

МЕТОДИКА ПОВЫШЕНИЯ ТОЧНОСТИ РАСПОЗНАВАНИЯ РАСТЕНИЙ И ИХ ЗАБОЛЕВАНИЙ С ИСПОЛЬЗОВАНИЕМ MULTILABEL КЛАССИФИКАЦИИ И БАЗЫ ЗНАНИЙ

Шиян А.С. (ИТМО), Баймуратов И.Р. (Ганноверский университет)
Научный руководитель – доктор технических наук, доцент Жукова Н.А.
(ИТМО)

Введение. Заболевания растений являются серьезной проблемой для продовольственной безопасности. Важно обнаруживать заболевания на как можно более ранней стадии, что в свою очередь требует привлечения экспертов и постоянного мониторинга. Для уменьшения трудозатрат при обнаружении заболеваний растений активно используется компьютерное зрение, но, поскольку симптомы многих заболеваний схожи между собой, требуется более комплексный подход, учитывающий связь между растением и его распространенными заболеваниями. Таким образом, в данной статье предложен гибридный метод, учитывающий как результаты распознавания болезней, так и взаимосвязи между растением и его распространенными заболеваниями и автоматически рассчитывающий коэффициенты этих факторов при помощи линейной регрессии. Ожидается, что предложенный метод позволит повысить точность распознавания заболеваний растений.

Основная часть. Своевременное обнаружение растений требует большого количества трудозатрат. Для решения этой проблемы активно используется компьютерное зрение [1]. Для повышения точности распознавания заболеваний растений необходимо учитывать связь между растением и его распространенными заболеваниями. Для решения данной задачи была разработана методика интеграции CNN-модели для распознавания растений и их заболеваний со знаниями о связях между растениями и их заболеваниями с помощью линейной регрессии. Основная идея состоит в том, чтобы использовать данные о наиболее распространенных заболеваниях для каждого растения и на их основе построить модель, которая количественно оценивает вклад распространенности заболеваний в общую точность распознавания.

Предлагаемый метод подразумевает следующие этапы:

- 1) Обучение CNN модели. Предобученные на наборе данных ImageNet модели хорошо зарекомендовали себя в распознавании растений [2]. В данном подходе предлагается использовать предобученную на ImageNet CNN-модель и дообучить ее multilabel классификации на наборе данных изображений растений и их заболеваний. Обученная модель используется для предсказания на тестовой выборке вероятностей вида растений и их заболеваний, которые в дальнейшем будут использованы при подсчете коэффициентов при помощи линейной регрессии.
- 2) Создание базы знаний растений и их распространенных заболеваний. Из открытых источников с помощью парсинга извлекаются данные о растениях и распространенных у них заболеваниях и записываются в онтологию. Онтология позволяет не только систематизировать информацию о растениях и их болезнях, но и в дальнейшем дополнять ее другими факторами, влияющими на распознавание. Например, в онтологию могут быть включены данные о патогенах, вызывающих заболевания, а также о факторах окружающей среды, таких как влажность, температура и тип почвы, которые могут способствовать развитию болезней.
- 3) Расчет коэффициентов. Коэффициенты для соединения точности распознавания растений и их заболеваний с информацией о распространенности заболеваний из онтологии рассчитываются при помощи линейной регрессии.
- 4) Подсчет совместной точности. На основе предсказанных значений линейной регрессии рассчитывается совместная точность, которая учитывает как точность распознавания растений и их заболеваний, так и распространенность заболеваний.

Преимущество предлагаемого метода заключается в его простоте, масштабируемости и

экономичности. Линейная регрессия позволяет быстро и эффективно оценить влияние распространенности заболеваний на точность распознавания, что делает этот метод применимым без экспертной оценки.

Выводы. Предложенный метод позволяет повысить качество оценки точности CNN-модели для распознавания растений и их заболеваний за счет учета взаимосвязей между растениями и их наиболее распространенными болезнями. Практическое применение данного метода может быть реализовано в системах автоматизированного мониторинга состояния растений, что особенно актуально для сельскохозяйственного сектора.

Список использованных источников:

1. Li L., Zhang S., Wang B. Plant disease detection and classification by deep learning—a review //IEEE Access. – 2021. – Т. 9. – С. 56683-56698.
2. Шиян А.С., Баймуратов И.Р. Эмпирическое исследование моделей глубокого обучения для классификации болезней растений // Альманах научных работ молодых ученых Университета ИТМО - 2024. - Т. 1. - С. 170-172