

УДК 004.891.3

## **Использование вероятностных нейронных сетей для предсказания локализации белков в клеточных компартментах**

**Назин П.С.<sup>\*1,2</sup>, Готовцев П.М.<sup>1</sup>**

<sup>1</sup>*НИЦ "Курчатовский институт", Москва, Россия*

<sup>2</sup>*Московский физико-технический институт, Москва, Россия*

**Аннотация.** В данной работе описано использование аппарата вероятностных нейронных сетей для решения проблем биоинформатики на примере задачи определения локализации белков по их первичной структуре. В качестве исходных данных использованы наборы характеристик аминокислотных последовательностей белков, полученные различными программными средствами, нацеленными на поиск специфических сигнальных последовательностей, а также данные о том, где в клетке локализуются эти белки в клетках двух микроорганизмов – бактерии *E. coli* и дрожжей *S. cerevisiae*. Источник данных – UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/datasets>). Показана возможность применения вероятностных нейронных сетей для решения данной задачи, так как получена точность классификации 57.5 % и 85.0 % для дрожжевых и бактериальных клеток соответственно. Полученные показатели точности классификации использованных данных превышают, те, которые по данным литературы были достигнуты при применении других методов распознавания. Отмечено, что высокая скорость обучения и возможность модификации делает вероятностные нейронные сети перспективным инструментом для анализа данных биоинформатики.

**Ключевые слова:** *вероятностные нейронные сети, сайты связывания белков, классификация данных, машинное обучение, протеомные базы данных.*

### **ВВЕДЕНИЕ**

#### **Искусственные нейронные сети**

Искусственные нейронные сети – класс моделей, построенных с использованием алгоритмов машинного обучения на основе принципа коннекционизма – предположения о том, что мыслительные явления могут быть описаны сетями из взаимосвязанных простых элементов, по аналогии с организацией биологических нейронных сетей. ИНС имитирует поведение системы, исходя из предоставленных экспериментальных или известных из других источников данных, позволяя пропустить этап создания алгоритмической/механической модели, необходимый для описания системы и решения связанных с ней задач при традиционном подходе и представляющий значительные трудности для сложных и нелинейных систем, часто встречающихся в задачах из области биологии [1]. К недостаткам нейронных сетей относятся в большинстве случаев длительный процесс обучения, потребность в большом объеме данных для обучения и работу по принципу "черного ящика" –

\* nazin@phystech.edu

построение алгоритмической/механической модели на основе нейронной сети представляет из себя отдельную задачу [2].

### **Вероятностные нейронные сети**

Вероятностные нейронные сети являются подвидом искусственных нейронных сетей на основе радиальной базисной функции, опирающимся на принцип непрерывности плотности вероятности. Используя известные точки данных, каждому возможному результату (классу) по принципу Парзеневского окна [3, 4], а именно сопоставляется аппроксимация функции плотности вероятности, построенная на соответствующих этому результату точках данных как суперпозиция Гауссовых ядер в качестве базовых функций [5]. Такой расчет проводится для всех возможных результатов и в качестве ответа выбирается вариант, отвечающий наибольшему значению функции плотности вероятности.

Вероятностная нейронная сеть состоит из четырех слоев: входной слой, скрытый слой с нейронами, соответствующими обучающим примерам, суммирующий слой, получающий выводы от нейронов скрытого слоя обучающих примеров соответствующего класса, и выходной слой, на котором проводится сравнение вероятностей и выводится результат.

Вероятностные нейронные сети широко применяются в задачах классификации и распознавания, таких как различные задачи в медицине [6], включая диагностику таких заболеваний, как аритмия [7], опухоли головного мозга [8], рак кожи [9] и прочие [10–12], диагностика в медицине [13, 14], распознавание голоса [15], определение личности по голосу [16], радужной оболочке глаза [17], изображению лица [18] и геометрии рук [19], распознавание, обработка и классификация изображений [20, 21]. В области биологии вероятностные нейронные сети находят применение в таких задачах, как классификация белков по суперсемействам [5].

В ряде работ также отмечается, что вероятностные нейронные сети выгодно отличаются при решении классификационных задач, способностью работать с нелинейными, сложно организованными системами, устойчивости к шумам, высокой скорости обучения и легкости модификации для включения новых данных [22–25], что является важным преимуществом при работе с постоянно дополняемыми базами данных. Примером таких баз являются быстро растущие базы данных, описывающие свойства и структуру различных белков [26]. В работе [5] уже была показана высокая эффективность классификации белков по суперсемействам с использованием вероятностных нейронных сетей, что подтверждает возможность использование подобных сетей для анализа данных, опубликованных в протеомных базах. С учетом легкости включения новых данных и высокой скоростью вероятностные нейронные сети могут быть перспективным инструментом для работы с данными из протеомных баз, причем с учетом описанных выше их свойств, возможно в будущем даже в качестве инструмента для анализа данных пользователями. В таком аспекте вероятностные нейронные сети могут быть даже более интересны, чем искусственные нейронные сети глубокого обучения, так как последним при каждом существенном изменении базы данных, необходимо будет проводить переобучение.

### **Задача определения внутриклеточной локализации белков**

Примером системы, сложной для экспериментального разрешения, является задача предсказания места локализации белков в клетке, включая как сайты их связывания, так и места накопления. Для правильного выполнения своих функций синтезируемые на рибосомах белки перемещаются механизмами внутриклеточного транспорта к соответствующим конкретным белкам местам накопления (локализации) – различные органеллы клетки: ядро, митохондрии, эндоплазматический ретикулум, аппарат

Гольджи, лизосомы, или выводятся из клетки [27, 28]. Это называется процессом внутриклеточной сортировки и локализации белков. Для организации этого процесса белки обладают специфическими метками – такими как сигнальные аминокислотные последовательности в случае ядерной или эндоплазматической локализации, или присоединенными олигосахаридами. Также сортировка может осуществляться за счет специфического взаимодействия белков с другими молекулами клетки, как мембранные липиды, взаимодействующие с некоторыми доменами белков [29, 30]. Понимание того, как белки распределяются внутри клетки, является важной задачей, стоящей перед современной клеточной биологией, так как оказывает влияние на физиологическое состояние клеток в нормальных условиях и при патологиях [31]. Внутриклеточные транспортные механизмы обеспечивают присутствие нужного белка на соответствующем сайте локализации в нужное время, позволяя организму достичь большего функционального разнообразия без дальнейшего усложнения структуры самих белков и обеспечивает их включение во внутриклеточные функциональные сети. Как отмечается в ряде работ, ошибки в работе транспортных механизмов, вызванные такими факторами как мутации, приводят к неправильному расположению белков внутри клетки, что может вызвать нарушение функционирования белков, что в свою очередь может привести к таким болезням, как рак, почечные камни, болезнь Альцгеймера [32–35]. Таким образом, сегодня определение сайтов локализации белков рассматривается как один из подходов, который позволит улучшить процесс выбора цели при разработке лекарств [31]. Также предсказание внутриклеточного расположения белков может быть важным, для понимания механизма болезней, в фармакокинетике [36], в поиске связей между новыми белками и генетическими заболеваниями [28] и поиске перспективных белков [37].

Так как определение расположения белков экспериментальными методами требует больших временных затрат и является дорогостоящим, а механизмы сортировки достаточно хорошо изучены только для небольшого количества белков и их возможных локализаций, в условиях поступления все большего количества данных, полученных в результате секвенирования, автоматические методы решения этой задачи становятся все более востребованными [28]. Одним из рассматриваемых сегодня в научной литературе методов является использование характеристик белков, таких как аминокислотная последовательность, дипептидный состав и др., в качестве опорных точек для автоматических систем предсказания [38].

## МАТЕРИАЛЫ И МЕТОДЫ

### Использованные данные

В работе использованы общедоступные наборы данных о локализации белков в дрожжах *Saccharomyces cerevisiae* (<https://archive.ics.uci.edu/ml/datasets/Yeast>) и кишечной палочке *Escherichia coli* (<http://archive.ics.uci.edu/ml/datasets/Ecoli>), объемом 1484 и 336 записей. Для каждого белка в этих базах данных, созданных авторами работ [39, 40], представлены перечисленные в таблицах 1–4 атрибуты, полученные исключительно из анализа аминокислотной последовательности различными программными средствами, а также соответствующее этому белку место локализации в клетке. Данная информация для дрожжей была собрана авторами по базе NBRF-PIR, а для бактерий – по SWISS-PROT для создания экспертной системы предсказания локализации в клетке белка с данной первичной структурой, основанной на правилах искусственного интеллекта (правилах «если-то»).

Мы использовали эти наборы дескрипторов белков при тестировании своей экспертной системы, основанной на применении вероятностных нейронных систем. Для обучения вероятностных нейронных сетей были использованы различные

разбиения данных на обучающий и тестовый наборы: 600/300, 800/350, 1000/400, 1200/200, 1400/80 случайно выбранных неповторяющихся точек в обучающем/тестовом наборе для дрожжей и 200/100, 210/100, 220/100, 230/100, 240/90 случайно выбранных неповторяющихся точек в обучающем/тестовом наборе для *E. coli*. Случайная выборка производилась следующим образом: для разбиения с  $x/y$  точек на обучающем/тестовом наборе при помощи функции `np.random.choice()` создавался массив из случайным образом перемешанных чисел от 0 до  $x + y$ , затем в обучающий набор выбирались примеры из базы данных с индексами, соответствующими первым  $x$  числам в массиве и в тестовый набор - оставшимся  $y$  числам.

**Таблица 1.** Атрибуты белков в дрожжах

№ п/п	Обозначение	Описание атрибута
1		Название последовательности: номер в базе NBRF-PIR
2	mcg	Распознавание сигнальных последовательностей методом McGeoch
3	gvh	Распознавание сигнальных последовательностей методом von Heijne
4	alm	Результат, полученный программой предсказания покрываемого региона мембраной ALOM
5	mit	Результат дескриминантного анализа аминокислотного содержания в N-терминальном (длиной в 20 остатков) в митохондриальных и немитохондриальных белках
6	erl	Наличие подпоследовательности "HDEL" (предположительно представляет собой сигнал для удержания в просвете эндоплазматической сети), двоичный атрибут
7	rox	Пероксисомный сигнал в C-терминальном участке
8	vac	Результат дескриминантного анализа аминокислотного содержания в вакуолярных и внеклеточных белках
9	nuc	Результат дескриминантного анализа аминокислотного содержания в ядерных и неядерных белках

**Таблица 2.** Типы локализации в дрожжах

№ п/п	Обозначение	Описание атрибута
1	cyt	Цитозольная или цитоскелетная
2	nuc	Ядерная
3	mit	Митохондриальная
4	me3	Мембранный белок без N-терминальной сигнальной последовательности
5	me2	Мембранный белок с сигнальной последовательностью
6	me1	Мембранный белок без сигнальной последовательности
7	exc	Внеклеточная
8	vac	Вакуольная
9	rox	Пероксисомная
10	erl	Локализация в эндоплазматическом ретикулуме

**Таблица 3.** Атрибуты белков в *E. coli*

№ п/п	Обозначение	Описание атрибута
1		Название последовательности: номер в базе SWISS-PROT
2	mcg	Распознавание сигнальных последовательностей методом McGeoch
3	gvh	Распознавание сигнальных последовательностей методом von Heijne
4	lip	Распознавание сигнальной цепи пептидазы II методом von Heijne, двоичный атрибут
5	chg	Наличие заряда на N-терминальном участке предсказанных липопротеинов, двоичный атрибут
6	aac	Результат дискриминантного анализа аминокислотного содержания в белках внешней мембраны и периплазматического пространства
7	alm1	Результат, полученный программой предсказания покрываемого региона мембраной ALOM
8	alm2	Результат, полученный программой предсказания покрываемого региона мембраной ALOM после исключения предполагаемых отщепляемых участков цепи

**Таблица 4.** Типы локализации в *E. coli*

№ п/п	Обозначение	Описание атрибута
1	cp	Цитоплазматическая
2	im	Внутренняя мембрана, без сигнальной последовательности
3	pp	Периплазматическая
4	imU	Внутренняя мембрана, необрезаемая сигнальная последовательность
5	om	Внешняя мембрана
6	omL	Липопротеин внешней мембраны
7	imL	Липопротеин внутренней мембраны
8	imS	Внутренняя мембрана, обрезаемая сигнальная последовательность

### Вероятностные нейронные сети

Для обработки данных были созданы две вероятностные нейронные сети. Для этого создана программа на языке программирования Python 3.5 с использованием подключаемого модуля для научных вычислений NumPy и программного пакета для работы с методами машинного обучения TensorFlow.

Вероятностная нейронная сеть, предназначенная для работы с набором данных по локализации белковых сайтов в дрожжах, имеет 8 нейронов на входном, 600/800/1000/1200/1400 нейронов на первом скрытом слое в соответствии с используемым количеством обучающих примеров и 10 нейронов на втором (выходном) скрытом слое. Вероятностная нейронная сеть, предназначенная для работы с набором данных о локализации белков в *E. coli*, имеет 7 нейронов на входном, 200/210/220/230/240 нейронов на первом скрытом слое в соответствии с используемым количеством обучающих примеров и 8 нейронов на втором скрытом слое. Схема устройства обоих вероятностных нейронных сетей представлена на рисунке 1.

В обоих случаях был сделан выбор в пользу прямой проекции объема имеющихся данных на первый скрытый слой нейронов без дополнительной оптимизации для упрощения алгоритма и скорости обработки ввиду небольшого объема данных и соответственно требующегося объема памяти.

В обоих случаях в качестве передаточной функции используется прямой расчет функции плотности вероятности через суперпозицию Гауссовых ядер в качестве базисных функций. Значение функции плотности вероятности для входного вектора  $x$ , создаваемое одиночным обучающим примером  $x_k$ , принадлежащим к  $i$ -му классу, равно:

$$g_i(x, x_k) = \frac{1}{(\sqrt{2\pi}\sigma)^p} \exp\left(-\frac{\|(x - x_k)\|^2}{2\sigma^2}\right)$$

где  $x$  – входной вектор (исследуемый пример),  $x_k$  – обучающие примеры соответствующего класса,  $p$  – размерность входного вектора и  $\sigma$  – параметр сглаживания. Параметр сглаживания был выбран равным 0.2, поскольку это значение обеспечивает наилучший результат для обеих вероятностных нейронных сетей, установлено это было экспериментально.

При суммировании вкладов всех обучающих примеров, принадлежащих одному классу, получаем значение функции плотности вероятности для данного класса в точке, задаваемой входным вектором  $x$ :

$$g_i(x) = \frac{1}{n(\sqrt{2\pi}\sigma)^p} \sum_{k=1}^n \exp\left(-\frac{\|(x - x_k)\|^2}{2\sigma^2}\right)$$

где  $n$  – число обучающих примеров соответствующего класса,  $x$  – исследуемый пример,  $x_k$  – обучающие примеры соответствующего класса,  $p$  – размерность входного вектора и  $\sigma$  – параметр сглаживания.

Для проведения перекрестной кросс-валидации каждое разбиение на обучающий/тестовый наборы для каждого значения  $\sigma$  было протестировано 10 раз с различными случайными выборками данных. Полученные значения были усреднены, отклонения от среднего составили 1–2 %.

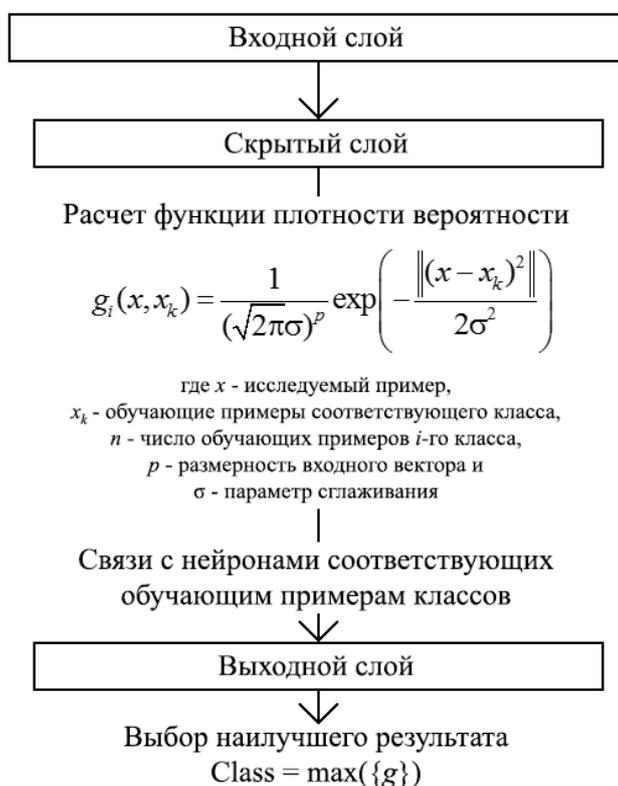
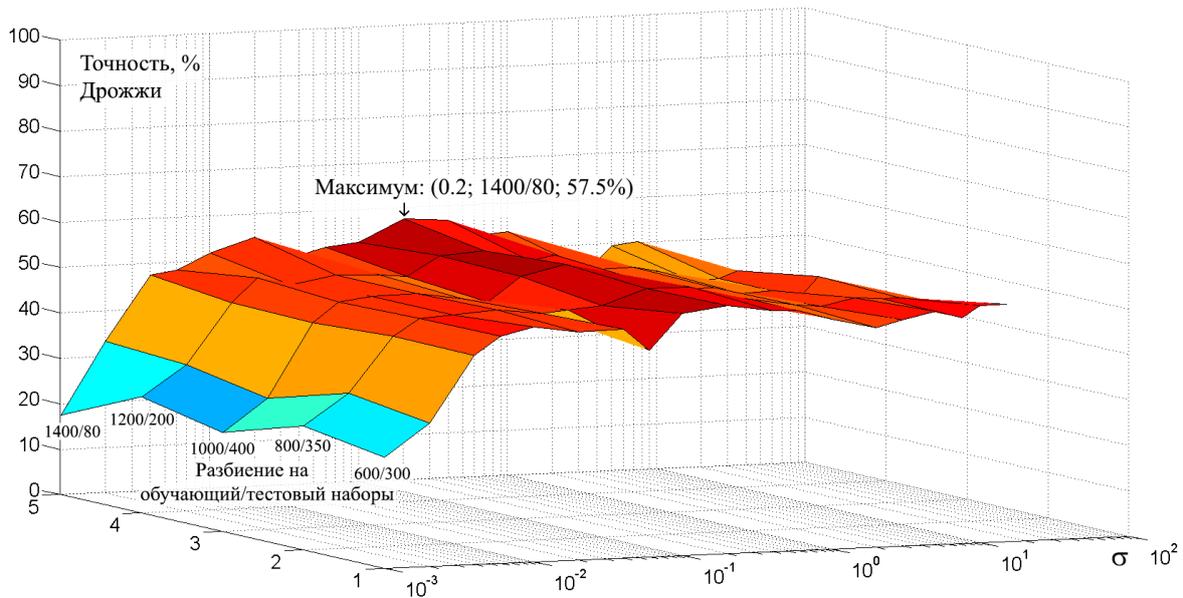


Рис. 1. Схема устройства используемых вероятностных нейронных сетей.

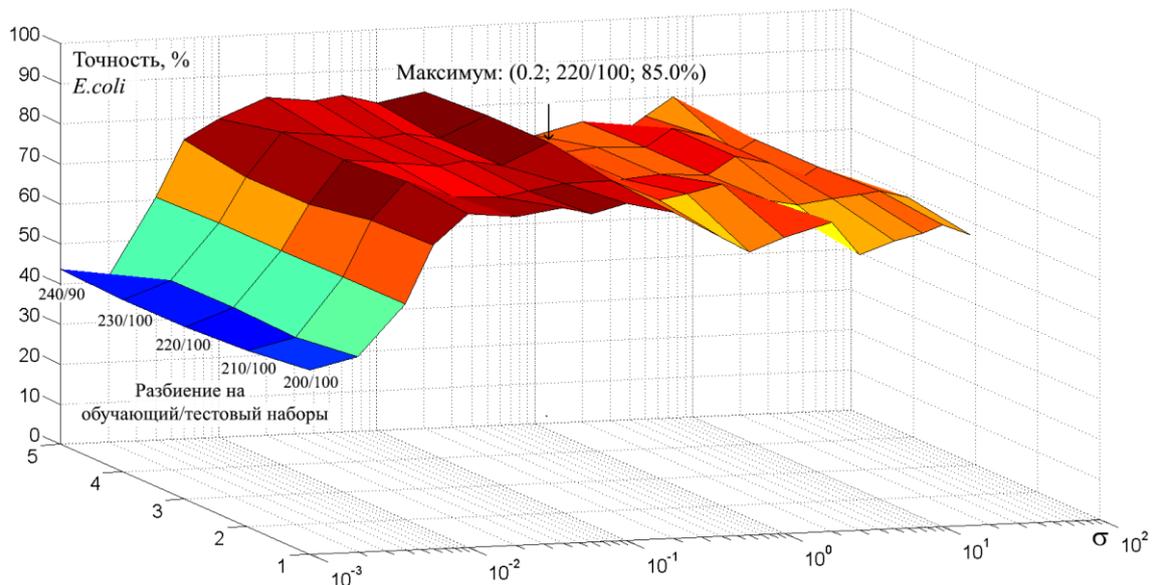
## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

При тестировании вероятностных нейронных сетей получена точность классификации в 57.5 % для задачи о локализации белковых сайтов в дрожжах и 85.0 % для задачи о классификации белковых сайтов в *E. coli*.

Экспериментально были получены зависимости точности классификации от значения параметра сглаживания  $\sigma$  и количества примеров из набора данных, использованных для обучения сети, представленные на рисунках 2 и 3. Из представленных на этих рисунках данных видно, что наибольшей точности классификации сети достигают в схожих диапазонах значения  $\sigma$  в обоих случаях.



**Рис. 2.** Зависимость точности классификации от параметра сглаживания  $\sigma$  и разбиения данных на обучающий/тестовый наборы для набора данных для клеток дрожжей.



**Рис. 3.** Зависимость точности классификации от параметра сглаживания  $\sigma$  и разбиения данных на обучающий/тестовый наборы для набора данных для клеток *E. coli*.

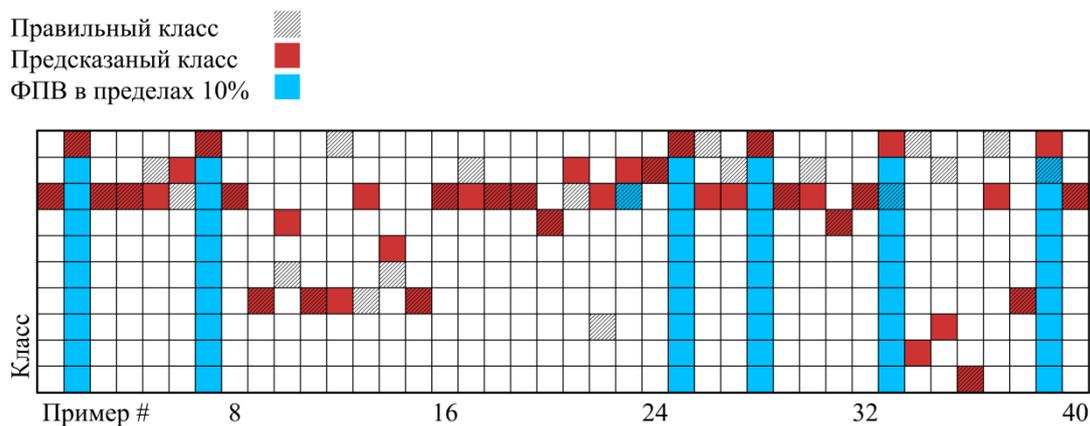


Рис. 4. Классификация случайной выборки примеров из тестового набора данных для клеток дрожжей.

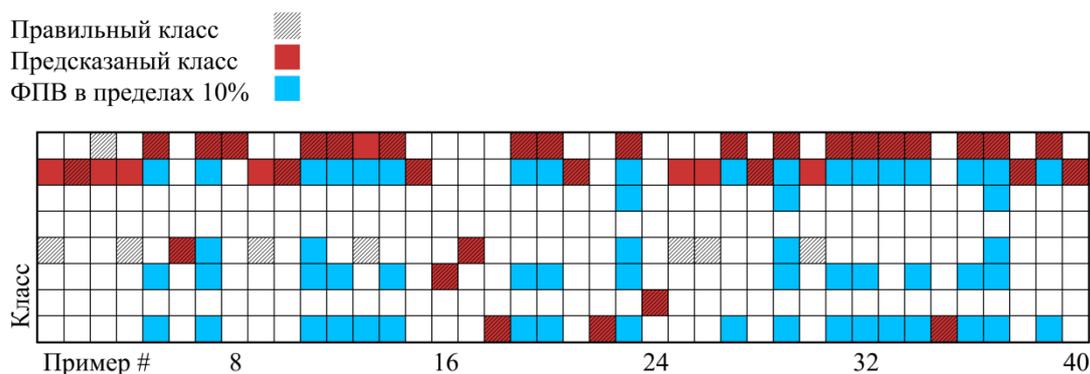


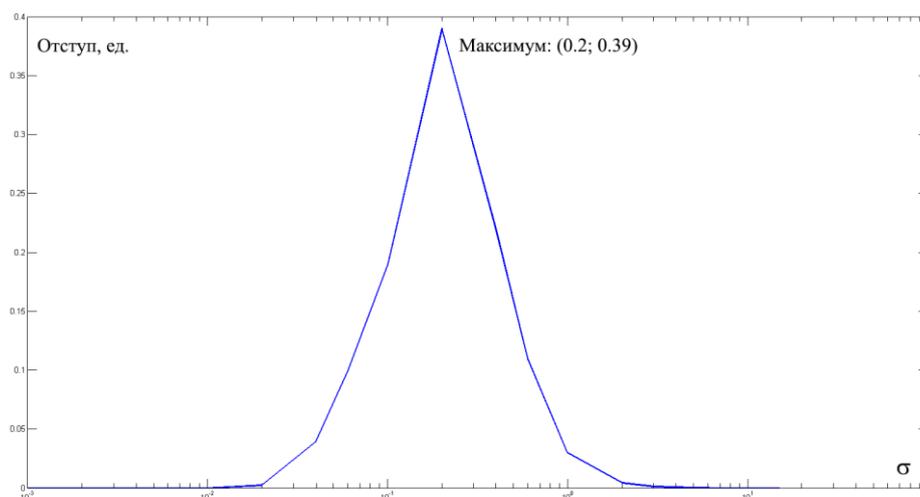
Рис. 5. Классификация случайной выборки примеров из тестового набора данных для клеток *E. coli*.

На рисунках 4 и 5 представлена визуализация классификации для 40 случайным образом выбранных примеров из тестового набора. Значения функции плотности вероятности (ФПВ) в пределах 10 % показаны синим, максимум этих значений – красным, заштрихованный красный квадратик обозначает совпадение предсказания с известным ответом. В 22 случаях для *S. cerevisiae* и в 32 случаях для *E. coli* была предсказана истинная локализация.

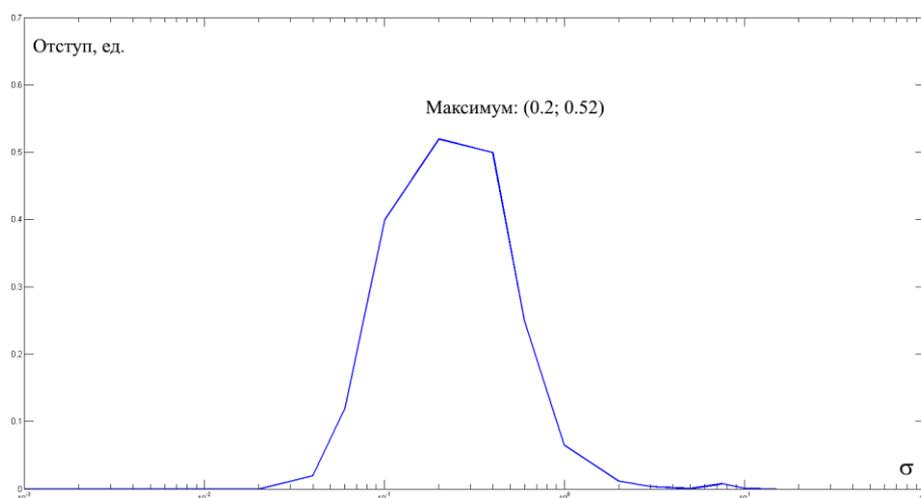
Также были получены зависимости значения отступа как усредненного значения разности ФПВ для некорректно выбранного класса и правильного класса. На рисунках 6 и 7 приведены графики зависимости значения отступа от сглаживающего параметра  $\sigma$  для разбиений данных на обучающий/тестовый набор 1000/400 для дрожжей и 220/100 для клеток *E. coli*. Наибольшее значение отступа характерно для оптимальных значений параметра сглаживания  $\sigma$ , при этом графики зависимости отступа от  $\sigma$  для обоих наборов данных имеют ярко выраженный экстремум в области этих значений, что говорит о надежности классификации.

В рамках данной работы было проведено исследование эффективности предсказания сайтов связывания и накопления белков в выбранных организмах путем обработки данных о характеристиках аминокислотных последовательностей белков при помощи классификаторов на основе вероятностных нейронных сетей. Полученные в результате точности классификации составили 57.5 % для дрожжей и 85.0 % для *E. coli*, что несколько превосходит результаты, показанные в работах [27, 41], использующих другие подходы, так в работе [27] для задачи классификации сайтов в дрожжах был получен результат 55 % и для бактериальных клеток результат 80 %. Более высокая эффективность предложенного метода обусловлена тем, что

вероятностные нейронные сети показывают способность выявлять связи внутри отдельных классов лучше, чем другие методы в похожих задачах [5] [8] [13] и [14].



**Рис. 6.** Зависимость разности значений функции плотности вероятности для правильного класса и предсказанного класса от параметра сглаживания  $\sigma$  для набора данных для клеток дрожжей при оптимальном разбиении данных на обучающий/тестовый наборы.



**Рис. 7.** Зависимость разности значений функции плотности вероятности для правильного класса и предсказанного класса от параметра сглаживания  $\sigma$  для набора данных для клеток *E. coli* при оптимальном разбиении данных на обучающий/тестовый наборы.

Такие результаты достигаются за счет двух основных особенностей вероятностных нейронных сетей и близких к ним по структуре сетей с общей регрессией, а именно большого количества нейронов в скрытом слое, которое близко или равно числу примеров в обучающем наборе, и вариативностью в подборе параметра сглаживания  $\sigma$  [42]. Как и в случае, например, искусственных нейронных сетей глубокого обучения, увеличение объема данных, как правило, позитивно сказывается на эффективности работы. Однако, в случае вероятностных нейронных сетей это ведет к росту обучающего набора и, следовательно, к увеличению количества нейронов в скрытом слое, что в свою очередь начинает негативно сказываться и на скорости обучения, и на скорости работы вероятностной нейронной сети. Поэтому на сегодняшний день такие алгоритмы интересны при объемах баз данных до 10 000 точек, что часто встречается в задачах, вроде той, что рассмотрена в данной статье.

Данная работа является одним из немногих примеров применения вероятностных нейронных сетей для анализа протеомных данных. Как и в работе [5], подтверждается возможность использования такого типа искусственных нейронных сетей для решения некоторых задач по анализу биоинформационных баз данных. Важными преимуществами выбранных алгоритмов является их способность к быстрому обучению и легкость интеграции новых данных, поэтому перспективным подходом может стать использование вероятностных нейронных сетей для решения задач классификации в условиях непрерывно пополняющейся базы данных.

Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках соглашения о предоставлении субсидии № 14.601.21.0017. Уникальный идентификатор прикладных научных исследований (проекта) RFMEFI60117X0017.

## СПИСОК ЛИТЕРАТУРЫ

1. Xie J., Dai G., Zhu F., Shao L., Fang Y. Deep Nonlinear Metric Learning for 3-D Shape Retrieval. *IEEE Transactions on Cybernetics*. 2016. V. 48. № 1. P. 412–422. doi: [10.1109/TCYB.2016.2638924](https://doi.org/10.1109/TCYB.2016.2638924).
2. García-Camacho F., López-Rosales L., Sánchez-Mirón A., Belarbi E.H., Chisti Y., Molina-Grima E. Artificial neural network modeling for predicting the growth of the microalga *Karlodinium veneficum*. *Algal Res.* 2016. V. 14. P. 58–64. doi: [10.1016/j.algal.2016.01.002](https://doi.org/10.1016/j.algal.2016.01.002).
3. Rosenblatt M. Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*. 1956. V. 27. № 3. P. 832–837. doi: [10.1214/aoms/1177728190](https://doi.org/10.1214/aoms/1177728190).
4. Parzen E. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*. 1962. V. 33. № 3. P. 1065–1076. doi: [10.1214/aoms/1177704472](https://doi.org/10.1214/aoms/1177704472).
5. Vipsita S. Protein Superfamily Classification Using Kernel Principal Component Analysis and Probabilistic Neural Networks. In: *Annu. IEEE India Conf.* Hyderabad: IEEE, 2011. doi: [10.1109/INDCON.2011.6139395](https://doi.org/10.1109/INDCON.2011.6139395).
6. Li Z. A Probabilistic Neural Network Approach For BCTTCM Classification. In: *IET International Conference on Information Science and Control Engineering 2012 (ICISCE 2012)*. Shenzhen: IET, 2012. doi: [10.1049/cp.2012.2324](https://doi.org/10.1049/cp.2012.2324).
7. Ghongade R., Deshmukh M., Joshi D. Arrhythmia Classification using Morphological Features and Probabilistic Neural Networks. In: *2014 Innovative Applications of Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH)*. Ghaziabad: IEEE, 2014. doi: [10.1109/CIPECH.2014.7019055](https://doi.org/10.1109/CIPECH.2014.7019055).
8. Sridhar D., Krishna I. M. Brain Tumor Classification Using Discrete Cosine Transform and Probabilistic Neural Network. In: *2013 International Conference on Signal Processing, Image Processing & Pattern Recognition*. Coimbatore: IEEE, 2013, doi: [10.1109/ICSIPR.2013.6497966](https://doi.org/10.1109/ICSIPR.2013.6497966).
9. Jain Y.K., Jain M. Skin Cancer Detection and Classification Using Wavelet Transform and Probabilistic Neural Network. In: *Fourth International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom2012)*. Bangalore: IET, 2012. doi: [10.1049/cp.2012.2540](https://doi.org/10.1049/cp.2012.2540).
10. Vazquez R.A., Condesa C., Condesa C. Classification of EEG Signals using Fractal Dimension Features and Artificial Neural Networks. In: *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*. Honolulu: IEEE, 2017. doi: [10.1109/SSCI.2017.8285230](https://doi.org/10.1109/SSCI.2017.8285230).
11. Novi M.M., Jankovi M.M., Kvaš G.S., Popovi M.B. Classification of forearm movements based on kinematic parameters using artificial neural networks. In: *2017*

- 25th Telecommunication Forum (TELFOR)*. Belgrade: IEEE, 2017. P. 25–28. doi: [10.1109/TELFOR.2017.8249359](https://doi.org/10.1109/TELFOR.2017.8249359).
12. Shimpi P., Shah S., Shroff M., Godbole A. An Artificial Neural Network approach for classification of Vector-Borne diseases. In: *2017 International Conference on Computing Methodologies and Communication (ICCMC)*. Erode: IEEE, 2017. P. 412–415. doi: [10.1109/ICCMC.2017.8282721](https://doi.org/10.1109/ICCMC.2017.8282721).
  13. Bemo E., Brambilla L., Canaparo R., Casale F., Costa M., Dellapepa C., Eandi M., Pasero E. Application of Probabilistic Neural Networks to Population Pharmacokinetics. In: *Proceedings of the International Joint Conference on Neural Networks*. Portland: IEEE, 2003. doi: [10.1109/IJCNN.2003.1223983](https://doi.org/10.1109/IJCNN.2003.1223983).
  14. Guo H., Wang B. Research on TCM Pulse Condition Identification Using Probabilistic Neural Networks. In: *3rd International Conference on Biomedical Engineering and Informatics*. 2010. V. 6. P. 2352–2355. doi: [10.1109/BMEI.2010.5639731](https://doi.org/10.1109/BMEI.2010.5639731).
  15. Li X., Chen J. An Improved Method of Speech Recognition Based on Probabilistic Neural Network Ensembles. In: *2015 11th International Conference on Natural Computation (ICNC)*. Zhangjiajie: IEEE, 2015. P. 650–654. doi: [10.1109/ICNC.2015.7378066](https://doi.org/10.1109/ICNC.2015.7378066).
  16. Ahmad K.S., Thosar A.S., Nirmal J.H., Pande V.S. A Unique Approach in Text Independent Speaker Recognition using MFCC Feature Sets and Probabilistic Neural Network. In: *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)*. Kolkata: IEEE, 2015. doi: [10.1109/ICAPR.2015.7050669](https://doi.org/10.1109/ICAPR.2015.7050669).
  17. Sharma R., Singh A., Joshi A., Gangwar A. Robust Iris Classification through a Combination of Kernel Discriminant Analysis and Parzen Based Probabilistic Neural Networks. In: *2014 Fourth International Conference of Emerging Applications of Information Technology*. Kolkata: IEEE, 2014. P. 268–272. doi: [10.1109/EAIT.2014.25](https://doi.org/10.1109/EAIT.2014.25).
  18. Qiakai N.I., Chao G.U.O., Jing Y. Research of Face Image Recognition Based on Probabilistic Neural Networks. In: *2012 24th Chinese Control and Decision Conference (CCDC)*. Taiyuan: IEEE, 2012. P. 3885–3888. doi: [10.1109/CCDC.2012.6243102](https://doi.org/10.1109/CCDC.2012.6243102).
  19. El-Alfy E.-S. Automatic Identification Based on Hand Geometry and Probabilistic Neural Networks. In: *2012 5th International Conference on New Technologies, Mobility and Security (NTMS)*. Istanbul: IEEE, 2012. P. 1–5. doi: [10.1109/NTMS.2012.6208758](https://doi.org/10.1109/NTMS.2012.6208758).
  20. Jiang X., Lin M., Zhao J. Woodland cover change assessment using decision trees, support vector machines and artificial neural networks classification algorithms. In: *2011 Fourth International Conference on Intelligent Computation Technology and Automation*. Shenzhen: IEEE, 2011. doi: [10.1109/ICICTA.2011.363](https://doi.org/10.1109/ICICTA.2011.363).
  21. Badariah N., Mustafa A., Arumugam K., Ahmed S.K., Sharrif Z.A. Classification of Fruits using Probabilistic Neural Networks - Improvement using Color Features. In: *TENCON 2011 – 2011 IEEE Region 10 Conference*. Bali: IEEE, 2011. P. 264–269. doi: [10.1109/TENCON.2011.6129105](https://doi.org/10.1109/TENCON.2011.6129105).
  22. Lotfi A., Benyettou A. Over-fitting avoidance in probabilistic neural networks. In: *World Congr. Inf. Technol. Comput. Appl.* Hammamet: IEEE, 2015. doi: [10.1109/WCITCA.2015.7367037](https://doi.org/10.1109/WCITCA.2015.7367037).
  23. Kokkinos Y. A Parallel Radial Basis Probabilistic Neural Network for scalable data mining in distributed memory machines. In: *IEEE 24th International Conference on Tools with Artificial Intelligence*. Athens: IEEE, 2012. P. 1094–1099. doi: [10.1109/ICTAI.2012.155](https://doi.org/10.1109/ICTAI.2012.155).
  24. Liu L., Li C., Lipu Q. Optimization of Probabilistic Neural Networks Based on Center Neighbor. In: *International Conference on E-Business and E-Government*. Guangzhou: IEEE, 2010. P. 1441–1444. doi: [10.1109/ICEE.2010.366](https://doi.org/10.1109/ICEE.2010.366).
  25. Kowalski P.A., Kusy M. Determining the significance of features with the use of Sobol method in probabilistic neural network classification tasks. In: *Federated Conference on*

- Computer Science and Information Systems (FedCSIS)*. Prague: IEEE, 2017. V. 11. P. 39–48. doi: [10.15439/2017F225](https://doi.org/10.15439/2017F225).
26. Sprenger J., Lynn Fink J., Karunaratne S., Hanson K., Hamilton NA., Teasdale RD. LOCATE: a mammalian protein subcellular localization database. *Nucleic Acids Res.* 2008. V. 36. P. D230–D233. doi: [10.1093/nar/gkm950](https://doi.org/10.1093/nar/gkm950).
  27. Horton P., Nakai K. A Probabilistic Classification System for Predicting the Cellular Localization Sites of Proteins. In: *International Conference on Intelligent Systems for Molecular Biology*. St. Louis: AAAI Press, 1996. P. 109–115.
  28. Al-mubaid H., Nguyen D.B. New Feature Weighting Technique for Predicting Protein Subcellular Localization. In: *2014 IEEE International Conference on Bioinformatics and Bioengineering*. Boca Raton: IEEE, 2014. doi: [10.1109/BIBE.2014.35](https://doi.org/10.1109/BIBE.2014.35).
  29. Mellman I., Nelson W. J. Coordinated protein sorting, targeting and distribution in polarized cells. *Nature Reviews Molecular Cell Biology*. 2008. V. 11. № 11. P. 833–845. doi: [10.1038/nrm2525](https://doi.org/10.1038/nrm2525).
  30. Razzini G., Brancaccio A., Lemmon M.A., Guarnieri S., Falasca M. The Role of the Pleckstrin Homology Domain in Membrane Targeting and Activation of Phospholipase C $\beta_1$ . *The Journal of Biological Chemistry*. 2000. V. 275. P. 14873–14881. doi: [10.1074/jbc.275.20.14873](https://doi.org/10.1074/jbc.275.20.14873)
  31. Hung M., Link W. Protein localization in disease and therapy. *Journal of Cell Science*. 2011. V. 124. P. 3381–3392. doi: [10.1242/jcs.089110](https://doi.org/10.1242/jcs.089110).
  32. Priya B., Chhabra A. Prediction of protein cellular localization site by using data mining techniques. In: *International Conference on Computing and Network Communications*. Trivandrum: IEEE. 2015, P. 731–736. doi: [10.1109/CoCoNet.2015.7411271](https://doi.org/10.1109/CoCoNet.2015.7411271).
  33. Satu S., Akter T., Uddin J. Performance Analysis of Classifying Localization Sites of Protein using Data Mining Techniques and. In: *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. Cox's Bazar: IEEE, 2017. P. 860–865. doi: [10.1109/ECACE.2017.7913023](https://doi.org/10.1109/ECACE.2017.7913023).
  34. Li W., Wang S., Xu D., Source A.D. Weighted Feature Dimensions According to Fisher's Linear Discriminant Rate and Its Application on Protein Sub-cellular Localization. In: *2015 4th International Conference on Computer Science and Network Technology (ICCSNT)*. Harbin: IEEE, 2015. P. 342–346. doi: [10.1109/ICCSNT.2015.7490765](https://doi.org/10.1109/ICCSNT.2015.7490765).
  35. Shibiao W., Mak M.-W., Sun-Yuan K. mLASSO-Hum: A LASSO-based interpretable human-protein subcellular localization predictor. *J. Theor. Biol.* 2015. V. 382. P. 223–234. doi: [10.1016/j.jtbi.2015.06.042](https://doi.org/10.1016/j.jtbi.2015.06.042).
  36. Hu Y., Li T., Sun J., Tang S., Xiong W., Li D., Chen G., Cong P. Predicting Gram-positive bacterial protein subcellular localization based on localization motifs. *J. Theor. Biol.* 2012. V. 308. P. 135–140. doi: [10.1016/j.jtbi.2012.05.031](https://doi.org/10.1016/j.jtbi.2012.05.031).
  37. Lubec G., Afjehi-sadat L., Yang J., Paul J., John P. Searching for hypothetical proteins: Theory and practice based upon original data and literature. *Prog. Neurobiol.* 2005. V. 77. P. 90–127.
  38. Tung T.Q., Lee D. A method to improve protein subcellular localization prediction by integrating various biological data sources. *BMC Bioinformatics*. 2009. V. 10. P. 1–10. doi: [10.1186/1471-2105-10-S1-S43](https://doi.org/10.1186/1471-2105-10-S1-S43).
  39. Nakai K., Kanehisa M. A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics*. 1992. V. 14. № 4. P. 897–911. doi: [10.1016/S0888-7543\(05\)80111-9](https://doi.org/10.1016/S0888-7543(05)80111-9).
  40. Nakai K., Kanehisa M. Expert system for predicting protein localization sites in gram-negative bacteria. *Proteins*. 1991. V. 11. № 2. P. 95–110. doi: [10.1002/prot.340110203](https://doi.org/10.1002/prot.340110203)

41. Sengur A. Prediction of protein cellular localization sites using a hybrid method based on artificial immune system and fuzzy  $k$ -NN algorithm. *Digital Signal Processing*. 2009. V. 19. № 5. P. 815–826. doi: [10.1016/j.dsp.2009.03.012](https://doi.org/10.1016/j.dsp.2009.03.012).
42. Gotovtsev P.M., Voronov V.N., Smetanin D.S. Analysis of the coolant condition with the help of artificial neural networks. *Therm. Eng.* 2008. V. 55. P. 552–557. doi: [10.1134/S0040601508070045](https://doi.org/10.1134/S0040601508070045)

Рукопись поступила в редакцию 29.11.2018, переработанный вариант поступил 17.05.2019.  
Дата опубликования 23.05.2019.