

ИДЕНТИФИКАЦИЯ ПОПУЛЯЦИЙ КЛЕТОК КРОВИ РЫБ НА ОСНОВЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ СОСТАВЛЕНИЯ ЛЕЙКОГРАММЫ

Г.И. ПРОНИНА, Д.В. БЫКОВ, А.В. УКОЛОВА, А.Е. УЛЬЯНКИН,
А.Н. КАРАСЕВ, М.А. ТУТРИКОВА, М.А. АКИМУШКИНА, К.А. КАНАЕВА

(Российский государственный аграрный университет – МСХА имени К.А. Тимирязева)

В аквакультуре требуется физиологическая оценка для контроля состояния здоровья рыб. Кровь является наиболее быстро реагирующей на изменения внешних факторов системой в организме гидробионтов. Изучение гематологических показателей рыб позволяет проводить раннюю диагностику заболеваний, обрабатывать технологический режим разведения и выращивания, селекцию. Типизация клеток циркулирующих жидкостей является важной для составления гемоцитарной и лейкоцитарной формул, характеризующих клеточное звено иммунного ответа организма.

В представленных исследованиях разрабатываются модели сверточной нейронной сети для классификации клеток крови карповых и осетровых рыб. Точность моделей оценивается на основе метрик Accuracy и Precision, Recall, F_1 при макроусреднении.

На основе обработки снимков крови подготовлено 1104 изображения клеток крови карповых и осетровых рыб, включающие в себя 15 популяций клеток: гемогистобласты, миелобласты, промиелоциты, миелоциты, метамиелоциты, палочкоядерные нейтрофилы, сегментоядерные нейтрофилы, эозинофилы, базофилы, моноциты, лимфоциты, эритробласты, нормобласты, зрелые эритроциты, тромбоциты.

Разработаны модели сверточной нейронной сети для распознавания популяций клеточных элементов крови (эритроцитов, лейкоцитов, тромбоцитов) карповых и осетровых рыб. Обучение моделей происходило на 80% подготовленных изображений. При этом удалось избежать проблемы переобучения, о чем свидетельствуют построенные графики изменения значений функции потерь (разреженной категориальной перекрестной энтропии – sparse categorical crossentropy) и точности (accuracy) в процессе обучения.

Построенные модели позволяют распознавать клетки крови карповых рыб с точностью 75,0% (метрика F_1 при макроусреднении равна 0,570) и клетки крови осетровых рыб с точностью 76,6% (F_1 при макроусреднении составляет 0,664).

Ключевые слова: клетки крови рыб, машинное обучение, сверточная нейронная сеть (CNN), классификация, обработка изображений.

Введение

В условиях искусственного разведения рыб важно проводить мониторинг их физиологического состояния и иммунного статуса. Это необходимо для профилактики и ранней диагностики заболеваний, а также для разработки и совершенствования технологий разведения и выращивания, так как среда обитания аквакультуры значительно отличается от естественной.

Кровь – циркулирующая жидкость организма, отражающая гомеостаз. Определен ряд иммунных и метаболических переменных крови животных для оценки и прогнозирования темпов роста [13].

Определение гематологических параметров является биоиндикаторным методом при загрязнении водоемов [2, 8]. Исследования на карпах показали, что оценка селекционного процесса по показателям крови является крайне необходимой [4].

Исследование лейкоцитарной формулы имеет большое значение в диагностике гематологических, инфекционных, воспалительных заболеваний, а также для оценки их тяжести [1]. Определение клеточного состава крови является одним из компонентов характеристики иммунного статуса особи [3].

Составление эритрограммы и лейкоцитарной формулы – довольно трудоемкий процесс, поэтому распознавание клеточных популяций лейкоцитов крови рыб с использованием технологий нейронных сетей является актуальной задачей.

Идентификация клеток крови относится к задаче мультiclassовой классификации с несбалансированными классами. Подобную задачу в различных областях решают многие исследователи [10, 20, 15, 16].

Поскольку необходимо классифицировать клетки в виде изображений, то одновременно решается вторая задача, связанная с распознаванием образов.

Обе обозначенные задачи с высокой эффективностью способны решать искусственные нейронные сети. Одним из наиболее эффективных и широко применяемых инструментов классификации изображений является сверточная нейронная сеть (Convolutional Neural Network, CNN) [12].

Цель исследований: создание модели сверточной нейронной сети, способной распознавать популяции лейкоцитов для их дальнейшего дифференциального подсчета.

Для достижения цели необходимо решить следующие задачи:

- подготовить изображения популяций клеток крови карповых и осетровых рыб;
- разработать и обучить модели сверточной нейронной сети;
- оценить точность распознавания популяций клеток крови карповых и осетровых рыб на основе разработанных моделей.

Объектом исследований являлись популяции клеток крови карповых и осетровых рыб, предметом – модели нейронных сетей для распознавания изображений клеток крови карповых и осетровых рыб.

В работе использованы методы нейросетевого моделирования, оценки качества обученной модели, табличный, графический методы и др.

Информационной базой исследований послужили 309 снимков окрашенных мазков крови карпа и 335 аналогичных снимков крови осетровых рыб, полученные на цифровом микроскопе Биолаб 11 Люм.

Материал и методы исследований

Снимки окрашенных мазков крови карповых и осетровых рыб были получены на цифровом микроскопе Биолаб 11 Люм. Кровь у рыб отбиралась прижизненно из хвостовой вены (рис. 1).

Состав лейкоцитов и долю незрелых форм клеток определяли в окрашенных по Паппенгейму мазках периферической крови [9] на цифровом микроскопе Биолаб Люм 11 (Россия), увеличение 100 – методом дифференциального подсчета.

При определении лейкоцитарной формулы крови рыб происходит подсчет каждой из 15 популяций лейкоцитов крови. Однако в связи с отсутствием изображений гемогистобласта на имеющихся снимках осетровых рыб число возможных классов при классификации их клеток крови составляет 14.

Список исследуемых типов лейкоцитов крови рыб представлен в таблице 1.

Таким образом, в исследованиях решалась задача мультiclassовой классификации с числом классов 15 (модель 1 для распознавания клеток крови карпа) и 14 (модель 2 для распознавания клеток крови осетра).

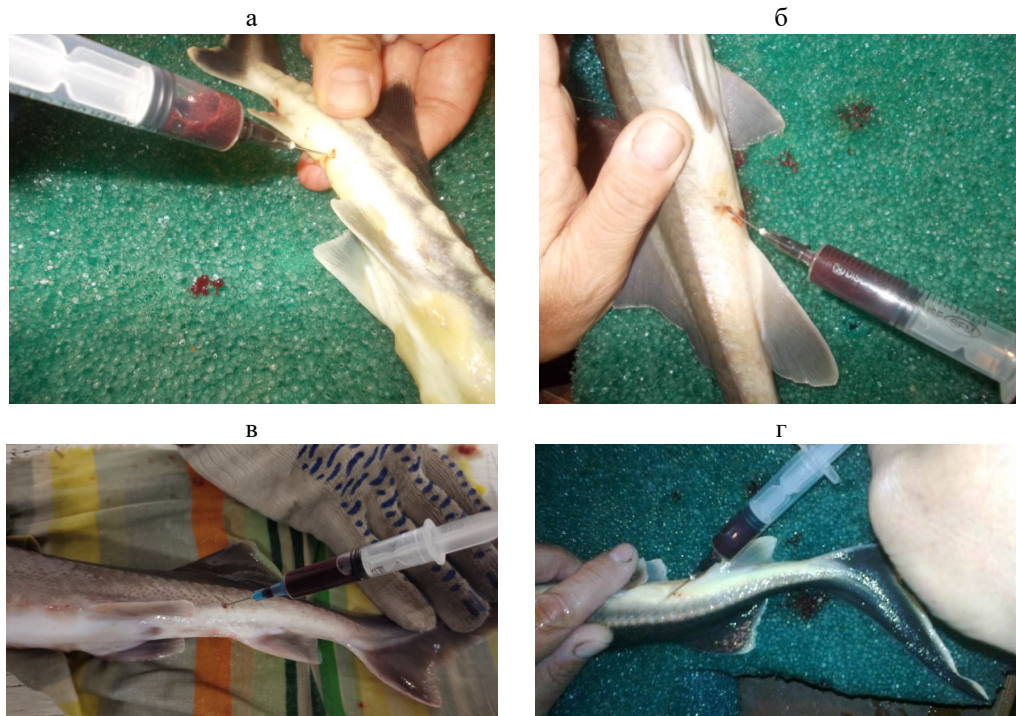


Рис. 1. Отбор крови у рыб:
а – русский осетр; б – белуга; в – сибирский (Ленский) осетр; г – стерлядь

Таблица 1

Популяции форменных элементов крови рыб и соответствующие им метки классов

Метка	Популяция	Метка	Популяция	Метка	Популяция
1	гемогистобласты	6	палочкоядерные нейтрофилы	11.	лимфоциты
2	миелобласты	7	сегментоядерные нейтрофилы	12	эритробласты
3	промиелоциты	8	эозинофилы	13	нормобласты
4	миелоциты	9	базофилы	14	зрелые эритроциты
5	метамиелоциты	10	моноциты	15	тромбоциты

Для возможности обучения нейронной сети распознавать популяции лейкоцитов крови карповых и осетровых рыб необходимо подготовить изображения каждой популяции и расположить их в соответствующие директории, причем число директорий должно соответствовать числу популяций.

На основе имеющихся снимков клеток крови карповых и осетровых рыб, каждый из которых содержит множество разных типов клеток, были сформированы изображения отдельных клеток, которые относятся к определенной популяции. Всего было подготовлено 464 таких изображения для карповых и 640 – для осетровых рыб. После этого они были распределены по 15 и 14 директориям соответственно, в зависимости от типа клетки на изображении.

Распределение подготовленных изображений клеток крови является неравномерным (рис. 2–4), что говорит о необходимости использования, помимо метрики *Accuracy*, более комплексных метрик мультиклассовой классификации, в том числе *Precision*, *Recall*, F_1 , которые позволяют учесть влияние малочисленных классов на итоговую точность распознавания клеток.

Меньше всего изображений в наборах данных (после гемогистобластов) оказалось метамиелоцитов: 10 ед. по карповым и 3 ед. по осетровым. По карповым в целом распределение клеток по популяциям более равномерное, чем по осетровым. Больше всего имеется изображений зрелого эритроцита (45 или 9,7%), эритробласта (63 или 13,6%), сегментоядерного нейтрофила (67 или 14,4%). По осетровым явно выделяются 3 наиболее многочисленных класса: изображений лимфоцитов насчитывается 88 (13,8%), тромбоцитов – 110 (17,2%), зрелых эритроцитов – 192 (30,0%).

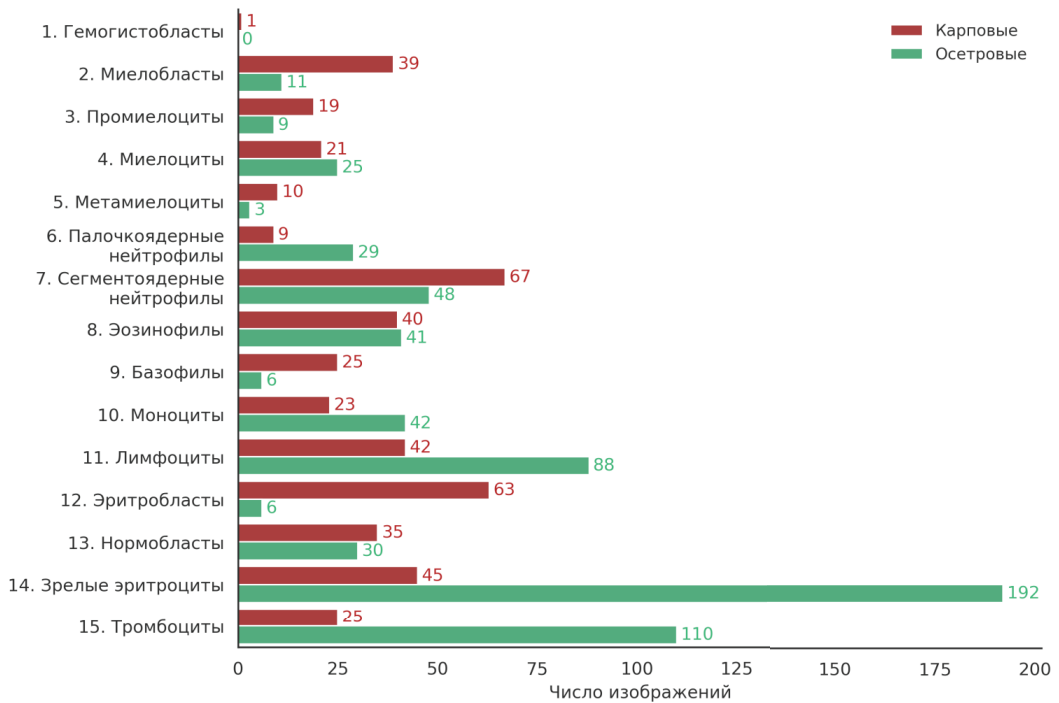


Рис. 2. Распределение изображений клеток крови карповых и осетровых рыб по популяциям

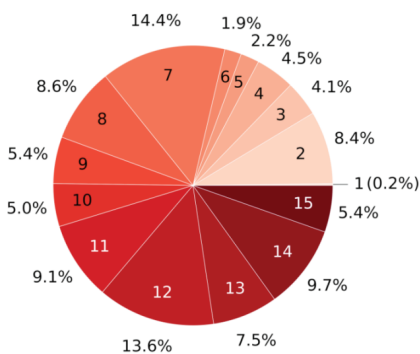


Рис. 3. Структура изображений клеток крови карповых рыб по 15 популяциям (1–15)

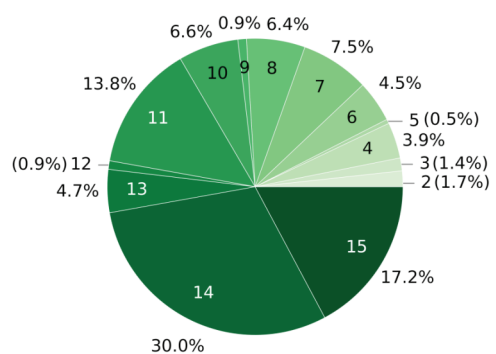


Рис. 4. Структура изображений клеток крови осетровых рыб по 14 популяциям (2–15)

В лейкоцитарной формуле рыб распределение популяций клеток другое. Кровь рыб, как правило, – лимфоцитарного типа. Доля лимфоцитов в лейкограмме рыб, в том числе осетровых рыб (сибирского осетра, терляди и др.), – более 65%, карпа – более 70% [8]. Лимфоциты являются иммунокомпетентными клетками специфического иммунитета.

Зрелые формы микрофагов – сегментоядерные нейтрофилы – обычно составляют около 1–20% лейкоцитов, что соответствует их удельному весу в полученных наборах изображений. Эти клетки играют ключевую роль во врожденной неспецифической клеточной защите.

В отличие от млекопитающих у рыб отсутствует красный костный мозг, и роль гемопоэза выполняют другие органы и ткани. В результате в норме в крови содержится определенное количество бластных форм: эритробласты, нормобласты, миелобласты, промиелоциты, миелоциты, метамиелоциты [3].

Некоторые случайно выбранные изображения, импортированные в среду разработки Python с подписанным номером соответствующей популяции клеток крови (фактической меткой класса), представлены на рисунке 5.

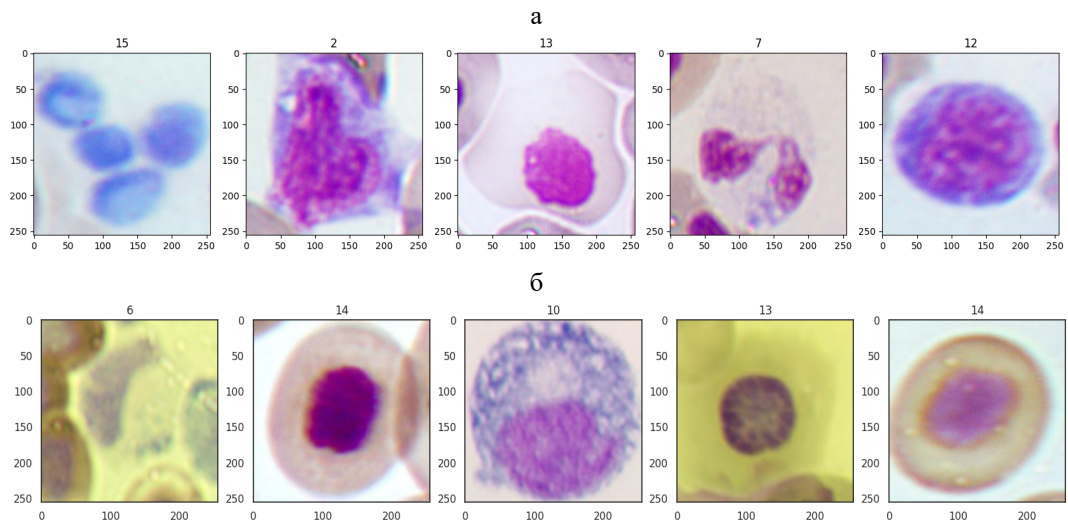


Рис. 5. Изображения клеток крови рыб и соответствующие им фактические метки классов: а – карповые рыбы; б – осетровые рыбы

Для разработки модели нейронной сети использовались язык программирования Python и специализированная библиотека TensorFlow.

Разработанная модель нейронной сети является сверточной. Сверточная нейронная сеть имеет как минимум один сверточный слой, на котором реализуется операция свертки, представляющая собой кодирование исходного изображения таким образом, чтобы выявить основные черты, особенности этого изображения, формируя в результате карты признаков. Число карт признаков зависит от числа каналов на выходе сверточного слоя, каждому из которых соответствует свой фильтр.

Операция свертки предполагает использование матрицы весов, называемой ядром свертки, для преобразования входного сигнала (входного пикселя) в выходной сигнал. При этом для получения выходного сигнала рассматривается лишь фрагмент изображения – пиксель и соседние пиксели. В результате преобразования каждого пикселя формируется карта признаков [5].

Для решения проблемы лишнего сжатия и потери информации о некоторых особенностях изображения используют операцию, предполагающую добавление

нулей по краям изображения, то есть 4 линии пикселей (*padding*). Например, если изображение имеет размер 10×10 пикселей, в результате операции *padding* его размер составляет 12×12 .

Построенные модели нейронной сети состоят из 5 слоев и отличаются лишь числом нейронов на последнем выходном слое:

- *Conv2D* – двумерный сверточный слой с размерностью выходного пространства (количество выходных фильтров в свертке), равной 16; размерностью окна свертки (ядра свертки) 3×3 ; шагом свертки по высоте и ширине 1; равномерным заполнением нулями по краям изображения при реализации операции *padding*; функцией активации *ReLU*, наиболее часто используемой в сверточных слоях.

- *MaxPool2D* – слой подвыборки (*pooling layer*), использующийся для уменьшения размерности карт признаков и предотвращения переобучения [5].

- *Flatten* – слой, представляющий собой результат преобразования многомерного массива сигналов нейронов предыдущего слоя в одномерный массив.

- *Dense* – полносвязный слой из 256 нейронов.

- *Dense* – полносвязный слой из k -нейронов, где k – количество имеющихся классов. Для модели 1, обрабатывающей изображения клеток крови карпа, $k = 15$; для модели 2, связанной с изображениями клеток крови осетра, $k = 14$.

В качестве функции ошибки была выбрана функция разреженной категориальной перекрестной энтропии (*sparse categorical crossentropy*), функции активации последнего слоя – многопеременная логистическая функция (*softmax*). Именно они используются при построении моделей нейронных сетей для решения задачи мультиклассовой классификации в случаях, когда каждый классифицируемый объект может быть отнесен лишь к одному классу [5].

Каждый из наборов изображений был подразделен на 3 выборки: обучающая выборка (80% изображений); валидационная выборка (10% изображений); тестовая выборка (10% изображений).

Процесс обучения регулировался процедурой *early stopping*, заключающейся в преждевременной остановке в случае прекращения существенного уменьшения функции потерь и увеличения метрики [19].

Помимо *Accuracy*, измерить точность бинарной классификации можно при помощи метрик *Precision*, *Recall*, F_1 . В их основе лежит матрица ошибок (*confusion matrix*) [17], отраженная в таблице 2. Она показывает долю правильных классификаций или долю объектов, для которых классификатор правильно предсказал класс. Как отмечено выше, не всегда может эффективно использоваться как показатель качества распознавания для модели при наличии малочисленных классов. В этом случае часто используются метрики *Precision*, *Recall*, F_1 [14].

Таблица 2

Матрица ошибок при бинарной классификации объектов

Класс	Объект отнесен к положительному классу (P)	Объект отнесен к отрицательному классу (N)
Положительный (P)	<i>TP</i> – правильно положительная классификация объекта	<i>FN</i> – ложно отрицательная классификация объекта
Отрицательный (N)	<i>FP</i> – ложно положительная классификация объекта	<i>TN</i> – правильно отрицательная классификация объекта

Метрика *Accuracy* может быть рассчитана по формуле:

$$Accuracy = \frac{TP + TN}{N} = \frac{TP + TN}{TP + FN + FP + TN}. \quad (1)$$

Precision – доля истинно положительных объектов среди объектов, отнесенных классификатором к положительному классу:

$$Precision = \frac{TP}{TP + FP}, \quad (2)$$

где $TP + FP$ – объекты, распознанные классификатором как положительные, образующие положительный класс, который сформировал классификатор.

Метрика отражает качество распознавания положительных объектов, не учитывая размер истинно положительного класса. Например, если объектов всего 20, положительный класс содержит 15 объектов, 10 истинно положительных объектов были отнесены к этому классу и 0 – истинно ложных, то $Precision = 1$.

Recall – доля истинно положительных объектов среди объектов положительного класса:

$$Recall = \frac{TP}{TP + FN}, \quad (3)$$

где $TP + FN$ – объекты, являющиеся истинно положительными, образующие истинный положительный класс.

Метрика отражает качество распознавания положительных объектов, учитывая размер истинно положительного класса. Например, если объектов всего 20, положительный класс содержит 15 объектов, 15 истинно положительных объектов были отнесены к этому классу, то $Recall = 1$.

F_1 – среднее гармоническое метрик *Precision* и *Recall*:

$$F_1 = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} = 2 \cdot \frac{Recall \cdot Precision}{Recall + Precision}. \quad (4)$$

Метрика отражает связь между истинно положительными метками и предсказанными классификатором как положительные [11, 17].

Мультиклассовая классификация с числом классов k представляется как совокупность из k -бинарных классификаций. В таком случае указанные метрики рассчитываются для каждой бинарной классификации, после чего реализуется одна из процедур их усреднения: микроусреднение или макроусреднение [17, 18].

Микроусреднение (*Micro-averaging*) предполагает усреднение матриц ошибок, полученных по k -бинарным классификациям, и последующий расчет метрик *Precision*, *Recall*, F_1 . При микроусреднении крупные классы сильнее влияют на значение метрики, тогда как вклад малых классов существенно ниже. Поэтому при несбалансированности классов предпочтение отдают макроусреднению, в котором и крупные, и малые классы в одинаковой степени влияют на итоговую метрику.

Макроусреднение (*Macro-averaging*) предполагает усреднение метрик *Precision* и *Recall*, полученных по k -бинарным классификациям. Для расчета $F_{1_{macro}}$ существуют

два подхода. Первый подход идентичен расчету $Precision_{macro}$ и $Recall_{macro}$: F_1 и рассчитывается для каждой бинарной классификации, затем из них находится среднее арифметическое:

$$F_{1_{macro}} = \frac{\sum_{i=1}^k F_{1_i}}{k}, \quad (5)$$

где k – число бинарных классификаций.

При втором подходе расчет производится по формуле (4) на основе [17]:

$$F_{1_{macro}} = 2 \cdot \frac{Recall_{macro} \cdot Precision_{macro}}{Recall_{macro} + Precision_{macro}}. \quad (6)$$

Библиотека scikit-learn позволяет реализовать макроусреднение метрики F_1 по формуле (5).

Результаты и их обсуждение

Модель 1 обучалась в течение 17 эпох обучения (рис. 6а, б). В конце первой эпохи обучения значение функции ошибки ($Loss$) для обучающей и валидационной выборок составило 22546,11 и 10611,38 соответственно, точность $Accuracy$ – 0,12 и 0,06. В конце обучения значение $Loss$ уменьшилось до 2,03 и 1,59, итоговая точность составила 0,85 и 0,89 по обучающему и валидационному набору данных соответственно.

Модель 2 обучилась быстрее: ей потребовалось 10 эпох обучения. Значения функции ошибки по двум выборкам к 10 эпохе уменьшились с 15866,42 и 6659,21 до 13,13 и 14,47 соответственно, при этом точность возросла с 0,23 и 0,19 до 0,71 и 0,75.

На основе графиков значений точности и ошибки можно констатировать отсутствие переобучения моделей, так как в целом точность по валидационной выборке превышает точность по обучающей выборке. Одновременно с этим ошибка по валидационной выборке не превышает ошибку по тренировочной выборке, что говорит о достаточно высокой способности моделей распознавать те изображения клеток крови, с которыми данные модели не сталкивались в процессе обучения.

Значения основных метрик, характеризующих качество построенных моделей с точки зрения классификации изображений клеток крови по обучающему (train), валидационному (validation) и тестовому (test) наборам данных, представлены в таблице 3.

Точность $Accuracy$ модели 1 по обучающей выборке составляет 0,85, по валидационной – 0,89, по тестовой – 0,75. Значения $Accuracy$ по модели 2 отличаются незначительно: по обучающей и валидационной выборкам они составляют 0,71 и 0,75 соответственно, однако по тестовой выборке точность выше и достигает 0,77.

Метрики $Precision$, $Recall$, F_1 при микроусреднении по обеим моделям идентичны $Accuracy$, что объясняется природой данных метрик. Несбалансированность наборов данных учтена в метриках $Precision$, $Recall$, F_1 при макроусреднении, о чем свидетельствуют их значения, заметно уступающие $Accuracy$, но более точно отражающие истинное качество моделей. Так, по модели 1 значение метрики $Precision_{macro}$ составляет 0,574, а по модели 2 – 0,707, то есть средний удельный вес правильно предсказанных клеток определенного типа среди всех предсказанных данного типа составляет 57,4 и 70,7% соответственно. Значения метрики $Recall_{macro}$ отличаются в меньшей степени и составляют 0,611 и 0,661, то есть средний удельный вес правильно предсказанных клеток среди всех имеющихся в наборе клеток данного типа составляет 61,1% по модели 1 и 66,1% по модели 2. Метрика $F_{1_{macro}}$ по двум моделям составляет 0,570 и 0,664.

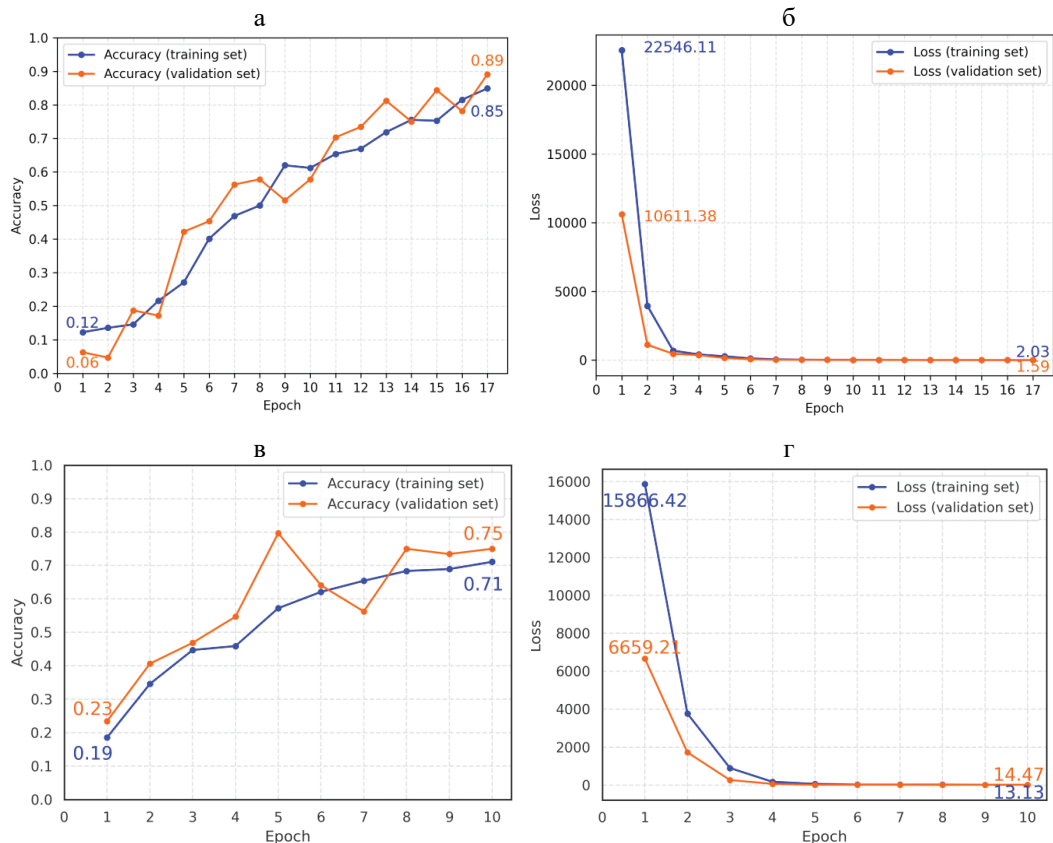


Рис. 6. Процесс обучения моделей сверточной нейронной сети:
 а, б – графики изменения точности *Accuracy* и функции ошибки *Loss* соответственно на каждой эпохе обучения модели 1 (по обучающей выборке – training set, по валидационной выборке – validation set); в, г – графики *Accuracy* и *Loss* в процессе обучения модели 2

Таблица 3

Метрики построенных моделей сверточной нейронной сети

Метрика	Модель 1			Модель 2		
	train	validation	test	train	validation	test
<i>Accuracy</i>	0,846	0,891	0,750	0,711	0,750	0,766
<i>Precision_{micro}</i>	0,846	0,891	0,750	0,711	0,750	0,766
<i>Precision_{macro}</i>	0,850	0,827	0,574	0,583	0,615	0,707
<i>Recall_{micro}</i>	0,846	0,891	0,750	0,711	0,750	0,766
<i>Recall_{macro}</i>	0,837	0,830	0,611	0,554	0,558	0,661
$F_{1_{micro}}$	0,846	0,891	0,750	0,711	0,750	0,766
$F_{1_{macro}}$	0,820	0,811	0,570	0,539	0,563	0,664

Модели способны распознавать клетки крови карповых и осетровых рыб при качестве выше среднего. При этом модель 2 на 16,49% лучше модели 1 распознает изображения клеток при сравнении по метрике $F_{1_{macro}}$.

На рисунке 7 представлена визуализация применения модели 1 для распознавания клеток крови карповых рыб: из 12 случайных изображений клеток крови осетровых рыб тестового набора модель правильно распознала 11 клеток. При этом в данной выборке больше всего зрелых эритроцитов (метка 14) – 4 клетки, и все их модель распознала без ошибки.

Вероятность соответствия (доля наибольшего сигнала на последнем слое нейронной сети по отношению к сумме k сигналов) прогнозных меток почти для всех правильно распознанных клеток близка к 100%, за исключением сегментоядерного нейтрофила (метка 7), для которого данная вероятность составляет 40,69%. Для неправильно распознанного сегментоядерного нейтрофила вероятность соответствия далека от максимума и равна 65,29%.

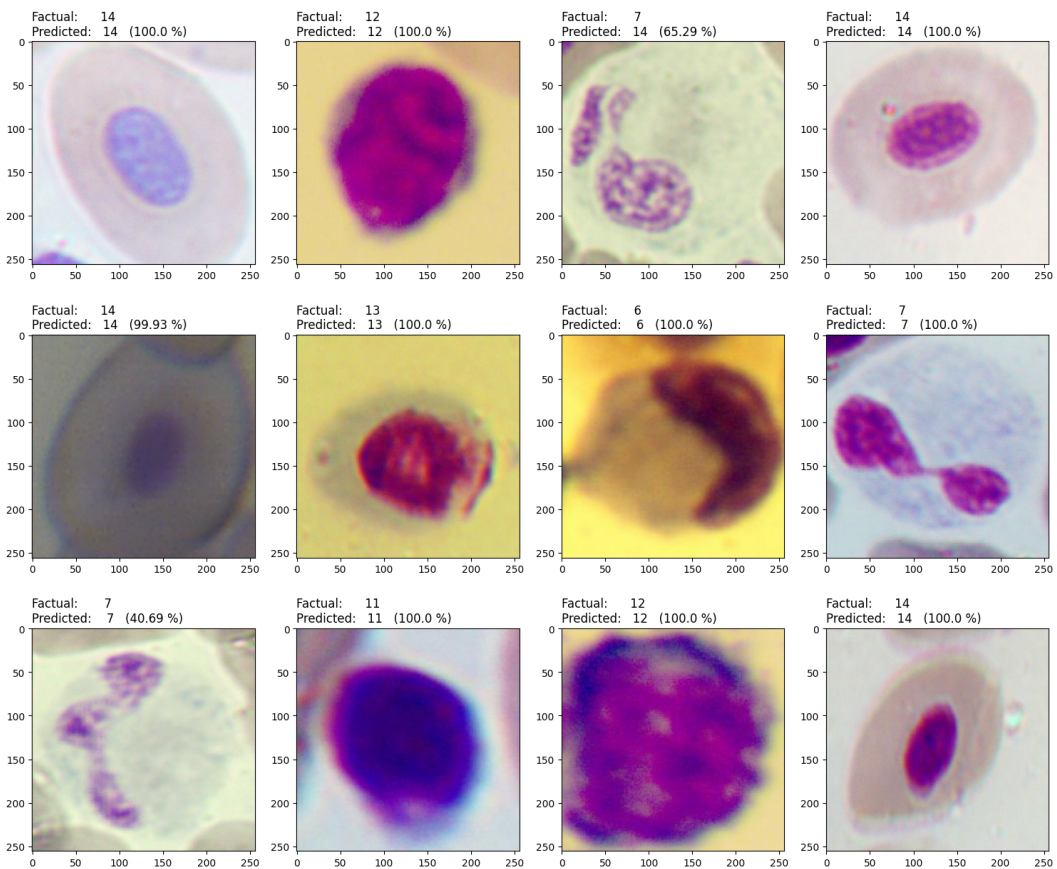


Рис. 7. Некоторые изображения клеток крови карповых рыб с соответствующими им фактическими метками классов (Factual) и спрогнозированными моделью 1 (Predicted)

На рисунке 8 отражены результаты применения модели 2 для распознавания клеток крови осетровых рыб. Моделью 2 также правильно были распознаны 11 клеток, большинство из которых (6 клеток) относится к классу зрелых эритроцитов (метка 14). При этом для 9 правильно распознанных клеток вероятность соответствия достигла 100%, за исключением одного зрелого эритроцита, для которого вероятность

соответствия была достаточно высокой и составила 87,16%. В распознавании данной выборки недостаток модели 2 связан с высокой вероятностью соответствия для неправильно распознанной клетки – палочкоядерного нейтрофила, и она составила 95,25%.

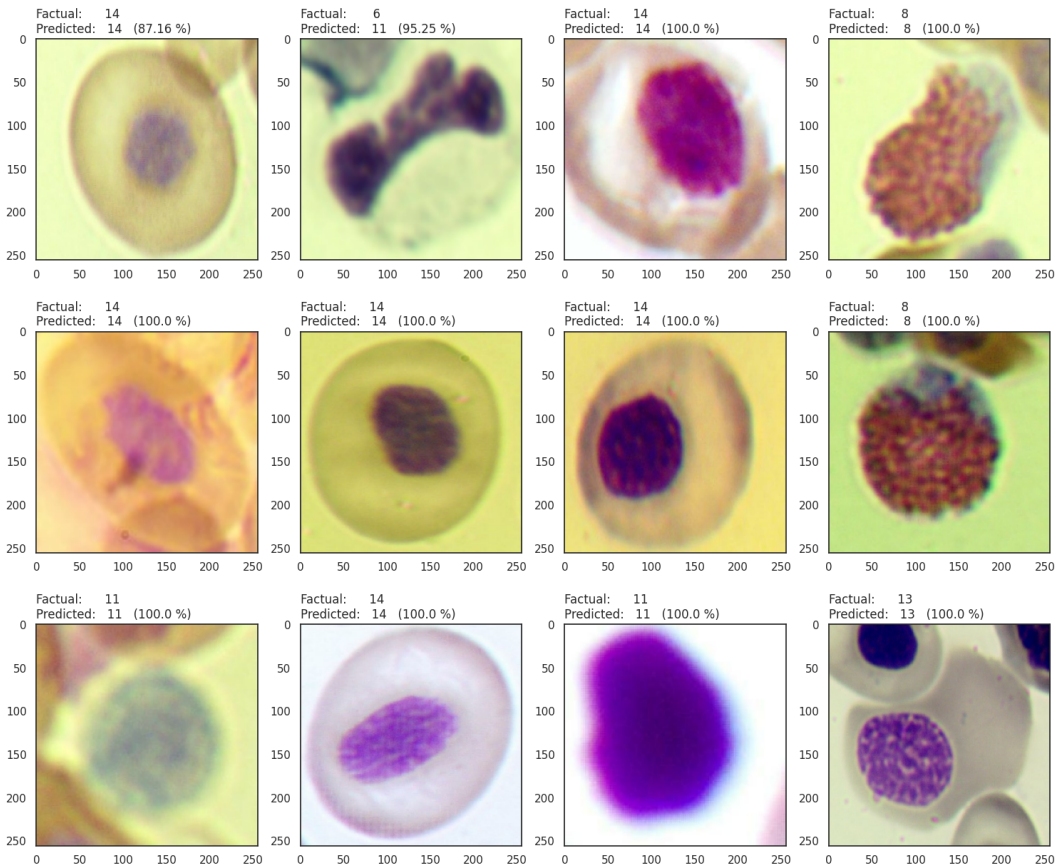


Рис. 8. Некоторые изображения клеток крови осетровых рыб с соответствующими им фактическими порядковыми номерами популяций (Factual) и спрогнозированными моделью 2 (Predicted)

Разработанные модели будут положены в основу дальнейших исследований, связанных с определением лейкоцитарной формулы крови рыб на основе технологий нейронных сетей. Поскольку точность обученных моделей составила 75,0 и 76,6%, значения метрики при макроусреднении – 0,570 и 0,664, в перспективе следует учесть размеры клеток и другие параметры, а также реализовать дополнительные процедуры машинного обучения для получения более точных результатов распознавания, в том числе аугментацию наборов данных (*data augmentation*) или процедуру вращения изображений [12].

Выводы

В результате исследований средствами языка Python и библиотеки TensorFlow разработаны модели сверточной нейронной сети для распознавания популяций лейкоцитов крови карповых и осетровых рыб. Обучение моделей происходило на 80% подготовленных изображений. При этом удалось избежать проблемы переобучения,

о чем свидетельствуют построенные графики изменения значений функции потерь (разреженной категориальной перекрестной энтропии – *sparse categorical crossentropy*) и точности (*accuracy*) в процессе обучения.

Построенные модели позволяют распознавать клетки крови рыб для автоматизации определения лейкоцитарной формулы крови со средней точностью 75,8% и средним значением метрики F_1 при макроусреднении 0,617. Повысить качество распознавания клеток крови можно путем модификации моделей и наборов данных, усложнения процесса обучения, увеличения числа и качества изображений клеток крови, учета размеров клеток, что требует продолжения исследований по данной тематике.

Библиографический список

1. Долгов В.В., Меньшиков В.В. Клиническая лабораторная диагностика // Национальное руководство. – М.: ГЭОТАР-Медиа, 2016. – 688 с.
2. Житенева Л.Д., Макаров Э.В., Рудницкая О.А. Основы ихтиогематологии (в сравнительном аспекте): М. – Ростов-на-Дону: Эверест, 2004. – 312 с.
3. Иванов А.А., Пронина Г.А., Корягина Н.Ю. Гематология пойкилотермных гидробионтов: Монография. – Иркутск: ООО «Мегапринт», 2018. – 133 с.
4. Маслова Н.И., Петрушин А.Б. Методические рекомендации по управлению селекционным процессом в рыбоводстве (на примере карпа). – М.: Россельхозакадемия, 2005. – 27 с.
5. Паттерсон Дж., Гибсон А. Глубокое обучение с точки зрения практика. – М.: ДМК Пресс, 2018. – 418 с.
6. Пищенко Е.В. Гематология пресноводной рыбы: Учебное пособие. – Новосибирск: Новосибирский государственный аграрный университет, 2002. – 48 с.
7. Пронина Г.И., Иванов А.А., Маннапов А.Г., Саная О.В. Иммунитет пойкилотермных гидробионтов // Известия ТСХА. – 2021. – № 2. – С. 71–91.
8. Пронина Г.И. Клиническая лабораторная диагностика. Практикум: Учебное пособие для вузов. – М., 2021. – 88 с.
9. Пронина Г.И., Корягина Н.Ю. Методология физиолого-иммунологической оценки гидробионтов: Учебное пособие. – СПб.: Лань, 2017. – 96 с.
10. Agustin R.I., Arif A., Sukorini, U. Classification of immature white blood cells in acute lymphoblastic leukemia L1 using neural networks particle swarm optimization // Neural Comput & Applic. – 2021. – № 33. – Pp. 10869–10880. <https://doi.org/10.1007/s00521-021-06245-7>
11. Fernández J.C., Carbonero M., Gutiérrez P.A. et al. Multi-objective evolutionary optimization using the relationship between F1 and accuracy metrics in classification tasks // Appl Intell. – 2019. – № 49. – С. 3447–3463. <https://doi.org/10.1007/s10489-019-01447-y>
12. Hameed Z., Garcia-Zapirain B., Aguirre J.J. et al. Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network // Sci Rep. – 2022. – № 12. – P. 15600. <https://doi.org/10.1038/s41598-022-19278-2>
13. Le Floc'h N., Gondret F., Resmond R. Identification of blood immune and metabolic indicators explaining the variability of growth of pigs under contrasted sanitary conditions // BMC Vet Res. – 2021. – № 17. – P. 166. <https://doi.org/10.1186/s12917-021-02872-3>
14. Mortaz E. Imbalance accuracy metric for model selection in multi-class imbalance classification problems // Knowledge-Based Systems. – 2020. – Vol. 210. – P. 106490. <https://doi.org/10.1016/j.knsys.2020.106490>
15. Sadafi A., Shen D. Multiclass Deep Active Learning for Detecting Red Blood Cell Subtypes in Brightfield Microscopy // Medical Image Computing and Computer

Assisted Intervention – MICCAI 2019. Lecture Notes in Computer Science. – 2019. – Vol. 11764. – Pp. 685–693. https://doi.org/10.1007/978-3-030-32239-7_76

16. *Shilaskar S., Ghatol A.* Diagnosis system for imbalanced multi-minority medical dataset // *Soft Comput.* – 2019. – № 23. – Pp. 4789–4799. <https://doi.org/10.1007/s00500-018-3133-x>

17. *Sokolova M., Lapalme G.* A systematic analysis of performance measures for classification tasks // *Information Processing & Management.* – 2009. – Vol. 45, Iss. 4. – Pp. 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>

18. *Takahashi K., Yamamoto K., Kuchiba A.* Confidence interval for micro-averaged F_1 and macro-averaged F_1 scores // *Appl Intell.* – 2022. – № 52. – Pp. 4961–4972. <https://doi.org/10.1007/s10489-021-02635-5>

19. Официальный сайт открытой программной библиотеки TensorFlow для машинного обучения. – URL: https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/EarlyStopping.

20. *Upadhyay V., Rathod D.* Location-Based Crime Prediction Using Multiclass Classification Data Mining Techniques // (eds) *IOT with Smart Systems. Smart Innovation, Systems and Technologies.* – 2022. – Vol. 251. – Pp. 619–626. https://doi.org/10.1007/978-981-16-3945-6_11

IDENTIFICATION OF FISH BLOOD CELL POPULATIONS ON THE BASIS OF A CONVOLUTIONARY NEURAL NETWORK FOR COMPILING A LEUKOGRAM

G.I. PRONINA, D.V. BYKOV, A.V. UKOLOVA, A.E. UL'YANKIN, A.N. KARASEV, M.A. TUTRIKOVA, M.A. AKIMUSHKINA, K.A. KANAeva

(Russian State Agrarian University – Moscow Timiryazev Agricultural Academy)

In aquaculture, physiological assessment is required to monitor the health status of fish. Blood is the most responsive system in the organism of hydrobionts to changes in external factors. The study of hematological parameters of fish allows for early diagnosis of diseases, working out the technological mode of breeding and rearing, and selection. The typing of cells in circulating fluids is important for compiling hemocytic and leukocyte formulas characterizing the cellular component of the organism's immune response.

In the present study, convolutional neural network models are developed to classify blood cells of carp and sturgeon fish. The quality of the models is estimated based on the metrics Accuracy and Precision, Recall, F_1 with macro-averaging.

Based on the processing of blood images, 1104 images of blood cells of carp and sturgeon fish were prepared, including 15 cell populations: hemohistoblasts, myeloblasts, promyelocytes, myelocytes, metamyelocytes, rod-shaped neutrophils, segmented neutrophils, eosinophils, basophils, monocytes, lymphocytes, erythroblasts, normoblasts, mature erythrocytes, and platelets.

Models of a convolutional neural network have been developed to recognize populations of blood cell elements (erythrocytes, leukocytes, platelets) of carp and sturgeon fish. The models were trained on 80% of the prepared images, avoiding the problem of overtraining, as evidenced by the constructed graphs of the loss function (sparse categorical cross entropy) and accuracy during the learning process.

The constructed models make it possible to recognize blood cells of carp fish with an accuracy of 75.0% (metric F_1 with macro-averaging is 0.570) and blood cells of sturgeon fish with an accuracy of 76.6% (F_1 with macro-averaging is 0.664).

Key words: fish blood cells, machine learning, convolutional neural network (CNN), classification, image processing.

References

1. Dolgov V.V., Men'shikov V.V. Clinical laboratory diagnostics. National guidelines. M.: GEOTAR-Media, 2016: 688. (In Rus.)
2. Zhiteneva L.D., Makarov E.V., Rudnitskaya O.A. Fundamentals of ichthyohaematology (in comparative aspect). Rostov-na-Donu: Everest, 2004: 312. (In Rus.)
3. Ivanov A.A., Pronina G.A., Koryagina N.Yu. Haematology of poikilothermic hydrobionts: Monograph. Irkutsk: OOO "Megaprint", 2018: 133. (In Rus.)
4. Maslova N.I., Petrushin A.B. Methodological recommendations on breeding process management in fish breeding (carp as an example). M.: Rossel'khozakademiya, 2005: 27. (In Rus.)
5. Patterson D., Gibson A. Deep learning from a practitioner's perspective. M.: DMK Press, 2018: 418. (In Rus.)
6. Pishchenko E.V. Haematology of freshwater fish: Textbook. Novosibirsk: Novosib. gos. agrar. un-t, 2002: 48. (In Rus.)
7. Pronina G.I., Ivanov A.A., Mannapov A.G., Sanaya O.V. Immunity of poikilothermic hydrobionts. Izvestiya of Timiryazev Agricultural Academy. 2021; 2: 71–91. (In Rus.)
8. Pronina G.I. Clinical Laboratory Diagnostics. Practicum: Textbook for Higher Education Institutions. Moscow, 2021: 88. (In Rus.)
9. Pronina G.I., Koryagina N.Yu. Methodology of physiological and immunological assessment of hydrobionts: Textbook. SPb: Lan', 2017: 96. (In Rus.)
10. Agustin R.I., Arif A., Sukorini U. Classification of immature white blood cells in acute lymphoblastic leukemia L1 using neural networks particle swarm optimization. Neural Comput & Applic. 2021; 33: 10869–10880. URL: <https://doi.org/10.1007/s00521-021-06245-7>
11. Fernández J.C., Carbonero M., Gutiérrez P.A. et al. Multi-objective evolutionary optimization using the relationship between F1 and accuracy metrics in classification tasks. Appl Intell. 2019; 49: 3447–3463. URL: <https://doi.org/10.1007/s10489-019-01447-y>
12. Hameed Z., Garcia-Zapirain B., Aguirre J.J. et al. Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network. Sci Rep. 2022; 12: 15600. URL: <https://doi.org/10.1038/s41598-022-19278-2>
13. Le Floc'h N., Gondret F., Resmond R. Identification of blood immune and metabolic indicators explaining the variability of growth of pigs under contrasted sanitary conditions. BMC Vet Res. 2021; 17: 166. <https://doi.org/10.1186/s12917-021-02872-3>
14. Mortaz E. Imbalance accuracy metric for model selection in multi-class imbalance classification problems. Knowledge-Based Systems. 2020; 210: 106490. URL: <https://doi.org/10.1016/j.knosys.2020.106490>
15. Sadafi A., Shen D. Multiclass Deep Active Learning for Detecting Red Blood Cell Subtypes in Brightfield Microscopy. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Lecture Notes in Computer Science. Springer, Cham. 2019; 11764: 685–693. URL: https://doi.org/10.1007/978-3-030-32239-7_76
16. Shilaskar S., Ghatol A. Diagnosis system for imbalanced multi-minority medical dataset. Soft Comput. 2019; 23: 4789–4799. URL: <https://doi.org/10.1007/s00500-018-3133-x>
17. Sokolova M., Lapalme G. A systematic analysis of performance measures for classification tasks. Information Processing & Management. 2009; 45; 4: 427–437. URL: <https://doi.org/10.1016/j.ipm.2009.03.002>
18. Takahashi K., Yamamoto K., Kuchiba A. Confidence interval for micro-averaged F_1 and macro-averaged F_1 scores. Appl Intell. 2022; 52: 4961–4972. URL: <https://doi.org/10.1007/s10489-021-02635-5>

19. `tf.keras.callbacks.EarlyStopping`. Official website of the TensorFlow – open source library for machine learning. URL: https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/EarlyStopping

20. *Upadhyay V., Rathod D.* Location-Based Crime Prediction Using Multiclass Classification Data Mining Techniques. (eds) IOT with Smart Systems. Smart Innovation, Systems and Technologies. Springer, Singapore. 2022; 251: 619–626. URL: https://doi.org/10.1007/978-981-16-3945-6_61

Пронина Галина Иозеповна, профессор, д-р биол. наук, доцент, Российский государственный аграрный университет – МСХА имени К.А. Тимирязева; 127434, Российская Федерация, г. Москва, Тимирязевская ул., 49; e-mail: gidrobiont4@yandex.ru

Быков Денис Витальевич, ассистент кафедры статистики и кибернетики, Российский государственный аграрный университет – МСХА имени К.А. Тимирязева; 127434, Российская Федерация, г. Москва, Тимирязевская ул., 49; e-mail: bykovdv@rgau-msha.ru

Уколова Анна Владимировна, канд. экон. наук, доцент, и.о. заведующего кафедрой статистики и кибернетики, Российский государственный аграрный университет – МСХА имени К.А. Тимирязева; 127434, Российская Федерация, г. Москва, Тимирязевская ул., 49; e-mail: statmsha@rgau-msha.ru

Ульянкин Александр Евгеньевич, ассистент кафедры статистики и кибернетики, Российский государственный аграрный университет – МСХА имени К.А. Тимирязева; 127434, Российская Федерация, г. Москва, Тимирязевская ул., 49; e-mail: aeulianckin@rgau-msha.ru

Карасев Андрей Николаевич, бакалавр 3-го года обучения, Российский государственный аграрный университет – МСХА имени К.А. Тимирязева; 127434, Российская Федерация, г. Москва, Тимирязевская ул., 49; e-mail: andrkar008@gmail.com

Мария Андреевна Тутрикова, бакалавр 3-го года обучения, Российский государственный аграрный университет – МСХА имени К.А. Тимирязева; 127434, Российская Федерация, г. Москва, Тимирязевская ул., 49; e-mail: maria.tutrikova@gmail.com

Магдалина Алексеевна Акимушкина, бакалавр 3-го года обучения, Российский государственный аграрный университет – МСХА имени К.А. Тимирязева; 127434, Российская Федерация, г. Москва, Тимирязевская ул., 49; e-mail: akimushkina.lina@gmail.com

Ксения Андреевна Канаева, бакалавр 3-го года обучения, Российский государственный аграрный университет – МСХА имени К.А. Тимирязева; 127434, Российская Федерация, г. Москва, Тимирязевская ул., 49; e-mail: kutk-ksusha@yandex.ru

Galina I. Pronina, DSc (Bio), Associate Professor, Professor of the Department of Aquaculture and Beekeeping, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy (49, Timiryazevskaya Str., Moscow, 127434, Russian Federation; E-mail: gidrobiont4@yandex.ru)

Denis V. Bykov, Assistant of the Department of Statistics and Cybernetics, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy (49, Timiryazevskaya Str., Moscow, 127434, Russian Federation; E-mail: bykovdv@rgau-msha.ru)

Anna V. Ukolova, CSc (Econ), Associate Professor, Acting Head of the Department of Statistics and Cybernetics, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy (49, Timiryazevskaya Str., Moscow, 127434, Russian Federation; E-mail: statmsha@rgau-msha.ru)

Aleksandr E. Ul'yankin, Assistant of the Department of Statistics and Cybernetics, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy (49, Timiryazevskaya Str., Moscow, 127434, Russian Federation; E-mail: aeulianckin@rgau-msha.ru)

Andrey N. Karasev, 3rd-year undergraduate student, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy (49, Timiryazevskaya Str., Moscow, 127434, Russian Federation; E-mail: andrkar008@gmail.com)

Maria A. Tutrikova, 3rd-year undergraduate student, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy (49, Timiryazevskaya Str., Moscow, 127434, Russian Federation; E-mail: maria.tutrikova@gmail.com)

Magdalina A. Akimushkina, 3rd-year undergraduate student, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy (49, Timiryazevskaya Str., Moscow, 127434, Russian Federation; E-mail: akimushkina.lina@gmail.com)

Ksenia A. Kanaeva, 3rd-year undergraduate student, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy (49, Timiryazevskaya Str., Moscow, 127434, Russian Federation; E-mail: kutk-ksusha@yandex.ru)