

ПРИМЕНЕНИЕ МЕТОДОВ НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ ДЛЯ ОПРЕДЕЛЕНИЯ ДИАГНОЗА ПАЦИЕНТОВ С КЛЕЩЕВЫМИ ИНФЕКЦИЯМИ

Сафронов В.С.¹, Сафронова Е.В.²

Научный руководитель: Аксёнов С.В.³

¹ ТГУ ИПМKN, аспирант, e-mail: vasilysafr@gmail.com

² ТПУ ИШИТР, аспирант, СибГМУ КМБК, ассистент, e-mail: ev.kashcheeva@mail.ru

³ ТПУ ИШИТР, ТГУ ИПМKN, СибГМУ КМБК, доцент, e-mail: axoenows@tpu.ru

Аннотация

Статья посвящена описанию проведенного кластерного анализа клинических и лабораторных данных пациентов с клещевым энцефалитом, иксодовым клещевым боррелиозом и микст-инфекцией. Приводится сравнение оценки работы различных моделей нечеткой кластеризации по нескольким метрикам качества. Также для отображения результатов кластеризации проведена визуализация данных.

Ключевые слова: кластерный анализ, нечеткая кластеризация, алгоритм Густаффсона-Кесселя, алгоритм С-средних, клещевой энцефалит, иксодовый клещевой боррелиоз, микст-инфекция.

Введение

Большинство субъектов Российской Федерации являются эндемичными по клещевым инфекциям. Широко распространены такие заболевания, как клещевой энцефалит (в 47 субъектах РФ) и иксодовый клещевой боррелиоз, также известный как болезнь Лайма (в 72 субъектах РФ). Последствия заболеваний варьируются от полного выздоровления до нарушений здоровья, приводящих к инвалидности или летальному исходу. Основной целью медицинских учреждений при поступлении пациентов с подозрением на наличие клещевой инфекции является проведение диагностики в кратчайшие сроки, чтобы как можно раньше начать необходимое лечение [1]. Инструменты искусственного интеллекта широко используются в различных сферах деятельности, в том числе и в здравоохранении. Алгоритмы машинного обучения решают задачи прогнозирования, способны в той или иной степени предоставлять рекомендации по принятию решений. Медицинские специалисты, используя инструменты машинного обучения, способны повысить качество, а также сократить время, затрачиваемое на постановку диагноза [2].

Кластерный анализ – это удобный инструмент для выявления скрытых зависимостей. Алгоритмы кластеризации позволяют из большой совокупности данных выявить группы, к которым так или иначе относится тот или иной объект. Помимо жесткой кластеризации, т. е. отнесения объекта исключительно к одной группе, существует и мягкая, также известная как нечеткая кластеризация, т. е. когда один объект с разной степенью может принадлежать к разным группам [3].

При постановке диагноза медицинские специалисты не редко сталкиваются с такими ситуациями, когда у пациента наблюдается не одно, а несколько заболеваний. Речь идет о микст-инфекциях, т.е. инфекциях, являющихся совокупностью нескольких. В таких случаях состояние пациента (течение заболевания) требует больше внимания, т. к. могут наблюдаться осложнения. Клещи могут быть переносчиками нескольких инфекций одновременно, т. е. организм человека быть поражен микст-инфекцией. Наиболее распространенной является микст-инфекция клещевого энцефалита (КЭ) и иксодового клещевого боррелиоза (ИКБ).

Целью данного исследования является проведение кластерного анализа клинических и лабораторных данных пациентов с заболеваниями, вызванными клещевыми инфекциями.

Основная часть

Сотрудниками кафедры инфекционных болезней и эпидемиологии были предоставлены деперсонализированные данные 204 пациентов, страдающих клещевыми инфекциями. Набор данных включает в себя 53 лабораторных и клинических показателя. Из 53 показателей были выбраны пары и тройка признаков, сочетание которых дает возможность наиболее полно оценить состояние здоровья пациента.

В качестве первой пары были выбраны такие показатели, как нейтрофильно-лимфоцитарный индекс и максимальное значение температуры тела. Нейтрофильно-лимфоцитарный индекс отражает соотношение нейтрофилов и лимфоцитов, что является чувствительным показателем воспалительной реакции на клеточном уровне [4]. Максимальная температура демонстрирует клиническую реакцию организма – наличие или отсутствие лихорадки. Сочетание этих двух признаков позволяет разделить пациентов на группы с выраженной или умеренной воспалительной реакцией, где лабораторная и клиническая информация взаимно дополняют друг друга.

Второй парой были выбраны лейкоцитарный индекс интоксикации (ЛИИ) по В.К. Островскому и пульс. ЛИИ по В.К. Островскому – это комплексный показатель, отражающий изменения в составе лейкоцитарной формулы, что может указывать на системный иммунный ответ. Пульс – это физиологический параметр, который часто увеличивается при стрессе и системном воспалении. В паре эти признаки помогают оценить, насколько выражен общий системный ответ организма на инфекцию или воспалительный процесс, что особенно важно при динамичном клиническом состоянии.

Третьей парой были выбраны скорость оседания эритроцитов (СОЭ) и систолическое давление (СД). СОЭ – традиционный лабораторный маркер воспаления, который может свидетельствовать о наличии острого или хронического воспалительного процесса [5]. Систолическое давление отражает состояние гемодинамики и может изменяться под влиянием воспалительного стресса и общего состояния пациента. Эта пара сочетает лабораторный показатель воспаления с измерением гемодинамических изменений, что позволяет выявить влияние воспалительного процесса на систему кровообращения и, соответственно, оценить тяжесть состояния.

Четвертой группой признаков стали нейтрофильно-лимфоцитарный индекс, максимальное значение температуры тела и пульс.

Кластеризация проводилась алгоритмами С-средних и Густаффсона-Кесселя. Эффективность проведенной кластеризации оценивалась следующими индексами: Фукуямо-Сугено, Си-Бени и нечеткого силуэта. Первые две метрики принимают наименьшие значения при наилучшей работе модели, наибольшее значение индекса нечеткого силуэта свидетельствует о более качественной кластеризации [6].

Результаты

Рассмотрим работу моделей в рамках отдельных групп признаков. Далее рядом со значениями метрик качества в круглых скобках будет указано количество кластеров. Для первой группы по индексу нечеткого силуэта наилучший результат – 0,2570 (10) показал алгоритм Густаффсона-Кесселя, по индексам Си-Бени и Фукуямо-Сугено со значениями 0,2137 (10) и 114,1412 (10), соответственно, лучше себя показал алгоритм С-средних. Аналогичные результаты были получены и для остальных групп. Алгоритм Густаффсона-Кесселя показал лучшие результаты по индексу нечеткого силуэта для второй группы – 0,3035 (10), для третьей группы – 0,2414 (10) и для четвертой группы 0,4368 (10). Кластеризация, проведенная алгоритмом С-средних, по индексам Си-Бени и Фукуямо-Сугено также показала лучшие результаты, соответственно, по второй группе – 0,2125 (8) и 118,4780 (10), третьей группе – 0,1896 (10) и 115,4750 (10), четвертой группе – 0,2526 (10) и 90,2721 (10). Видим, что

оптимальное число кластеров 8 и 10, однако в большинстве случаев метрики принимали свои наилучшие значения при числе кластеров равном 10.

На Рис. 1 представлены точечные графики распределения значений систолического давления и СОЭ. На графике (а) форма и цвет точек зависит от диагноза, который был поставлен медицинскими специалистами. На графике (б) кластеры, определенные методом С-средних, представлены в виде эллипсов разных цветов. Центроиды кластеров обозначены крупными разноцветными точками с белыми границами.

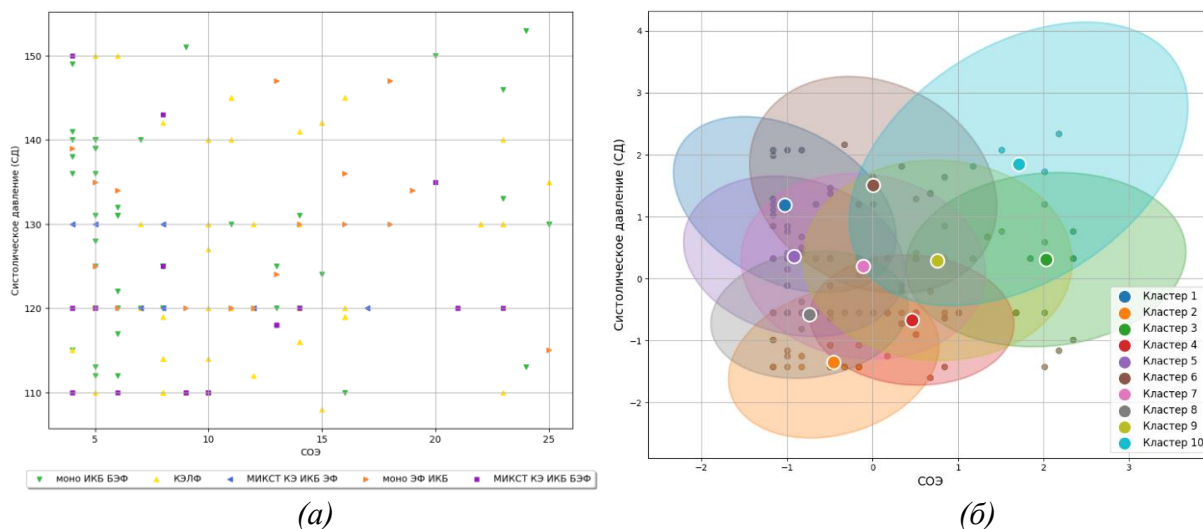


Рис. 1. Распределение значений СОЭ и систолического давления с учетом диагноза (а) и кластеризации алгоритмом С-средних (б)

При сопоставлении сведений из двух графиков, можно сделать следующие выводы:

- Безэритемная форма ИКБ (моно ИКБ БЭФ). Три точки, принадлежащие исключительно 10 кластеру; одна - 6 кластеру; одна – выброс; 10 точек – на пересечении 1, 5, 6, 7 кластеров; 4 точки – на пересечении 8 и 2 кластеров.
- Лихорадочная форма клещевого энцефалита (КЭЛФ). 13 точек принадлежат 6 кластеру, 11 точек – 4 кластеру, причем 2 из них на пересечении этих двух кластеров.
- Эритемная форма микст-инфекции (МИКСТ КЭ ИКБ ЭФ). Все 7 значений входят в 7 кластер, и 6 из них также входят в 8 кластер, четыре точки – в четвертый кластер, четыре точки также принадлежат 9 кластеру.
- Эритемная форма ИКБ (моно ЭФ ИКБ). 11 точек принадлежат 6 кластеру, 5 точек входят в 8 кластер.
- Безэритемная форма микст-инфекции (МИКСТ КЭ ИКБ БЭФ). 11 точек распределились между 1, 3, 4 и 8 кластерами.

Соотношение значений нейтрофильно-лимфоцитарного индекса, максимального значения температуры и пульса по результатам кластеризации алгоритмами Густаффсона-Кесселя и С-средних представлено на рисунке 2 (а) и (б), соответственно. Большей степенью принадлежности к кластеру соответствующего цвета центроида обладают более яркие точки, серые же точки характеризуются принадлежностью к большему числу кластеров.

Визуальное сопоставление результатов кластеризации с фактически поставленными диагнозами в трехмерном пространстве менее наглядно, чем в двумерном.

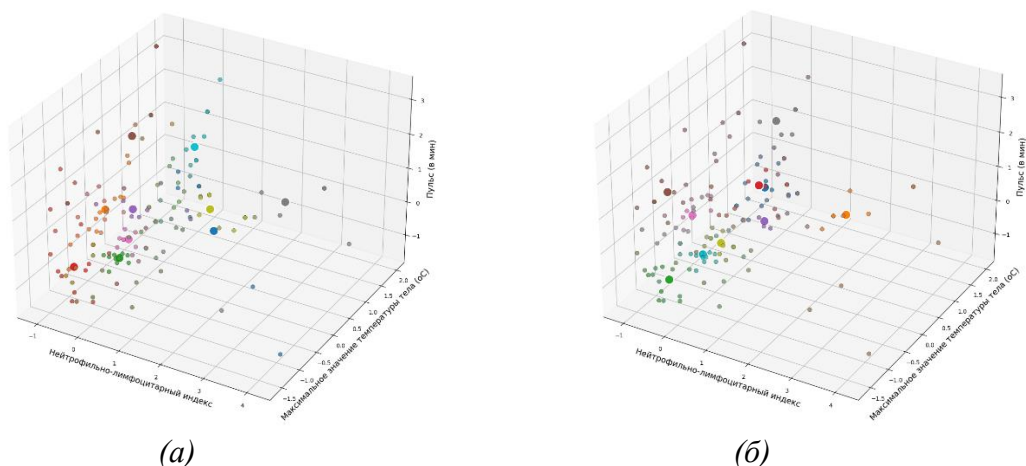


Рис. 2 – Распределение значений Нейтрофильно-лимфоцитарного индекса, Максимального значения температуры тела и Пульса с результатами кластеризации алгоритмами Густафсона-Кесселя (а) и С-средних (б)

В целом, явной закономерности принадлежности объектов к тому или иному кластеру в зависимости от фактически поставленного диагноза выявлено не было.

Заключение

На протяжении данного исследования была проведена нечеткая кластеризация клинических и лабораторных признаков пациентов с заболеваниями, передающимися иксодовыми клещами. По результатам проведенной кластеризации также была проведена оценка качества работы используемых моделей нечеткой кластеризации. Более эффективное разделение на кластеры было проведено на группе следующих показателей: пульс, максимальное значение температуры тела и нейтрофильно-лимфоцитарный индекс. Из двух моделей кластеризации выявить наилучшую не удалось, т.к. результаты варьируются в зависимости от различных метрик качества. По визуальному анализу можно сказать, что значения фактически поставленных диагнозов распределены хаотично и сопоставить их с полученными кластерами не удалось.

Список использованной литературы

1. Клещевой энцефалит – острое природно-очаговое инфекционное заболевание // «ДИАВАКС» центр вакцинопрофилактики. – [Электронный ресурс]. – URL: diavax.ru/infections/kleshchevoj-ehncefalit/ (дата обращения: 15.03.2025).
2. Аверьянова О.А., Коршак В.И. Искусственный интеллект в условиях современной медицины // Естественные и математические науки в современном мире. – 2016. – №5(40). [Электронный ресурс]. – URL: cyberleninka.ru/article/n/iskusstvennyy-intellekt-v-usloviyah-sovremennoy-medsiny (дата обращения: 10.03.2025).
3. Егоров А.В., Куприянова Н.И. Особенности методов кластеризации данных // Известия ЮФУ. Технические науки. – 2011. – № 11. [Электронный ресурс]. – URL: cyberleninka.ru/article/n/osobennosti-metodov-klasterizatsii-dannyh (дата обращения: 4.03.2025).
4. Нейтрофильно-лимфоцитарный коэффициент // Медицинская лаборатория полного цикла «Express Medical Laboratory». [Электронный ресурс]. – URL: e-m-l.ru/lib/biomarker/neytrofilno-limfotsitarnyy-koeffitsient/ (дата обращения: 19.03.2025).
5. Скорость оседания эритроцитов (СОЭ) // Helix. [Электронный ресурс]. – URL: <https://helix.ru/kb/item/02-007> (дата обращения: 9.03.2025).
6. Кудинов Ю.И., Кудинов И.Ю. Нечеткое моделирование и кластеризация // Проблемы управления. – 2008. – № 6. [Электронный ресурс]. – URL: cyberleninka.ru/article/n/nechetkoe-modelirovanie-i-klasterizatsiya (дата обращения: 12.03.2025).