

И.А. Беляев, С.А. Фоминых

Построение математических моделей для оценки значения коэффициента светопропускания спонгиозы по характеристикам ее изображений

I.A. Belyaev, S.A. Fominykh

The Construction of Mathematical Models to Estimate Optical Transmission Coefficient for Spongy Bone Using Image Features

Дано краткое описание проблемы оценки возраста в судебной медицине и одного из способов ее решения, основанного на измерении коэффициента светопропускания спонгиозы. Предложена математическая модель, позволяющая получать значения этой величины при помощи анализа цифровых изображений губчатой кости.

Ключевые слова: обработка и анализ изображений, формальное моделирование, фрактальная размерность, регрессия на опорных векторах.

The article gives a synopsis of age estimation problem in forensic medicine and a description of one of quantity methods that solve this problem based on optical transmission measurements. The authors propose a mathematical model that allows obtaining optical transmission coefficient using analysis of digital images of sponge tissue.

Key words: image processing and analysis, soft modeling, fractal dimension, support vector regression.

1. Введение. Установление личности неизвестных людей – одна из актуальных проблем судебной медицины. Когда отождествление личности по признакам внешности и папиллярным узорам пальцев кистей невозможно, возникает необходимость в проведении судебно-медицинских идентификационных исследований, где основными объектами будут выступать кости скелета, которые наиболее устойчивы к различным внешним воздействиям и состояние которых, согласно мнению многочисленных исследователей [1–3], отражает общие процессы развития организма и является информативным показателем возрастных изменений.

В настоящее время в судебно-медицинской практике наиболее распространен описательный подход к определению возраста по макроскопическим параметрам костных останков [4–5]. Но в последние годы предпринимается все больше попыток перехода к количественным методам. Один из таких методов, разрабатываемый на кафедре судебной медицины с основами права Алтайского государственного медицинского университета (АГМУ) [6–8], основан на измерении коэффициента светопропускания губчатого костного вещества (спонгиозы) и построении регрессионной модели, описывающей связь между получаемой при измерении величиной и хронологическим возрастом.

Однако не всегда лаборатории судебно-медицинской экспертизы оснащены фотоколориметрами, что может существенно ограничить распространение указанного метода. Кроме того, процедура проводится в ручном режиме, что нередко является

источником ошибок, а результат может зависеть от опыта эксперта.

Известно, что в случаях, когда прямые измерения требуют больших затрат труда, времени, уникального оборудования, а иногда и вовсе могут быть опасны или даже невозможны, проводят косвенные измерения. При таком походе измеряются некоторые величины, связанные с определяемым параметром математическими соотношениями.

По характеру построения функциональной связи между целевыми параметрами объекта исследования со значениями информационных данных, полученных в процессе измерений, можно выделить *содержательные* (теоретические) и *формальные* (эмпирические) модели [9, с. 16]. Содержательное моделирование базируется на физических принципах и позволяет экстраполировать значения искомой величины в новых условиях. Параметры такой модели имеют физический смысл и их значения могут помочь при интерпретации найденной зависимости. Однако такой подход может быть использован только тогда, когда модель известна априори, на основе некоторой формализованной теории.

При использовании формальных моделей данные описываются эмпирической зависимостью, справедливой в ограниченном диапазоне условий. В этом случае не нужно знать, как устроен механизм исследуемого явления, однако такой метод не позволяет решать задачи экстраполяции. Стоит заметить, что формальный подход позволяет проводить ненаучное моделирование сколь угодно сложных объектов и явлений [10, с. 25], но пара-

метры таких моделей лишены физического смысла и должны интерпретироваться соответствующими математическими методами.

Один из способов получения измерительной информации об объекте – построение его изображения. Применение вычислительной техники и математических методов для анализа изображений позволяет автоматизировать процесс измерений, повысить точность результатов исследований. В настоящей работе рассматривается построение формальной математической модели для оценки коэффициента светопропускания спонгиозы на основе характеристик ее изображения. Проводится сравнение точности прогноза при различных подходах к описанию изображения.

2. Теория. В общем случае процедура анализа изображения включает в себя четыре этапа: получение изображения, его предварительную обработку, построение вектора признаков и регрессионной или классификационной модели, описывающей взаимосвязь дескрипторов изображения и характеристик объекта. Одна из главных проблем при таком подходе – поиск описателей изображения, коррелирующих с определяемыми свойствами объекта.

Губчатое вещество кости может описываться как изотропная (рис. 1а–б) или, в случае сильных изменений, квазиизотропная (рис. 1в) текстура. Такое представление позволяет применять различные количественные дескрипторы текстур для построения вектора признаков. Во-первых, это классические статистические описатели текстур на основе матрицы совместной встречаемости интенсивностей [11,

с. 951–952]. Не менее эффективно в текстурном анализе применение спектральных характеристик на основе статистик коэффициентов вейвлет-преобразования [11, с. 567–577], а также метод средних угловых (АМТ, Angle Measure Technique) спектров [12], показавший свою высокую эффективность в ряде задач, связанных с анализом изображений неоднородных структур [13, 14]. Кроме того, была предложена процедура построения вектора признаков на основе фрактальной размерности. Для этого исходное изображение было подвергнуто процедуре бинаризации по порогу яркости с его значением $T \in [0, 1; 0 < 2, \dots; 0, 9]$ и последующим инвертированием, как это показано на рисунке 2. Здесь черные пиксели соответствуют фону, белые – структурным элементам. Затем для каждого бинарного изображения рассчитывалась клеточная фрактальная размерность [15, с. 21–23].

Согласно алгоритму, изображение покрывается сеткой с шагом δ (рис. 3 слева) и подсчитывается количество ячеек, занятых структурными элементами. Затем шаг сетки увеличивается, и подсчет повторяется снова. После чего в дважды логарифмических координатах строилась зависимость количества занятых ячеек от шага разбиения (рис. 3 справа). Фрактальная размерность рассматриваемого изображения численно равна тангенсу угла наклона прямой, аппроксимирующей полученную зависимость. Далее рассчитанные значения фрактальной размерности объединялись в вектор признаков.

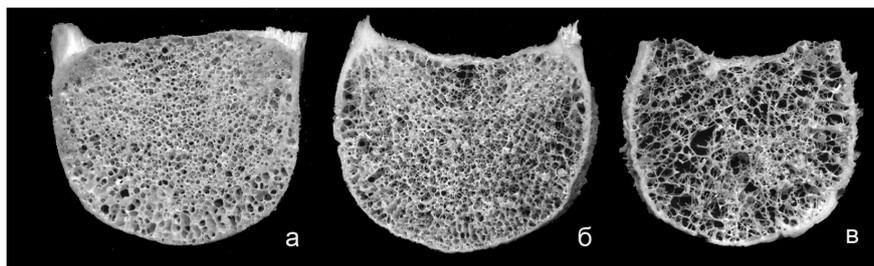


Рис. 1. Изображения губчатых костей с разной степенью изменений пористой структуры

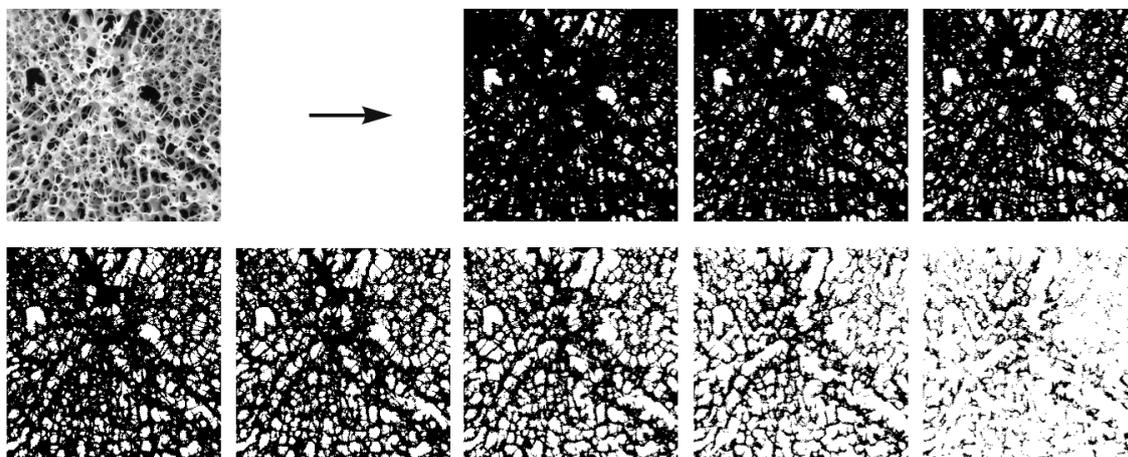


Рис. 2. Бинаризация изображения среза позвонка с разными значениями порога яркости

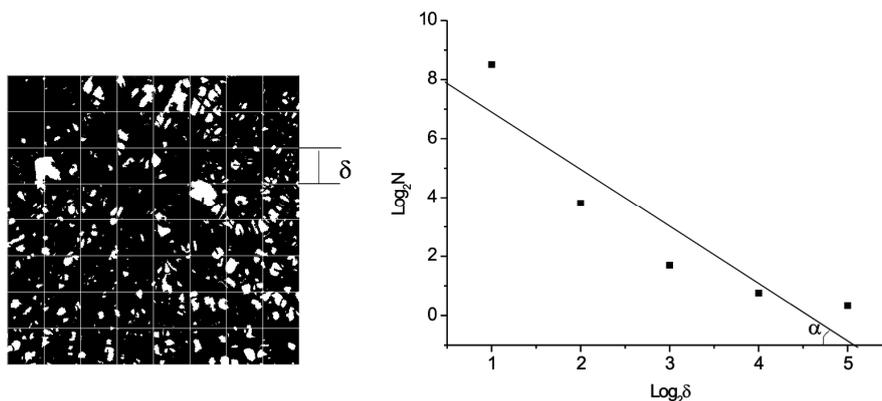


Рис. 3. Процедура расчета клеточной фрактальной размерности для бинарного изображения

Для построения моделей была использована регрессия на опорных векторах [16] с ядром в виде радиальной базисной функции. Для каждого типа векторов рассчитывалась собственная модель.

3. Эксперимент и некоторые результаты. Для экспериментов был использован набор из 210 костных пластинок I–V позвонков поясничного отдела, изъятых у 42 лиц мужского пола. Процедура предварительной подготовки образцов детально описана в [7, 8]. Для получения цифровых изображений при помощи планшетного сканера каждую пластинку сканировали с двух сторон.

Изображения срезов, принадлежащие одному объекту, сохранялись внутри каталога, имя которого состояло из буквы, обозначающей пол (М), паспортного возраста умершего и номера записи в журнале регистрации трупов. Основной каталог содержал пять подкаталогов с именами от 1 до 5, соответствующих номерам срезов. Затем из полученных снимков «вырезался» фрагмент, содержащий изображение губчатого вещества размером 2048x2048 точек, что составляло 70–80% площади спонгиозы на снимке (рис. 4а). После чего для каждого изображения рассчитывались четыре различных вектора признаков: на основе статистик матрицы совместной встречаемости (далее – TF), статистик коэффициентов вейвлет-преобразования (WT), среднеугловых спектров (AMT), фрактальной размерности (FD). Полученные значения векторов признаков составляли матрицы информационных

данных (матрица предикторов X), где каждая строка соответствовала образцу, столбец – переменной.

Кроме того, для этих же срезов определялся интегральный показатель светопропускания. Измерения проводились в 10 областях пластинки при помощи фотокориметра КФК-3 на длине волны 550 нм. Затем полученные значения усреднялись. Схема расположения точек измерения приведена на рисунке 4б, здесь же показан контур области, для которой рассчитывался вектор признаков. Значения интегральных показателей светопропускания объединялись в матрицу целевых значений (матрицу откликов Y). Затем данные были разбиты на два множества. Калибровочный набор составили 110 образцов со значением коэффициента светопропускания от 2,87 до 53,74%, проверочный – 100 образцов с коэффициентом пропускания от 3,68 до 41,94%. Разбиение на калибровочный и тестовый набор проводилось экспертом в области судебной медицины. Для построения регрессионных моделей, описывающих связь коэффициента светопропускания и характеристик изображений срезов, была использована регрессия на опорные вектора с ядром в виде радиальной базисной функции.

При построении регрессионных соотношений возникает проблема сбалансированного описания данных – необходимо сбалансировать точность описания калибровочного набора и качество прогнозирования для новых образцов.

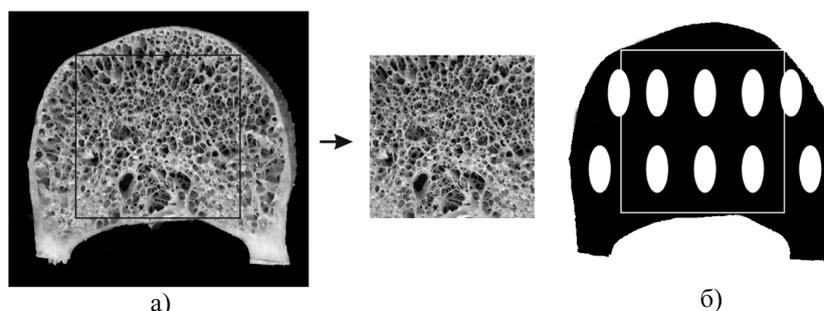


Рис. 4. а – изображение среза тела I поясничного позвонка (справа – вырезанный фрагмент, на основе которого производится расчет вектора признаков); б – примерная схема расположения точек измерения светопропускания на костной пластинке

Точность калибровки модели принято характеризовать величиной *среднеквадратичного остатка калибровки* (root-mean square error of calibration, *RMSEC*) [9, с. 32], который вычисляется по формуле

$$RMSEC = \sqrt{\sum_{i=1}^I (y_i - \hat{y}_i)^2 / I}, \quad (1)$$

где y_i и \hat{y}_i соответственно, измеренные и предсказанные значения светопропускания для образцов сравнения $i = 1, \dots, I$. Ясно, что чем меньше *RMSEC*, тем точнее описываются исходные данные.

Прогнозные свойства модели можно оценить, используя величину, называемую *среднеквадратичным остатком прогноза* (root-mean square error of prediction, *RMSEP*), определяемую по формуле, аналогичной (1):

$$RMSEP = \sqrt{\sum_{i=1}^M (y_i - \hat{y}_i)^2 / M}, \quad (2)$$

где y_i и \hat{y}_i – соответственно, измеренные и предсказанные значения светопропускания; M – количество образцов проверочного набора.

Параметры *RMSEC* и *RMSEP* связаны между собой – улучшение одного из них ведет к ухудшению другого. Анализируя их совместно, можно подобрать оптимальные параметры модели. Кроме того, эти величины являются удобным инструментом для сравнения разных моделей, так как имеют размерность исходных данных.

Для каждого типа вектора признаков обучался спектр моделей с различными значениями параметра ядра и начальными значениями весовых коэффициентов, и для каждой из них вычислялись средне-

квадратичные остатки калибровки и прогноза, используя выражения (1) и (2). Выбиралась модель, которая имела минимальные значения *RMSEC* и *RMSEP* с наименьшей разностью между ними. В таблице представлены значения этих характеристик для окончательных моделей. Здесь же указан еще один параметр, характеризующий качество модели – коэффициент корреляции R между измеренными y_i и предсказанными \hat{y}_i значениями светопропускания. Для идеальной модели $R = 1$.

Точность калибровки и прогноза для различных моделей (R_c – коэффициент корреляции между предсказанными и измеренными значениями для калибровочного набора; R_p – для тестового)

| | RMSEC | R_c | RMSEP | R_p |
|-----|-------|-------|-------|-------|
| AMT | 0,039 | 0,920 | 0,055 | 0,840 |
| WT | 0,038 | 0,940 | 0,055 | 0,820 |
| TF | 0,400 | 0,920 | 0,038 | 0,920 |
| FD | 0,038 | 0,940 | 0,038 | 0,920 |

Из таблицы видно, что наиболее точный прогноз как для калибровочного, так и для тестового набора дает модель с вектором признаков на основе значений фрактальной размерности (FD). Сравнимыми характеристиками обладает модель с вектором признаков на основе статистик матрицы совместной встречаемости интенсивностей (TF). Проведение дополнительных экспериментов по исследованию устойчивости этих моделей к различным факторам, искажающим изображения, таким как шум и изменение разрешения, позволит выделить наилучшую модель и определить условия ее применения на практике.

Библиографический список

1. Попов Н.В. Судебная медицина: учебник для юридических институтов. – М., 1938.
2. Никитюк Б.А. Старение скелета // Проблемы геронтологии и гериатрии в ортопедии: сб. науч. ст. – Киев, 1966.
3. Пиголкин Ю.И., Федулова М.В., Гончарова Н.Н. Судебно-медицинское определение возраста. – М., 2006.
4. Hansen G. Die Altersbestimmung am proximalen Humerus und Femurende im Rahmen der Identifizierung menschlicher Skelettreste // Wissenschaftliche Zeitschrift der Humboldt-Universit. – 1953. – Vol. 3.
5. Неклюдов Ю.А. Экспертная оценка возрастных изменений скелета верхней конечности. – Саратов, 1992.
6. Эрлих Э.Р. Комплексное исследование тел поясничных позвонков для определения пола и возраста человека: автореф. дис. ... канд. мед. наук. – Барнаул, 1993.
7. Янковский В.Э., Фоминых С.А., Пивоваров А.В. Сравнительная оценка остеопоротических изменений губчатого вещества тел поясничных позвонков в возрастном аспекте // Актуальные вопросы судебной и клинической медицины: сб. науч. ст. – Ханты-Мансийск, 2008. – Вып. 10.
8. Янковский В.Э., Фоминых С.А., Пивоваров А.В. Применение регрессионных уравнений при диагностике возраста по инволютивным изменениям губчатой

кости // Актуальные вопросы судебной медицины и экспертной практики: сб. науч. ст. – Новосибирск; Томск, 2008. – Вып. 13.

9. Родионова О.Е., Померанцев А.Л. Хемометрика в аналитической химии. – М., 2008.

10. Зиновьев А.Ю. Визуализация многомерных данных. – Красноярск, 2000.

11. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М., 2005.

12. Andrieu R. The Angle Measure Technique: a new method for characterizing the complexity of geomorphic lines // Mathematical Geology. – 1994. – Vol. 26, №1.

13. Casper K. Dahl, Kim H. Esbensen, Image analytical determination of particle size distribution characteristics of natural and industrial bulk aggregates // Chemometrics and Intelligent Laboratory Systems. – 2007. – Vol. 89(1).

14. Kucheryavski S., Belyaev I. Classification and analysis of non-isotropic images by Angle Measure Technique (AMT) with contour unfolding // Analytica Chimica Acta. – 2009. – Vol. 642, Issues 1–2.

15. Федер Е. Фракталы. – М., 1991.

16. Harris Drucker, Chris J.C. Burges, Linda Kaufman, Alex Smola and Vladimir Vapnik. Support Vector Regression Machines // Advances in Neural Information Processing Systems. – 1996. – Vol. 9.