




СЕКЦІЯ 11

**ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ
У ФАРМАЦІЇ ТА МЕДИЦИНІ**



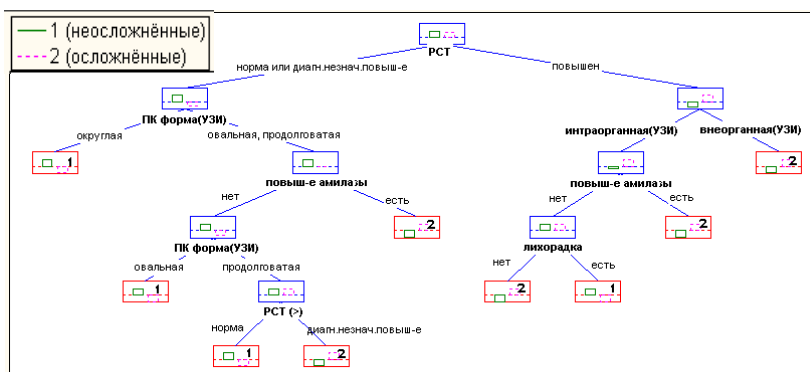
ПРИМЕНЕНИЕ ЛОГИЧЕСКИХ АЛГОРИТМОВ ДЛЯ ИЗВЛЕЧЕНИЯ ЗНАНИЙ ИЗ МЕДИЦИНСКИХ БАЗ ДАННЫХ

Носулич М.С., асс. Нессонова М.Н., асс. Балака С.Н*

Национальный фармацевтический университет, г.Харьков

*Харьковский национальный медицинский университет

В основе обширного класса алгоритмов классификации лежит принцип индуктивного вывода логических закономерностей или индукции правил. Особую ценность представляют закономерности, которые описываются простой логической формулой (правила). Процесс поиска правил по выборке называют добычей данных (data mining) или извлечением знаний из данных (knowledge discovery). На практике логические закономерности часто ищут в виде конъюнкций небольшого числа элементарных высказываний. Именно в такой форме люди привыкли выражать свой житейский и профессиональный опыт, т.е. знания, представленные таким образом, являются интерпретируемыми. Одним из наиболее популярных методов решения задач классификации и прогнозирования является метод деревьев решений (decision trees). Иногда этот метод data mining также называют деревьями решающих правил или деревьями классификации (если целевая переменная принимает дискретные значения). В данной работе представлены результаты использования метода деревьев решений для добычи данных о развитии осложнений жидкостных полостных образований поджелудочной железы (ПОПЖ) в различные сроки после приступа деструктивного панкреатита.



Метод деревьев решений позволил представить классификацию осложнённых и неосложнённых случаев течения ПОПЖ в виде простой иерархической структуры (рис.). Дерево классификации было построено методом полного перебора деревьев с одномерным ветвлением по алгоритму CART с мерой Джини (Gini) для оценки качества согласия модели с исходными данными; для остановки ветвления выбрана прямая остановка по методу FACT с установленной долей неклассифицированных наблюдений менее 5%. Для оценки качества обучения алгоритма проведена V-кратная кросс-проверка (V-fold cross-validation), что позволило получить оптимальное дерево решений. На каждом этапе данный диагностический алгоритм имеет два варианта решения, приводящих в итоге к заключению о наличии осложнений. На обучающей выборке получена 100%-но правильная классификация объектов, а при глобальной кросс-проверке 4 объекта из группы больных с отсутствием осложнений были ошибочно отнесены к пациентам с осложненным течением, а 2 случая из группы больных с наличием осложнений течением – как неосложнённые. Таким образом, общая точность классификации составила 96,5%.

Полученные результаты позволили выявить и формализовать закономерности возникновения осложнений у пациентов с ПОПЖ, что является основным необходимым условием для их прогнозирования и предупреждения.