1480 and 1240-1000 cm^{-1} . The finding that the most important changes occurs in the nucleic acids may has therapeutic implication.

Keywords: diagnosis, fourier transform infrared, machine learning analysis, non-hodgkin lymphoma, prognosis, support vector machine

Volume 11 Issue 1 - 2023

Mohamed Mabed,¹ Maged Saafan,¹ Mohamed H Ahmed,² Eman Abdel Maksoud,³ Noha Eisa,³ Farid Badria⁴

¹Hematology Unit, Oncology Center, Faculty of Medicine, Mansoura University, Egypt

²Department of Biochemistry, Faculty of Science, Mansoura University, Egypt

³Department of Information Systems, Faculty of Computers and Information, Mansoura University, Egypt

⁴Department of Pharmacognosy, Faculty of Pharmacy, Mansoura University, Egypt

Correspondence: Mohamed Mabed, Hematology Unit, Oncology Center, Faculty of Medicine, Mansoura University, Egypt, Email mohmabe@mans.edu.eg

Received: September 27, 2022 | **Published:** January 20, 2023

Introduction

Lymphoma is a solid tumor of immune system that develop from lymphocytes.^{1,2} Non-Hodgkin's lymphomas (NHL) are a group of cancers that most of them arise from B lymphocytes and some from NK- or T-lymphocytes. The tumors are usually in lymph nodes, however, NHL can occur in any tissue with different degrees of aggressiveness such as indolent follicular lymphoma, or aggressive Burkett's lymphomas and diffuse large B-cell (DLBCL).¹ DLBCL type of NHL represents 30% of NHL cases.^{3,4} DLBCL can occur in any tissue organ with the gastrointestinal track as the most common involved organ.⁵ Computerized tomography (CT) scan and magnetic resonance imaging (MRI) are the standard methods for the detection and the assessment of NHL tumors with the increasing use of positron emission tomography (PET) scan with 2-[18F] fluoro-2-deoxy-D-glucose (FDG) in the management of the disease.¹ Despite the use of different staging systems such as Ann Arbor staging system and the International Prognostic Index (IPI) as prognostic model, the primary assessment of NHL patients is done by the physician building on the CT scan and MRI.^{1,3}

Most techniques used for the detection of cancer are not accurate, usually detect the disease at late stages, and depend on the estimation of the physician.⁶ The absence of enough reliable methods for early detection of cancers requires a search for new and more effective techniques for screening and prevention. Finding of an appropriate technique to test risk groups would increase the chances of successful treatment and subsequently reduce mortality. Accurate diagnostic, minimally invasive, rapid, sensitive and cheap procedures for early detection of cancer are a main requirement to reduce the mortality rates caused by the disease.⁷ The characterization of cancer on the genomic

level is promising, however, the techniques are faced with difficulties such as the tumor extended heterogeneity and the techniques expensive cost.⁸ One of the new techniques that might be promising for the early and effective detection of cancer tumor is Fourier Transform Infrared (FTIR).

FTIR can detect the slight changes in the biochemical global composite on of samples obtained from the patients including changes in carbohydrates, proteins, nucleic acid, lipids or even water. These small changes could be detected before any morphological changes occur.⁹⁻¹¹ Using FTIR as imaging tool could be a revolutionized technique for rapid and qualitative analysis of bio fluids, cells and tissue in the clinical routine with no reagents and cheap advantages.⁶ FTIR instruments cover a range of spectra from 4000-400 cm^{-1} . These frequencies correspond to specific bonds in the analyzed compound or samples. The spectrum resulted from the bond vibration frequencies evidences the presence of different functional groups and chemical bonds in the sample. FTIR is useful for the identification of organic molecular compounds and groups due to the range of cross-links, side chains and functional groups, all of which have characteristic vibrational frequencies in the infra-red range.¹²

Classification turn out to be a specific task of machine learning that can be used commonly in the medical diagnosis. It can expect the class label for unobserved instances. In classification, objects are classified relied on the extracted features. For example, the object could be normal or abnormal, benign or malignant. Classification can be a single or multi labeled. The former comprises of binary and multi-class classification. If each instance is linked to an unrivaled and only one class label, that is named a single classification. If a number of these class labels equivalents to two, at that time the classification is

well-defined as binary. On the other hand, if the class labels are extra than two then the classification is labeled as multi-class. In binary classification, the prediction is 0 or 1 (positive or negative, pathology or normal).¹³

We herein, investigate the potential use of this technique as a developing clinical tool for the identification and correlation with prognosis in patients with Non-Hodgkin lymphoma.

Patients and methods

Study Design

It is an observational study in which FTIR spectroscopy was used to test the serum of DLBCL type of NHL patients and the serum of healthy persons. In order to identify the patients' serum and healthy persons' serum, the absorbance (wave numbers) ranges were measured. Also, FTIR spectroscopy was used to test serum samples of patients after receiving treatment to determine the efficiency of the therapy. The study approved by the Institutional Review Board at Mansoura Faculty of Medicine, Mansoura, Egypt. Written informed consents were obtained from each patient. The study included samples from (30) patients with Non-Hodgkin lymphoma (DLBCL type) and samples from (33) healthy controls. The study carried out June 2019 to July, 2020.

Study population

Inclusion criteria:

- Newly diagnosed Non-Hodgkin lymphoma (DLBCL type) patients.
- Patients aged 18 years or above.
- Both sexes are eligible.
- Sign a written informed consent.

Exclusion criteria:

- Patients with other types of Non-Hodgkin lymphoma.
- Patients already received chemotherapy.

Investigational plan: The following investigations were performed. The patients were on follow up in the outpatient clinic of Oncology Center, Mansoura University, Egypt.

- Detailed history, with attention to the presence or absence of systemic symptoms, and a careful physical examination.
- Adequate biopsy reviewed by a hemato-pathologist.
- Core-needle biopsy of bone marrow, biopsy of any suspicious extra nodal Lesions; and cytological examination of any effusion.
- Complete blood count (CBC), differential leukocyte counts and blood film.
- Lactate dehydrogenase (LDH) levels and erythrocyte sedimentation rate (ESR).
- Computerized tomography (CT) scans for neck, chest, abdomen and pelvis.
- Positron emission tomography-computed tomography (PET/CT).
- Liver function tests
- Kidney function tests

j. Virology markers; hepatitis C virus (HCV), hepatitis B virus (HBV) and human immunodeficiency virus (HIV).

k. Samples for FTIR spectroscopy.

Sampling: Five mL of whole blood was collected in a sterile empty tube following standard procedures. The samples were allowed to clot in a standing position for about 20-30 minutes followed by centrifugation for 10 minutes at approximately 1000g. Using clean pipette, aliquots of serum was collected in a plastic screw-cap vials and stored at -80°C till the FTIR analysis.

Assessment of attenuated total reflectance (ATR)-FTIR spectroscopy: FTIR spectroscopy mode was estimated using a Thermo-Nicolet 6700 FTIR spectrophotometer provided with ATR and with full integration with the OMNIC software (USA). Serum samples were lyophilized then directly examined with standard procedures of ATR-FTIR. The FTIR spectra was obtained in the range from 400 to 4000 cm^{-1} at room temperature and a resolution of 4 cm^{-1} with the co-addition of 32 scans for sample spectra and 128 scans for the background spectra. The peaks were plotted as the wave number (cm^{-1}) on the Y-axis and transmittance percentage on the X-axis.

Analysis of FT-infrared results: In our binary classification, there are two major phases, training and testing. The classifier was trained using the training set putting into account the cross validation such as 10-fold or 5-fold. On the other hand, the confusion matrix was computed to evaluate classifier due to accuracy, specificity and sensitivity.

The output of the classifier or the classification technique is the classification model which can be used to predict the class labels in testing phase. Support vector machine (SVM), Knearest neighbors (KNN), and artificial neural network (ANN) and decision tree are different types of classification techniques.¹⁴ In this work, SVM was used.

From Figure 1, the image on the top left reveals three classes of data (1, 2 and 3). They used Multi SVM (MSVM) with the methodology of OVR. From (b) to (d), the classifier works with every class against the rest. It considered every class is positive, while the rest were negative.

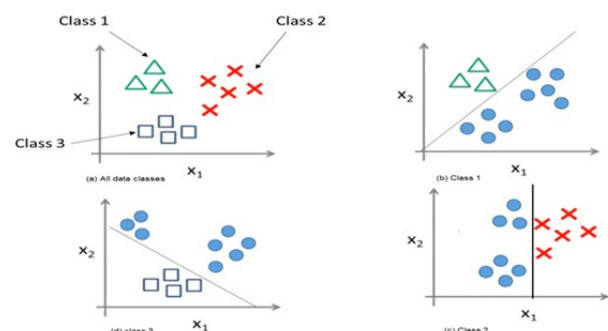


Figure 1 One-versus-rest (OVR) (a) all data classes (b) class 1 (c) class 2 and (d) class 3.

The proposed framework

Balanced dataset was used which means that the number of normal cases is nearly equal to the number of pathologies that are previously known and used in training the classifier. The entire dataset is relational. It consists of attributes for each instance or case (Figure 2).

The classifier was trained with the training set data and validated the accuracy, sensitivity and specificity by using confusion matrix and Receiver Operating Characteristic (ROC) curve will be illustrated

below. Finally, the established classification model was used in predicting the testing dataset to predict if the cases responded to the used treatment plan and drugs or not. The clinical data and the physician observations were used to validate and make the comparisons for the results in order to evaluate the established classification model. The dataset that resulting of the used pathology cases in training was tested but after treatment and other different dataset after treatment that have no training data. The dataset will be illustrated in detail in the next section. The confusion matrix formulas as illustrated in Figure 3.

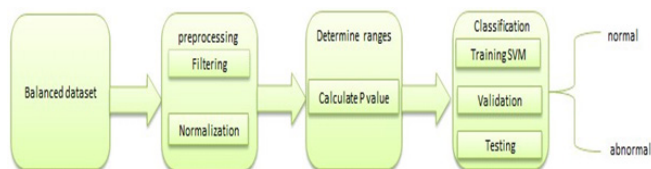


Figure 2 The classification and predicting the cases.

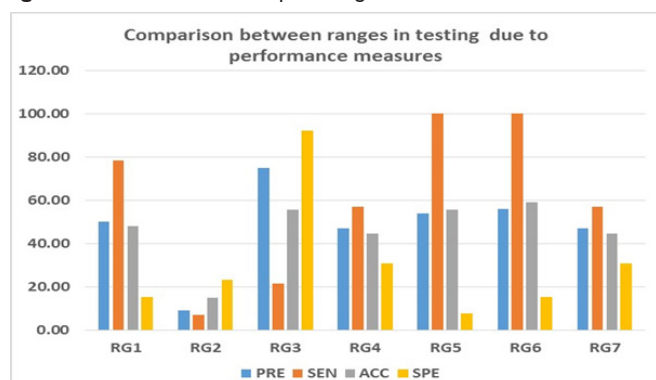


Figure 3 Comparison between ranges in testing due to performance measures.

Three important meanings in the confusion matrix should be defined: sensitivity, specificity and accuracy. Sensitivity or true positive rate (TPR) is defined as probability that a test result will be positive when the disease is present. Specificity or true negative rate (TNR) is a probability that a test result will be negative when the disease is not present. Both sensitivity and specificity are expressed as percentages. Accuracy is the true predicted results while the disease is present or absent (true positive rate and true negative rate). These parameters are calculated according to the following formulas.

1. True Positive (TP) = $\frac{\text{No of resulted diseased records}}{\text{total No of records}}$
2. True Negative (TN) = $\frac{\text{No of records that haven't disease}}{\text{total No of records}}$
3. False Positive = $\frac{\text{No of records that haven't disease and detected positive}}{\text{total No of records}}$
4. False Negative = $\frac{\text{No of records have disease and not detected}}{\text{total No of records}}$
5. Accuracy = $\frac{TP + TN}{TP + TN + FP + FN}$

The ROC curve is a fundamental tool for diagnostic test evaluation. In a ROC curve the TPR (sensitivity) is plotted in function of the FPR (100-Specificity) for different cut-off points of a parameter. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. The area under the ROC curve (AUC) is a measure of how well a parameter can distinguish between two diagnostic groups (diseased/normal). A test with perfect discrimination (no overlap in the two distributions, 100% sensitivity, 100% specificity) has a ROC curve that passes through the upper left corner (Figure 4). Therefore, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test.

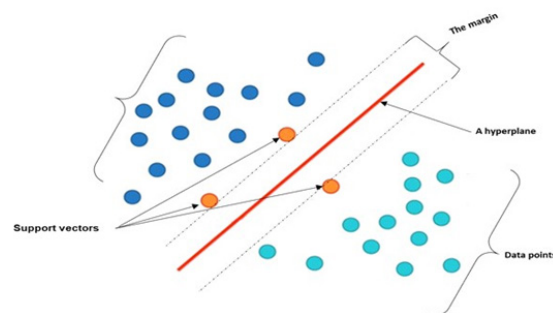


Figure 4 The classification using linear SVM.

Results

The descriptive data for NHL patients at diagnosis

The descriptive data and laboratory data of NHL patients at diagnosis are summarized in Table 1. In this study 30 patients were examined at Diagnosis, twenty-two males (73.3%) and eight females (26.7%), mean age at diagnosis was 54.6±12.3 years. Regarding laboratory data of NHL patients at diagnosis, hemoglobin median was 12±2.6 g/dL, white blood cells (WBCs) count median was 6×10⁹ cell/liter (ranges from 1.4 to 14 ×10⁹ cell/liter), platelets median was 165×10⁹ cell/liter (ranges from 2 to 549×10⁹ cell/liter), Lymphocytes median was 1.7×10⁹ cell/liter (ranges from 0.3 to 8.9×10⁹ cell/liter). ESR 1st hour median was 20 mm (ranges from 5 to 60 mm), alanine aminotransferase (ALT) median was 21 U/L (ranges from 5 to 109 U/L), aspartate transaminase (AST) median was 30.5 U/L (ranges 8 to 83 U/L), total bilirubin median was 0.8 mg/dL (ranges from 0.2 to 2 mg/dL), direct bilirubin median was 0.3 mg/dL (ranges from 0.1 to 0.7 mg/dL), serum uric acid median was 5 mg/dL (ranges from 1.2 to 21 mg/dL), serum creatinine median was 1.1 mg/dL (ranges from 0.6 to 3 mg/dL), serum albumin was 3.3±0.8 g/dL. Regarding viral markers of patient of NHL patients at diagnosis, HIV infection was zero, HCV infected patients were twenty (66.7%), HBV infected patients were one (3.3%).

At Diagnosis, elevated LDH was detected in 18 patients (16%), nine patients with age > 60 years (31.0%), stage III or IV was diagnosed in 25 patients (83.3%), performance status was > 1 in 18 patients (10.6%), and more than one extra-nodal site was found in 11 patients (36.7%) as shown in Table 2. International Prognostic Index (IPI) score, low risk patients were 7 (23.3%), low intermediate risk patients were 7 (23.3%), high intermediate risk patients were 13 (43.3%), high risk patients were 3 (10.0%) as shown in Table 2.

Staging results of NHL patients were as the following, stage I in one patient (3.3%), stage II in two patients (6.7%), stage IIE in two patients (6.7%), stage III in six patients (20.0%), stage IIIs in three patients (10%), stage IV in 16 patients (53.3%). At diagnosis, examination revealed lymphadenopathy in twenty-seven patients (90.0%), splenomegaly in seventeen patients (56.7%), hepatomegaly in twelve patients (40.0%). By history twelve patients were suffering from B symptoms. There data are summarized in Table 3.

Patients received 4 cycles chemotherapy before re-evaluation of response, three patients received cyclophosphamide, vincristine, and prednisone (COP) (10.3%), eleven patients received rituximab (Rituxan), cyclophosphamide, doxorubicin hydrochloride, vincristine

(Oncovin) and prednisolone (standard R-CHOP) regimen (36.0%), three patients received rituximab and reduced dose CHOP (R-mini-

CHOP) (10.3%), twelve patients received dose adjusted etoposide, vincristine, doxorubicin, cyclophosphamide and oral prednisone (DAEPOCH) (41.4%) as shown in Table 4.

After 4 cycles chemotherapy, patient's response to treatment was evaluated. Thirteen patients achieved complete response (48.2%), six patients achieved partial response (22.2%), seven patients showed no response or stable disease (25.9%), two patients showed progression of the disease (7.4%) as shown in Table 4.

The descriptive data for NHL patients' post-therapy

Patient achieve complete response: Re-evaluation of patients' condition after receiving chemotherapy revealed that 13 patients achieved complete response, all of them were free from B-symptoms (0.0%), by examination all of them showed no lymphadenopathy (0.0%), no splenomegaly (0.0%), no hepatomegaly (0.0%). Performance status (PS) of nine patients from those who achieved complete response were zero (69.2%), four patients were one (30.8%), PS median was 1. All the patients with complete response were free from central nervous system (CNS) infiltration and extranodal sites infiltration. Three patients show elevated LDH (10.7%) as shown in Table 5.

Patient achieve partial response: Re-evaluation of patients' condition after chemotherapy revealed that 6 patients achieve partial response, four of them had B-symptoms (66.7%). By examination, three of them showed lymphadenopathy (50.0%), no splenomegaly (0.0%), and no hepatomegaly (0.0%). PS of one patient from those who achieved partial response was zero (16.7%), five patients were one (83.3%), PS median was one. All the patients with partial response

were free from CNS or extranodal infiltration (0.0%). Five patients showed elevated LDH (83.3%) as shown in Table 5.

Patient show no response or stable disease: Re-evaluation of patients' condition after chemotherapy revealed 7 patients with no response or stable disease, three of them had B-symptoms (42.8%). By examination, four of them showed lymphadenopathy (57.0%). Two patients had splenomegaly (28.5%), and all of them had no hepatomegaly (0.0%). PS of two patients from those showed no response or stable disease was zero (28.5%), five patients were one (71.4%), PS median was 1. All patients were free from CNS or extranodal infiltration (0.0%), and three patients showed elevated LDH (42.8%) as shown in Table 5.

Patient show progression: Re-evaluation of patients' condition after chemotherapy revealed 2 patients with progression, one of them was still suffering from B-symptoms (50.0%). By examination, both showed lymphadenopathy (100.0%), one of them showed splenomegaly (50.0%). The two patients were free from hepatomegaly (0.0%). PS of one patient from those with progressive response was zero (50.0%), and the other was one (50.0%), PS median was 1.5. All patient with progressive response were free from CNS infiltration (0.0%), one patient has extranodal site infiltration, and one patient shows elevated LDH (50.0%) as shown in Table 5.

Analysis of FTIR results and its correlation with clinical data (Tables 6,7)

Using *P* values, ranges 3500-2800, 1700-1600, 1580-1480, 1380-1325, 1240-1190, 1140-1000, and 1800-900 cm^{-1} were elected.¹⁵⁻¹⁷ Example of FTIR spectra that have measured is illustrated in Figure 5.

Table 1 Demographic and Laboratory data of NHL patients at Diagnosis

Parameters	Number of patients ⁱ	Percentage of patients ⁱ
Gender	Male	22
	Female	8
HIV infection	0	0
HCV infection	20	66.7
HBV infection	1	3.3
	Median	Range
WBCs ($\times 10^9$ cell/liter)	6	1.4-14.0
Lymphocytes ($\times 10^9$ cell/liter) ⁱⁱ	1.7	0.3-8.9
Platelets ($\times 10^9$ cell/liter)	165	2.0-549.0
ESR 1st hour (mm) ⁱⁱⁱ	20	5.0-60.0
ALT (U/L)	21	5.0-109.0
AST (U/L)	30.5	8.0-83.0
Total bilirubin (mg/dL)	0.8	0.2-2.0
Direct bilirubin (mg/dL) ^{iv}	0.3	0.1-0.7
Uric acid (mg/dL)	5	1.2-21.0
Serum. creatinine (mg/dL)	1.1	0.6-3.0
	Median	Standard deviation (SD)
Age at diagnosis (years)	54.6	12.3
Hemoglobin (g/dL)	12	2.6
Serum albumin (g/dL) ^v	3.3	0.8

ⁱTotal number of patients is 30.

ⁱⁱTotal number of patients is 29.

ⁱⁱⁱTotal number of patients is 21.

^{iv}Total number of patients is 16.

^vTotal number of patients is 18.

Table 2 Risk score of NHL patients at Diagnosis

Parameters	Number of patients ⁱ	Percentage of patients (%) ⁱ
Elevated LDH	18	60
Age > 60 years (n=29)	9	31
Stage III or IV	25	83.3
Performance status > I	18	10.6
More than one extra-nodal site	11	36.7
IPL score	Low risk	7
	Low intermediate risk	7
	High intermediate risk	13
	High risk	3

ⁱTotal number of patients is 30.

Table 3 Clinical data (Ann arbor staging and organomegaly) of NHL patients at Diagnosis

Parameters	Number of patients ⁱ	Percentage of patients (%) ⁱ
I S	1	3.3
II	2	6.7
II E	2	6.7
III	6	20
III S	3	10
IV	16	53.3
B symptoms (n=29)	12	41.4
Lymphadenopathy	27	90
Splenomegaly	17	56.7
Hepatomegaly	12	40

ⁱTotal number of patients is 30.

Table 4 Treatment type and response assessment

Parameters	Number of patients	Percentage of patients (%)
Type of treatment ⁱ	COP	3
	R. CHOP	11
	R. mini CHOP	3
	DA EPOCH	12
Re-evaluation of response ⁱⁱ	Complete response	13
	Partial response	6
	Progressive	2
	No response or stable disease	7

ⁱTotal number of patients is 29.

ⁱⁱTotal number of patients is 28.

Table 5 Clinical parameters of patients after treatment

Parameters	Complete response ⁱ		Partial response ⁱⁱ		Progressive ⁱⁱⁱ		Stable disease ^{iv}		
	Number of patients	Percentage of patients (%)	Number of patients	Percentage of patients (%)	Number of patients	Percentage of patients (%)	Number of patients	Percentage of patients (%)	
B- symptoms	0	0.0	4	66.7	1	50.0	3	42.8	
Lymphadenopathy	0	0.0	3	50.0	2	100.0	4	57.0	
Splenomegaly	0	0.0	0	0.0	1	50.0	2	28.5	
Hepatomegaly	0	0.0	0	0.0	0	0.0	0	0.07	
Performance status	0	9	69.2	1	16.7	1	50.0	2	28.5
	I	4	30.8	5	83.3	1	50.0	5	71.4
	Median (range)	1 (0-1)		1 (0-1)		1.5 (0-1)		1(0-1)	
CNS infiltration	0	0.0	0	0.0	0	0.0	0	0.0	
Elevated LDH	3	10.7	5	83.3	1	50.0	3	42.8	
Extra nodal sites infiltration	0	0.0	4	66.7	1	50.0	0	0.0	

ⁱTotal number of patients is 13.

ⁱⁱTotal number of patients is 6.

ⁱⁱⁱTotal number of patients is 2.

^{iv}Total number of patients is 7.

Table 6 Confusion matrix parameters results of each range

Ranges (cm ⁻¹)	Range No.	True positive (TP) (%)	True negative (TN) (%)	False positive (FP) (%)	False negative (FN) (%)	Specificity (%)	Sensitivity (%)	Precision (%)	Accuracy (%)
3500-2800	RG1	40.74	7.41	40.74	11.11	15.38	78.57	50.00	48.15
1700-1600	RG2	3.70	11.11	37.04	48.15	23.08	7.14	9.09	14.81
1580-1480	RG3	11.11	44.44	3.70	40.74	92.31	21.43	75.00	55.56
1380-1325	RG4	29.63	14.81	33.33	22.22	30.77	57.14	47.06	44.44
1240-1190	RG5	51.85	3.70	44.44	0.00	7.69	100	53.85	55.56
1140-1000	RG6	51.85	7.41	40.74	0.00	15.38	100	56.00	59.26
1800-900	RG7	29.63	14.81	33.33	22.22	30.77	57.14	47.06	44.44

Table 7 Biomolecular assignments of ATR–FTIR spectrum bands of a whole serum dried film. Adapted from¹⁵⁻¹⁸

Wave number ranges (cm ⁻¹)	Range no	Biomolecular assignment
3500-2800	RG1	<ul style="list-style-type: none"> • N—H and O—H symmetric stretching of amide A band. • CH₃ asymmetric stretching of lipid acyl chains. • CH₂ asymmetric stretching of lipids. • C—H stretching of lipid acyl chains.
1700-1600	RG2	<ul style="list-style-type: none"> • C=O, C—N and C—N—N stretching of amide I of proteins; α-helical, β-pleated sheet, β-turns, random coils and side-chain structures. • C=O stretching of lipids.
1580-1480	RG3	<ul style="list-style-type: none"> • N—H bending and C—N stretching of amide II of proteins; α-helical, β-pleated sheet, unordered conformation structures. • CH₃ bending of methyl groups of proteins.
1380-1325	RG4	□ C—H deformation due to CH ₃ /CH ₂ bending of groups in α and β anomers of lipids and proteins.
1240-1190	RG5	□ Phosphate I asymmetric stretching of PO ² of phospholipids, nucleic acid and phosphate.
1140-1000	RG6	<ul style="list-style-type: none"> □ C—O, C—C stretching, C—H bending, PO² symmetric stretching of carbohydrates; deoxyribose/ribose, and nucleic acids; DNA, RNA. • Left-handed DNA helix DNA (Z form). • C—O stretch of glucose.
1800-900	RG7	<ul style="list-style-type: none"> • C—C—N backbone. • C—C stretch. • Glucose. • Biomolecular assignment of RG2-6.

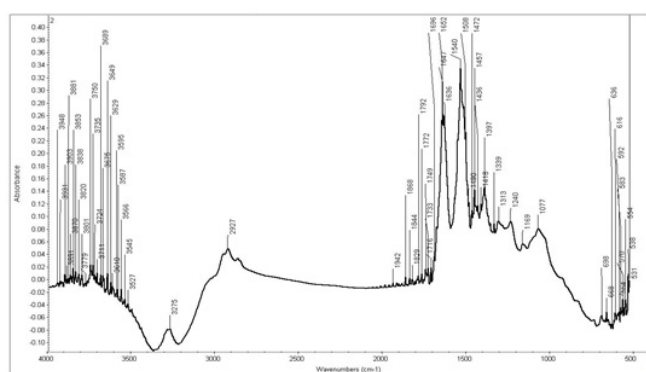


Figure 5 Example FTIR spectra that have measured.

All datasets

The ensemble discriminant classifier with 5-fold cross validation was used and the accuracy was 100%. The model predicted two normal cases which, in the clinical data, were in a complete response.

Range 3500-2800 cm⁻¹ (RG1): By selecting the range of 3500 to 2800, the linear SVM classifier was trained with 5-fold cross validation. It

gave accuracy 48.15%, sensitivity 78.57%, specificity 15.38% . The model predicted only two normal cases of the after-treatment data that were complete response from the clinical data. On the other hand, it defined the other cases of the after treatment as abnormal cases. In details, three cases were detected as normal cases after treatment, two of them were partial response and one was progressive in clinical data. By using the same range, data were trained by 4- and 10-fold cross validation in training the linear SVM classifier. It gave the same results. However, by using the 2-fold cross validation, the results were as follows; accuracy 87.1%, TPR 89.3%, and FNR 85.3%.

Range 1700-1600 cm⁻¹ (RG2): SVM classifier with 5-fold cross validation was used and It gave accuracy 14.81%, sensitivity 7.14%, specificity 23.08%. The model predicted 16 cases as normal, three of them were truly normal (true classes) while the others were false classes (not normal but between stationary, stable and partial response).

Range 1580-1480 cm⁻¹ (RG3): In this range, accuracy was 55.56%, sensitivity 21.43%, specificity 92.31. The model detected three samples as abnormal and the remaining were normal. While one case was a complete response, the remaining two were stable disease. On the other hand, in the clinical data, 12 cases were normal (complete response), the other 9 cases are not normal cases.

Range 1380-1325 cm⁻¹ (RG4): The range was trained by the classifier SVM with 5-fold cross validation. The confusion matrix results were as follows; accuracy 44.44%, sensitivity 57.14%, specificity 30.77%. The model predicted only 4 truly complete response cases and 2 partial responses as normal cases. It detected two cases, which were stable disease at clinical data, as normal.

Moreover, it detected one case, which is partial response in clinical data, as normal.

Range 1240-1190 cm⁻¹ (RG5): The range was trained by the ensemble subspace discriminant with 5-fold cross validation. It gave accuracy 55.56%, sensitivity 100%, specificity 7.69%.

The overall error was 11.3%. The model predicted only one normal case after therapy, which was actually a complete response in clinical data.

Range 1140-1000 cm⁻¹ (RG 6): By using the ensemble discriminant with 5-fold cross validation; accuracy was 59.26%, sensitivity was 100%, specificity was 15.38%. The model detected only two truly normal cases and detected the remaining as abnormal.

Range 1800-900 cm⁻¹ (RG7): By using the ensemble discriminant with 5-fold cross validation, model gave accuracy 44.44%, sensitivity 57.14%, specificity 30.77%. The SVM gave the same accuracy.

The model predicted ten after therapy cases as normal while the true cases that are actually normal from the clinical data were only five. On the other hand, it predicted one case as normal while it was progressive. The error in this model was in three cases which are stable disease. Regarding accuracy, the best result seen in ranges 6 (RG6), 5 (RG5), 3 (RG3), and 1 (RG1) with percentages 59.26%, 55.56%, 55.56% and 48.51%, respectively. Least accuracy seen in range 2 with percentage 14.81% as illustrated in Figure 3.

Regarding sensitivity, the best 2 ranges are range 6 (RG6) and 5 (RG5) with sensitivity equals to 100% for both ranges, while range 2 (RG2) had the least sensitivity (7.14%) as illustrated in Figure 3.

Discussion

MicroRNAs (miRNAs) are less susceptible to degradation due to their small size. They act as regulators for gene expression by binding to mRNA containing their complementary sequence and suppressing the translation process.^{18,20} miRNAs were found to be a prognostic marker for DLBCL. miR21, miR23A, miR27A, miR34A and miR18A were found to be downregulated in patients with lower overall survival (OS), while low miR19, miR195, miRLET7G and miR181A were associated with shorter event free survival (EFS).⁴

In our study, RG5 and RG6 had the best TP, FN and accuracy results among all tested spectral ranges, TN, FP, specificity and precision were the best results for RG3, and sensitivity was the best for RG1. RG5 and RG6 represent frequencies that are assigned to bonds found in nucleic acids. These findings indicate that, in the whole serum sample of lymphoma patients, one of the most important changes are in the nucleic acids according to Figure 4 which illustrates the main idea of SVM.

Serum lactate dehydrogenase (LDH) and beta-2-microglobulin are often increased in DLBCL patients [5]. LDH level is used as one of the five scoring factors used in IPI [21]. LDH structure is comprised of 40% alpha helices and 23% beta sheets²² with amide I region containing a substantial amount of C=O stretch of the polypeptide chain.²³

RG2 represents the amide I region and C=O stretches that are major structures of LDH. However, RG2 confusion matrix parameters results were not as good as expected. This might due to the interference of C=O stretching of lipids along with other common protein secondary structures that are characteristic of RG2 spectra.

Beta-2-microglobulin (β_2M) gene mutation was reported along with other immune-surveillance genes in DLBCL patients.^{24,25} About half of the amino acid residues in β_2M participate in two large beta structures, one of four strands and the other of three, linked by a central disulfide bond.²⁶ RG1 and RG3 significant confusion matrix parameters results might be due to the direct correlation between DLBCL and β_2M .

Conclusion

FTIR spectra ranges that could be recommended to be used as diagnostic and prognostic tool for DLBCL patients are 3500-2800, 1580-1480 and 1240-1000 cm⁻¹. Those ranges had the most promising confusion matrix parameters results. The finding that the most important changes occurs in the nucleic acids may has therapeutic implication. However, further interdisciplinary studies on larger numbers of patients are required to determine the major biomolecular changes that are related to lymphoma despite the difference in biological markers between patients according to the diseased organ.

Acknowledgments

None.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Shankland KR, Armitage JO, Hancock BW. Non-Hodgkin lymphoma. *The Lancet*. 2012;380(9844):848–857.
2. Taylor EJ. *Dorland's Illustrated medical dictionary*. 29th ed. Philadelphia: Saunders; 2000.
3. Ansell SM. Non-Hodgkin Lymphoma: Diagnosis and Treatment. *Mayo Clin Proc*. 2015;90(8):1152–1163.
4. Jamil MO, Mehta A. Diffuse Large B-cell lymphoma: Prognostic markers and their impact on therapy. *Expert Rev Hematol*. 2016;9(5):471–7.
5. Li S, Young KH, Medeiros LJ. Diffuse large B-cell lymphoma. *Pathology*. 2018;50(1):74–87.
6. Kumar S, Srinivasan A, Nikolajeff F. Role of Infrared Spectroscopy and Imaging in Cancer Diagnosis. *Curr Med Chem*. 2018;25(9):1055–1072.
7. Smith RA, Manassaram-Baptiste, Brooks D, et al. Cancer screening in the United States: a review of current American cancer society guidelines and current issues in cancer screening. *CA Cancer J Clin*. 2019;69(3):184–210.
8. Holton SE, Walsh MJ, Kajdacsy-Balla A, et al. R Label-free characterization of cancer-activated fibroblasts using infrared spectroscopic imaging. *Biophys J*. 201;101(6):1513–1521.
9. Bhargava R. Towards a practical Fourier transform infrared chemical imaging protocol for cancer histopathology. *Anal Bioanal Chem*. 2007;389(4):1155–1169.
10. Kumar S, Desmedt C, Larsimont D, et al. Change in the microenvironment of breast cancer studied by FTIR imaging. *Analyst*. 2013;138(14):4058–4065.

11. Shetty G, Kendall C, Shepherd N, et al. Raman spectroscopy: elucidation of biochemical changes in carcinogenesis of oesophagus. *BrJ cancer*. 2006;94(10):1460–1464.
12. Senoretta BA, Sumathy JH. Analysis of beta carotene from flowers using fourier transform infrared spectroscopy. *Research Journal of Pharmaceutical Biological and Chemical Sciences*. 2016;4(5):1–4.
13. Alazaidah R, Kabir F. Trending Challenges in Multi Label Classification. *International Journal of Advanced Computer Science and Applications*. 2016;7(10):127–131.
14. Aldrees A, Chikh A. Comparative evaluation of four multi-label classification algorithms in classifying learning objects. *Comput Appl Eng Educ*. 2016;24:651–660.
15. Hands JR, Abel P, Ashton K, et al. Investigating the rapid diagnosis of gliomas from serum samples using infrared spectroscopy and cytokine and angiogenesis factors. *Anal Bioanal Chem*. 2013;405(23):7347–7355.
16. GÖK S. Attenuated total reflectance fourier transform infrared spectroscopy of fluid systems: Case study applications to diagnosis and screening in biomedical and food areas. *Biology*. 2013;103:85.
17. Ghimire H, Venkataramani M, Bian Z, et al. ATR-FTIR spectral discrimination between normal and tumorous mouse models of lymphoma and melanoma from serum samples. *Sci Rep*. 2017;7:16993.
18. Wang Q, He H, Li B, et al. UV-Vis and ATR-FTIR spectroscopic investigations of postmortem interval based on the changes in rabbit plasma. *PLoS One*. 2017;12(7):e0182161.
19. Alencar AJ, Malumbres R, Kozloski GA, et al. MicroRNAs are independent predictors of outcome in diffuse large B-cell lymphoma patients treated with R-CHOP. *Clin cancer res*. 2011;17(12):4125–4135.
20. Lawrie CH, Soneji S, Marafioti T, et al. MicroRNA expression distinguishes between germinal center B cell-like and activated B cell-like subtypes of diffuse large B cell lymphoma. *Int J Cancer*. 2007;121(5):1156–1161.
21. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. A Predictive Model for Aggressive Non-Hodgkin's Lymphoma. *New England Journal of Medicine*. 1993;329:987–994.
22. Singh SN, Kanungo MS. Alterations in Lactate Dehydrogenase of the Brain, Heart, Skeletal Muscle, and Liver of Rats of Various Ages. *J Biol Chem*. 1968;243(17):4526–4529.
23. Qiu L, Gulotta M, Callender R. Lactate Dehydrogenase Undergoes a Substantial Structural Change to Bind its Substrate. *Biophysical Journal*. 2007;93(5):1677–1686.
24. Karube K, Enjuanes A, Dlouhy I, et al. Integrating genomic alterations in diffuse large B-cell lymphoma identifies new relevant pathways and potential therapeutic targets. *Leukemia*. 2018;32(3):675–684.
25. Young KH, Weisenburger DD, Dave BJ, et al. Mutations in the DNA-binding codons of TP53, which are associated with decreased expression of TRAILreceptor-2, predict for poor survival in diffuse large B-cell lymphoma. *Blood*. 2007;110(13):4396–4405.
26. Becker JW, Reeke GN Jr. Three-dimensional structure of beta 2-microglobulin. *Proc Natl Acad Sci U S A*. 1985;82(12):4225–4229.