

# Информационные системы для анализа медицинских диагнозов

А.М. Доровской<sup>1</sup>

<sup>1</sup>Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

**Аннотация.** Сегодня информационные технологии позволяют нам собирать множество статистических данных и вычислительных мощностей достаточно для выполнения самых сложных операций с собранными данными. Объединив эти факторы, мы можем использовать их для улучшения работы системы здравоохранения. В качестве примера в данной статье рассматривается применение регрессионного анализа данных в области медицины. Идентификация данных подтверждает корреляцию на основе медицинских диагнозов и позволяет установить чему будет равен медицинский показатель на основе известного значения иного показателя.

## 1. Введение

В течение последних 10 лет в здравоохранении идет внедрение информационных систем, которые по проблематике разделяются на несколько типов: хранение данных, анализ данных, ввод/вывод данных. При этом наблюдается неуклонный рост числа врачей, использующих информационные системы ежедневно.

**Таблица 1.** Статистика информационных систем в США[3].

Год	Врачи (%)
2013	48
2014	50
2015	63
2016	66
2017	67
2018	70

Важно отметить, что, по статистике, российские врачи используют информационные системы в полтора раза реже, но при этом наблюдается та же динамика. Применение некоторых информационных систем может ускорять прием у врача на 20-25%. Исходя из этого, если поликлиники в России принимают в среднем около 500 человек в день, то становится возможным прием 600-650 человек ежедневно[4].

## 2. Информационные системы интеллектуального анализа данных

В разделе описаны подходы, применяемые в информационных системах, которые внедряются в лечебно-профилактических учреждениях.

### 2.1. Типы информационных систем

Внедряемые в медицине информационные системы, реализующие интеллектуальный анализ данных, можно разделить на четыре группы, каждая из которых решает определенные задачи. Разделение производится следующим образом: прогнозирование (течения болезни, воздействия препарата или группы препаратов, уровня смертности), обследование (постановка диагноза на основе совокупности симптомов), классификация (уточнение диагноза), поиск ассоциаций (поиск скрытых зависимостей между различными показателями здоровья пациентов). Какие именно методы интеллектуального анализа применяются для каждой группы описано в таблице 2.

**Таблица 2.** Задачи интеллектуального анализа данных в медицине и методы, применяемые для их решения.

Цель анализа	Методы с учителем	Методы без учителя
Прогнозирование	Метод наименьших квадратов Логистическая регрессия Нейронные сети Деревья принятия решений Сплайны	—
Обследование	Деревья принятия решений	Метод главных компонент Кластеризация Анализ ссылок
Классификация	Деревья принятия решений Нейронные сети Дискриминантный анализ Бустинг Наивный Байесовский классификатор	Кластеризация Самоорганизующиеся карты Кохонена
Поиск ассоциаций	—	Факторный анализ Априорный алгоритм

### 2.2. Примеры использования информационных систем с интеллектуальным анализом

Одним из наиболее ярких примеров результативности подобных систем является работа систем классификации, которые способны правильно определить тип опухоли с уровнем точности 95%, в то время как оценка даже самого квалифицированного специалиста не будет превышать 80%. Эта информация позволит назначить более правильное лечение, что существенно увеличивает шансы на выздоровление пациента.

Системы прогнозирования позволяют заранее предсказать вспышки заболеваний или вероятное повышение смертности от тех или иных причин, что позволит подготовить заранее необходимые медикаменты и уменьшить число фатальных случаев.

## 3. Применение информационной системы прогнозирования смертности

В данном разделе описан пример использования прогнозирования смертности на основе данных Самарской области за 2010-2014 года. На данный момент одной из самых частых причин смертности являются болезни системы кровообращения (БСК). С помощью математического анализа мы можем предсказать повышение смертности от БСК в зависимости от какого-либо признака.

### 3.1. Способ анализа

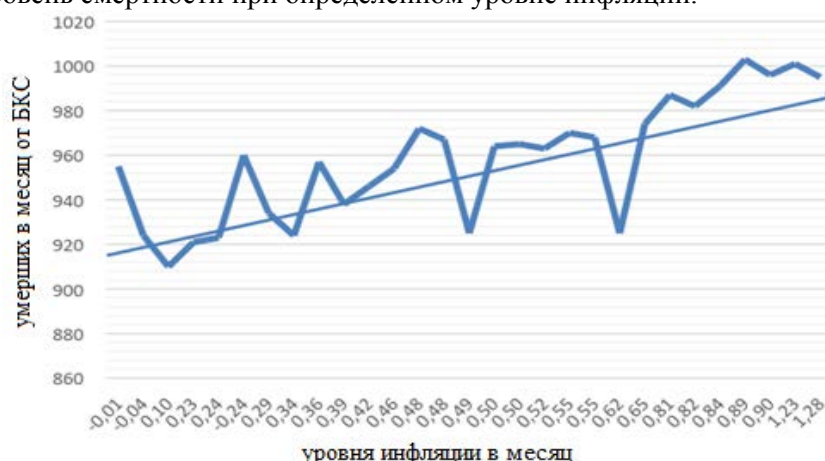
В качестве способа анализа наиболее подходящим является регрессионный анализ, который позволяет определить значение одного признака в зависимости от другого. Попробуем предсказать смертность от БСК в зависимости от уровня инфляции. Известным признаком является значение инфляции. Задача восстановления линии регрессии заключается в построении некоторой функции  $y(x_1, x_2, \dots, x_m)$ , которая наилучшим образом описывает данные из некоторой обучающей выборки  $U$ , в которой каждому вектору предикторов  $x$  ставится в соответствие зависимая переменная  $y$ . Понятно, что эту функцию следует искать в некотором конкретном классе функций.

$$y(x_1, x_2, \dots, x_m) = a^T x + b + \varepsilon \tag{1}$$

где  $m, a \in R$  и  $b \in R$  – это неизвестные коэффициенты в уравнении регрессии,  $\varepsilon$  – гауссовский белый шум с нулевым средним и некоторой дисперсией  $\sigma^2$ . Под  $R$  понимается множество вещественных чисел. В этом случае задача восстановления линейной регрессии состоит в определении параметров  $a$  и  $b$ .

### 3.2. Результаты анализа

В результате проведенных расчетов был построен график, отображенный на рисунке 1, на который были выведены данные, а также построена регрессионная прямая, позволяющая предсказать уровень смертности при определенном уровне инфляции.



**Рисунок 1.** График соотношения уровня инфляции в месяц и умерших в этот месяц от БСК

## 4. Заключение

В данной работе приведены возможности внедрения информационных систем и методов анализа, в лечебных учреждениях. Разнообразие способов анализа позволяет говорить о возможности применения их во все аспектах медицины. Статистика внедрения показала, что есть неоспоримая польза от подобных систем. В качестве примера была продемонстрирована возможность применения регрессионного анализа, с помощью которого удалось установить на каком уровне окажется смертность пациентов с БСК при известном уровне инфляции. Полученный результат не позволяет утверждать, что эти признаки связаны напрямую. Однако выяснение того, какие еще показатели могут влиять на ситуацию, не является темой данной статьи.

## 5. Литература

- [1] Баранов, А.А. Системный анализ в медицине и биологии / А.А. Баранов, Л.С. Намазова-Баранова, И.В. Смирнов, Д.А. Девяткин, А.О. Шелманов, Е.А. Вишнева, Е.В. Антонова, В.И. Смирнов, А.В. Латышев // Труды ИСА РАН. – 2015. – Т. 65. – С. 93.
- [2] Гайдель, А.В. Лабораторные работы по курсу: методические указания / А.В. Гайдель, А.Г. Храмов // Интеллектуальный анализ данных: Самарский университет, 2017. – 102 с.

- [3] Информационные технологии в здравоохранении Health IT & CIO report [Электронный ресурс]. – Режим доступа: <https://www.beckershospitalreview.com/healthcare-information-technology/10-ehr-implementation-statistics.html> (02.12.2018).
- [4] Новостной портал U24.ru [Электронный ресурс]. – Режим доступа: <https://u24.ru/news/15510/vizit-k-vrachu-nachinaetsya-s-registratury-v-poliklinike> (22.11.2018).
- [5] Таблица инфляции [Электронный ресурс]. – Режим доступа: [уровень-инфляции.рф](http://уровень-инфляции.рф) (01.11.2018).

## Information systems for analyzing medical diagnoses

A.M. Dorovskoy<sup>1</sup>

<sup>1</sup>Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

**Abstract.** Today technology allows us to collect a lot of statistical data. And we have enough computing power to perform the most complex operations with the collected statistical data. By combining these factors we can use this to improve health. The paper discusses considering how it is possible to apply regression analysis of data in the medical service industry. Identification of the data validate correlation based on medical diagnoses. In which areas of the service is this already used and what are the results. The result indicating the level of conclusions of medical diagnoses in the time period of interest. The abstract presents an example of the information system. The goal of the information system: Use medical diagnosis cases for analysis in order to obtain a forecast of the disease of the population according to various criteria: the type of diagnosis, age of patient, date when diagnosis is installed.