ИСПОЛЬЗОВАНИЕ МЕТОДА КОМБИНАЦИОННОГО РАССЕЯНИЯ СВЕТА ПРИ ДИАГНОСТИКЕ ОПУХОЛЕВЫХ ЗАБОЛЕВАНИЙ ЧЕЛОВЕКА

Павлов В.Н.[1], Билялов А.Р.[1], Ковтуненко А.С.[2], Гильманова Р.Ф.[1], Бикмеев А.Т.[2], Карчевский С.Г.[3], Халилов Р.Р.[3]

[1] ФГБОУ ВО «Башкирский государственный медицинский университет» Минздрава России
г. Уфа, РФ
[2] ФГБОУ ВО «Уфимский государственный авиационный технический университет»
г. Уфа, РФ
[3] ГУП «Институт нефтехимпереработки Республики Башкортостан»
г. Уфа, РФ

Поступила в редакцию: 01.08.2018.

Аннотация. Данное исследование посвящено разработке универсального алгоритма анализа спектрограмм комбинационного рассеяния света (раман-спектроскопии) при помощи интеллектуальных методов обработки данных. Выявлено что оптимальным алгоритмом является вычисление базовой линии спектра итерационной полиномиальной регрессией, последующим частотным анализом спектра, определением главных компонент спектра и машинным обучением. Достоверность идентификации и классификации нормальной ткани и ткани злокачественной опухоли составила от 97,5 до 98 %.

Ключевые слова: *комбинационное рассеяние света, раман-спектроскопия, злокачественная опухоль, частотный анализ, deep learning.*

Ранняя диагностика рака имеет решающее значение для своевременного, эффективного и, в конечном итоге, успешного лечения. Изменения структуры и концентрации основных биохимических веществ клеток и тканей начинаются задолго до появления клинических симптомов злокачественной опухоли. В связи с этим спектроскопия, позволяющая обнаруживать изменения химических связей в молекулах является потенциальным инструментом ранней диагностики опухолей. В качестве метода молекулярной спектроскопии спектроскопия комбинационного рассеяния света (раман-спектроскопия) может обнаруживать индуцированные раком изменения молекулярной структуры и состава ткани [1].

Спектроскопия комбинационного рассеяния света (раман-спектроскопия) представляет собой оптический метод, который сегодня рассматривается для характеристики множества заболеваний, в том числе в приложениях *in vivo*, демонстрирующих различия между доброкачественными и злокачественными опухолевыми тканями [2]. В связи с этим становится актуальной проблема разработки высокоточной методики регистрации спектрограмм, ведущей к разработке точных моделей, которые впоследствии могут быть использованы вместе с инструментами *in vivo* для проведения оценки состояния тканей и определения границ опухоли.

Интерес к спектроскопии биологических тканей быстро растет, поскольку как клинические, так и неклинические исследователи признают, что вибрационные спектроскопические методы, инфракрасные (ИК) и спектроскопические методы комбинационного рассеяния потенциально могут стать неинвазивными инструментами для диагностики заболеваний. Однако существует значительный пробел в разработке методов анализа спектрограмм, поскольку, как представляется, детали характерных пиковых частот и их определения, которые могут быть отнесены к конкретным функциональным химическим группам, присутствующим в биологических тканях, не полностью поняты. Кроме того, на сегодняшний день не существует единого источника, который учитывал бы как ИК, так и комбинаторные спектроскопические исследования биологических тканей, поскольку исследователи должны полагаться на ряд источников исследований, и в большинстве случаев интерпретация спектральных данных существенно различается [3].

Целью данного исследования является разработка методики анализа спектрограмм комбинационного рассеяния света (раман-спектроскопии) при помощи интеллектуальных методов обработки данных.

МАТЕРИАЛЫ И МЕТОДЫ

В работе использовали ткани опухолевой и нормальной ткани, полученные после оперативного удаления злокачественных и доброкачественных опухолей у 20 пациентов в отделениях урологии и онкологии Клиники БГМУ. Все пациенты находились на стационарном лечении. В процессе проведения работы от всех пациентов, включенных в исследование, было получено информированное согласие. Патологическая постановка диагноза первичной опухоли (pT) определялась из образца опухолевой ткани в соответствии с системой классификации TNM (7-е издание) на основании клинических данных.

В операционном отделении проводилась секция послеоперационного препарата, и его маркировка в соответствии со стандартами Клиники БГМУ. Образцы тканей переносили в физиологический раствор и доставляли в лабораторию оптической спектроскопии в течение 2 часов после окончания операции.

Для получения рамановских спектров использовался аппарат Horiba XploRA plus, Model BX 41 TF (Horiba, Ltd., Япония). Для исследования биологических тканей использовался лазер с длиной волны 785 нм, мощностью

до 100 мВт. Всего было получено 370 спектрограмм комбинационного рассеяния света.

Данные со спектрографа обрабатывали на программном комплексе пакет LabSpec v.6.4.4. Последующий графический анализ проводился с применением программного обеспечения Spectragryph v. 1.2.8.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

На сегодняшний день имеются попытки использования спектроскопии комбинационного рассеяния Фурье, конфокальной рамановской микроспектроскопии, резонансной спектроскопии комбинационного рассеяния, спектроскопии комбинационного рассеяния на поверхности для ранней диагностики опухолевой трансформации клеток [4].

Рамановская спектроскопия представляет собой вибрационно-спектроскопический метод, который используется для оптического зондирования молекулярных изменений, связанных с пораженными тканями.

Спектры комбинационного рассеяния представляют собой графики заданной интенсивности в зависимости от разности энергий падающего и рассеянного фотонов и получены путем указания монохроматического лазерного луча на образец. Когда свет взаимодействует с молекулой, большая часть света рассеивается с той же частотой, что и падающий свет (упругое рассеяние). Как уже упоминалось, только небольшая фракция рассеивается на другой длине волны (неупругое или комбинационное рассеяние) из-за энергии света, изменяющей колебательное состояние молекулы. Потери (или усиления) в энергии фотонов соответствуют разности в окончательном и начальном уровнях колебательной энергии молекул, участвующих во взаимодействии. Полученные спектры характеризуются сдвигами в волновых числах (обратных длине волны в $см^{-1}$) от частоты падающего. Разность частот между рассеянным и рамановским рассеянием называется рамановским сдвигом, который уникален для отдельных молекул и измеряется детектором и представлен как $^{1}/_{см}$. Рамановские пики спектрально узкие и во многих случаях могут быть связаны с вибрацией определенной химической связи (или одной функциональной группы) в молекуле. Поскольку эти колебательные переходы связаны с соответствующими молекулярными связями, они уникальны для молекулы и генерируют четкие спектры комбинационного рассеяния (как отпечатки пальцев) [5, 6].

В спектроскопических исследованиях точное определение пиков может оказать большое влияние на надежность результатов. В процессе изучения современных публикаций по данной тематике стало очевидным, что большинство ученых в основном использовали для расчетов данные спектров. Однако, отсутствие надежной и подробной базы данных, охватывающей большинство известных спектральных пиков, точная идентификация химического состава биологической ткани является чересчур трудоемкой задачей.

Анализ литературы показывает, что эффективность анализа спектрограмм и обнаружения типовых для опухолевой ткани изменений, достигаемая с помощью различных методов обработки сигналов различается. Можно заметить, что метод главных компонент (PCA) в сочетании с методом опорных векторов (SVM) дает наивысшую точность (99,9 %), за которой следуют комбинация PCA и искусственных нейронные сети (98 %) и логистический регрессионный анализ (LRA) – 97 %. Методы обработки PCA + LDA (латентное размещение Дирихле) является самым популярным (45 %). PCA-LDA является самым популярным (45 %), который сопровождается PCA-ANN (33 % и SVM (22 %).

В настоящей работе предлагается новый алгоритм обработки сигналов для автоматического распознавания раман-спектров, в общем случае, включающий в себя следующие этапы: предварительная обработка сигнала, представление сигнала в требуемом виде, выявление особенностей, удаление ошибок, отбор и классификация особенностей. Алгоритм реализован на языке Python в виде оригинального программного обеспечения.

Предварительная обработка спектра проводилась методом корректировки базовой линии итерационной полиномиальной аппроксимацией. Это один из наиболее часто используемых методов для корректировки базовой линии в спектроскопии. На рисунке 1 показано использование аппроксимации различных порядков.

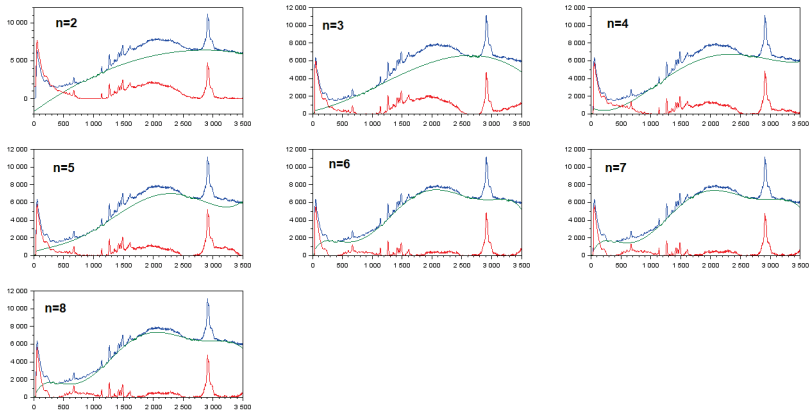

Рисунок 1. Исходная спектрограмма (синяя кривая), найденная базовая линия (зеленая кривая) и скорректированный спектр (красная линия), n – степень полинома

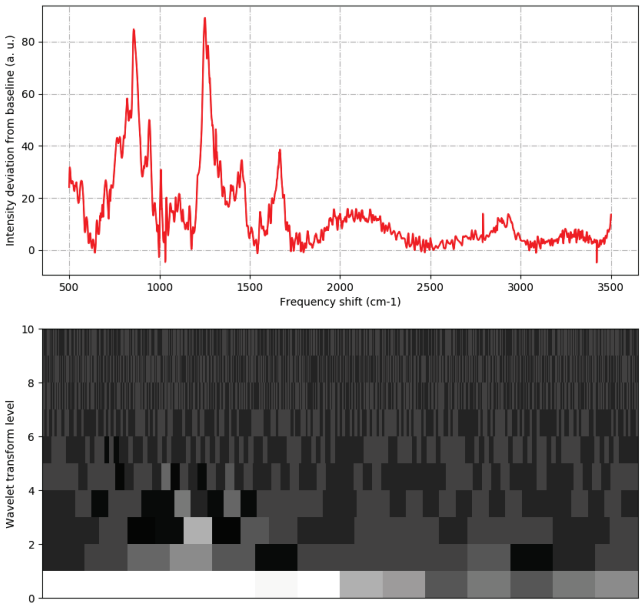

Рисунок 2. Графическое представление вейвлет-преобразования раман-спектрограмм

Первичное изображение на основе спектра строилось с использованием интегральных преобразований. Для выявления особенностей спектра, в качестве которых мы будем рассматривать пики большой высоты, и устранения шума в экспериментальных данных мы воспользовались разложением скорректированного спектра в ограниченный функциональный ряд по функциям Добеши четвертого порядка. Данное разложение позволяет проводить многоуровневый частотный анализ спектра абстрагируясь от физического смысла пиков на спектре.

На рисунке 2 показано графическое представление разложения спектрограмм по функциям Добеши после корректировки базовой линии. Низкочастотные составляющие (расположены в самом низу картинки) имеют значительные отличия, тогда как в средней части спектра можно наблюдать повторяющиеся картины. Таким образом можно сделать вывод о теоретической возможности классификации различных тканей на основе более глубокого анализа частотного разложения их раман-спектров. Результатом разложения каждого спектра является набор коэффициентов разложения, совокупность которых представляет собой образ в многомерном признаковом пространстве.

Для снижения размерности признакового пространства и выявления наиболее характерных абстрактных свойств спектров, на основе корреляционного анализа всей выборки производился выбор главных компонент и выработка проецирующего линейного преобразования образов из исходного признакового пространства в редуцированное. Для этого мы использовали анализ методом главных компонент (PCA). Во-первых, PCA преобразует данные входных функций в ортогональное пространство с использованием ортогонального линейного преобразования. Результатом являются ортогональные компоненты, известные как главные компоненты. Во-вторых, главные компоненты организованы в соответствии с их дисперсией. Разница является мерой вариабельности распределения выборок и выражается как среднее квадратическое отклонение каждого образца от его среднего значения.

Метод главных компонент состоит в разложении (с помощью ортогонального преобразования) k-мерного случайного вектора $\boldsymbol{X} = (X_1, X_2, \dots, X_k)^T$ по системе линейно независимых векторов, в качестве которой выбирается ортонормированная система собственных векторов, отвечающих собственным значениям ковариационной матрицы вектора $\boldsymbol{X}$.

Линейная модель главных компонент для центрированного вектора-столбца $\dot{\boldsymbol{X}} = \boldsymbol{X} - E\boldsymbol{X}$ записывается в виде $\dot{\boldsymbol{X}} = \boldsymbol{AF}$, где $\boldsymbol{F} = (F_1, F_2, \dots, F_k)^T$ — центрированный и нормированный случайный вектор-столбец некоррелированных главных компонент $F_j,\ j = \overline{1,k}$; $\boldsymbol{A} = (a_{ij}) \in R^{kxk}$— неслучайная матрица нагрузок случайных величин X_i на компоненты $F_j,\ i = \overline{1,k}, j = \overline{1,k}$.

Метод главных компонент допускает следующую геометрическую интерпретацию:

- сначала происходит центрирование исходного вектора $\boldsymbol{X}$, т.е. фактически производится перенос начала координат в точку $E\boldsymbol{X}$, являющуюся центром эллипсоида рассеяния случайного вектора $\boldsymbol{X}$ (рис. 2);
- затем производится поворот осей координат таким образом, чтобы новые оси координат $Of^{(1)}, Of^{(2)}$, ...были направлены вдоль осей эллипсоида рассеяния, причем разброс точек вдоль оси $Of^{(1)}$должен быть не меньше, чем вдоль оси $Of^{(2)}$ и т.д. При этом разброс наблюдений вдоль новой оси $Of^{(1)}$для исследователя наиболее важен, менее важен разброс вдоль оси $Of^{(2)}$, а разбросом вдоль нескольких последних осей можно пренебречь. Графически это иллюстрирует рисунок 3.

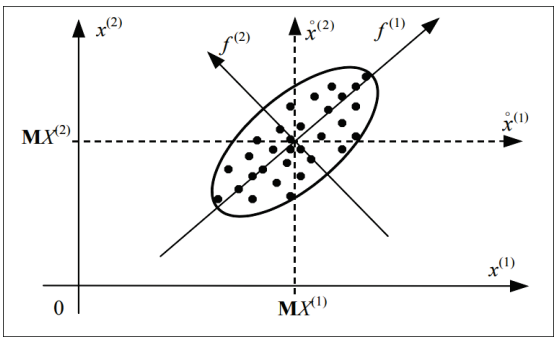

Рисунок 3. Иллюстрация метода главных компонент

Ниже изложен алгоритм построения вектора $\boldsymbol{F}$ и матрицы $\boldsymbol{A}$.

Пусть $\boldsymbol{\Sigma} = E(\dot{\boldsymbol{X}}\dot{\boldsymbol{X}}^T)$ - ковариационная матрица вектора $\boldsymbol{X}$. Будучи симметричной и неотрицательно определенной, она имеет k вещественных неотрицательных собственных значений $\lambda_1, \lambda_2, \dots, \lambda_k$. Пусть $\lambda_1 > \lambda_2 > \dots > \lambda_k$.

Обозначим за $\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{pmatrix}$ и $\boldsymbol{v}_j = (v_{1j}, v_{2j}, \dots, v_{kj})^T$–нормированные собственные векторы-столбцы матрицы $\boldsymbol{\Sigma}$, соответствующие λ_j, $j = \overline{1,k}$, а также введем матрицу $\boldsymbol{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$.

Тогда для всех $j = \overline{1,k}$ справедливы следующие равенства:

$$det|\boldsymbol{\Sigma} - \lambda_j \boldsymbol{I}|,$$

где $\boldsymbol{I}$– единичная матрица;

$$\boldsymbol{\Sigma}\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_j;$$

$$\boldsymbol{v}_p{}^T \boldsymbol{v}_j = \sum_{i=1}^{k} v_{ip} v_{ij} = \delta_{pj} = \begin{cases} 1, & p = j \\ 0, & p \neq j \end{cases}, \; p = \overline{1,k}$$

С учетом этого

$$\boldsymbol{v}_j{}^T \boldsymbol{\Sigma} \boldsymbol{v}_p = \lambda_j \boldsymbol{v}_j{}^T \boldsymbol{v}_p = \begin{cases} \lambda_j, & p = j \\ 0, & p \neq j \end{cases}, \; p = \overline{1,k},$$

следовательно,

$$\boldsymbol{V}^T \boldsymbol{\Sigma} \boldsymbol{V} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{pmatrix} = \boldsymbol{\Lambda}.$$

Пусть $\dot{\boldsymbol{F}} = \boldsymbol{V}^T \dot{\boldsymbol{X}}$. Этот вектор будет центрированным, т.к. $E\dot{\boldsymbol{F}} = E\left(\boldsymbol{V}^T \dot{\boldsymbol{X}}\right) = \boldsymbol{V}^T E\dot{\boldsymbol{X}} = 0$.

Поскольку

$$E\left(\dot{\boldsymbol{F}}\dot{\boldsymbol{F}}^T\right) = E\left(\boldsymbol{V}^T \dot{\boldsymbol{X}}\dot{\boldsymbol{X}}^T \boldsymbol{V}\right) = \boldsymbol{V}^T E\left(\dot{\boldsymbol{X}}\dot{\boldsymbol{X}}^T\right)\boldsymbol{V} = \boldsymbol{V}^T \boldsymbol{\Sigma} \boldsymbol{V},$$

то компоненты вектора $\dot{\boldsymbol{F}}$ некоррелированные и $D\dot{F}_j = \lambda_j, j = \overline{1,k}$.

Поэтому

$$\boldsymbol{F} = \boldsymbol{\Lambda}^{-\frac{1}{2}} \dot{\boldsymbol{F}} = \boldsymbol{\Lambda}^{-\frac{1}{2}} \boldsymbol{V}^T \dot{\boldsymbol{X}}.$$

Стоит заметить, что $tr\boldsymbol{\Sigma} = tr\boldsymbol{\Lambda}$, откуда $\sum_{i=1}^{k} D\dot{X}_i = \sum_{i=1}^{k} DX_i = tr\boldsymbol{\Sigma} = tr\boldsymbol{\Lambda} = \sum_{j=1}^{k} \lambda_j = \sum_{j=1}^{k} D\dot{F}_j$, т.е. дисперсия исходных случайных величин $X_1, X_2, \dots, X_k$ полностью исчерпывается дисперсией компонент $\dot{F}_1, \dot{F}_2, \dots, \dot{F}_k$. А поскольку $D\dot{F}_1 > D\dot{F}_2 > \dots > D\dot{F}_k$, дисперсией каждой следующей компоненты объясняется меньшая доля дисперсии исходных случайных величин, чем дисперсией предыдущей.

Кроме того, так как $E(\boldsymbol{F}^T\boldsymbol{F}) = \boldsymbol{I}$, то $\boldsymbol{\Sigma} = E\left(\dot{\boldsymbol{X}}\dot{\boldsymbol{X}}^T\right) = E(\boldsymbol{A}\boldsymbol{F}^T\boldsymbol{F}A^T) = \boldsymbol{A}E(\boldsymbol{F}^T\boldsymbol{F})\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^T$ или $cov\left(X_i, X_p\right) = cov\left(\dot{X}_i, \dot{X}_p\right) = \sum_{j=1}^{k} a_{ij} a_{pj}$, $j = \overline{1,k}$, $p = \overline{1,k}$. В частности, $D\dot{X}_i = DX_i = \sum_{j=1}^{k} a_{ij}{}^2$, $i = \overline{1,k}$, т.е. ковариационная матрица вектора $\boldsymbol{X}$ полностью воспроизводится матрицей нагрузок $\boldsymbol{A}$.

Также $E\left(\dot{\boldsymbol{X}}\boldsymbol{F}^T\right) = E(\boldsymbol{A}\boldsymbol{F}\boldsymbol{F}^T) = \boldsymbol{A}E(\boldsymbol{F}\boldsymbol{F}^T) = \boldsymbol{A}$, следовательно, $cov\left(X_i, F_j\right) = a_{ij}$, $i = \overline{1,k}$, $j = \overline{1,k}$, т. е. ковариация случайной величины X_i и компоненты F_j равна нагрузке a_{ij}.

Используя ортогональность матрицы $\boldsymbol{V}$, можно получить:

$$\boldsymbol{V}\dot{\boldsymbol{F}} = \boldsymbol{V}\boldsymbol{V}^T \dot{\boldsymbol{X}} = \boldsymbol{V}\boldsymbol{V}^{-1} \dot{\boldsymbol{X}} = \dot{\boldsymbol{X}}.$$

Кроме того,

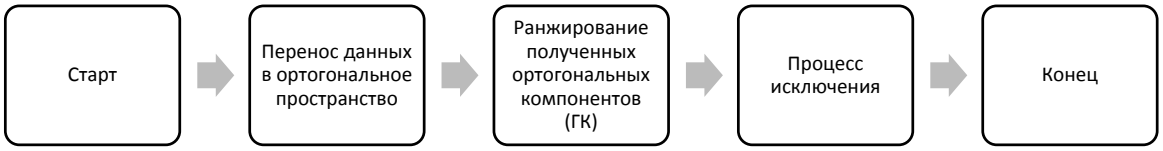

Рисунок 4. Алгоритм метода анализа главных компонент

$$\dot{\boldsymbol{X}} = \boldsymbol{V}\dot{\boldsymbol{F}} = \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{F},$$

откуда

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}}, \qquad F_j = \frac{\sum_{i=1}^{k} a_{ij}\dot{X}_i}{\lambda_j} = \frac{\sum_{i=1}^{k} v_{ij}\dot{X}_i}{\sqrt{\lambda_j}}, j = \overline{1,k}$$

Как правило, для анализа используют $k' < k$ первых главных компонент, которыми исчерпывается не менее 70% дисперсии исходных случайных величин.

На рисунке 4 описаны этапы разработки алгоритма PCA. Во-первых, PCA преобразует данные входных функций в ортогональное пространство с использованием ортогонального линейного преобразования. Результатом являются ортогональные компоненты, известные как главные компоненты. Во-вторых, главные компоненты организованы в соответствии с их дисперсией. Разница является мерой вариабельности распределения выборок и выражается как среднее квадратическое отклонение каждого образца от его среднего значения следующим образом:

$$Var = \frac{\sum(sample - var)^2}{\sum|sample|}.$$

Заключительный шаг исключает ГК с наименьшим вкладом в дисперсию в наборе данных. В принципе, выбор ГК достаточно для учета полной дисперсии наблюдаемых переменных. ГК с наибольшей дисперсией занимает первое место, а те, у кого наименьшая дисперсия, занимают последнее место.

Для решения задачи распознавания в редуцированном признаковом пространстве использовались современные вычислительные технологии и технологии искусственного интеллекта deep learning, основанные на методах машинного обучения и глубокого анализа обучающих выборок. Нейроны искусственной нейронной сети (ANN) функционируют как переключатели для приема сигналов от других нейронов. Состояние выхода нейрона либо «активировано», либо «неактивно», в зависимости от суммы умножения входов и весов, подающих нейрон. Вес, при котором умножается вход, соответствует силе синапса. Было найдено, что АНН успешно решает проблемы, начиная от распознавания речи, кластеризации, системы прогнозирования, распознавания образов и классификации заболеваний.

На рисунке 5 показан общий процедурный поток для ANN. Во-первых, исходные спектральные данные проходят стадию предварительной обработки. Этот этап используется для подавления фона. Например, могут быть лишние признаки в рассеянных данных, которые необходимо обрезать с помощью метода выбора признаков. Выбранные функции затем используются как входные данные для классификатора ANN. Наконец, классификатор интерпретирует результат классификации шаблонов обучения.

В данной работе мы использовали открытую программную библиотеку для машинного обучения TensorFlow, разработанную компанией Google для решения задач построения и тренировки нейронной сети с целью автоматического нахождения и классификации образов. Вычисления в TensorFlow выражаются в виде потоков данных через граф состояний.

На рисунке 6 показаны результаты обучения при 10000 эпохах (50 эпох на точку графика). При этом оценка корректности распознавания на наборе данных для обучения составила 97,5 %.

Наилучшие результаты обучения системы были δ=0,00968478, при обучении 5000 эпох. Корректность распознавания злокачественной ткани при кросс-валидации системы составила 98 %.

Рисунок 5. Алгоритм искусственной нейронной сети

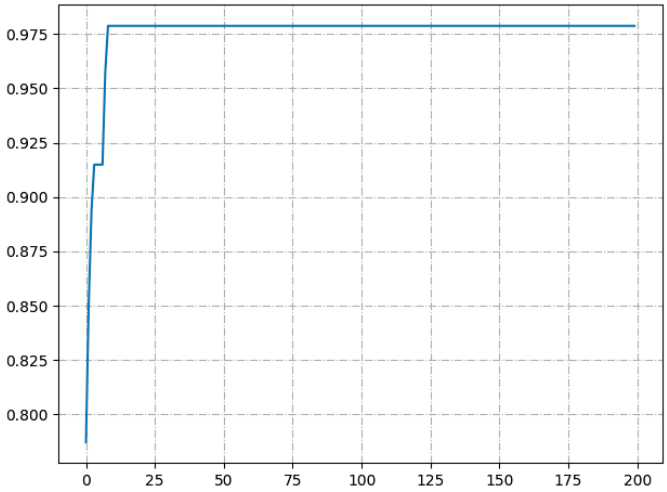

Рисунок 6. Изменение ошибки в процессе обучения (δ=0,00976077)

ЗАКЛЮЧЕНИЕ

В данной работе предложен интеллектуальный алгоритм распознавания злокачественной ткани на основе глубокого анализа спектров комбинационного рассеяния света. Алгоритм был реализован на языке Python с использованием технологий Deep Learning. Достоверность идентификации злокачественной опухолевой ткани составила от 97,5 % до 98 %.

Список литературы / References:

1. Li Q.-B., Wang W., Liu Ch.-H., Zhanga G.-J. Discrimination of breast cancer from normal tissue with Raman spectroscopy and chemometrics. *J. Applied Spectroscopy*, 2015, vol. 82, pp. 450-455.

2. Jermyn M., Desroches J., Aubertin K. [et al.] A review of Raman spectroscopy advances with an emphasis on clinical translation challenges in oncology. *Phys. Med. Biol.*, 2016, vol. 61, no. 23, pp. R370-R400.

3. ur Rehman I., Movasaghi Z., Rehman Sh. *Vibrational spectroscopy for tissue analysis*. CRC Press, 2012, 356 p.

4. Zhu J., Zhou J., Guo J. [et al.]Surface-enhanced Raman spectroscopy investigation on human breast cancer cells. *Chemistry Central Journal*., 2013, pp. 7-37.

5. Harmsen S., Wall M.A., Huang R., Kircher M.F. Cancer imaging using surface-enhanced resonance Raman scattering nanoparticles. *Nat. Protoc.*, 2017, vol. 12, pp. 1400-14.

6. Harris A.T., Lungari A., Needham C.J. [et al.] Potential for Raman spectroscopy to provide cancer screening using a peripheral blood sample. *Head Neck Oncol*., 2009, vol. 1, no. 1, art. 34.

USING THE METHOD OF COMBINATION LIGHT SCATTERING IN DIAGNOSIS OF HUMAN TUMORS

Pavlov V.N., Bilyalov A.R., Kovtunenko A.S., Gilmanova R.F., Bikmeev A.T., Karchevsky S.G., Khalilov R.R.

Bashkir State Medical University
Ufa, Russia
Ufa State Aviation Technical University
Ufa, Russia
Institute of Petrochemical Processing of The Republic of Bashkortostan
Ufa, Russia

Abstract. The purpose of this study is to develop a universal algorithm for analyzing the spectrograms of Raman scattering (Raman-spectroscopy) using intelligent data processing methods. It is found that the optimal algorithm is the calculation of the baseline spectrum by iterative polynomial regression, subsequent frequency analysis of the spectrum, determination of the main components of the spectrum, and machine learning. The reliability of identification and classification of normal tissue and tissue of a malignant tumor was 97.5 to 98 %.

***Key words**: Raman scattering, Raman spectroscopy, malignant tumor, frequency analysis, deep learning.*