

МЕТОДЫ СТАТИСТИЧЕСКОГО АНАЛИЗА И МАШИННОГО ОБУЧЕНИЯ В СУДЕБНО-БАЛЛИСТИЧЕСКОЙ ЭКСПЕРТИЗЕ

Федоренко В.А., Сорокина К.О.

*Саратовский государственный университет им. Н.Г. Чернышевского
Саратов, Россия*

В настоящее время в отдельных экспертных исследованиях применяют методы статистического анализа и машинного обучения. Их широкое внедрение в экспертную практику осложняется отсутствием у экспертов компетенций в области современных методов статистического анализа, к которым можно отнести и методы машинного обучения. В работе рассмотрены современные методик, основанные на статистическом анализе и методах машинного обучения, адаптированные для решения задач судебной баллистики и баллистической идентификации. Внедрение таких методик в экспертную практику ведет к изменению парадигмы формирования категорических выводов. Суть изменений заключается в переходе от модели, когда категорические выводы эксперта опираются в основном на его внутреннее убеждение, к модели, где внутреннее убеждение формируется на основе количественной оценки доказываемого факта, которая должна быть предъявлена суду и может быть перепроверена другими независимыми специалистами. Для широкого внедрения в экспертную практику математических методов сравнения и оценки схожести следов, предварительно необходимо обеспечить возможность приобретения экспертами «соответствующих» компетенций. С этой целью была разработана дисциплина «Математические методы в судебно-баллистической экспертизе», которая может быть взята за основу курсов повышения квалификации экспертов-баллистов, имеющих естественно-научное или техническое высшее образование.

Ключевые слова: *идентификация огнестрельного оружия, методы машинного обучения, компетенции, судебно-баллистическая экспертиза, курсы повышения квалификации.*

В ряде экспертных баллистических лабораторий полиции России и других стран широко применяются автоматизированные баллистические идентификационные системы (АБИС). Кроме этого, практически все лаборатории оснащены цифровыми микроскопами, которые используются для визуального сравнения следов. Применение таких микроскопов позволяет формировать региональные базы данных цифровых изображений следов на стреляных гильзах и выстреленных пулях, изъятых с мест происшествий. Эффективное использование

Адрес для корреспонденции: Федоренко Владимир Александрович, кандидат физико-математических наук, доцент, заведующий учебно-научной лабораторией криминалистического материаловедения Саратовского государственного университета им. Н.Г. Чернышевского, 410012, г.Саратов, ул. Б. Казачья, 112, 8 корпус, тел. 89061547642, fed77@yandex.ru

баз данных для проведения автоматизированных криминалистических проверок по ним может быть достигнуто за счет применения методов искусственного интеллекта для сравнения следов. Однако, для внедрения современных информационных технологий в процедуру сравнения изображений следов на выстреленных пулях и стреляных гильзах и для проведения автоматизированных проверок по базам данных таких цифровых изображений, эксперты должны обладать соответствующими компетенциями в области методов статистического анализа и машинного обучения.

Еще одной актуальной проблемой судебной экспертизы в целом, и судебной баллистики в частности, является необходимость количественного обоснования экспертом категорических выводов, сформированных им на основе проведенных экспертных исследований. Вопрос надежности и достоверности выводов является очень важным для всех видов экспертных исследований, особенно для судебно-баллистической идентификации огнестрельного оружия. Оценка уникальности совпадающих признаков, основанная только на субъективных суждениях эксперта (на внутреннем убеждении эксперта), иногда приводит к ошибочным выводам и, как следствие, к ошибкам правосудия. Достаточно подробно эта проблема рассмотрена в докладе научно-технической комиссии при президенте США [1], посвященном разбору экспертных ошибок, допущенных в США. Проблема количественного обоснования категорических выводов экспертом предопределяется отсутствием практики применения в экспертных исследованиях математических методов статистической оценки результатов сравнения совпадающих и различающихся признаков исследуемых следов, недостаточной апробацией в судах уже существующих методов количественного анализа.

Внедрение в судебно-баллистическую экспертизу современных методов статистического анализа и машинного обучения ведет к изменению парадигмы экспертных исследований в этой области. Суть изменений заключается в переходе от модели, когда категорические выводы эксперта опираются только на его внутреннее убеждение, к модели, где внутреннее убеждение формируется на основе количественной оценки доказываемого факта, которая должна быть предъявлена суду и может быть перепроверена другими независимыми специалистами. Однако, обязательность количественного обоснования категорических выводов невозможно обеспечить без приобретения экспертами новых компетенций. Таким образом, без решения компетентностной проблемы, разработанные методы будут оставаться чисто научными, оторванными от экспертной практики.

Информационные технологии, применяемые в судебно-баллистической экспертизе.

Кратко рассмотрим некоторые современные разработки в области статистического анализа и методов машинного обучения, адаптированные для решения задач судебной баллистики и баллистической идентификации. Данные разработки в настоящее время могут быть использованы для количественного обоснования выводов в статусе частной экспертной инициативы, но широкого применения в экспертной практике пока не получили.

1. Одним из первых способов количественного обоснования категорических выводов, применяемых при баллистической идентификации, является метод последовательно совпадающих трасс, разработанный А. Биазотти [2]. В соответствии с данной методикой

производится оценка схожести совмещенных следов полей нарезков, основанная на статистическом анализе максимально длинных серий последовательно совпадающих трасс, построении распределения Пуассона и оценке коэффициента правдоподобия гипотезы о совпадении (несовпадении) сравниваемых следов.

2. Для оценки схожести изображений статических следов бойков и следов патронного упора разработаны методы конгруэнтно совпадающих ячеек и метод коррелирующих ячеек [3–5]. Они позволяют выделить на анализируемых изображениях конгруэнтно расположенные области, имеющие высокие значения коэффициента корреляции, оценить коэффициент правдоподобия выдвигаемой гипотезы («следы совпадающие» или «следы несовпадающие»).

3. Для оценки схожести вторичных следов на деформированных пулях Дж. Сонгом с соавторами разработан метод конгруэнтно совпадающих сегментов профилей следов [6, 7]. Данный метод позволяет анализировать следы на деформированных пулях. Как показано в работе [8] его эффективность может быть увеличена за счет применения на завершающей стадии одного из методов машинного обучения, например, метода k-ближайших соседей.

4. Разработан метод, позволяющий оценивать вероятность случайного совпадения комплекса трасс в совмещенных динамических следах [9, 10]. При оценке вероятности учитывается число совпадающих трасс, серии последовательно совпадающих трасс.

5. Разработана методика бинарной классификации сравниваемых следов по категориям «совпадающие» и «несовпадающие» с помощью методов машинного обучения k-ближайших соседей [8] и «случайный лес» [11].

6. Апробированы методики классификации изображений статических следов по экземплярам оружия с помощью нейронных сетей (полносвязная, сверточная и сиамская) [12, 13]. В работе [12] показано, что точность классификации полносвязной нейронной сетью изображений следов бойков достигает 85%.

7. Методы машинного обучения применяются и в других областях судебной баллистики. Например: при оценке скорости пули по глубине огнестрельного повреждения, деформации снаряда и по дефекту массы снаряда рассчитывается многомерная функция регрессии; при определении механизма ранения по пятнам брызг крови применяется алгоритм «случайного леса» [14] и методы обработки цифровых изображений.

Таким образом, в настоящее время методы статистического анализа и машинного обучения достаточно интенсивно адаптируются под решение судебно-баллистических задач. Однако образовательные программы ВУЗов МВД России и курсов повышения квалификации экспертов-баллистов МВД России не включают дисциплины соответствующего математического профиля, адаптированные под задачи судебной баллистики в целом и судебно-баллистической идентификации в частности.

Содержание дисциплины для подготовки экспертов-баллистов нового поколения.

В Саратовском государственном университете им. Н.Г. Чернышевского ведется обучение студентов по магистерской образовательной программе по направлению «Материаловедение, технологии материалов» по профилю «Криминалистическое материаловедение». В процессе преподавания таких дисциплин, как «Общенаучный семинар», «Учебная ознакомительная

практика», «Методы количественного обоснования категорических выводов в судебной экспертизе», «Методы обработки цифровых изображений», «Основы внутренней и внешней баллистики» были отработаны вопросы преподавания методов статистического анализа и машинного обучения применительно к решению задач судебной баллистики. Данные наработки были объединены в одну дисциплину «Математические методы в судебно-баллистической экспертизе», которая рассчитана на 108 часов аудиторных занятий. Данная дисциплина может быть положена в основу курсов повышения квалификации экспертов-баллистов, имеющих естественно-научное или техническое высшее образование.

Дисциплина включает два модуля: методы математической статистики в криминалистике и судебной баллистике; методы машинного обучения в судебно-баллистической экспертизе. В первый модуль вошли следующие темы:

- Статистический анализ случайных погрешностей. В рамках данной темы изучаются вопросы: оценка погрешности прямых и косвенных измерений, погрешности измерения функции; понятие стандартного отклонения и стандартного отклонения среднего; нормальное и равномерное распределение, оценка случайных погрешностей; понятие доверительного интервала; коэффициенты Стьюдента и оценка доверительного интервала при малом числе измерений.
- Корреляционные методы. В данной теме изучаются вопросы: смешанный второй момент и коэффициент линейной корреляции; количественный критерий значимости коэффициента корреляции; функция взаимной корреляции двух изображений и оценка схожести изображений с ее помощью; оценка схожести статических следов методом конгруэнтно совпадающих ячеек; метод конгруэнтно совпадающих сегментов профилей вторичных следов на деформированных пулях и его применение для оценки схожести сравниваемых динамических следов.
- Вероятностные методы оценки схожести следов. В данную тему входят следующие вопросы: распределение Пуассона; коэффициент правдоподобия выдвинутой гипотезы, обоснование категорических выводов в соответствии с методом А. Биазотти на примере совмещенных вторичных следов; оценка вероятности совпадения комплекса трасс в совмещенных динамических следах с учетом ширины трасс в сериях последовательно совпадающих трасс; теорема Байеса, примеры ее применения в криминалистике.

Во второй модуль входят темы, связанные с машинным обучением:

- Введение в методы машинного обучения. В данном разделе рассматриваются вопросы: основные метрики оценок и результатов прогнозирования; проблема переобучения в машинном обучении, принцип бритвы Оккама в криминалистике; методы формирования валидационной обучающей и тестовой выборок; метод наименьшей квадратичной ошибки; минимизация функции потерь методом градиентного спуска.
- Регрессия. В данную тему входят следующие вопросы: регрессия непрерывной и дискретной целевой переменной; регрессия по нескольким независимым переменным; методика оценки скорости пули по глубине пробойны в преграде, деформации пули и дефекту массы пули; понятие бинарной регрессии.

- Метод k-ближайших соседей. Данный раздел включает в себя: измерение расстояний в многомерном пространстве признаков; кросс-валидационный способ определения оптимального k; программная реализации алгоритма k-ближайших соседей; примеры применения алгоритма k-ближайших соседей в баллистической идентификации.
- Метод «случайный лес». Основные вопросы: построение дерева принятия решений для одной и для нескольких объясняющих переменных; понятие информационной энтропии Шеннона и индекса Джинни; построение дерева принятия решений на принципе минимальной энтропии Шеннона; формирование ансамблей, реализация метода «случайный лес»; программная реализации алгоритма «случайный лес», представленная на сайте MathWorks.com.; примеры применения метода «случайный лес» в судебной баллистике и баллистической идентификации.
- Полносвязные нейронные сети. Тема включает вопросы: архитектура полносвязной нейронной сети (НС); корректировка весовых коэффициентов методом обратного распространения ошибки; обучение полносвязной НС на примере изображений следов бойков; программная реализации алгоритма полносвязной НС в среде программирования MathLab.
- Сверточные нейронные сети. Основные вопросы по теме: элементы архитектуры сверточной нейронной сети; выделение признаков; метод обратного распространения ошибки в сверточных и подвыборочных слоях; корректировка коэффициентов маски свертки; программная реализации алгоритма сверточной нейронной сети; принципы формирования аугментированной обучающей выборки; пример классификации следов бойков по экзemplарам оружия с помощью сверточной нейронной сети.
- Сиамские сверточные нейронные сети. Основные вопросы по теме: архитектура сиамской сверточной нейронной сети; методика корректировки коэффициентов фильтров; обучение сиамской сверточной нейросети; бинарная классификация изображений следов бойков сиамской нейронной сетью.

При необходимости дисциплина может быть расширена другими темами, например, «Метод опорных векторов», «Понижение размерности», которые пока не нашли широкого применения в судебной баллистике.

Часто возникает вопрос: «Зачем дополнительно обучать экспертов, когда уже есть готовые математики и программисты?». Во-первых, как уже отмечалось, эксперты должны иметь достаточные компетенции для количественного обоснования категорических выводов. Во-вторых, разработчики методов машинного обучения на основе собственного практического опыта указывают, что успешное решение задач и качественный анализ результатов расчетов возможны только в случае, если специалисты, применяющие методы машинного обучения, обладают достаточно глубокими и обширными знаниями в предметной области, в которой применяются данные методы. Это особенно важно с учетом того, что эксперт является участником судопроизводства и, при необходимости, он обязан дать показания в суде. В качестве иллюстрации можно привести следующие ситуации.

1. Например, при совпадении 50% трасс с учетом серий последовательно совпадающих трасс в совмещенных вторичных следах вероятность случайного совпадения данного

комплекса трасс может составить 0.01%. В определенных ситуациях этого может быть достаточно для обоснования категорического положительного вывода. Однако, корректно ответить на вопрос: «Почему 50% совпадающих трасс бывает достаточно для положительного категорического вывода, а 50% оставшихся несовпадающих трасс оказалось недостаточно для формирования хотя бы вероятного отрицательного вывода?» без знаний о механизмах формирования и вариативности следов на выстреленных пулях невозможно.

2. Индивидуализирующие признаки могут по-разному восприниматься компьютером и человеком, что может привести к недопониманию результатов расчетов. Например, на рисунке 1 представлено изображение следа бойка, полученное на растровом электронном микроскопе. Цифрами 1 и 2 обозначены признаки, которые воспринимаются человеком, как темная и светлая дуга. Однако, на самом деле они имеют одну яркость, что легко увидеть, если провести широкую линию с такой же яркостью пикселей. Поэтому эти дуги не будут восприниматься как разные признаки, например, программой бинаризации изображения по глобальному порогу яркости. Можно привести множество и других аналогичных примеров.

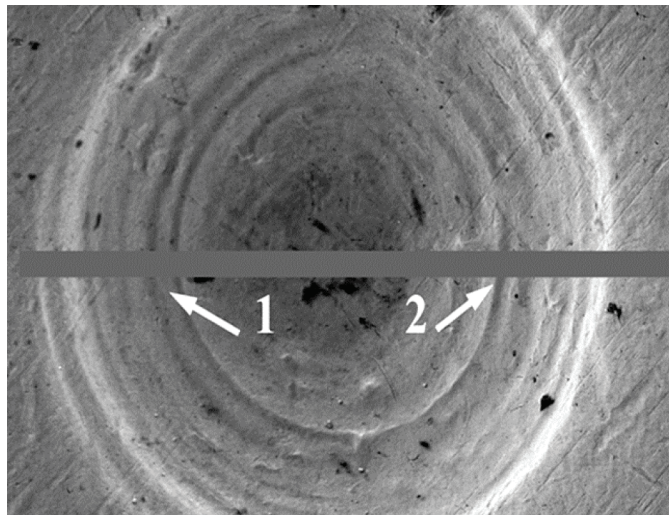


Рис.1. След бойка, полученный на растровом электронном микроскопе:
1. «светлая дуга»; 2. «темная дуга».

В последнее время в ряде стран обсуждается идея формирования категорических выводов только по результатам оценки вероятности события по аналогии с ДНК-анализом биоматериалов. Однако, представленные примеры показывают, что в судебно-баллистической идентификационной экспертизе нельзя ориентироваться только на результаты расчетов. На «выходе» любой программы должен быть эксперт, как лицо, обладающее процессуальным статусом в суде и которое должно взять ответственность за экспертные выводы на себя. Авторы считают, что ссылка на опыт формирования выводов при производстве ДНК экспертизы как пример для судебно-баллистической идентификации не правомерна сразу по нескольким причинам: из-за присутствия в следах на пулях и стреляных гильзах паразитных неоднородностей, порой неотличимых от самих индивидуализирующих признаков; достаточно высокой вариативности отображения признаков в следах каждого экземпляра оружия; возможного присутствия так называемых подгрупповых признаков (*subclass characteristics*),

которые могут относиться одновременно к нескольким десяткам, а иногда и сотням экземпляров оружия.

Таким образом, подытоживая изложенное, сформулируем ряд следующих основных выводов:

1. Категорические выводы эксперта должны опираться не только на его субъективное внутреннее убеждение, но и на количественное обоснование данного убеждения.
2. Для широкого внедрения в экспертную практику современных методов статистического анализа и машинного обучения необходимо ввести специализированные курсы повышения квалификации для экспертов-баллистов.
3. Разработанная учебная дисциплина «Математические методы в судебно-баллистической экспертизе», адаптированная под решение судебно-баллистических задач, может быть взята за основу курсов повышения квалификации экспертов-баллистов.

Список литературы:

1. PCAST. Report to the President on Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. September 2016. Available at: <https://www.documentcloud.org/documents/3121011-Pcast-Forensic-Science-Report-Final.html> Accessed May 25, 2023.
2. Biasotti A. 1959. A Statistical Study of the Individual Characteristics of Fired Bullets. *Journal Forensic Sciences*. 4 (1):34-50.
3. Song J. 2013. Proposed NIST Ballistics Identification System (NBIS) Based on 3D Topographic Measurements on Correlation Cells. *AFTE Journal*. 45 (2):184-194.
4. Song J. 2015. Proposed Congruent Matching Cells (CMC) Method for Ballistic Identification and Error Rate Estimation. *AFTE Journal*. 47 (3):177-185.
5. Сорокина, К. О. Оценка схожести изображений следов патронного упора методом корреляционных ячеек / К. О. Сорокина, В. А. Федоренко, П. В. Гиверц // Информационные технологии и вычислительные системы. 2019. № 3. С. 3-15. DOI: 10.14357/20718632190301
6. Chen Z., Chu W., Soons J. A., Thompson R. M., Song J., Zhao X. 2019. Fired bullet signature correlation using the Congruent Matching Profile Segments (CMPS) method. *Forensic Science International*. 305:10-19. DOI: 10.1016/j.forsciint.2019.109964
7. Chen Z., Song J., Soons J. A., Thompson R. M., Zhao X. 2020. Pilot study on deformed bullet correlation. *Forensic Science International*. 306:1-11. DOI:10.1016/j.forsciint.2019.110098
8. Федоренко, В. А. Анализ следов на выстреленных пулях методами конгруэнтно совпадающих сегментов профилей и k-ближайших соседей / В. А. Федоренко, К. О. Сорокина, П. В. Гиверц // Информационные технологии и вычислительные системы. 2021. № 1. С. 70-82. DOI: 10.14357/20718632210108
9. Федоренко, В. А. Критерии и алгоритм оценки уникальности комплексов совпадающих трасс в следах на выстреленных пулях / В. А. Федоренко, Е. В. Навроцкая //

Информационные технологии и вычислительные системы. ш2019. ш№ 1. шС. 110-120. ш DOI: 10.14357/20718632190110

10. Федоренко, В. А. Количественные критерии обоснования тождественности вторичных следов на пулях, выстреленных из АК-74 / В. А. Федоренко // Теория и практика судебной экспертизы. ш2019. шТ. 14, № 3.ш С. 54-62. шDOI:10.30764/1819-2785-2019-14-3-54-62
11. Carriquiry A., Hofmann H., Tai X. H., VanderPlas S. 2019. Machine Learning in Forensic Applications. Significance. 16(2):29-35. DOI:10.1111/j.1740-9713.2019.01252.x
12. Федоренко, В. А. Многогрупповая классификация следов бойков с помощью полносвязной нейронной сети / В. А. Федоренко, К. О. Сорокина, П. В. Гиверц // Информационные технологии и вычислительные системы.ш 2022. ш№ 3. ш С. 43-57. ш DOI: 10.14357/20718632220305
13. Giverts P., Sorokina K., Fedorenko V. 2022. Examination of the possibility to use Siamese networks for the comparison of firing pin marks. Journal of Forensic Sciences. 67(6):2416-2424. DOI:10.1111/1556-4029.15143
14. Liu Y., Attinger D., Brabanter K. 2020. Automatic Classification of Bloodstain Patterns Caused by Gunshot and Blunt Impact at Various Distances Journal of Forensic Sciences 65(3):729-743. DOI:10.1111/1556-4029.14262

ՎԻՃԱԿԱԳՐԱԿԱՆ ՎԵՐԼՈՒԾՈՒԹՅԱՆ ԵՎ ՄԵՔԵՆԱՅԱԿԱՆ ՈՒՍՈՒՑՄԱՆ ՄԵԹՈԴՆԵՐԸ ԴԱՏԱԿԱՆ ԲԱԼԻՍՏԻԿԱՅԻ ՓՈՐՁԱՔՆՆՈՒԹՅՈՒՆՆԵՐՈՒՄ

Ֆեդորենկո Վ.Ա .,Սորոկինա Կ.Օ.

Ներկայումս որոշ փորձագիտական ուսումնասիրություններում օգտագործվում են վիճակագրական վերլուծության և մեքենայական ուսուցման մեթոդներ: Դրանց լայնածավալ ներդրումը փորձագիտական պրակտիկայում բարդանում է վիճակագրական վերլուծության ժամանակակից մեթոդների ոլորտում փորձագետների կոմպետենտության բացակայությամբ, որոնք ներառում են նաև մեքենայական ուսուցման մեթոդներ: Աշխատանքում դիտարկվում են ժամանակակից մեթոդիկաներ, որոնք հիմնված են վիճակագրական վերլուծության և մեքենայական ուսուցման մեթոդների վրա, և հարմարեցված են դատափորձագիտական բալիստիկ և բալիստիկ նույնականացման խնդիրներ լուծելու համար: Նման մեթոդիկաների ներդրումը փորձագիտական պրակտիկայում հանգեցնում է պարադիգմի փոփոխության՝ կարեգորիկ եզրակացությունների ձևավորման մեջ: Փոփոխությունների էությունը՝ մոդելից անցում է կատարել, երբ փորձագետի կարեգորիկ եզրակացությունները հիմնված են հիմնականում նրա ներքին համոզմունքի վրա, այն մոդելին, որտեղ ներքին համոզմունքը ձևավորվում է ապացուցվող փաստի քանակական գնահատման հիման վրա, որը և պետք է ներկայացվի դատարան և կարող է վերստուգվել այլ անկախ մասնագետների կողմից: Փորձագիտական պրակտիկայում հետքերի նմանությունը համեմատելու և գնահատելու մաթեմատիկական մեթոդների համատարած ներդրման համար նախ անհրաժեշտ է

ապահովել փորձագետների համար «համապատասխան» իրավասություններ ձեռք բերելու հնարավորությունը: Այդ նպատակով մշակվել է «Դատական բալիստիկ փորձաքննության մաթեմատիկական մեթոդները» կարգը, որը կարող է հիմք հանդիսանալ բնագիտական կամ տեխնիկական բարձրագույն կրթությամբ բալիստիկ փորձագետների խորացված վերապատրաստման դասընթացների համար:

Բանալի բառեր. հրազենի նույնականացում, մեքենայական ուսուցման մեթոդներ, իրավասություններ, դատական բալիստիկա, խորացված վերապատրաստման դասընթացներ:

METHODS OF STATISTICAL ANALYSIS AND MACHINE LEARNING IN FORENSIC BALLISTICS

Fedorenko V., Sorokina K.

Currently, in some expert studies, methods of statistical analysis and machine learning are used. Their widespread introduction into expert practice is complicated by the lack of competencies among experts in the field of modern methods of statistical analysis, which include machine learning methods. The paper considers modern methods based on statistical analysis and machine learning methods, adapted to solve problems of forensic ballistics and ballistic identification. The introduction of such methods into expert practice leads to a change in the paradigm of the formation of categorical conclusions. The essence of the changes lies in the transition from the model, when the categorical conclusions of the expert are based mainly on his inner conviction, to the model, where the inner conviction is formed on the basis of a quantitative assessment of the fact being proved, which must be presented to the court and can be rechecked by other independent specialists. For the widespread introduction of mathematical methods for comparing and evaluating the similarity of traces into expert practice, it is first necessary to ensure the possibility for experts to acquire «appropriate» competencies. For this purpose, the discipline «Mathematical methods in forensic ballistic examination» was developed, which can be taken as the basis for advanced training courses for ballistic experts with a natural science or technical higher education.

Keywords: firearms identification, machine learning methods, competencies, forensic ballistics, advanced training courses.

Статья поступила: 05.04.2023

Принята к печати: 14.06.2023