

**АНАЛИЗ МНОГОМЕРНЫХ СТАТИСТИЧЕСКИХ МЕТОДОВ
ИСПОЛЬЗУЕМЫХ ДЛЯ РАСПОЗНАНИЯ ИЗОБРАЖЕНИЙ
ЗАБОЛЕВАНИЙ ДЕРЕВЬЕВ В МАШИННОМ ОБУЧЕНИИ**

Титова Елизавета Павловна, студентка 1 курса магистратуры института экономики и управления АПК, ФГБОУ ВО РГАУ – МСХА имени К. А. Тимирязева, titova.eli.002@mail.ru

Научный руководитель – Маслакова Веста Владимировна, к.э.н., доцент кафедры статистики и кибернетики института экономики и управления АПК, ФГБОУ ВО РГАУ – МСХА имени К. А. Тимирязева

***Аннотация.** Разработана информационная система на основе алгоритмов машинного обучения для деревообрабатывающего предприятия. По результатам анализа были определены методы использования машинного обучения для распознавания изображений заболеваний деревьев в машинном обучении.*

***Ключевые слова:** машинное обучение, распознавание изображений, деревья, многомерные методы, болезни.*

**ANALYSIS OF MULTIDIMENSIONAL METHODS USED FOR IMAGE
RECOGNITION OF TREE DISEASES IN MACHINE LEARNING**

Titova Elizaveta Pavlovna, 1st year graduate student of the Institute of Economics and Management of the Agroindustrial Complex, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, titova.eli.002@mail.ru

Scientific supervisor – Maslakova Vesta Vladimirovna, Ph.D. in Economics, Associate Professor of the Department of Statistics and Cybernetics of the Institute of Economics and Management of the Agroindustrial Complex, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy,

***Annotation.** An information system has been developed based on machine learning algorithms for a woodworking enterprise. As a result of the analysis, methods for using machine learning for image recognition of tree diseases were identified.*

***Key words:** machine learning, image recognition, trees, multidimensional methods, diseases.*

Распознавание изображений в лесной промышленности – мощный инструмент для выявления болезней деревьев. Алгоритмы искусственного интеллекта анализируют фотографии листьев, стволов и корней, чтобы

определить наличие патогенов, болезней или вредителей. Это позволяет автоматизировать диагностику, сократить расходы на ручной труд, своевременно обнаружить проблемы и увеличить эффективность мониторинга.

Использование распознавания изображений для многолетних насаждений растений (МНП) имеет большой потенциал. Во-первых, это позволяет предотвратить распространение инфекций за счет раннего выявления болезней. Во-вторых, это оптимизирует работу агрономов и специалистов по защите растений.

Цель разработки – создать эффективную модель, которая быстро выявляет признаки заболеваний деревьев. Использование глубоких нейронных сетей и методов обработки изображений обеспечит высокую точность и надежность диагностики, что улучшит здоровье лесов и садов, а также внесет вклад в устойчивое развитие сельского хозяйства.

В анализе состояния деревьев используются методы машинного обучения: классификация (определение типа заболевания), кластеризация (выявление схожих паттернов) и регрессия (предсказание степени развития заболевания).

Для внедрения машинного обучения на деревообрабатывающем предприятии, необходимо:

1. Собрать данные: фотографии больных и здоровых деревьев.
2. Подготовить данные: очистить, заполнить пробелы, закодировать категориальные признаки.
3. Выбрать метод: нейронные сети, случайный лес и т.д.
4. Оценить производительность: проверить модель на отложенных данных.
5. Развернуть модель: использовать ее для анализа заболеваний деревьев, регулярно мониторить и обновлять.

В таблице 1 представлены технологии машинного обучения и типичные задачи для них.

Таблица 1

Методы обучения

Название	Задача
Обучение с учителем	Регрессия Классификация
Обучение без учителя	Кластеризация Уменьшение размерности
Обучение с подкреплением	Управление роботом Оптимизация в играх
Обучение с частичным привлечением учителя	Кластеризация
Глубокое обучение	Классификация изображений

Обучение с учителем:

Преимущества: высокая точность, понятная интерпретация, широкий спектр применения.

Недостатки: трудоемкость, дороговизна, риск переобучения.

Обучение без учителя:

Преимущества: обнаружение структур и паттернов, упрощенный сбор данных.

Недостатки: отсутствие четких меток для оценки, трудность оценки производительности.

Обучение с подкреплением:

Преимущества: адаптация к меняющимся условиям, эффективность в принятии решений.

Недостатки: требует много времени и ресурсов, чувствительность к настройкам.

Обучение с частичным привлечением учителя:

Преимущества: использует как помеченные, так и не помеченные данные, снижает потребность в большом количестве размеченных данных.

Недостатки: трудности в настройке, результаты зависят от качества помеченных данных.

Глубокое обучение:

Преимущества: высокая точность, автоматическое извлечение признаков.

Недостатки: не обнаружено

Выбор метода машинного обучения для распознавания болезней деревьев зависит от конкретной задачи и данных. После выбора метода, нужно оценить его эффективность с помощью метрик, таких как точность, полнота и F1-мера.

Для описания этих метрик используется **матрица ошибок**, которая показывает количество правильных и ошибочных предсказаний модели.

В таком случае матрица ошибок классификации будет выглядеть так, как показано на рисунке 1.

Реальность	
П р о г н о з	TP
	FP
П р о г н о з	FN
	TN

Рисунок 1 – Матрица ошибок классификации

Для каждого элемента в наборе данных возможны четыре сценария:

- TP (True Positive) – правильно предсказанные положительные случаи;
- TN (True Negative) – правильно предсказанные отрицательные случаи;
- FP (False Positive) – неправильно предсказанные положительные случаи, также известные как ошибка первого рода;
- FN (False Negative) – неправильно предсказанные отрицательные случаи, также известные как ошибка второго рода.

True – предсказание верно, False – предсказание неверно.

Positive – предсказание положительного результата, Negative – предсказание отрицательного результата.

В первом случае, мы верно предсказали положительный результат, во втором случае верно предсказали отрицательный результат, в третьем случае предсказали положительный результат неверно, в четвертом случае предсказали отрицательный результат неверно

Для применения методов машинного обучения на деревообрабатывающем предприятии важно следовать определенным шагам:

1. Сбор данных о состоянии деревьев, в нашем случае фотографии больных и здоровых деревьев.

2. Очистка и подготовка данных для обучения модели, включая заполнение пропущенных значений, кодирование категориальных признаков и масштабирование числовых признаков.

3. Выбор подходящий метод (например, нейронные сети, случайный лес).

4. Оценить производительность модели на отложенной выборке данных с использованием метрик F1.

5. Развернуть обученную модель в производственной среде предприятия и начать использовать для распознавания и анализа заболеваний деревьев. Регулярно отслеживать производительность модели и обновлять ее при необходимости.

Для данной работы был произведен анализ широкого спектра многомерных статистических методов, которые применялись для идентификации заболеваний деревьев на основе собранных и имеющихся изображений. Исследование проводилось в несколько этапов: сбора данных изображений, на которых были видимые повреждения листьев от заболеваний и полностью здоровых, количество которых составляло 1500 изображений. Для второго этапа сделана предварительная обработка изображений, на которой были произведены нормализующие действия, такие как поворот изображений, отображение, изменение размера, с целью увеличения объема обучающей выборки и повышения устойчивости модели. Далее были рассмотрены различные многомерные методы статистики, такие как : метод главных компонент (PCA), для репродуктивности размерности; линейный дискриминантный анализ (LDA), для классификации; кластеризация (например, K-средние) для средней группировки изображений по признакам; обучение модели с процентами 80% к обучению и 20% на основную часть; модели машинного обучения (например, SVM, случайные леса, нейронные сети); обучились на подготовленных моделях с 1200 изображений; оценка качества: Для оценки качества моделей использовались метрики точности, полноты и F-меры. Также проводился кросс-валидационный анализ для проверки стабильности результатов.

Таким образом, наилучшие результаты были достигнуты при комбинации PCA и LDA с нейронными сетями. Предварительная обработка данных оказала значительное влияние на качество распознавания. Аугментация изображений способствовала улучшению общей производительности моделей. Результаты

исследования могут быть использованы в агрономии для разработки систем раннего выявления заболеваний деревьев, что позволит фермерам принимать более эффективные меры по защите растений.

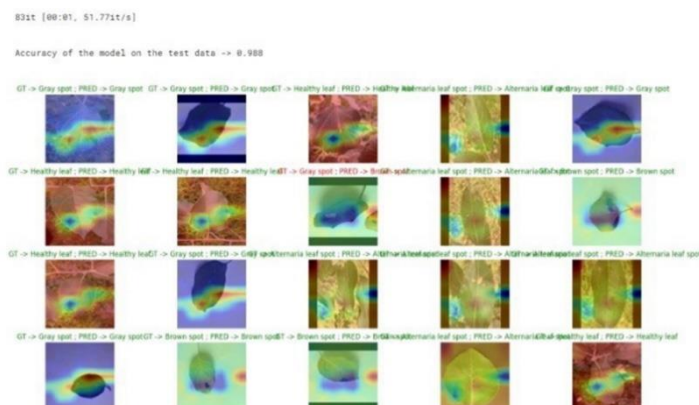


Рисунок 2 – Итоговый результат работы анализа многомерных статистических методов, используемых для распознавания изображения заболеваний деревьев в машинном обучении изображения

Библиографический список

1. Демичев, В. В. Применение методов машинного обучения для прогнозирования эффективности сельского хозяйства с учетом последствий глобального потепления / В. В. Демичев, В. В. Маслакова, И. И. Филатов // Информационно-аналитическое обеспечение устойчивого развития сельского хозяйства. – Москва : Российский государственный аграрный университет – МСХА им. К.А. Тимирязева, 2023. – С. 255-281. – EDN LHLOGR.
2. Кириллов, Р. С. Методы машинного обучения для анализа изображений в ботаническом исследовании / Р. С. Кириллов // Вестник Ботанического института. – 2020. – Т. 17, № 3. – С. 123-145.
3. Иванова, А. Г., Сидоров, Н. А. Многомерные статистические методы обработки изображений в агрономии / А. Г. Иванова, Н. А. Сидоров // Научные аспекты агрономии. – 2021. – № 2. – С. 85-102.
4. Петров, И. В. Применение нейронных сетей для диагностики заболеваний деревьев на основе изображений / И. В. Петров // Журнал компьютерных наук. – 2019. – Т. 14, № 4. – С. 321-335.
5. Смирнова, Т. П., Лебедев, П. Н. Машинное обучение в профилактике заболеваний растений / Т. П. Смирнова, П. Н. Лебедев // Аграрная наука. – 2022. – Т. 45, № 1. – С. 19-36.
6. Михайлов, Е. С. Анализ изображений в агрономии: статистические методы и алгоритмы / Е. С. Михайлов // Известия высших учебных заведений. Серия: Плодовые культуры. – 2021. – Т. 60, № 2. – С. 48-60.
7. Литвинов, А. В., Кузнецова, М. И. Сравнительный анализ методов машинного обучения для классификации заболеваний деревьев / А. В. Литвинов, М. И. Кузнецова // Современные технологии в науке. – 2023. – Т. 78, № 5. – С. 101-115.