



## Модель оптической локализации железнодорожных подвижных единиц на основе искусственной иммунной системы отрицательного отбора

*И.С. Артемьев, А.И. Долгий, А.И. Лебедев, А.Е. Хатламаджиян*

*Ростовский государственный университет путей сообщения*

**Аннотация:** Предложен новый подход к локализации подвижных единиц, основанный на распознавании автосцепки вагона на графических изображениях, получаемых с помощью видеокамер универсальной системы автоматического распознавания номеров вагонов (УС АРНВ), в качестве маркера межвагонного пространства.

Разработана модель распознавания автосцепки на основе вещественного алгоритма отрицательного отбора с детекторами переменного размера (МВОП). Проведены вычислительные эксперименты, показавшие эффективность модели МВОП в сравнении с классическим методом опорных векторов на реальных данных, полученных на одном из объектов внедрения УС АРНВ.

**Ключевые слова:** искусственная иммунная система, отрицательный отбор, мягкие вычисления, локализация подвижных единиц, идентификация номеров вагонов, система технического зрения.

### Введение

Задача локализации подвижных единиц в железнодорожном составе является крайне актуальной и распространенной в системах полуавтоматической блокировки, ограждения пешеходных переходов, повагонного моделирования, а также для различных систем диагностирования подвижного состава. Самым известным и часто используемым методом решения задачи является метод, основанный на принципе фиксации прохождения колеса с помощью индуктивных датчиков. Однако, и он не свободен от некоторых принципиальных недостатков в работе. Так одними из основных проблем ложного срабатывания этих датчиков является перемещение вблизи них технологического металлического инструмента, помехи от контактной сети, сбой в работе аппаратной части и др. Ввиду чего возникает необходимость альтернативного способа счета подвижных единиц, имеющего другую природу входных данных и способа их получения.

В этой статье мы предлагаем подход, основанный на распознавании автосцепки вагона на графических изображениях, в качестве маркера межвагонного пространства. Все эксперименты были проведены на базе универсальной системы автоматического распознавания номеров вагонов (УС АРНВ) [1-3], где в качестве основных датчиков используются видеокамеры.

Задача распознавания автосцепки вагона, по сути, является задачей классификации на два класса: «автосцепка» и «не автосцепка». Однако, такая постановка представляется достаточно некорректной и сложной для практической реализации, т.к., в общем случае, в класс «не автосцепка» входит бесконечное множество примеров. В таких случаях логичнее использовать методы для так называемой «одноклассовой» классификации, в которых отсутствует необходимость в отрицательных примерах при обучении. Наиболее перспективным из них, по мнению авторов, является искусственные иммунные системы отрицательного отбора (ИИС ОО), основанные на базовом принципе функционирования иммунной системы человека, заключающейся в задаче разделения антигенов на «своих» и «чужих» [4].

Тем не менее, в экспериментальной части данной статьи мы сравним результаты классификации с помощью ИИС ОО с другим классическим методом одноклассовой классификации: одноклассовым методом опорных векторов (англ. OSVM, one-class support vector machine) [5].

### **Принцип отрицательного отбора**

Основная задача иммунной системы состоит в защите организма от заболеваний, вызванных различными патогенными микроорганизмами. Для обнаружения и устранения болезнетворных микроорганизмов, иммунная система содержит определенные типы белых кровяных клеток, называемых лимфоцитами, которые могут распознавать патогенные объекты, называемые

антигенами. Лимфоциты можно рассматривать в качестве детекторов, так как они несут на своей поверхности, так называемые антитела, которые имеют возможность распознавать и классифицировать белки на те которые производятся внутри организма («свои») и вне его («чужие»). Во избежание неправильной классификации собственных белков лимфоцитами, иммунная система уничтожает реактивные лимфоциты в процессе отрицательного отбора. После процесса отрицательного отбора, иммунная система содержит лимфоциты, которые распознают чужеродные белки. Поскольку количество лимфоцитов в любой момент времени ограничено, лимфоциты, которые не участвуют в процессе распознавания, уничтожаются, а вместо них добавляются новые. Постоянный оборот новых и старых лимфоцитов позволяет иммунной системе эффективно распознавать всевозможные чужеродные белки с течением времени с ограниченным количеством антител.

На основе описанного принципа в работе [4] предложен алгоритм отрицательного отбора для обнаружения вирусов. Основная идея этого алгоритма заключается в создании ряда детекторов в дополнительном пространстве с последующим их применением для классификации новых примеров на «своих» и «чужих».

Алгоритм отрицательного отбора можно представить в следующем обобщённом виде. Обозначим все пространство примеров –  $U$ , множество «своих» клеток –  $S$  и множество «чужих» клеток –  $N$ , где  $U = S \cup N$ , а  $S \cap N = \emptyset$ .

1. Определить множество «своих» клеток как множество  $S$  элементов мощностью  $l$  в пространстве  $U$ .
  2. Создать множество детекторов  $D$ , таких, что  $D \cap S = \emptyset$ .
  3. Контролировать множество  $S$  на предмет изменений, непрерывным сопоставлением множества детекторов  $D$  и множества  $S$ .
-

Алгоритм был подвержен критике [6] из-за высоких требований ко времени работы и низкой точности классификации. Однако, некоторые его модификации, в частности алгоритм отрицательного отбора с детекторами переменного размера [7], показали хорошие результаты.

### **Вещественный алгоритм отрицательного отбора с детекторами переменного размера**

В работах [8, 9] предложен алгоритм отрицательного отбора, который оперирует в пределах гиперкуба  $[0, 1]^n$ . Детектор  $d = (c, r_{ns})$  имеет центр  $c \in [0, 1]^n$  и радиус распознавания «чужих» клеток  $r_{ns} \in (0, \sqrt{n})$ .

Кроме того, каждый «свой» элемент  $s = (c, r_s)$  имеет центр и радиус  $r_s$ , который нужен для того, чтобы новые примеры, располагающиеся достаточно близко к «своему» элементу, не распознавались как «чужие».

Если элемент лежит в пределах детектора (гиперсферы), то он классифицируется как «чужой», в противном случае, как «свой». Элемент  $e$  лежит в пределах детектора  $d = (c, r_{ns})$ , если евклидово расстояние

$$\text{dist}(c, e) = \left( \sum_{i=1}^n (c_i - e_i)^2 \right)^{1/2} < r_{ns} .$$

В работе [10] предложен вещественный алгоритм отрицательного отбора с детекторами переменного размера (называемые V-детекторами), где центры детекторов, располагаются случайным образом и не должны лежать в пределах гиперсфер «своих» элементов.

Радиус детектора динамически изменяется до касания с границами гиперсфер «своих» элементов. Алгоритм завершается, если сгенерировано заданное количество детекторов или ранее, если predetermined часть пространства примеров накрыта детекторами. Этот алгоритм продемонстрировал хорошие результаты классификации на различных

стандартных наборах данных в сравнении с другими алгоритмами отрицательного отбора.

На рисунке 1 проиллюстрирован принцип отрицательного отбора с детекторами переменного размера в случае двумерного пространства. Пространство «чужих» элементов, покрыто детекторами переменного размера, изображенных в виде серых окружностей. «Свои» элементы изображены в виде серых кругов с черным центром.

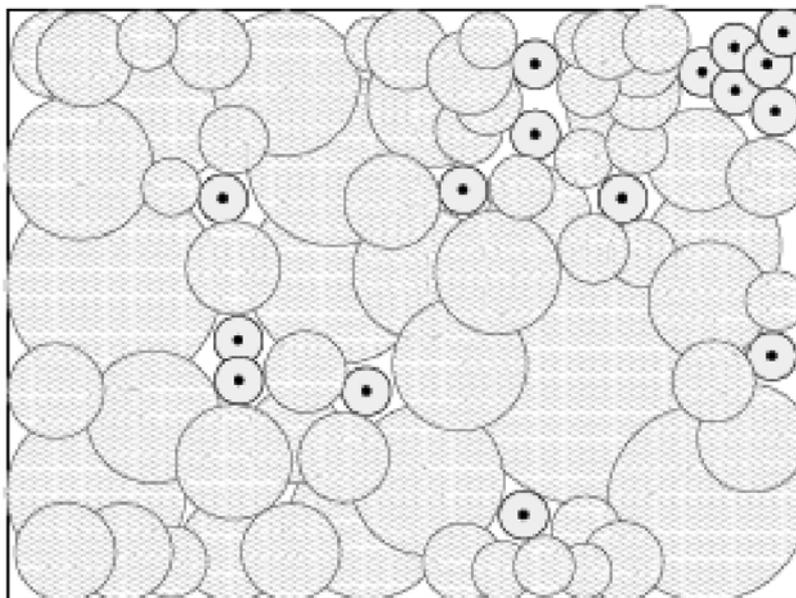


Рис. 1. – Принцип отрицательного отбора с детекторами переменного размера

### **Решение задачи локализации подвижных единиц**

Целью предлагаемого подхода, является распознавание автосцепки вагона на графических изображениях в качестве маркера межвагонного пространства, для повышения надежности локализации подвижных единиц в системе УС АРНВ. Во время движения состава видеокамеры генерируют последовательный набор видеокадров (изображений), на которых в определенной стационарной области (обычно в середине кадра по горизонтали) и в определенные моменты времени можно наблюдать автосцепку. Факт ее наблюдения в текущий момент времени позволяет

однозначно идентифицировать межвагонное пространство и, следовательно, локализовать сами вагоны.

На рисунке 2 в качестве примера представлены два кадра с проследующим составом. В кадре а) показан момент времени, в котором автосцепка находится в заданной области интереса, а в кадре б) – автосцепка в заданной области отсутствует.

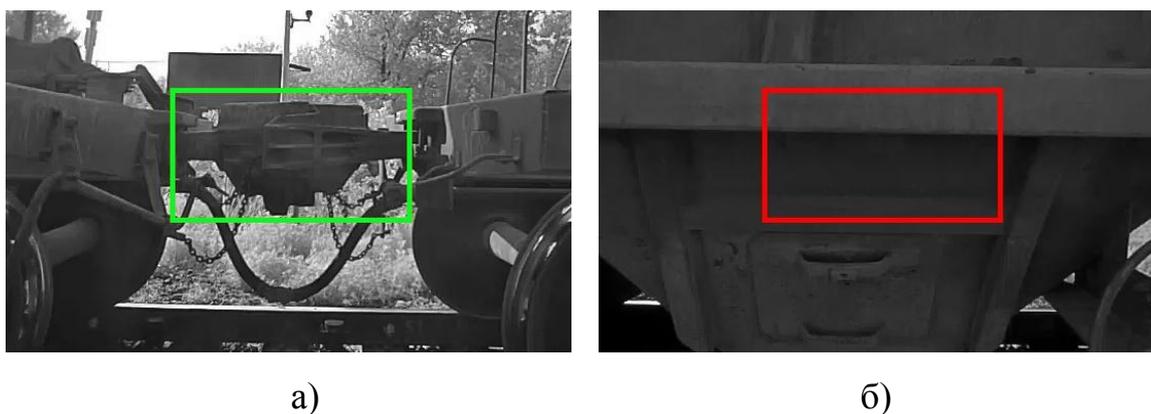


Рис. 2. – Примеры кадров с выделенной заданной областью интереса: а) автосцепка находится в заданной области интереса, б) автосцепка в заданной области отсутствует

Приняв за примеры подлежащие классификации, фрагменты видеок кадров с постоянными координатами, соответствующими стационарной области интереса для наблюдения автосцепки, мы получаем возможность сформировать множество положительных примеров класса «свой» из изображений, на которых присутствует автосцепка.

### **Модель распознавания автосцепки**

Опыт авторов, полученный при разработке нейроиммунной модели классификации (НИМК) показал эффективность объединения неустойчивых классификаторов в так называемые «ансамбли» [11]. Исходя из этого, в предлагаемой модели распознавания автосцепки на основе вещественного алгоритма отрицательного отбора с детекторами переменного размера (МВОП) нами используется аналогичный подход.

Таким образом, предлагаемая модель МВОП состоит из следующих этапов:

1. Сокращение размерности данных. Имеется некоторая выборка положительных примеров (класса «свой»)  $X = \{x_n\}_{n=1}^N, x_n \in R^D$ . В целях уменьшения вычислительных затрат на этапе обучения, сжатия размерности данных и формирования компактных областей расположения данных, относящихся к одному классу, в признаковом пространстве, необходимо представить выборку в пространстве меньшей размерности  $d < D$ . Для достижения поставленной цели применяется широкоизвестный метод главных компонент (Principal component analysis – PCA) [12], заключающийся в обнаружении проекции на гиперплоскость с сохранением основной части дисперсии в данных:

$$x = \mu + t_1 w_1 + \dots + t_d w_d = Wt + \mu.$$

Здесь  $x \in R^D, t \in R^d$  – новые координаты объекта,  $W = (w_1 | \dots | w_d) \in R^{D \times d}, \mu \in R^D; w_1, \dots, w_d$  – базис гиперплоскости.

2. Формирование ансамбля ИИС ОО. Необходимо сформировать высокое разнообразие классификаторов, обученных на различных выборках данных. Для достижения поставленной цели применяется подход, использующий бутстреп-выборки [13] из обучающего множества – «бэггинг» [14]. Он состоит из следующих этапов:

– Возьмем первоначальное обучающее множество примеров  $Y$  размером  $n$ .

– Сгенерируем  $m$  новых обучающих множеств примеров  $\hat{Y}_i (i = \overline{1, m})$  размером  $n' \leq n$  путем случайной выборки с возвратом из множества  $Y$ . Т.к. выборка производится случайным образом, то количество одинаковых примеров в этих обучающих множествах  $\hat{Y}_i$  будет различным: некоторые из них могут быть отобраны несколько раз, а другие – ни разу. Если  $n' \leq n$ , то при больших  $n$  каждое множество  $\hat{Y}_i$  ожидаемо будет иметь

$1-1/e$  ( $\approx 63,2\%$ ) уникальных примеров множества  $Y$ , остальные из них будут повторяться [15].

– Обучим  $M$  классификаторов на  $m$  выборках.

3. Классификация. Результат работы ансамбля ИИС ОО, полученного на предыдущем шаге, определяется путем голосования выходов классификаторов:

$$Class = \begin{cases} "self", n^{(self)} \geq n^{(non-self)} \\ "non-self", n^{(self)} < n^{(non-self)} \end{cases}$$

где  $n^{(self)}$  и  $n^{(non-self)}$  – количество классификаторов ансамбля, проголосовавших за принадлежность текущего примера к классу «свой» и к классу «чужой» соответственно.

### Вычислительные эксперименты

Программный прототип предлагаемого алгоритма распознавания автосцепки был реализован для проведения вычислительных экспериментов. В качестве примеров обучающего и тестового множества использовались монохромные изображения автосцепок ж.-д. вагонов, полученные на одном из объектов внедрения УС АРНВ (рис. 3).

В обучающем множестве содержится 10 тыс. уникальных положительных примеров, а в тестовом – 1 тыс. положительных и 1 тыс. отрицательных примеров размерностью 86 на 53 пикселей (4558 признаков).



Рис. 3 – Примеры обучающей выборки – класса «свой»

Программа экспериментов состояла из трех этапов:

1. Обучение и классификация с помощью модели МВОП;
2. Обучение и классификация с помощью модели МВОП, в которой ансамбль ИИС ОО заменен на классификатор OSVM.

Каждый из экспериментов проводился по пять раз, а полученные результаты усреднялись.

Эксперименты выполнялись на специализированном компьютере серверного исполнения с установленными двумя процессорами Intel Xeon E5-2660V2 частотой 2660 МГц.

Размерность пространства поиска была снижена до 20 в результате применения метода главных компонент.

Ансамбль ИИС, используемый в экспериментах первого этапа состоял из десяти ИИС ОО. Количество детекторов каждой ИИС ОО было ограничено значением 10 тысяч. Радиус  $r_s$  был принят равным 0,025.

Полученные результаты распознавания автосцепок приведены в таблице 1.

Таблица № 1

Результаты распознавания автосцепок

Этап	Правильное распознавание ( $P_{TR}$ ), %	Ложное распознавание ( $P_{FR}$ ), %	Время обучения, ч	Время распознавания, с
1	98,7	2,2	734	2
2	96,2	3,4	0,23	6

### Обсуждение результатов

Как видно из таблицы 1 результат правильного распознавания автосцепки моделью МВОП выше, а процент ложного распознавания ниже. Время распознавания у модели МВОП меньше, однако, время обучения намного больше. Этот факт объясняется известной проблемой ИИС ОО заключающейся в необходимости больших временных затратах на процесс обучения, который содержит в себе элементы случайного поиска.

В поставленной прикладной задаче время камерального процесса обучения не является критичным параметром и его значение не снижает эффективности модели МВОП в целом.

Так как модель МВОП планируется реализовать как минимум для двух видеокамер, расположенных по разные стороны ж.д. пути, ориентировочные значения правильного  $\bar{P}_{TR}$  и ложного  $\bar{P}_{FR}$  распознавания для этого случая можно рассчитать по формулам:

$$\bar{P}_{TR} = P_{TR}^2 = 0,987^2 = 0,974,$$
$$\bar{P}_{FR} = P_{FR}^2 = 0,022^2 = 4,8 \cdot 10^{-4}.$$

Как видно из полученных значений, при несущественном снижении правильного распознавания  $\bar{P}_{TR}$  мы получили ощутимое снижение ложного распознавания  $\bar{P}_{FR}$  (более чем в 45 раз).

Кроме того, предлагаемый подход планируется использовать в совокупности с методом на основе индуктивных датчиков прохождения колеса (ДПК), где приоритет МВОП может отдаваться в нехарактерных для нормальной работы технологических ситуациях, таких как, несоответствие пройденных осей «шаблону» подвижной единицы, расхождение количества вагонов с натурным листом поезда, появлению мешающих электромагнитных возмущений и др.

Для примерной оценки вероятности ошибок в работе такой системы, рассмотрим пример. Из статистических данных, также полученных на одном из объектов внедрения УС АРНВ, вероятность отказа метода ДПК из-за несрабатывания датчиков составляет примерно  $P_{ErrDPK} = 5,7 \cdot 10^{-4}$ . Тогда итоговая вероятность неудачной локализации подвижной единицы этой системы  $P_{NL}$  будет равна:

$$P_{NL} = P_{ErrDPK} * (1 - \bar{P}_{TR}) = 5,7 \cdot 10^{-4} * (1 - 0,974) = 1,5 \cdot 10^{-5}.$$

Таким образом, количество случаев неудачной локализации при комплексном использовании двух методов уменьшится приблизительно в 40 раз.

### **Заключение**

В работе предложен новый подход локализации подвижных единиц, основанный на распознавании автосцепки вагона на графических изображениях, получаемых с помощью видеокамер УС АРНВ, в качестве маркера межвагонного пространства.

Разработана модель распознавания автосцепки на основе вещественного алгоритма отрицательного отбора с детекторами переменного размера (МВОП). Были проведены вычислительные эксперименты, показавшие эффективность модели МВОП в сравнении с классическим методом опорных векторов на реальных данных, полученных на одном из объектов внедрения УС АРНВ.

Показана эффективность предлагаемого подхода в дублированном режиме работы (с использованием двух видеокамер) и в совокупности с базовым методом локализации подвижных единиц на основе индуктивных датчиков прохождения колеса.

Работа выполнена при финансовой поддержке РФФИ, проект № 13-07-00226 А, проект № 13-07-13109 офи\_м\_РЖД.

### **Литература**

1. Долгий А.И., Хатламаджиян А.Е. и др. Новые подходы к разработке систем автоматической идентификации номеров вагонов // Бюллетень Объединённого ученого совета ОАО «РЖД», Москва, № 3, 2014 г. – с. 23-40.
2. Суханов А.В., Артемьев И.С., Долгий А.И., Хатламаджиян А.Е. Метод оптической идентификации железнодорожных подвижных единиц на основе

интегральных устойчивых признаков // Инженерный вестник Дона, 2013, № 4  
URL: [ivdon.ru/magazine/archive/n4y2013/2217](http://ivdon.ru/magazine/archive/n4y2013/2217).

3. Артемьев И.С., Лебедев А.И., Долгий А.И., Хатламаджиян А.Е., Меерович В.Д. Метод блочного оптического распознавания инвентарных номеров железнодорожных подвижных единиц на основе комитетной нейроиммунной модели классификации // Инженерный вестник Дона, 2014, № 1 URL: [ivdon.ru/magazine/archive/n1y2014/2259](http://ivdon.ru/magazine/archive/n1y2014/2259).

4. S. Forrest, A.S. Perelson, L. Allen, R. Cherukuri. Self-Nonself Discrimination in a Computer. In Proc. of the 1994 IEEE Symposium on Research in Security and Privacy. IEEE Computer Society Press, 1994. – pp. 202-212.

5. Ma J., Perkins S. Time-series novelty detection using one-class support vector machines // Proceedings of the International Joint Conference on Neural Networks. – July 2003. – Vol. 3. – P. 1741–1745.

6. Freitas A.A., Timmis J. Revisiting the foundation of artificial immune systems: A problem-oriented perspective // Second International Conference on Artificial Immune System (ICARIS 2003): Proceedings. – Berlin–Heidelberg: Springer-Verlag, 2003. – Ser. LNCS 3239. – pp. 229-241.

7. Ji Z., Dasgupta D. Real-valued negative selection algorithm with variable-sized detectors // Genetic and Evolutionary Computation (GECCO 2004): Proceedings. – Berlin–Heidelberg: Springer-Verlag, 2004. – Ser. LNCS 3102. – Part I. – pp. 287–298.

8. F. Gonzalez, D. Dasgupta, and R. Kozma. Combining negative selection and classification techniques for anomaly detection. In Congress on Evolutionary Computation, pages 705–710. IEEE, May 2002.

9. F.A. Gonzalez, D. Dasgupta, and L.F. Niño. A Randomized Real-Valued Negative Selection Algorithm. In Proc. of the 2nd Int. Conf. on Artificial Immune Systems (ICARIS), volume 2787 of LNCS, pages 261–272, Edinburgh, UK, 2003. Springer-Verlag.

---



10. Z. Ji and D. Dasgupta. Real-valued negative selection algorithm with variable-sized detectors. In Genetic and Evolutionary Computation – GECCO-2004, Part I, volume 3102 of LNCS, pages 287–298, Seattle, WA, USA, 26-30 June 2004. Springer-Verlag.

11. Артемьев И.С., Долгий А.И., Суханов А.В., Хатламаджиян А.Е. Нейроиммунная модель классификации в задачах идентификации на транспорте // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сборник научных трудов VII-й Международной научно-технической конференции (Коломна, 20-22 мая 2013 г.). В 3-х томах. Т.3. – М.: Физматлит, 2013. с. 980 – 987.

12. K. Pearson. On Lines and Planes of Closest Fit to Systems of Points in Space. Philosophical Magazine, vol. 2, pp. 559–572, 1901.

13. Hastie T., Tibshirani R., Friedman J., The elements of statistical learning: Data mining, inference, and prediction. Springer, 2001. – p. 533.

14. Breiman L., Bagging predictors. Univ. California Technical Report No. 421, September 1994. – pp. 123-140.

15. Aslam, Javed A., Popa, Raluca A., Rivest, Ronald L., On Estimating the Size and Confidence of a Statistical Audit. Proceedings of the Electronic Voting Technology Workshop (EVT '07), Boston, 2007. – pp. 8-10.

### References

1. Dolgiy A.I., Khatlamadzhiyan A.E. i dr. Byulleten' Ob"edinennogo uchenogo soveta OAO «RZhD», Moskva, № 3, 2014 – pp. 23-40.

2. Sukhanov A.V., Artem'ev I.S., Dolgiy A.I., Khatlamadzhiyan A.E. Inženernyj vestnik Dona (Rus), 2013, № 4 URL: ivdon.ru/magazine/archive/n4y2013/2217.

3. Artem'ev I.S., Lebedev A.I., Dolgiy A.I., Khatlamadzhiyan A.E., Meerovich V.D. Inženernyj vestnik Dona (Rus), 2014, № 1 URL: ivdon.ru/magazine/archive/n1y2014/2259.

4. S. Forrest, A.S. Perelson, L. Allen, R. Cherukuri. In Proc. of the 1994 IEEE Symposium on Research in Security and Privacy. IEEE Computer Society Press, 1994. pp. 202-212.

5. S. Forrest, A.S. Perelson, L. Allen, R. Cherukuri. In Proc. of the 1994 IEEE Symposium on Research in Security and Privacy. IEEE Computer Society Press, 1994. pp. 202-212.

6. Freitas A.A., Timmis J. Second International Conference on Artificial Immune System (ICARIS 2003): Proceedings. – Berlin–Heidelberg: Springer-Verlag, 2003. – Ser. LNCS 3239. pp. 229-241.

7. Ji Z., Dasgupta D. Genetic and Evolutionary Computation (GECCO 2004): Proceedings. – Berlin–Heidelberg: Springer-Verlag, 2004. – Ser. LNCS 3102. – Part I. pp. 287–298.

8. F. Gonzalez, D. Dasgupta, and R. Kozma. In Congress on Evolutionary Computation, pages 705–710. IEEE, May 2002.

9. F.A. Gonzalez, D. Dasgupta, and L.F. Niño. In Proc. of the 2nd Int. Conf. on Artificial Immune Systems (ICARIS), volume 2787 of LNCS, pages 261–272, Edinburgh, UK, 2003. Springer-Verlag.

10. Z. Ji and D. Dasgupta. In Genetic and Evolutionary Computation – GECCO-2004, Part I, volume 3102 of LNCS, pages 287–298, Seattle, WA, USA, 26-30 June 2004. Springer-Verlag.

11. Artem'ev I.S., Dolgiy A.I., Sukhanov A.V., Khatlamadzhyan A.E. Nejroimmunnaja model' klassifikacii v zadachah identifikacii na transporte [Neuroimmune classification model in identification problems in transport]. Sbornik nauchnykh trudov VII-y Mezhdunarodnoy nauchno-tehnicheskoy konferentsii (Kolomna, 20-22 maya 2013 g.). V 3-kh tomakh. T.3. – M.: Fizmatlit, 2013. pp. 980 – 987.

12. K. Pearson. Philosophical Magazine, vol. 2, pp. 559–572, 1901.

---



13. Hastie T., Tibshirani R., Friedman J., The elements of statistical learning: Data mining, inference, and prediction. Springer, 2001. p. 533.
14. Breiman L. Univ. California Technical Report No. 421, September 1994. pp. 123-140.
15. Aslam, Javed A., Popa, Raluca A., Rivest, Ronald L. Proceedings of the Electronic Voting Technology Workshop (EVT '07), Boston, 2007. pp. 8-10.