Медицинские изображения – это структурно-функциональный образ органов человека, предназначенный для диагностики заболеваний и изучения анатомо-физиологической картины организма.

Аналоги. Для автоматического построение признаков можно использовать деревья решений, индуктивное логическое программирование и аннотации.

Актуальность выбранной темы обосновывается тем, что своевременное распознавание патологических процессов в организме человека приведет к оказанию необходимой медицинской помощи. Проблема классификации патологических процессов по данным медицинских изображений не может быть правильно решена, если важные взаимодействия и отношения между оригинальными признаками, не принимаются во внимание.

Проблема. Эффективность работы классификатора сильно зависит от входного множества признаков. Как выбрать оптимальное множество признаков для классификатора?

Объектом исследования выступает классификация изображений. Предметом исследования является алгоритм построения признаков для классификации.

Цель – нахождение и отработка методики построения признаков для решения задачи классификации.

Для достижения поставленной цели необходимо выполнить следующие задачи:

* изучить соответствующую литературу;
* разработать алгоритм построения признаков;
* реализовать программу по данному алгоритму;
* оценить эффективность работы алгоритма и сравнить с результатами классификации без построения признаков;
* обобщить полученные результаты и сделать соответствующие выводы.

Классификация заключается в прогнозировании значения категориального атрибута (класса) на основе значений признаков объекта при известном множестве обучающих примеров.

Концептуально любой метод построения признака можно рассматривать как выполнение следующих действий:

1) Выбор начального пространства признаков F0 (ручное построение признаков).

2) Преобразование F0 для построения нового пространства признаков FN (преобразование признаков).

3) Выбор подмножества признаков Fi из FN (выбор признака).

a. Определение полезности Fi для задачи классификации.

b. Если некоторые критерии завершения не достигнуты:

i. Вернитесь к шагу 3.

c. Иначе множество FT = Fi.

4) FT – это сконструированное пространство признаков.

Для решения задачи автоматического построения признаков будет использоваться генетическое программирование. Данный метод основывается на принципах биологической эволюции: естественный отбор, скрещивание и мутация.

В качестве популяции будут выступать леса – наборы признаков. Каждое дерево леса представляет собой один сконструированный признак, состоящий из функций, исходных признаков и констант.

Будем параллельно развивать несколько наборов лесов. Каждому такому набору назначим свой уникальный классификатор.

Оценивать пригодность особей будем через ошибку классификации. Подаем классификатору набор сконструированных признаков и вычисляем отношение неправильно предсказанных значений класса к числу всех объектов.

В качестве меры пригодности используется квадратичная ошибка. Мы вычисляем ее значение для каждой особи и впоследствии нормализуем. Выбор родителей происходит пропорционально нормализованным значениям пригодности.