Санкт-Петербургский государственный университет

**Кафедра технологии программирования**

**Сун Мин**

**Магистерская диссертация**

**Обнаружение и классификация рака с помощью микроскопических изображений образцов биопсии**

Направление 010300

Фундаментальная информатика и информационные технологии

Магистерская программа «Технологии баз данных»

|  |  |
| --- | --- |
|  |  |
| Научный руководитель,доктор физ.-мат. наук,профессор | Андрианов С.Н. |
|  |  |

Санкт-Петербург

2016

# Содержание

[Содержание 2](#_Toc451368623)

[Введение 4](#_Toc451368624)

[Актуальность темы 4](#_Toc451368625)

[Состояние исследования микроскопического изображения биопсии 6](#_Toc451368626)

[Цель данного исследования 7](#_Toc451368627)

[Практическая значимость исследования 7](#_Toc451368628)

[Задачи данного исследования 7](#_Toc451368629)

[Глава 1. Постановка и формализация задачи 9](#_Toc451368630)

[Глава 2. Методы анализа изображения раковых клеток 10](#_Toc451368631)

[2.1 Сегментация цветного изображения 10](#_Toc451368632)

[Сегментация цветного изображения рака желудка 10](#_Toc451368633)

[Оттенки серого патологического изображения 11](#_Toc451368634)

[Метод Оцу и быстрый алгоритм выбора порогов Оцу 12](#_Toc451368635)

[Алгоритм k-means 15](#_Toc451368636)

[Алгоритм сегментации по водоразделам (WaterShed) 18](#_Toc451368637)

[Принцип алгоритма 18](#_Toc451368638)

[Традиционное математическое описание водораздельного преобразования 18](#_Toc451368639)

[Водораздел Сегментация на основе расстояния преобразования 20](#_Toc451368640)

[2.2 Извлечения признаков микроскопическое изображение клеток 23](#_Toc451368641)

[Функция представления изображения и способ описания 24](#_Toc451368642)

[Описание характеристики изображения 27](#_Toc451368643)

[Описание особенности формы раковых клеток 31](#_Toc451368644)

[Глава 3 основные средства 39](#_Toc451368645)

[Вывод 43](#_Toc451368646)

[Заключение 44](#_Toc451368647)

[Список литературы 46](#_Toc451368648)

[Приложение 47](#_Toc451368649)

# Введение

### Актуальность темы

На современном этапе развития человечества онкологические заболевания являются одной из основных причин заболеваемости и смертности во всем мире — С ростом заболеваемости и смертности, рак является ведущей причиной смерти по всему мир и в Китае, что является серьезной проблемой общественного здравоохранения. Из-за огромного населения Китая, на данный момент, которое составляет 1,37 миллиардов человек, и ужасающего состояния загрязнения окружающей среды в Китайской Народной Республике[1], каждый год заболевает три тысячи миллионов человек, что составляет весомую часть от всех жителей республики. О такие статистике свидетельствует ежегодный отчет о распространении онкологических заболеваний на территории страны.

В настоящее время, большинство врачей для диагностики онкологических заболеваний используют, визуальные методы диагностики, что с помощью определённых приборов, помогает им выявить наличие пораженных клеток, в их число входят такие методы[2], как:

Биопсия – это самый распространённый метод диагностирования раковых клеток, который включает в себя, исследование с использованием полой иглы, и других хирургических приборов, которые помогают отделить частицы ткани, которые вызвали подозрение у врача и рассмотреть их более детально, для постановления более точного диагноза.

Компьютерная томография – это более современный метод диагностирования рака, который включает в себя послойный рентген. При помощи компьютера, который управляет съемкой, строится трехмерное изображение, которое позволяет разглядеть даже самые маленькие пораженные клетки.

Следующем методом диагностики является эндоскопия – это исследование полых органов при помощи эндоскопического гибкого зонда, который позволяет высветить внутренность органа и при необходимости удалить полипы.

Компьютерная томография с использованием магнитных колебаний – этот метод схож по принципу с обычной томографией, но в отличие от рентгена используется магнитное поле, которое обладает большим проникающим действием, что обеспечивает более надежную диагностику.

Также, не мало важными считаются методы диагностики такие как, маммография, которая является частью томографии, сонография, где используют для диагностики звуковые колебания, сцинтиграфия, где используется радиоактивные вещества.

Все эти методы позволяют более точно диагностировать онкологические заболевания, но, к сожалению, на современном этапе развития социума, такие методы не являются достаточными, клетки мутируют, что в значительной степени затрудняет диагностику и распознавание заболевания.

Но также такого методы такого рода, имеют ряд отрицательных моментов, такие как,

1. Результаты исследования носят строго субъективный характер;
2. Неэффективность, постановка диагноза иногда является неверной;
3. Присутствует человеческий фактор, например, усталость.
4. В средних или маленьких городах Китая из-за неравномерного распределения медицинских ресурсов некоторые врачи не имеют клинического опыта и знаний, что также сказывается на постановке диагноза.

 Эти недостатки будут влиять на точность диагностического анализа и решения, как было отмечено американским учеными – существует необходимость компьютерной диагностики (САПР), чтобы уменьшить нагрузку на специалистов в процессе постановки диагнозов, и более серьёзного диагностирования более сложных случаев.

### Обзор методов анализа микроскопических изображений биопсии

С использованием компьютера сбор, обработка и хранение микроскопического изображения медицинской биопсии является актуальным направлением исследования медицинских работников в восьмидесятых и начале девяностых годов прошлого века. С быстрым развитием компьютерных технологий, люди связывают вычислительную мощность компьютера с субъективной способностью диагностики самим врачом, в то же время компьютер может сделать как и медицинские эксперты, он четко распознаёт все типы клеток и быстро выявляет соответствующий диагноз, что в значительной степени повышает эффективность и точность клинической диагностики[3].

Использование четырех различных классификаторов, обученных на 25-мерным вектором признаков, повышает производительности в 98% на 737 изображениях. Аналогично в работах [4], [5], [6] предлагается система диагностики для рака молочной железы, на основе информации об ядрах сегментации цитологического изображения. При использовании различных моделей машинного обучения, такие как нейронные сети и метод опорных векторов, утверждается об уровне диагностической точности в пределах от 76% до 94% на наборе из 92 изображений. Чжан [7] предложил каскадный подход с опцией отказа. На первом уровне каскада, автор предлагает решать простые случаи, в то время как сложные направляются на второй уровень, где выполняется более тщательная диагностика, а также используется система классификации. Они оценивают при помощи метода, который был предложен Израильским технологическим институтом, который состоит из 361 изображений (40 ×). При этом методе достоверность данных составляет 97%.

Cao [8] представил классификацию методов гистологии изображений на основе особенностей текстуры и с помощью Adaboost-BP максимальную точность классификации, которая была получена, равно 91,5%.

### Цель данного исследования

Предложить структуру автоматического обнаружения и классификации рака для микроскопического изображения биопсии с использованием клинически значимых и биологически интерпретируемых особенностей.

### Практическая значимость исследования

Применение технологии компьютерной обработки изображений для микроскопического исследования и распознавания изображений, необходимо, для того чтобы уменьшить интенсивность труда врача при «чтении» биопсии, а также для того, чтобы повысить точность постановки диагноза. Традиционные методы идентификации предполагают исключительно обнаружение раковых клеток в желудке под микроскопом, при помощи опыта каждого врача онколога, а функция распознавания клеток рака желудка не имеют количественных параметров в качестве эталона, полностью зависит от опыта врача, в результате чего скорость распознавания не высока. Для быстрого и точного количественного анализа клеток, они должны быть охарактеризованы и выделены определёнными признаками клеточной морфологии, а также измерены для выявления больных клеток.

Так называемое выделение признаков является количественным описанием клеток, оно играет очень важную роль в процессе идентификации клетки, и непосредственно влияет на точность распознавания. Из-за своей морфологической сложности рак желудка можно выделить, используя признаки, одновременно извлекая их различными способами, таким образом, ключевым моментом является поиск характеристики, которая могла бы выделять признаки и могла бы быть реализована в качестве универсального критерия характеристических параметров. Основываясь на большом количестве биопсии желудочных клеток, изображениях, и медицинских книгах, можно создать основную базу для эффективной классификации. Например, можно использовать как характеристические параметры компьютерного распознавания раковых клеток периметр, площадь, округлость.

Задачи данного исследования

1. Сглаживание изображений с помощью фильтра шумов изображения, для улучшения качества изображения, для тонкой сегментации клеток в области перекрытия.
2. Для получения признака микроскопических изображений биопсии, требуется анализ контура, и обнаружение как можно большей информации о клетке.
3. Для выделения значимых показателей структур, будет использоваться подход, который включает в себя ограниченное адаптивное уравнивание гистограммы.
4. Для сегментации фоновых клеток, предлагается комплексный алгоритм k-средних алгоритм и алгоритм сегментации по водоразделам (WaterShed).
5. На этапе выделения признаков, предлагается, извлечь различные биологически интерпретируемые и клинически значимые формы, а также особенности, основанные на морфологии из сегментированных изображений. К ним относятся периметр, площадь, прямоугольность, округлость и эксцентриситет.
6. Наконец, используется метод BP\_Adaboost для классификации изображений в нормальных условиях, так как он работает лучше, по сравнению с другими широко используемыми методами обратного распространения ошибки для данной задачи.

# Глава 1. Постановка и формализация задачи

Основная цель данной работы заключается в разработке структуры и программного инструмента для автоматического обнаружения и классификации рака по микроскопическим изображениям биопсии с помощью вышеуказанных клинически значимых и биологически интерпретируемых особенностей. В данной диссертации в разделе сегментации изображения основное внимание уделяется комплексному алгоритму, включающему k-средних алгоритм и алгоритм сегментации по водоразделам (WaterShed). Еще используя внедрение метода BP\_Adaboost для классификации изображений, для повышения точности предсказания нейронной сети, что имеет большую практическую ценность.

Остальная часть работы структурирована следующим образом. В Главе 2 описывает соответствующие работы с представлением методов и моделей. В Главе 3 описывает основные средства и соответствующее программное обеспечение. В Главе 4 описываются результаты и проводится обсуждение. И наконец, Глава 5 описывается результаты, полученные в процессе подготовки работы.

# Глава 2. Методы анализа изображения раковых клеток

## 2.1 Сегментация цветного изображения

При анализе изображений возникает необходимость их сегментации, то есть разделение пикселей изображений на группы по некоторым признакам[].

 Целью же процесса сегментации является упрощение или изменение полученного изображения, для более продуктивного последующего анализирования. Сегментация помогает более чётко выделить границы объекта, что значительно упрощает их интерпретацию.

Результатом сегментации является изображение, которое разделено на множество сегментов, сегменты же в свою очередь разделены на пиксели, которые схожи между собой некоторыми особенностями, это может быть цвет или яркость.

Для получения информации о желаемых признаках изображения в стадии анализа, обработки изображения, а также распознавания изображения, должны сначала пройти стадию сегментации изображений. Сегментация изображения является одной из основных технологий компьютерного зрения, которая является ключевым моментом от процесса обработки изображений до анализа изображений. Качество сегментации изображения оказывает непосредственное влияние на выделение признаков, распознавание образов и точность измерений и восприятия, и является ключевым шагом в анализе изображений, а также является ключевым фактором в научно-исследовательской деятельности.

### Сегментация цветного изображения рака желудка

Сегментация цветного изображения состоит из видов текстур регионов и может быть трудной задачей, а именно в моменте вычисления точного поля текстуры и определения оптимального количества областей сегментации изображения, если оно содержит подобные и / или необычные текстуры полей. Многими методами можно решать проблему сегментации изображении, например методы с использованием гистограммы, методы, основанные на кластеризации и методы разрастания областей и так далее.

В данной работе, изображение обрабатывается с помощью HE (hematoxylin-eosin staining) витражного окрашенного изображения клеток тканей желудка, один вариант — это преобразовать гистологическое изображение в оттенки серого и двоичное изображение, которое компьютер может обработать и распознавать. Второй вариант - с помощью метод сегментации цветного изображения, например алгоритм K-средних (метод k-means).



 Рис. 1. Окраска гематоксилином и эозином.

### Оттенки серого патологического изображения

Как и в оттенках серого, здесь содержится только информацию о яркости, патологическое изображение не содержит информацию о цвете изображения, значение яркости можно разделить на 256 уровней. Бинарное изображение можно получить через выбор порога обработки изображения. Преобразование цветного изображения в оттенки серого. Изображения обычно представлены в цветовом пространстве RGB, RGB цветовое пространство является наиболее распространенной цветовой системой, каждый пиксель представлен своими компонентами RGB. RGB изображение преобразуется в черно-белое изображение, нам нужно вычислить значение каждого цвета градацией.

Gray = R\*0.299 + G\*0.587 + B\*0.114

Цветное изображение, в соответствии с приведенной выше формулой, может быть преобразовано в черно-белое изображение.



Рис. 2. Преобразование цветного изображения черно-белое.

### Метод Оцу и быстрый алгоритм выбора порогов Оцу

Метод Thresholding [9] представляет собой простой и эффективный метод сегментации изображений, который использует один или несколько пороговых уровней серого, изображение делится на несколько частей, принцип сегментации – пиксели похожего цвета в одном объекте. Наиболее важной особенностью этого метода является простота вычисления. Основная проблема технологии сегментации - как правильно выбрать порог, наиболее распространенный и простой способ основан на гистограмме.

Пусть изображение поделено на уровни , Число пикселей на уровне обозначается , а общее число пикселей обозначается , Вероятность уровня серого

 (1)

В методе с двумя уровнями, пиксели изображения разделены на два класса с уровнями серого и с уровнями серого где –порогом. Вероятности серого уровня распределение на двух классов

 (2)

 (3)

Среднее значение серого обозначается

 (4)

 (5)

общее среднее уровней серых обозначается

 (6)

дисперсии класса обозначается

 (7)

 (8)

Внутри- -class дисперсии

 (9)

дисперсия между-класса

 (10)

общая дисперсия уровней Грея

 (11)

Метод Оцу выбирает оптимальный порог , c помощью вычисления максимизации между классом-дисперсии

 (12)

Метод Оцу может быть расширен до многоуровневого метода пороговой. Пусть всего порог разделили пиксели на классы изображении,

 (13)

где

 (14)

 (15)

 (16)

 (17)

 (18)

После того, как было проведено много исследований, сравнивая результаты, можем делать вывод, метод Оцу (*Otsu's method*) результат сегментации в ошибке классификации пиксельной частоты и делится лучшей однородности области производительности, алгоритм сегментации Оцу имеет преимущества в случае сегментации изображений, мы используем улучшенный алгоритм Оцу для сегментации изображения клетки рака желудка с предварительной обработкой [10].



Рис. 3. Otsu technique T=145



Рис. 4. Fast Otsu technique T=143 [11]

### Алгоритм k-means

Алгоритм K-средних(метод k-means)[12], это алгоритм кластеризации, который был изобретен математиком Гуго Штейнгаузом, одновременно с Стартом Ллойдом, но данный метод стал популярным только после работ Маккуина. Суть метода заключается в том, что он стремится минимизировать суммарное квадратное отклонение точек кластеров от центра самих этих кластеров.

 На каждой интеграции центров кластеров перевычисляется центры масс, которые были получены раннее, затем векторы разбиваются на кластеры вновь, согласно концепции, по которой они выбирают центр который оказался ближе всего по выбранной ими метрике. Алгоритм является завершенным тогда, когда не произошло изменение центра масс каждого кластера

Основная идея K- средних алгоритма

1. Случайно выбрать k объектов данных из общего количества n, означает в качестве исходных центров кластеров K-алгоритма.
2. На основе принципа минимального расстояния, каждой выборки данных набор выборок х, , разделить выборок x на близком классе.
3. Использовать среднее значение каждого новообразованных кластера для обновления центра в новом кластерном
4. Если изменяется значение кластерных центров, вернуться к шагу 2, чтобы продолжить вычислять среднее значение текущего класса в качестве нового центра кластера; Если новый центр кластера не изменяется, то вычисление останавливается,  сохранить ток класс означает, как кластеризация центра;
5. получаем никаких изменений средней категорий K является ли кластерные центры K мы спрашиваем, что классы K, которые мы хотим разделить класс.

Лучший критерий функции

пусть выброки есть выборки, каждая точка является вектором, который характеризуется d-вектором набора данных. С помощью алгоритма K-средних точки данных n делиться на классы , состоящие из новых наборов данных , пусть это каждый центр кластера , определение расстояния между точкой набора данных и центром кластера

 (19)

тогда, сумма евклидового расстояния между точкой в разграниченном кластере и центром данного кластера

 (20)

в виде вычисления суммы евклидового расстояния между точкой под-кластера и множеством кластеров . тогда выполнить статистики каждого подкласса в множестве кластера, получаем сумму евклидово расстояния между каждыми точками набора данных и центром разграниченного кластера, это Лучший критерий функции

 (21)

И . Тогда видно, чтобы получать лучший критерий минимума функции , необходимо положить равной нулю производную от каждого центра кластера, тога получаем минимальное значение ,

 (22)

Тогда , только будет минимальным значением функции, этот центра кластера является средним значением всех выборок каждого класса, эффект кластеризации является лучшим

Производительность алгоритма

Несколько моментов о преимуществах:

(1) Идея алгоритма является наиболее простой и простота реализации, нужно меньше всего времени для обработки больших массивов данных;

(2) для обработки больших массивов данных, алгоритм кластеризации имеет некоторую масштабируемость и своевременность. Его сложность O(n\*k\*t), где n является общим количеством элементов в наборе данных, k представляет количество кластеров, t есть число итераций процесса кластеризации. Обычно K << п, т << п;

(3) Когда кластеризуются данные интенсивного или непрерывного типа, если для кластеризации использовать расстояние между классами, то разница между ними гораздо более очевидна, получаются лучшие результаты классификации.



Рис. 5. K=3 результат сегментации изображении

### Алгоритм сегментации по водоразделам (WaterShed)

Алгоритм сегментации изображения по водоразделам (водоразделов, также известный как разделительная линия / ватерлинии) алгоритм сегментации изображения[13][14], это основанный на областях метод математической морфологии, широко используемый в последние годы. Алгоритм является классической технологией сегментации с помощью использования математической морфологии, генерированные результаты сегментации более стабильны. Алгоритм сегментации по водоразделам был введен в 1979 С. Беукэром и К. Лэнтуеджуло, в области обработки изображений для анализа простого бинарного изображения; в результате научно-исследовательской работы Beucher, Винсент и др., была установлена теория алгоритма сегментации по водоразделам, и с начала 1980-х годов этот алгоритм широко используется для отработка полутонового сегментации изображений.

### Принцип алгоритма

Основная идея состоит из размещения водного источника в каждом лакальном минимуме в области, чтобы затопить все область при источниках и построить барьеры, когда различные водные источники встречаются. Получающийся набор барьеров составляет водораздел области.

### Традиционное математическое описание водораздельного преобразования

Перед тем как дать строгое математическое определение алгоритм водораздела, давайте рассмотрим понятие геодезической геодезического расстояния и пораженной области:

В множестве есть две точки, геодезическое расстояние является минимальной длиной всех путей, соединяющих эти две точки :

, это путь связывающий

Пусть В это подмножество А, В и Геодезическое расстояние между точкой х в множестве:

 (23)

Пусть ; это компоненты связности , геодезическая зона влияния множества в множестве определяется как :

 (24)

Множество является объединением геодезических зон влияния соединяемых компонентов т.е.

 (25)

Дополнение множества в множестве называется SKIZ (скелет с помощью зон влияния):

 (26)

Таким образом, математическое описание:

Пусть функция в области является изображением, которое будет обрабатываться. Изображение пороговое максимальное и минимальное, соответственно

Множество представляет собой пороговую функцию :..

Полуводный бассейн (т.е. область водораздела) состоит из , получается в результате следующей итерации:

Водоразделом функции называют дополнение множества , соответствует самой низкой точке изображения Глобального, Итерационный процесс соответствует разливу воды, устанавливание поля между водой бассейна водосбора, является процессом записи геодезической зоны влияния точечного скелета, все разграничительные линии соответствую водоразделу функции .

### Сегментация по водоразделу на основе расстояния преобразования

Принципами сегментации по водоразделу на основе расстояния преобразования являются: во-первых, с помощью расстояния трансформации клетки различной информации о местоположении пикселя в различной серой информации, повторное с использованием адаптивное алгоритм двоичного водоразделом различать пиксели различной природы, а затем отметить граничную точку, граница будет разделена клетки на отдельную клетку.

Расстояние преобразование арифметических операций для бинарного изображения, бинарное изображение будет преобразовано в черно-белое изображение. В этом черно-белом изображении, оттенки серого каждого пикселя — это расстояние от того пикселя до ближайшего фона. Теоретически, для расчета кратчайшего расстояния пикселя к фону пиксели, требуется выполнить глобальную операцию изображения действий, то есть вычисляется расстояние между этим пикселем и всеми фоновыми пикселями, а затем принимается минимум. Если размер цифрового изображения был очень мал, в противном случае будет массивный расчет. В практическом применении алгоритм с помощью расстояния преобразования должен исходить из соседних пикселей, каждый рассчитывает только минимальное расстояние с несколькими местными соседними пикселями, в соответствии с принципами Глобального расстояния из локального расстояния масштабируется наложения изображения до и после двух сканирований, окончательно приближенного изображения расстояния. Исходя из приведенного выше принципа, в практике использования алгоритма расстояния преобразования в произвольных размеров для преобразования, выполняется просто и быстро, рассчитанное расстояние разумное приближение к истинному расстоянию евклидовой геометрии.

В двумерном случае алгоритм расстояния преобразования в произвольных размеров, дает шаблон, похожий на свертку, в процессе работы, подобно свертку движения во всем изображении. Сначала пусть фон серого бинарного изображения устанавливается в 0, мишень (ячейка) градация устанавливается в 1, а затем сделанное изображение сканируется дважды для слева направо, сверху вниз; как только центр шаблона перемещается в заданное положение, каждый элемент шаблонна на его соответствующее значение пикселя изображения в положении суммы, в результате получается набор из двух, и их используют вместо минимального значения значений пикселов изображения в центре шаблона.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | \* | 5 |  |
| 11 | 7 | 5 | 7 | 11 |
|  | 11 | \* | 11 |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 11 |  | 11 |  |
| 11 | 7 | 5 | 7 | 11 |
|  | 5 | \* |  |  |

Вперед

назад

Рис. 6. 5 × 5 template pair used in chamfer algorithm realizing two-dimensional distance transformation

Разница между расстоянием с помощью расчет шаблона с 5 × 5 матрицы и реальном геометрическом расстоянием, как правило, рассчитывается на 1,96% - 2,02%. Конечно можно используя шаблоны более высокого порядка, каждый раз надо рассматривать пиксели большей окрестности, такие как шаблон 7 × 7, но это повышает точность не на много, примерно между 1,28% ~ 1,97%. С учетом пробелов и эффективности вычислений, выбираем 5 × 5 шаблона, расчетное расстояние в 5 раз превышает евклидово расстояние. \* Представляет центр шаблона.

шаги по реализации алгоритма Сегментации по Водоразделу на основе расстояния преобразования:

* + 1. бинаризация изображения
		2. Операция расстояние преобразования для бинарного изображения с использованием 5 × 5 шаблонов фаски и сохранение изображения в оттенках серого, подсчитывание высокого уровеня серого изображения ;
		3. Сегментация с использованием алгоритма водораздела сегментации

Во-первых, установить массив флаг, значений каждого элемента массива (значение по умолчанию равно 0), в свою очередь, представитель флаг каждого пикселя изображения, установить начальный порог Сегментирование , отмечены начальные объекты частиц, следуя снижение порога каждый раз, алгоритм выполняется в отмеченный круг. затем выполнятся пересечение всего пикселей изображения, когда расстояния пиксели F, определить, является ли новая частица:

1. Если марка окрестности этого пикселя вокруг 20 являются '0', а затем переведен (б) суждения, в противном случае идти (с) суд.
2. Если марка окрестности этого пиксель вокруг 48 являются '0’, это была бы новая частица, семена плюс один, в это время значение числа семян, присвоенных марка пикселя; если не полностью в '0', и это уже помечены частицы в расширенной области, значение марка не равно '0' присваивается марка пикселя;
3. Если эта марка окрестности вокруг 8 пикселя одинаковы, и тех же частиц, то это значение марки пикселя присваивается той же окрестности около 8, в противном случае пиксель будет точкой границы, и значение пикселя присваивается значению '0'.

Повторив шаги (3) содержание, до F значение будет 0.

* + 1. В соответствии с вышеизложенным после раскола для того, чтобы показать более ясно, анти-цвет и определение порога (порог 255 в это время, так как с помощью значений пикселей изменений в клеточной адгезии и задние достопримечательности находятся в 255, то значения пикселей меньше, чем 255 клеток) результаты



Рис. 7. Водораздел гряда линии Рис. 8. результат Сегментация изображения на основе отметки расстояния

## 2.2 Извлечения признаков микроскопическое изображение клеток

Для того, чтобы выполнять идентифицирование микроскопического изображения, основываясь на компьютерный технологий, сначала нужно установить характеристики изображения раковых клеток, таких как периметр, площадь, глубина, округлость характеристик вблизи особенностей изображения, компьютер может быть идентифицированы, следовательно, в этой главе желудочных клеток аденокарциномы в выделения признаков изображения, чтобы сделать конкретное описание.

Обзор изображения

особенность изображения, характеризующее основные атрибуты или характеристики, характеристики изображения могут быть либо естественного, который человеческой визуальной особенность, может быть идентифицирована, его можно определить с помощью определенных характеристик человека. Анализ изображений является основой распознавания элементов изображений, только приобретает черты изображения на основании описанного, чтобы иметь возможность анализа изображений и идентификации.

В принципе, есть два способа представлять регион: Во-первых, вы можете использовать характеристики области для его представления (например, границ ячеек, размер, округлость), вторая характеризуется своей текстурой, чтобы представлять регион (например, распределение неровности серого, в основном отражающие характеристики распределения хроматина)

### Функция представления изображения и способ описания

выражение цепного кода изображения[15]

Есть много методов представлении граничного клеточного изображения, в которых код представляет собой способ кодирования цепи для представления граничной точки, которая характеризуется использованием ряда соединенных отрезков конкретной длины и ориентацией отображения границ мишени.

Цепной код является представление граничной точки кодирования, которая характеризуется использованием серии соединенных отрезков конкретной длины и ориентации, чтобы представлять границы мишени. Поскольку длина каждого сегмента в фиксированном направлении принимается ограниченной, поэтому только для представления начальной точки границы необходимо использовать абсолютные координаты, остальные точки могут продолжать представлять только направление смещения.

Так как длина каждого сегмента соединения фиксирована, и число направлений также ограниченное, так что только начальную точку границы необходимо использовать для представления координаты, остальные точки могут быть представлены только смещением направления изменения. Цифровые изображения, как правило, коллекция на равном расстоянии сетки путем приобретения, Таким образом, самый простой код граница цепи присваивается каждому двух смежных пикселей контура в одном направлении проводного значения. Так как число битов указывает направление небольшого числа и точку для каждого номера только одно направление может быть заменено двумя значениями координат, код цепи может значительно уменьшить объем данных, требуемых для представления границы, и он представляет собой значение требуемой координаты.

 1 3 2 1

0

4

2 223234

5 6 7

 3

Рис. 9. 4-направлениях кодирование цепи и 8-направлениях кодирование цепи

Для точки изображения, окрестность имеет восемь направлений, против часовой стрелки для каждого направления обозначим восемь направлений соответственно 0, l, 2, 3, 4, 5, 6 7, этот код называется коднаправления.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | S |  |  |  |
|  |  | 5 |  |  | 3 |  |  |
|  | 5 |  |  |  |  | 3 |  |
|  | 6 |  |  |  |  |  | 3 |
| 5 |  |  |  |  |  |  | 2 |
| 7 |  |  |  |  |  |  | 2 |
|  |  | 0 |  |  |  |  | 1 |
|  |  |  | 7 |  |  |  |  |

Рис. 10. Пример 8-направленного кодирования цепи

Предположим, что из исходной точки области графика на границе обозначенной выше кодирующие последовательности могут быть записаны следующим образом: их значения 0,1,2,3,4,5,6,7, Последовательность направления кода называется кодом цепи. Из приведенной выше диаграммы мы видим, что четный сегмент кода цепи будет вертикальным или горизонтальным участком кода; нечетные сегменты кода цепи соответствуют наклонному направлению кода.

**нормализация цепного кода**

Учитывая цепной код от любой возникающей точки, это можно было бы рассматривать как натуральное число, составленное по количеству направлений. Эти направления в соответствии с количеством циклов в одном направлении (число этих направлений по циркуляциям направления) создают минимальное значение. Таким образом, преобразование соответствует исходной точке кода цепи в качестве отправной точки этой границы нормализованного цепного кода

1. 0

 1 Исходном код цепи нормализации 1 3

 0 10103322 01033221

 1 1 3

 2 2 2 2

Рис. 11. Цепь кодирования трамплина точки нормализации

С кодом цепи представление границы данной области, если будет перевод области, изменение кода цепи не происходит, если область вращается, код цепи изменяется. Для того, чтобы решить эту проблему, вы можете воспользоваться дифференциальным кодом цепи первого порядка, чтобы восстановить последовательность. Это соответствует нормализация вращающегося кода цепи. Эту разницу можно использовать вычитание два числа (против часовой стрелки) направления соседний.

1 3 Новый код 211 0033 1 3

 3 Исходный код 10103322 2 3

2 2 2

Рис. 12. Револьверная нормализация кодирования цепи

## Описание характеристики изображения

Поскольку для достижения точного описания при идентификации раковых клеток часто приходится полагаться на размер ядра, так что важны геометрические характеристики описания изображения.

Геометрические особенности изображений[16] в анализе изображений играют важную роль во многих прикладных системах обработки изображений, часто используются некоторые из заданных характеристик области, такие как размер, расположение и ориентация, для определения местоположения объектов и их идентификации. Потому что только забота о геометрии изображения, для того, чтобы облегчить выделение признаков, перед извлечением набора функций изображения, необходимо обработать двоичное изображение, а именно в обработке черно-белых изображений только 0 и 1 выполняются оба значения. Кроме того, поскольку бинарное изображение легко приобрести, доступна и обработка, может быть обнаружена контурная информация об объекте, и поэтому этот метод подходит для извлечения из изображений геометрических особенностей, в этом разделе, если не указано иное, обсуждается бинарное изображение.



Рис. 12. Маркировка целевой области.

Ниже несколько основных аспектов рака от положения и ориентации, периметра, площади и расстояния вводит геометрические характеристики изображения.

1. При анализе изображения положение и ориентация, как правило, рассматривается не только случай пиксела в изображении, но и случай геометрии области изображения, поэтому, как геометрические характеристики области изображения, область часто используется центральной точки в качестве представляющей области центральной точки. Так как бинарного изображения, и следовательно, то же региональное центральное положение как и центр масс объекта.

Кроме того, мы должны не только знать расположение объектов в изображении, но и знать направление объекта в изображении. Определение направление объекта, чем вычисление положения объекта сложнее, некоторая форма направления не является уникальным. Для того чтобы определить единственное направление, то, как правило, берется продолговатый предмет, который длиннее в направлении оси и определяется как направление объекта.

1. Периметр области, который является длиной граничной области, область добавляется внутренней граничной точки и точек. Для области R, его каждая граничная точка Р должна удовлетворять двум условиям, а именно: Р сама принадлежит области R, несколько пикселей окрестности P не принадлежит к области Р.

Граница B области R состоит из всех граничной точки R с помощью джойстика или восьми направлениях перехода, другие точки в области называются внутренними точками области. если внутренняя точка области R определяет восьми двустороннюю связь, результирующее краевое с четырьмя направлениями связи. Если внутренняя точка область R находится в связи с четырьмя, чтобы определить направление, направление результирующей граничном восемь связи.

Для указанной выше граничной область точки и внутренней точки, надо использовать различные соединения, чтобы определить, не было бы несоответствие

Представление периметра обычно используется, являются следующие:

1. Когда пиксели изображения на единицу площади считается небольшой ящик, область изображения и фона формируется с помощью маленьких квадратиков. Периметром области является длина области зазора и фона.
2. рассмотрение пикселей в качестве точки, то периметр с кодом цепи, найти периметр, как рассчитанную длину кода цепи. В это время, в точке центра к верном, нежном левом и правом четыре очка на расстояние 1, а расстояние скоса (1,414), поэтому с использованием последовательности кодовой цепи границы можно рассчитать периметр клеток. Это периметр Р выражается как

Здесь число код цепи направления с четным и нечетным номером.

1. Различные методы расчета периметра, периметр будет приводить к различным результатам.

Площадь является мерой размера объекта. Приграничная территория объекта связана с имеющимся расчетным методом площади пикселя.

Самый простой метод для вычисления количества статистической Площадь в пределах граничных пикселей. В этом методе расчета площади очень прост, просто найти сумму пикселей в граничной область, это может быть рассчитана как

Здесь в бинарном изображении значение 1 указывает объект, значение 0 указывает фон, тогда площадь равна сумме числа .

1. Расстояние во многих областях применения, необходимо рассчитать расстояние между двумя пикселями в изображении. В определении расстояния, евклидово расстояние известно, но в преобразовании расстояния, принимая простоту расчета и значение расстояния представляет собой целое число, и т.д., часто с помощью соседей с поля и восемь соседей расстояния, пусть Р координаты точки (т, п), Q представляет координаты (K, L), рассчитывается как удельная
2. Евклидово расстояние (расстояние является нормой. 2), определяемое как
3. 4-район расстояние, определяемое как
4. 8-район расстояние, определяемое как

Очевидно, что в качестве отправной точки Р на расстояние меньше или равно соседей домена т (т = 1,2, ...) точки Р формируется в центре ромба. Когда T ≤ 2 балла из представления. Видна от Р до Q есть длина кратчайшего пути из четырех. Точно так же, в качестве отправной точки Р на восьми окрестности расстояние меньше или равно T (T = 1,2, ...) из точки Р формируется в центре квадрата. Например, когда T ≤ 2, с расстоянием между точкой показывая эти точки, как показано на рисунке. Как можно видеть, длина от Р до Q кратчайшему пути из восемь

Рис. 13. Рис. 14.

### Описание особенности формы раковых клеток

По медицинских предварительных знаний экспертов патологоанатома обнаружили, что после канцерогенеза клеток будет существенным изменениям в форме, поэтому для того, чтобы добиться признания раковых клеток необходимо использовать компьютер для извлечения признаков формы изображения

1. **Прямоугольность**

Прямоугольность отражает полный уровень объекта ограниченного прямоугольника, прямоугольная фитинга фактором является параметр, отражающий степень объекта прямоугольника

Здесь площадь объекта , Его площадь ограниченного прямоугольника .

 отражает полный уровень вписанного в объект прямоугольника. Для прямоугольных объектов получается максимальное значение 0,1, для кругового объекта значение будет , для тонких или изогнутых объектов значение становится меньше. Прямоугольная фитинга фактор имеет значение в диапазоне от 0 до 10.

Другим связанным с особенностью форма является соотношениям сторон

Это отношение ширины и длины вписанного прямоугольника. Эта функция может отделить более тонкие объекты и квадратный или круглый объект.

1. **округлость**

Отношение площади ядра к площади окружности, соответствующей самому длинному диаметру ядра, известено как ядро компактности, показано

При тех же условиях области, граница ячейки области является гладкой и круглой, то будет самая короткая длина окружности, округлость C = 1; чем больше форма площади ячейки отходит от круговой, тем меньшие значения С. Это квадрат окружности соотношения площадей. Для особенности круглой формы минимальное значение 4л. Чем более вытянутая форма объекта, тем больше ее значение*.*

1. **Эксцентриситет**

Эксцентриситет также можно назвать относительным удлинением, что описывается компактность области. Сначала определить координаты центра ядер и клеток: то же самое может быть получено путем вычисления расстояния

Эксцентриситет: расстояние от центра клетки и центра ядра клетки, деленному на радиус ячейки. Чем меньше значение, тем меньше эксцентриситет. Когда значение 0, ядро клетки и также концентрическими. Региональные и ось ячейки длина малой оси отношение длины также может быть использован, также использовано отношение длинышпинделя площади клеток и длины шпинделя

 A

 B

 Рис. 15. Эксцентриситет измерения A / B

Главные и второстепенные оси взаимно перпендикулярны, и что длина является максимальной двух направлений. Но будет относительно велико влияние от шума и вычисленного формы объекта.

**число формы**

число формы представляет собой способ описания формы границы, основанное на коде цепи, оно зависит от начальной точки кода цепи, с выражением граничного кода цепи может иметь множество дифференциала первого порядка. Число формы является рядом минимального значения дифференциальной границы. Другими словами, число форма является минимальным значением (цепи код) дифференциального кода.

Число формы обеспечивает метод измерение формы. На границе двух регионов, с помощью сходство между ними может воспользоваться их число, чтобы описать формы. От мала до велика постепенно вычислить число формы каждого ранг две формы границы и сравнить друг с другом, до тех пор найдено похожее число формы максимального порядка, что сходство границы между двумя регионами с наибольшим числом из них, чтобы измерить форму общественности. Таким образом, чтобы получить измерение расстояния между двумя формами, и она обратно пропорциональна сходству между двумя формами. Обратите внимание, что расстояние не общий смысл этого, но супер расстояние.

**Экспериментальные данные рака**

Вышеуказанные характеристики раковых клеток могут быть получены для создания информационной базы данных рака. В сравнении с базами данных рака данные клеточного изображения могут быть диагностированы, получаем подозрительность для данного клеток быть раковым клеткам.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| площадь | периметр | округлость | прямоугольность | Eccentricity |
| 6668 | 827.1097 | 0.122484 | 0.484875 | 0.944584 |
| 434 | 108.1838 | 0.465989 | 0.560724 | 0.945526 |
| 77 | 48.14214 | 0.417494 | 0.427778 | 0.916569 |
| 71 | 40.38478 | 0.547057 | 0.464052 | 0.954755 |
| 6309 | 1568.83 | 0.032212 | 0.215965 | 0.9345 |
| 242 | 85.35534 | 0.417411 | 0.606516 | 0.748583 |
| 102 | 51.55635 | 0.482221 | 0.53125 | 0.89444 |
| 79 | 35.55635 | 0.78524 | 0.675214 | 0.821316 |
| 67 | 41.79899 | 0.481896 | 0.429487 | 0.92582 |
| 510 | 158.267 | 0.255858 | 0.428571 | 0.485372 |
| 51 | 26.72792 | 0.897118 | 0.62963 | 0.699576 |
| 163 | 68.18377 | 0.440591 | 0.411616 | 0.807144 |
| 241 | 64.28427 | 0.732854 | 0.616368 | 0.872675 |
| 50 | 29.89949 | 0.702833 | 0.641026 | 0.939258 |
| 83 | 43.45584 | 0.552321 | 0.532051 | 0.623009 |
| 88 | 37.79899 | 0.773984 | 0.564103 | 0.648491 |
| 124 | 48.62742 | 0.658975 | 0.561086 | 0.801009 |
| 73 | 38.38478 | 0.622608 | 0.579365 | 0.845984 |
| 173 | 82.08326 | 0.322661 | 0.303509 | 0.973358 |
| 107 | 50.87006 | 0.5196 | 0.636905 | 0.936207 |
| 1743 | 220.066 | 0.452274 | 0.620726 | 0.926967 |
| 117 | 52.62742 | 0.53085 | 0.531818 | 0.917262 |
| 55 | 31.55635 | 0.694063 | 0.572917 | 0.88025 |
| 157 | 82.42641 | 0.290387 | 0.402564 | 0.957254 |
| 230 | 77.01219 | 0.487325 | 0.501089 | 0.855842 |
| 82 | 46.62742 | 0.473959 | 0.488095 | 0.83207 |
| 101 | 53.69848 | 0.440157 | 0.590643 | 0.932668 |
| 251 | 83.1127 | 0.456614 | 0.550439 | 0.700651 |
| 159 | 56.87006 | 0.617788 | 0.602273 | 0.912803 |
| 67 | 36.97056 | 0.615988 | 0.553719 | 0.777706 |
| 76 | 58.28427 | 0.281139 | 0.262976 | 0.852645 |
| 199 | 111.6396 | 0.200644 | 0.197421 | 0.991243 |
| 277 | 94.08326 | 0.393247 | 0.360677 | 0.928211 |
| 50 | 25.31371 | 0.980547 | 0.714286 | 0.705105 |
| 104 | 70.42641 | 0.263495 | 0.330159 | 0.908169 |
| 113 | 47.79899 | 0.621514 | 0.570707 | 0.854998 |
| 78 | 47.45584 | 0.435236 | 0.464286 | 0.805499 |
| 125 | 60.18377 | 0.433672 | 0.438596 | 0.818035 |
| 90 | 46.04163 | 0.53352 | 0.428571 | 0.911205 |
| 62 | 36.97056 | 0.570019 | 0.433566 | 0.930865 |
| 159 | 70.76955 | 0.398946 | 0.519608 | 0.633494 |
| 453 | 183.9239 | 0.16828 | 0.431429 | 0.765645 |
| 1497 | 267.7645 | 0.262377 | 0.317699 | 0.907204 |
| 7419 | 752.524 | 0.164632 | 0.317798 | 0.905103 |
| 1982 | 191.9239 | 0.676169 | 0.664655 | 0.809365 |
| 75 | 45.2132 | 0.461042 | 0.524476 | 0.590024 |
| 216 | 89.01219 | 0.342582 | 0.397059 | 0.946816 |
| 84 | 42.62742 | 0.580913 | 0.4375 | 0.951998 |
| 155 | 92.91169 | 0.225632 | 0.249597 | 0.961194 |
| 65 | 46.97056 | 0.37023 | 0.65 | 0.98299 |
| 57 | 25.89949 | 1.06783 | 0.791667 | 0.499965 |
| 131 | 58.28427 | 0.484594 | 0.655 | 0.95741 |
| 66 | 30.97056 | 0.86468 | 0.611111 | 0.862372 |
| 393 | 186.4092 | 0.142124 | 0.23227 | 0.942847 |
| 344 | 173.5807 | 0.143471 | 0.199768 | 0.940349 |
| 74 | 38.97056 | 0.612306 | 0.544118 | 0.911926 |
| 69 | 29.31371 | 1.009061 | 0.784091 | 0.715422 |
| 52 | 26.14214 | 0.956161 | 0.65 | 0.758239 |
| 1060 | 401.3036 | 0.082712 | 0.438198 | 0.83581 |
| 89 | 62.52691 | 0.286066 | 0.327206 | 0.906723 |
| 52 | 29.55635 | 0.748017 | 0.541667 | 0.923546 |
| 184 | 83.35534 | 0.332783 | 0.32 | 0.835937 |
| 63 | 30.72792 | 0.838463 | 0.65625 | 0.791316 |
| 82 | 48.28427 | 0.44199 | 0.465909 | 0.846863 |
| 296 | 138.468 | 0.194 | 0.434018 | 0.818171 |
| 201 | 71.69848 | 0.491344 | 0.67 | 0.94741 |
| 55 | 40.14214 | 0.428915 | 0.407407 | 0.938593 |
| 239 | 114.5685 | 0.228811 | 0.341429 | 0.938862 |
| 53 | 36.97056 | 0.487274 | 0.438017 | 0.823321 |
| 62 | 32.97056 | 0.716719 | 0.529915 | 0.863012 |
| 240 | 89.25483 | 0.37858 | 0.417391 | 0.892301 |
| 61 | 37.31371 | 0.550558 | 0.469231 | 0.817904 |
| 127 | 51.69848 | 0.597115 | 0.498039 | 0.847472 |
| 364 | 146.1249 | 0.214221 | 0.495238 | 0.771612 |
| 105 | 48.04163 | 0.571694 | 0.648148 | 0.905855 |
| 173 | 105.2548 | 0.196233 | 0.376906 | 0.575162 |
| 68 | 29.55635 | 0.978176 | 0.686869 | 0.681232 |

Таб.1. Экспериментальные данные

Из таб.1, мы можем видеть, что поскольку разные ячейках округлость отличается из-за разной площади клеток. Округлость клетки большая, территория большая, небольшая округлость, площадь мала. Округлость отражает область клеток с возможной неточностью.

**Идентификация рака с помощью AdaBoost**

В предыдущей работе извлечены признаки клетки, в дальнейшем будет описана классифиция клеток в соответствии с признаками клеток.

В данной работе на основе сильного классификатора выполняется классификация изображения и идентификация рака. Сначала с помощью алгоритма adaboost оптимизируются отдельные нейронные сети BP (backpropagation), получается несколько слабых классификаторов, наконец с помощью метода голосования сливаются несколько сильных классификаторов в интегрированную нейронную сеть. Наконец с помощью MATLAB строится система извлечения признаков и распознавания рака.

**Распознавание изображения рака**

Распознавание изображения рака (pattern recognition), объекты на различных раковых клеток классифицируются в соответствии с определенными характеристиками, то исследовано новые и неизвестные объекты рака определяют их соответствующие категории

**идентификация рака, Основана на сильном классификаторе алгоритма AdaBoost**

Из-за недостатка возможностей обобщения алгоритма индивидуальной нейронной сети, в данной работе используя алгоритм AdaBoost оптимизировать BP[17], для улучшения их возможности обобщения, и вводится новый алгоритм стратегии распределения веса дерева, получаются три сильных классификатор нейронной сети для идентификации раковых клеток.

AdaBoost представляет собой итеративный алгоритм, основная идея одинакова: для обучающего набора обучить различные классификаторы (то есть слабых классификаторов), а затем установить эти слабые классификаторы, чтобы построить более сильный классификатор. Алгоритм является путем изменения распределения данных для достижения верной классификации каждой обучающей выборки в каждом образце, а также точности предыдущей общей классификации, чтобы определить вес каждого образца. Отправить модифицированный новый набор данных весов в нижней подготовки классификаторов, и, наконец, объединить каждый обучающий классификатор вместе, в качестве окончательного решения классификатора. Используя AdaBoost классификатор может исключить некоторые ненужные функции обучающих данных, а также обучение по ключевым данным.

В классическом алгоритме AdaBoost, используется одинаковый вес в процессе инициализации образца, такой вес, устанавливающий на одинаковое количество образцов библиотеки образца является очень эффективным, но для разного количества образцов библиотеки образца, результаты в следующих случаях: так как некоторого типа количество образца относительно мало, поэтому слабый классификатор трудно такие образцы, но этот слабый классификатор может эффективно отличать образцы с больше количеством образцов, слабый классификатор получает большой вес, не может сделать небольшое количество образец получать достаточно внимания. В данной работе вводится новую политику распределения веса, в соответствии с отношением количества образцов различных типов.

Улучшенный алгоритм AdaBoost

Инициализируем  образцы , где ,

в классическом алгоритме вес каждого образца .

В данной работе инициализируем  в соответствии с отношением количества образцов различных типов. Пусть данные образцы должны быть классифицированы на кластеров , если количество каждого образца , то вес каждого образца

 (27)

 (28)

Шаг 2 Тест слабого классификатора, для каждого

В соответствии с правилами обучения нейронной сети, обучать образцы, получить сумму ошибки выхода предсказания:

 (29)

Здесь выход слабого классификатора; ожидаемый выход.

Шаг 3 вычисление веса слабого классификатора. Форма слабого классификатора:

 (30)

Шаг 4 Следующая итерация веса тестовых данных

 (31)

Здесь для того чтобы сумма распределения весов равно 1, должно установить фактор нормализации .

Шаг 5 прогноз сильного классификатора

Через т раз итеративного обучения, получаем т кластеров слабый классификатор нейронной сети , сильный классификатор получен из т кластеров слабого классификатора

 (32)

**Обучение по методу обратного распространения ошибки (BP - Back Propagation)**

Алгоритм обратного распространения ошибки является одним из методов обучения многослойных нейронных сетей прямого распространения, называемых также многослойными персептронами. Метод использует ошибку на выходе сети для расчета частных производных по весам последнего слоя обучаемых связей, затем по весам последнего слоя и ошибке сети определяется ошибка на выходе предпоследнего слоя и процесс повторяется.

Процесс построения алгоритма включает в себя создание нейронной сети, обучение нейронной сети и прогнозирование нейронной сети. Наиболее важным является процесс обучения нейронной сети, с обучением постепенно настраиваются весы нейронной сети и пороги, до того как получаем идеальный результат.

Шаг 1 инициализируем алгоритм обратного распространения ошибки. В соответствии входному сигналу и выходному сигналу, определяется число узлов входных слоя и выходного слоя нейронной сети. Инициализируется три веса соединения между нейронами , Порог скрытого слоя , порог выходного слоя . Определить число узлов скрытого слоя, функцию обучения нейронной сети и функцию возбуждения.

Шаг 2 вычисление выхода скрытого слоя нейронной сети.

 (33)

Здесь -число узлов скрытого слоя, - функция возбуждения скрытого слоя.

 (34)

Шаг 3 вычисление выходного слоя нейронной сети.

 (35)

Шаг 4 вычисление ошибки нейронной сети

 (36)

Шаг 5 обновление веса

 (37)

 (38)

Здесь - коэффициент «скорости обучения», позволяет управлять средней величиной изменения весов;

Шаг 6 обновление порога. Обновление порога a, b узлов

 (39)

 (40)

Шаг 7 завершение итерационного процесса, если не достигнуто ожидаемого выхода в пределах параметра, то возвращаемся в шаг 2.

Проблем выбрать количество скрытого слоя, доказано, что нейронная сеть включает только один скрытый слой и выполняет задачу подхода для всех функции, поэтому в данной работе используется нейронная сеть одного скрытого слоя, на входе матрица данных на 124\*9, на выходе матрица данных на 124\*1, переходная функция узлов является касательном S типа, функция обучения является traingd по градиентным спуском, функция обучения нейронной сети является learngdm. Для количества нейронов формула выглядит как

В данной работе входное количество узлов равно 9, выходное количество узлов равно 1, номер опыта x на , количество нейронов s на . Через несколько тестов и сравнений, наконец определять лучшее количество скрытого слоя на 6. Максимальное число итерации на 20, ошибка функции обучения на 0,0001.

Используя алгоритм AdaBoost, улучшаем алгоритм обратного распространения ошибки, определяем число итерации на 10, чтобы выполнялся прогноз итеративного обучения на образцах библиотеки рака изображений, после завершения итерационного процесса, получаем 10 слабых классификаторов алгоритма обратного распространения ошибки, через взвешенное суммирование получаем сильный классификатор BP\_Adaboost.

# Глава 3 основные средства

Предложенные методики были реализованы с помощью MATLAB 2013a, на наборе данных в цифровую форму с 5-кратным увеличением на ПК с 3,4 ГГц процессором IntelCorei7, 8 Гб оперативной памяти, а также на операционной системе Windows 10.

Для целей тестирования и экспериментов, выбираем50 гистологических изображений рака и 50 гистологических изображений без рака из набора данных изображений гистологии. Распределения изображения на основе структур основных тканей в гистологии набор данных включает микроскопические изображения биопсии с увеличениях 2.5x, 5x.

Предложенная методика для обнаружения и диагностики обнаружения рака от микроскопических изображений биопсии состоит из этапов повышения качества изображения, сегментации, выделения признаков и классификации.

Для того, чтобы лучше сохранить нужную информацию в микроскопических биопсии изображений в процессе сегментации, были рассмотрены различные подходы кластеризации и текстуры, основанные на сегментации. Для получения микроскопических изображений биопсии требуется, чтобы обнаружить как можно больше информации о клетке, из результатов и анализа [18] представленных в разделе 2, k-средних алгоритм сегментации был использован для сегментирования биопсии изображения микроскопических, так как он работает лучше по сравнению с другими методами. В процессе сегментации метода кластеризации k-средних, количество кластеров было установлено. Для поиска центры кластеров используются квадрат евклидова расстояния в качестве меры подобия. На практике, выбираем комплексный алгоритм k-средних и алгоритм сегментации по водоразделам (WaterShed), то есть если с использованием алгоритм k-средних, получено изображение хорошее, тогда не требуется использоваться алгоритма сегментации по водоразделам (WaterShed).

Выбранные особенности микроскопические изображения биопсии включают площадь, периметр, округлости, прямоугольность, эксцентриситет, ширину, высоту и большую полуось. Ряд результата: злокачественный - 1, нормальная клетка - 0. Рак желудка и незлокачественные данные желудка клеток случайным образом разделены на две группы, эти случайные выборки используются в рамках тренировочного набора, а остальное в виде набора тестового образца.

При создании модели, с помощью нейронной сети входной слой BP содержит 9 узлов, соответствующие шести клеткам записи индекса морфологии, выходного слоя в качестве узла, доброкачественным и злокачественным клетки, соответствующим результатам дискриминации. Перед тренировкой, сеть инициализируется коэффициентом рандомизации, начиная от подключения весов и порогов в [-1,1], скорость обучения устанавливается в , Итерационный обучение нейронной сети выполняется, сеть сходится, суммарная погрешность выходной мощности составляет менее 0,0001.



Рис.16. обучение, аттестации и тестирования производительности.

# Вывод

Использован тест проверки данных для обучения нейронной сети, Таблица 2 содержит результаты работы слабых классификаторов (BP 1-10) и сильного классификатора (AdBP) испытанного на тестовом наборе. Мы можем видеть из таблицы 2, BP\_Adaboost алгоритм выполняется с повышением точности для одной нейронной сети

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| BP1 | BP2 | BP3 | BP4 | BP5 | BP6 | BP7 | BP8 | BP9 | BP10 | **AdBP** |
| 89.2% | 89.2% | 91.9% | 89.2% | 89.2% | 94.6% | 91.9% | 89.2% | 91.9% | 89.2% | 90.5% |

Таб. 2

**предположение**

в связи с особенностями, рассматриваемых без дальнейшего анализа эффективности, точность классификации была бы снижена, но результаты должны соответствовать требованиям классификации.

# ****Заключение****

С быстрым развитием компьютерных технологий, технологии обработки изображений можно применять для компьютерной сегментации аденокарциномы микроскопических клеток и распознавания изображений, этого необходимо для того, чтобы облегчить медицинскую трудоемкость и повысить точность постановки диагноза, поэтому эти технологии имеют большое значение для научной деятельности.

Основная положения данной исследовательской работы и инновации данной работы:

1. в данной работе с помощью метод Оцу определяется оптимальный порог, который обеспечивает подготовку анализа изображения к алгоритму сегментации по водоразделам (WaterShed).
2. Для получения микроскопических изображений биопсии требуется обнаружить как можно больше информации о клетке для того, чтобы сделать надежное и точное обнаружение и диагностику, основанную на клетках и параметре ядер. С помощью алгоритма k-means, получена грубая сегментация изображений клетки.
3. Для проблемы перекрытия клеток и проблем с адгезией, мы используем алгоритм водораздела на основе расстояния преобразования для сегментации перекрывающихся разделения клеток и предлагается алгоритм водораздела на основе улучшений маркера расстояния сегментации, что позволяет избежать некоторой чрезмерной сегментации алгоритма водораздела.
4. Для описания характеристики раковых клеток, после тщательного анализа, в данной работе выбраны следующие характеристики площадь ячейки, периметр, длина большой оси, длина малой оси, округлость, правила формы, эти параметры информации заложили хорошую основу для последующей изображения идентификации изображений клеток.
5. На основе извлеченных параметров информации о раковых клетках, с помощью BP\_Adaboost классификации нейронной сети для классификации аденокарциномы клетки получены хорошие результаты распознавания.

По результатам данной работы, выполнены все поставленные цели и задачи и проделана научно-исследовательская работа, по результатам работы, были выявления не доработки, которые необходимо исследовать в будущем с целью улучшения работы:

1. Из-за отсутствия времени, в исследовании разделения перекрывания клеток, существуют некоторые недостатки в качестве чрезмерной сегментации, которые влияют на общий эффект сегментации. Для клеток сильной адгезии и перекрывшихся клеток выполняется простое деления клеток, клетка не может изображать исходный контур перекрытия участков. Поэтому, перекрывая деление клеток, проблема деления клеток является одним из направлений для будущих исследований.
2. Сложность в выявлении проблем, желудочные раковые клетки, раковые клетки из-за этой неопределенности, делают конечную цель отбора признаков и распознавания очень высокой. Как выбрать более надежные характерные параметры – одно из направлений для улучшения скорости распознавания клеток рака желудка, что может быть рассмотрено в будущих исследованиях.

# Список литературы

1. Chen W, Zheng R, Baade P D, et al. Cancer statistics in China, 2015[J]. CA: a cancer journal for clinicians, 2016, 66(2): 115-132.
2. Методы диагностики злокачественных опухолей. http://www.cancer.ic.ck.ua/index\_2\_5.htm.
3. M. N. Gurcan et al., “Histopathological image analysis: A review,” IEEE Reviews in Biomedical Engineering, vol. 2, pp. 147–171, 2009.
4. M. Kowal et al., “Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images,” Computers in Biology and Medicine, vol. 43, no. 10, pp. 1563–1572, 2013.
5. P. Filipczuk et al., “Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies,” IEEE Transactions on Medical Imaging, vol. 32, no. 12, pp. 2169–2178, 2013.
6. Y. M. George et al., “Remote computer-aided breast cancer detection and diagnosis system based on cytological images,” IEEE Systems Journal, vol. 8, no. 3, pp. 949–964, 2014.
7. Y. Zhang et al., “Breast cancer diagnosis from biopsy images with highly reliable random subspace classifier ensembles,” Machine Vision and Applications, vol. 24, no. 7, pp. 1405–1420, 2013.
8. Cao J, Chen J, Li H. An adaboost-backpropagation neural network for automated image sentiment classification[J]. The Scientific World Journal, 2014, 2014.
9. Otsu N. A threshold selection method from gray-level histograms[J]. Automatica, 1975, 11(285-296): 23-27.
10. Al-Bayati M, El-Zaart A. Mammogram Images Thresholding for Breast Cancer Detection Using Different Thresholding Methods[J]. 2013.
11. Liao P S, Chen T S, Chung P C. A fast algorithm for multilevel thresholding[J]. J. Inf. Sci. Eng., 2001, 17(5): 713-727.
12. Giannakeas N, Tsipouras M G, Tzallas A T, et al. A clustering based method for collagen proportional area extraction in liver biopsy images[C]//Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE. IEEE, 2015: 3097-3100.
13. Beucher S. The watershed transformation applied to image segmentation[J]. SCANNING MICROSCOPY-SUPPLEMENT-, 1992: 299-299.
14. Belaid L J, Mourou W. Image segmentation: a watershed transformation algorithm[J]. Image Analysis and Stereology, 2009, 28: 93-103.
15. Freeman H. On the encoding of arbitrary geometric configurations[J]. Electronic Computers, IRE Transactions on, 1961 (2): 260-268.
16. Thiran J P, Macq B. Morphological feature extraction for the classification of digital images of cancerous tissues[J]. Biomedical Engineering, IEEE Transactions on, 1996, 43(10): 1011-1020.
17. Li N, Cheng X, Zhang S, et al. Recognizing human actions by BP-AdaBoost algorithm under a hierarchical recognition framework[C]//Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013: 3407-3411.
18. Kumar R, Srivastava R, Srivastava S. Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features[J]. Journal of Medical Engineering, 2015, 2015.

# Приложение

clc; clear all;

I= imread('cencer.jpg');

I\_rgb = rgb2gray(I);

figure, imshow(I,[]), title('original image');

R=I\_rgb(:,:,1);

G=I\_rgb(:,:,2);

B=I\_rgb(:,:,3);

figure,subplot(221),imshow(R);

subplot(222),imshow(G);

subplot(223),imshow(B);

%% kmeans\_1

filter=ones(5,5);

filter=filter/sum(filter(:));

denoised\_r=conv2(I(:,:,1),filter,'same');

denoised\_g=conv2(I(:,:,2),filter,'same');

denoised\_b=conv2(I(:,:,3),filter,'same');

denoised\_rgb=cat(3, denoised\_r, denoised\_g, denoised\_b);

D\_rgb=uint8(denoised\_rgb);

figure();imshow(D\_rgb);title('');

C =makecform('srgb2lab');

I\_lab= applycform(D\_rgb, C);

%K-mean

ab =double(I\_lab(:,:,2:3));

nrows= size(ab,1);

ncols= size(ab,2);

ab =reshape(ab,nrows\*ncols,2);

nColors= 3;

[cluster\_idx,cluster\_center] =kmeans(ab,nColors,'distance','sqEuclidean','Replicates',3);

pixel\_labels= reshape(cluster\_idx,nrows,ncols);

imshow(pixel\_labels,[]), title('image labeled by cluster index');

segmented\_images= cell(1,3);

rgb\_label= repmat(pixel\_labels,[1 1 3]);

for k= 1:nColors

 color = I;

 color(rgb\_label ~= k) = 0;

 segmented\_images{k} = color;

end

figure(),imshow(segmented\_images{1}),title('1');

figure(),imshow(segmented\_images{2}),title('2');

figure(),imshow(segmented\_images{3}),title('3');

hy = fspecial('sobel');

hx = hy';

Iy = imfilter(double(I\_se), hy, 'replicate');

Ix = imfilter(double(I\_se), hx, 'replicate');

g = sqrt(Ix.^2 + Iy.^2);

l=watershed(g);

wr =l==0;

g2=imclose(imopen(g,ones(5,5)),ones(5,5));

l2=watershed(g2);

wr2=l2==0;f2=I\_se;

f2(wr2)=255;

figure,imshow(I\_se);

figure,imshow(g,[]);title('ÌÝ¶ÈÍ¼Ïó')

figure,imshow(l,[]);

figure,imshow(g2,[]);

figure,imshow(l2,[]);

figure,imshow(f2,[]);

I\_rgb = rgb2gray(I\_se);

bw\_img = bwareaopen(f2,50);

[L,num]=bwlabel(bw\_img);

img\_reg=regionprops(L,'area','boundingbox','Perimeter','MajorAxisLength','MinorAxisLength','Eccentricity');

A = [img\_reg.Area];

P=[img\_reg.Perimeter];

MajorAxisLength=[img\_reg.MajorAxisLength];

MinorAxisLength=[img\_reg.MinorAxisLength];

Eccentricity=[img\_reg.Eccentricity];

metric\_string=sprintf('%2.2f',A);

plot(MajorAxisLength,'\*')

rects = cat(1,img\_reg.BoundingBox); figure(1),imshow(bw\_img);

%BoundingBox is the bounding box (x,y,width,height)

for i = 1:size(rects, 1)

 rectangle('position', rects(i, :), 'EdgeColor', 'r');

end

[~, max\_id] = max(A);

max\_rect = rects(max\_id, :);

figure(2),

imshow(bw\_img);

rectangle('position', max\_rect, 'EdgeColor', 'r');

clear

data=load('input\_train.csv');

[mm,nn]=size(data);

D(1,:)=ones(1,nn)/nn;

k=rand(1,77);

[m,n]=sort(k);

input=data(:,1:9);

output=data(:,10);

input\_train=input(n(1:52),:)';

output\_train=output(n(1:52),:)';

input\_test=input(n(53:77),:)';

output\_test=output(n(53:77),:)';

K=10;

for i=1:K

 [inputn,inputps]=mapminmax(input\_train);

 [outputn,outputps]=mapminmax(output\_train);

 error(i)=0;

 net=newff(inputn,outputn,8);

 net.trainParam.epochs=20;

net.trainParam.lr=0.1

net.trainParam.goal=0.0001;

 net=train(net,inputn,outputn);

 an1=sim(net,inputn);

 test\_simu1(i,:)=mapminmax('reverse',an1,outputps);

 inputn\_test =mapminmax('apply',input\_test,inputps);

 an=sim(net,inputn\_test);

 test\_simu(i,:)=mapminmax('reverse',an,outputps);

 kk1=find(test\_simu1(i,:)>0);

 kk2=find(test\_simu1(i,:)<0);

 aa(kk1)=1;

 aa(kk2)=-1;

 for j=1:nn

 if aa(j)~=output\_train(j);

 error(i)=error(i)+D(i,j);

 end

 end

 at(i)=0.5\*log((1-error(i))/error(i));

 for j=1:nn

 D(i+1,j)=D(i,j)\*exp(-at(i)\*aa(j)\*test\_simu1(i,j));

 end

 Dsum=sum(D(i+1,:));

 D(i+1,:)=D(i+1,:)/Dsum;

end

output=sign(at\*test\_simu);

kkk1=0;

kkk2=0;

for j=1:25

 if output(j)==1

 if output(j)~=output\_test(j)

 kkk1=kkk1+1;

 end

 end

 if output(j)==-1

 if output(j)~=output\_test(j)

 kkk2=kkk2+1;

 end

 end

end

kkk1

kkk2

disp();

disp([kkk1 kkk2 kkk1+kkk2]);

plot(output)

hold on

plot(output\_test,'g')

%Í³¼ÆÈõ·ÖÀëÆ÷Ð§¹û

for i=1:K

 error1(i)=0;

 kk1=find(test\_simu(i,:)>0);

 kk2=find(test\_simu(i,:)<0);

 aa(kk1)=1;

 aa(kk2)=-1;

 for j=1:25

 if aa(j)~=output\_test(j);

 error1(i)=error1(i)+1;

 end

 end

end

disp(‘error1’);

error1

disp()

(kkk1+kkk2)/77

disp()

(sum(error1)/(K\*25))