УДК 004.89
МРНТИ 28.23.15

https://doi.org/10.55452/1998-6688-2025-22-3-149-160

[1]Джураев Д.Б.,
бакалавр, ORCID ID: 0009-0008-5800-9960,
e-mail: doninadirov@mail.ru
[1]*Уалиева И.М.,
к.ф.-м.н., ассоциированный профессор, ORCID ID 0000-0003-3853-8896,
*e-mail: i.ualiyeva@gmail.com
[2]Акжалова А.Ж.,
доктор PhD, профессор, ORCID ID 0000-0002-1141-7595,
e-mail: a.akzhalova@kbtu.kz

[1]Казахский национальный университет им. аль-Фараби, г. Алматы, Казахстан
[2]Казахстанско-Британский технический университет, г. Алматы, Казахстан

АНАЛИЗ ЭФФЕКТИВНОСТИ ТРАНСФОРМЕРОВ И СВЕРТОЧНЫХ НЕЙРОСЕТЕЙ ДЛЯ КЛАССИФИКАЦИИ ЗАБОЛЕВАНИЙ РИСА

Аннотация

В данной статье представлен сравнительный анализ современных архитектур нейросетей, сверточных нейронных сетей (CNN) и трансформеров для автоматической диагностики заболеваний листьев риса. В рамках экспериментов были обучены и протестированы модели DenseNet121, ResNet, Vision Transformer (ViT) и MaxViT, после чего проведена их оценка по точности и вычислительной эффективности. Исследование выполнено на основе крупномасштабного датасета, включающего реальные изображения больных и здоровых листьев риса, что делает результаты актуальными для сельскохозяйственной науки и практики. Эксперименты включали оптимизацию гиперпараметров, применение методов аугментации данных, а также использование функций потерь и методов регуляризации с целью повышения обобщающей способности моделей. Для оценки качества использовались метрики точности классификации, F1-мера, а также показатели вычислительной эффективности, такие как время предсказания и объем потребляемых ресурсов. Полученные результаты показали, что модели на основе трансформеров, в частности MaxViT, достигают точности до 94,10%. Это связано с их способностью эффективно моделировать как локальные, так и глобальные признаки изображений за счет механизмов внимания и глубокой контекстуализации. В то же время CNN-архитектуры, такие как DenseNet121 и ResNet, демонстрируют высокую скорость работы и устойчивость в условиях ограниченных вычислительных ресурсов.

Ключевые слова: глубокое обучение, сверточные нейронные сети, Vision Transformer, MaxViT, диагностика растений, рис, классификация изображений.

Введение

Задача эффективной диагностики такой сельскохозяйственной культуры, как рис, играет важную роль в мировом продовольственном обеспечении. Болезни листьев риса наносят значительный урон урожаям, что ведет к экономическим потерям. Традиционные методы диагностики характеризуются высокой временной затратностью и значительными потребностями в человеческих ресурсах, что обусловливает актуальность применения автоматизированных систем, основанных на нейронных сетях, в качестве перспективного направления развития.

Классификация заболеваний растений в сельском хозяйстве с использованием методов машинного обучения широко распространена [2], но и применение нейронных сетей также растет постоянно. Сверточные нейронные сети (CNN), включая такие архитектуры, как DenseNet121, Inceptionv3, MobileNetV2, ResNeXt101, ResNet152V и SE-ResNeXt101, проде-

монстрировали высокую результативность при классификации заболеваний риса. Кроме того, в исследованиях рассматриваются ансамблевые решения, например модель DEX, объединяющая DenseNet121, EfficientNetB7 и Xception [3].

Трансформеры стали мощной альтернативой CNN, особенно благодаря способности учитывать дальнодействующие зависимости в изображениях [4].

Ashraful Haque и соавт. [5] продемонстрировали эффективность применения архитектуры Vision Transformer для задачи классификации изображений риса. Авторы использовали датасеты, собранные в различных регионах, и внедрили методы аугментации данных, что позволило повысить обобщающую способность и устойчивость модели. Предложенная авторами сеть, состоящая из четырех энкодеров трансформера с механизмом многоголовой самовнимательности и плотными слоями, показала точность 98.83% и F1-меру 98.87%.

Ali Mhaned и соавт. [6] показали, что архитектуры Vision Transformer (ViT) эффективно классифицируют 55 классов заболеваний растений, что делает их перспективным инструментом для точного земледелия.

Улучшенная модель ViT с механизмами самовнимательности, предложенная Zhang и соавт. [7], достигла точности 99.21% при классификации заболеваний риса.

Tharani Pavithra P. и соавт. [8] предложили гибридную модель ViT-CNN для классификации заболеваний листьев риса, обученную на ~8000 изображениях, включая пирикулез (Blast), бурую пятнистость (Brown Spot), тунгро (Tungro), ложную головню (False Smut) и бактериальный ожог влагалища листа (Bacterial Sheath Blight). Предложенная архитектура достигла точности (precision) 93.84% и точности top-5 100%.

Sumaya Mustofa и соавт. [9] продемонстрировали, что архитектура MaxViT является наиболее эффективной трансформерной моделью для классификации заболеваний листьев томата, что связано с более стабильной и плавной кривой обучения по сравнению с другими трансформерными архитектурами.

В рамках данного исследования проведено сравнение сверточных (CNN) и трансформерных (ViT, MaxViT) нейронных сетей для классификации заболеваний листьев риса на крупном клиническом датасете, содержащем изображения наиболее распространенных заболеваний риса.

Эксперименты показали, что все модели достигают высокой точности, однако трансформеры в определенных сценариях демонстрируют преимущества в точности благодаря возможности учета глобальных зависимостей. Результаты подтверждены графиками изменения точности и потерь во время обучения.

Особенно эффективной оказалась модель MaxViT, которая смогла не только обеспечить высокую точность, но и продемонстрировать стабильную работу в условиях реальных данных. Традиционные CNN, такие как DenseNet121 и ResNet, также показали хорошие результаты и имеют преимущества в вычислительной скорости.

Постановка задачи. Основной целью работы являлась разработка и оценка модели автоматической классификации заболеваний риса на базе архитектур CNN и трансформеров, включая DenseNet121, ResNet, ViT и MaxViT. Для достижения поставленной цели необходимо было: 1) сделать анализ существующих методов классификации изображений; 2) сделать анализ ключевых особенностей выбранных архитектур для распознавания болезней риса; 3) провести сравнительный эксперимент по точности и эффективности каждого метода; 4) оценить практическую применимость данных решений.

Материалы и методы

Сверточные нейронные сети. Сверточные нейронные сети [10] являются одним из наиболее эффективных и наиболее часто используемых инструментов при работе с изображениями, их обработке и анализе. Они были разработаны специально для извлечения высокоуровневых и низкоуровневых признаков из изображений и продемонстрировали выдающуюся

производительность в задачах компьютерного зрения, таких как классификация, детекция объектов и сегментация.

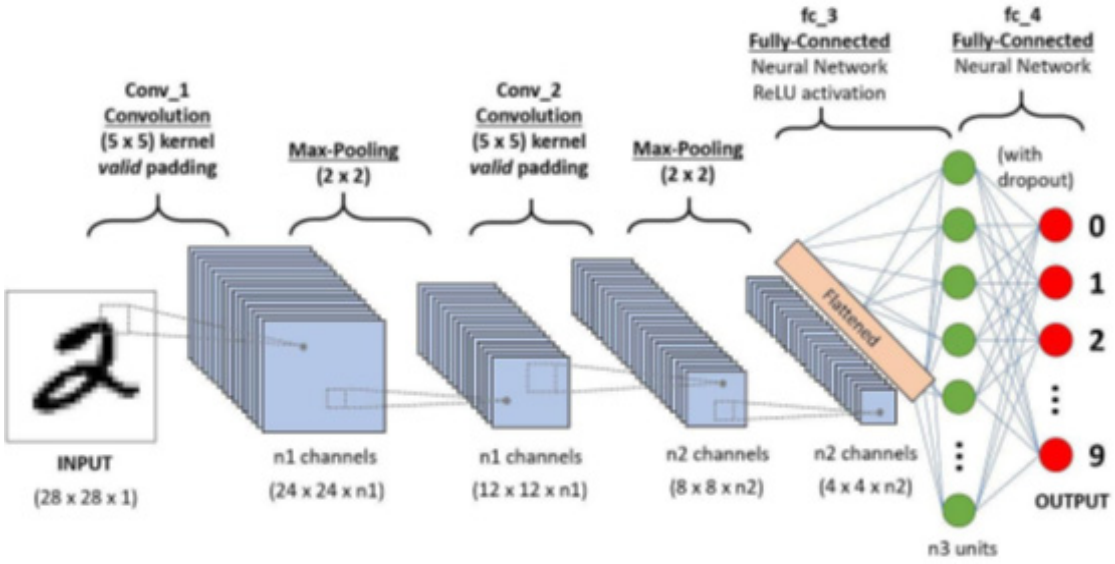

Рисунок 1 – Архитектура сверточных нейронных сетей

В отличие от классических полносвязных архитектур, CNN используют сверточные слои, в которых обучаемые фильтры (ядра) позволяют автоматически извлекать информативные признаки, начиная от простых элементов, таких как линии и границы, до более сложных структур – текстур и форм. Количество сверточных слоев в архитектуре варьируется от 5 до 30 и зависит от конкретной модели (рисунок 1). Первые слои преимущественно выделяют низкоуровневые характеристики изображения, тогда как более глубокие слои фиксируют высокоуровневые признаки. Оптимальный выбор числа слоев свертки является важным фактором, влияющим на итоговую производительность модели. После этапа свертки признаки преобразуются в векторное представление и передаются на полносвязанные слои, где выполняется окончательная классификация. На рисунке 2 показан процесс свертки на примере наложения ядра свертки размером 3 на 3.

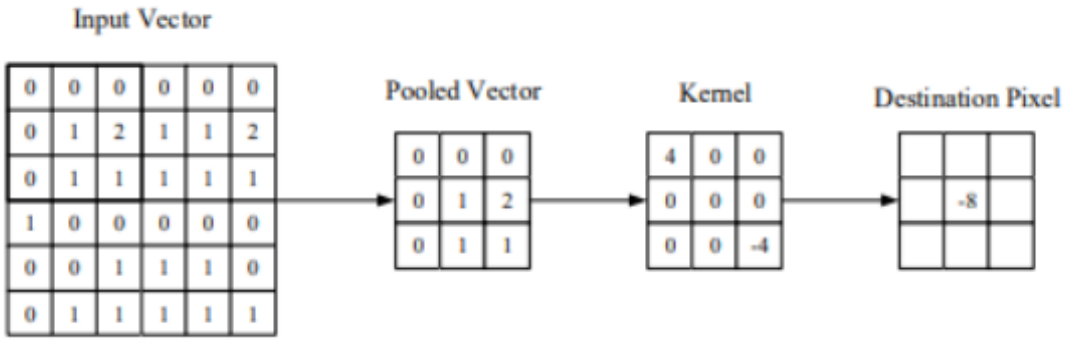

Рисунок 2 – Процесс свертки

Ключевые компоненты сверточного слоя включают: ядра свертки, padding, stride, pooling и функцию активации. Ядра свертки представляют собой небольшие матрицы (обычно 3×3 или 5×5), веса которых оптимизируются с использованием алгоритма обратного распространения ошибки. При их применении осуществляется операция скользящего окна с поэлемент-

ным умножением и суммированием. Чтобы сохранить размерность входных данных после операции свертки, применяют padding, который представляет собой дополнение краевых областей изображения нулями. При перемещении по изображению можно определить величину шага фильтра. Этот параметр называют stride. Операция Pooling понижает размерность карты признаков и уменьшает чувствительность к шумам. Наиболее распространены методы максимального, минимального и усредняющего пулинга. Функция активации вносит нелинейность в модель. На практике чаще всего используется ReLU, которая зануляет отрицательные значения, ускоряет обучение и позволяет избежать проблемы исчезающих градиентов.

Модель Vision Transformer (ViT). Архитектура глубокой нейронной сети Vision Transformer (ViT) [11] использует Transformer encoder. Модель сегментирует изображение на патчи фиксированного размера и эффективно выявляет сложные зависимости между ними, обеспечивая высокую точность в задачах классификации изображений (рисунок 3).

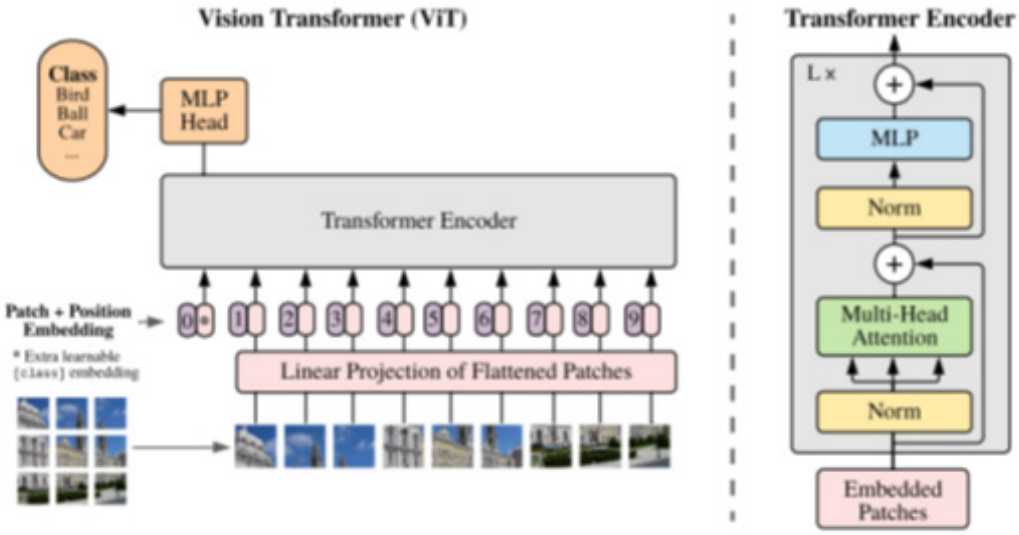

Рисунок 3 – Обзор Vision Transformer

В модели ViT изображение разбивается на непересекающиеся фрагменты фиксированного размера, которые преобразуются в векторы признаков. Для сохранения пространственной информации к ним добавляются позиционные эмбеддинги. Далее векторы обрабатываются в модуле Transformer Encoder, включающем механизмы многоголового самовнимания (MHSA), нормализацию слоев (LN) и многослойный перцептрон (MLP). Самовнимание обеспечивает взаимодействие между всеми патчами изображения, выполняя роль, аналогичную сверткам. Итоговое представление поступает в полносвязный слой, где осуществляется классификация с использованием функции softmax.

Модель MaxViT. В качестве основы для решения задачи классификации заболеваний листьев риса в данной работе использовалась архитектура MaxViT (Multi-Axis Vision Transformer), представляющая собой усовершенствованную модификацию Vision Transformer. В отличие от классического ViT, MaxViT сочетает механизмы сверток и трансформеров, что позволяет учитывать как локальные, так и глобальные признаки изображения при сниженных вычислительных затратах.

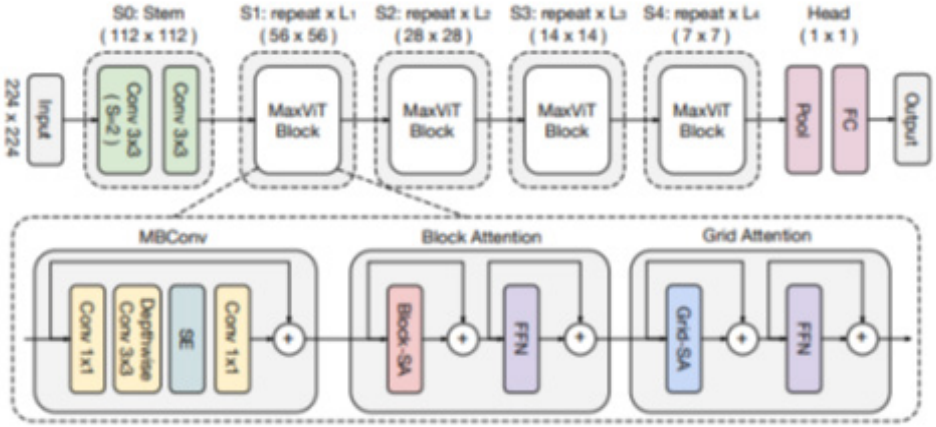

Рисунок 4 – Архитектура MaxViT

Механизм многоосевого внимания (multi-axis attention), как ключевой компонент MaxViT, включает два уровня: Block Attention, извлекающий локальные зависимости в пределах ограниченных областей изображения, и Axial Attention, обеспечивающий эффективный обмен информацией между патчами по строкам и столбцам. Такое сочетание позволяет одновременно анализировать мелкие детали и общий контекст изображения, что критически важно при распознавании визуально сложных патологий растений.

Дополнительной особенностью архитектуры является использование сверточных слоев совместно с трансформерными блоками. Интеграция сверток способствует более качественной обработке низкоуровневых признаков, уменьшает вычислительные затраты и снижает требования к ресурсам. Благодаря этому MaxViT демонстрирует высокую эффективность и масштабируемость при работе с изображениями высокого разрешения, обеспечивая улучшенную точность классификации по сравнению с классическим ViT.

В контексте задач точного земледелия, и в частности анализа заболеваний риса, MaxViT формирует иерархическое представление изображений, позволяя выявлять даже небольшие и неравномерно распределенные повреждения листьев, связанные с патологическими изменениями.

Датасет

Для проведения экспериментов были взяты изображения наиболее распространенных заболеваний листьев риса, любезно предоставленные нам авторами статьи [12]. Датасет содержит девять классов заболеваний листьев риса, где в каждом заболевании представлено сто изображений (рисунок 5).

Рисунок 5 – Изображения больных листьев риса

В датасет включены наиболее опасные заболевания листьев, такие как Bacterial Leaf Blight, вызываемое Xanthomonas oryzae и способное снижать урожайность до 50–80%, а также Bacterial Panicle Blight, вызванное Burkholderia glumae и B. gladioli, которое приводит к потемнению метелок и гниению зерна. Заболевание Brown Spot, вызываемое Bipolaris oryzae, проявляется темными овальными пятнами на листьях, тогда как насекомое-вредитель Hispa (Dicladispa armigera) повреждает эпидермис листьев. Leaf Scald (Microdochium oryzae), вызывает ожогоподобные поражения, и Leaf Smut (Entyloma oryzae), характеризуется темными пятнами и ложными спорообразованиями. Leaf Blast аналогично Sheath Blight вызывается Rhizoctonia solani и может приводить к потерям урожая до 15%, при этом Sheath Blight преимущественно поражает влагалища листьев и широко распространено в Бангладеш. И последний тип – это вирусное заболевание Tungro, переносимое насекомыми (Nephotettix virescens), которое приводит к замедлению роста и деформации листьев.

После проведенной аугментации изображений объем датасета вырос до 42 876 изображений, которые мы разделили на обучающие и валидационные.

Результаты и обсуждение

В ходе обучения модели была выбрана вычислительная среда Kaggle Notebooks благодаря ее удобству для работы с данными в формате TFRecord и наличию бесплатных ресурсов на базе графических процессоров NVIDIA T4. Все модели были обучены на предварительно обученных весах на датасете ImageNet. В качестве основного фреймворка был выбран PyTorch за счет его гибкости и широкой функциональности для построения и обучения моделей компьютерного зрения. Для унификации процесса обработки данных были разработаны функции для парсинга TFRecord файлов, создания собственных классов Dataset и организации загрузки данных через DataLoader, что обеспечило корректное сравнение эффективности различных архитектур.

Для модели Vision Transformer (ViT) была использована предобученная версия ViT-BasePatch16-224, модифицированная заменой выходного слоя на полностью связанный слой с девятью нейронами для решения задачи многоклассовой классификации заболеваний риса (рисунок 6).

```
# Загрузка и настройка модели ViT
model = ViTForImageClassification.from_pretrained("google/vit-base-patch16-224")
model.classifier = torch.nn.Linear(model.config.hidden_size, 9)  # 9 классов
model.to(device)

# Функция потерь и оптимизатор
criterion = torch.nn.CrossEntropyLoss()
optimizer = torch.optim.AdamW(model.parameters(), lr=1e-5, weight_decay=1e-4)
```

Рисунок 6 – Загрузка модели ViT. Функции потерь и оптимизатор

В процессе обучения проводился систематический подбор гиперпараметров, таких как скорость обучения, размер батча, количество эпох, выбор оптимизатора и параметры регуляризации, с целью улучшения точности модели и ее устойчивости к переобучению (рисунок 7).

```python
for epoch in range(epochs):
    model.train()
    running_loss, correct, total = 0.0, 0, 0
    for images, labels in tqdm(train_loader, desc=f"Epoch {epoch+1}/{epochs}"):
        images, labels = images.to(device), labels.to(device)
        optimizer.zero_grad()
        outputs = model(images).logits
        loss = criterion(outputs, labels)
        loss.backward()
        optimizer.step()
        running_loss += loss.item()
        _, predicted = outputs.max(1)
        total += labels.size(0)
        correct += predicted.eq(labels).sum().item()
    train_acc = 100 * correct / total
    train_losses.append(running_loss / len(train_loader))
    train_accuracies.append(train_acc)
    print(f"Epoch {epoch+1}, Train Loss: {train_losses[-1]:.4f}, Train Accuracy: {train_acc:.2f}%")
```

Рисунок 7 – Обучение модели ViT

На рисунке 8 представлена диаграмма изменения функции потерь и точности. Из графиков видно, что сохранение наилучшей модели по результатам на валидационной выборке обеспечивает значительное преимущество, позволяя избежать переобучения и повысить обобщающую способность. Максимальная достигнутая точность составила 91.21%, а минимальное значение функции потерь – 0.2292.

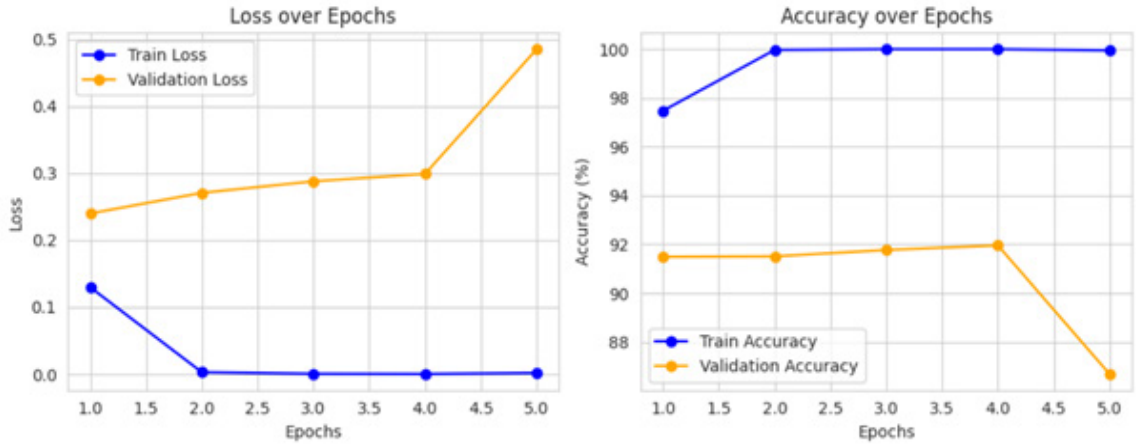

Рисунок 8 – Изменение функции потерь и точности
(лучшие значения: Accuracy = 91.21%, Loss = 0.2292).

В качестве функции потерь применялась кросс-энтропия, а для оптимизации весов использовался алгоритм AdamW с L2-регуляризацией. Проведенные эксперименты позволили повысить точность модели с первоначальных 87% до 91.21%.

Для обучения модели использовалась предварительно натренированная архитектура MaxViT-T (версия IMAGENET1K_V1), в которой выходной полносвязанный слой был модифицирован для решения задачи классификации девяти классов заболеваний. Общая численность параметров модели составила 32,1 млн. Процесс обучения проводился в течение 10

эпох с использованием L2-регуляризации и функции потерь на основе кросс-энтропии. Для оптимизации весов применялся алгоритм AdamW. Дополнительно были реализованы механизмы ранней остановки и сохранения лучшей версии модели по результатам на валидационной выборке. По итогам экспериментов с различными модификациями наилучшая точность на валидационном наборе данных составила 94.10%.

Результаты следующего эксперимента продемонстрировали более стабильное обучение и сбалансированные значения метрик Precision и Recall для большинства классов, что свидетельствует о повышенной устойчивости обобщающих свойств модели. Несмотря на то что разница в 0.36% выглядит незначительной, она имеет существенное значение при решении задачи классификации заболеваний риса. На рисунке 9 представлена динамика изменения функции потерь и точности в процессе обучения модели на архитектуре MaxViT.

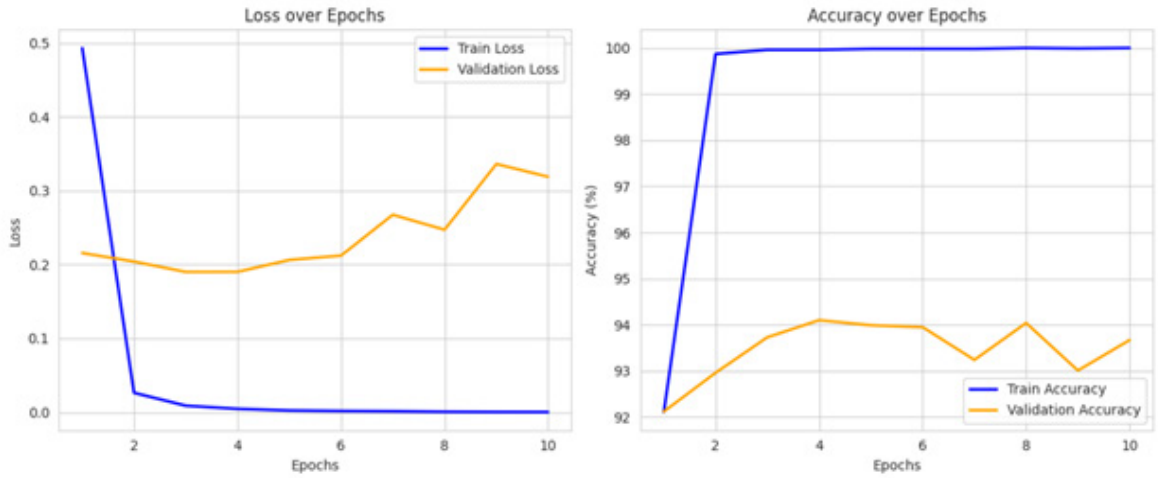

Рисунок 9 – Изменение точности и функции потерь при обучении модели MaxViT
(разница 0.36% имеет значение для классификации заболеваний риса)

Для сравнения с трансформерами были обучены архитектуры сверточных нейронных сетей ResNet и DenseNet121.

Модель ResNet в среднем продемонстрировала точность положительных классификаций на уровне 90.26%, полноту 87% и значение F1-меры 88.93% по всем классам заболеваний, что свидетельствует о достаточно хорошей способности к обобщению. Наиболее высокие результаты получены для заболеваний Blight, Bacterial Leaf Blight и Tungro, где все три метрики близки к оптимальным значениям. Однако на заболевании Leaf Smut модель показала низкий уровень точности (58.96%), что можно считать неприемлемым для задач, связанных с выявлением патологий растений. На рисунке 10 представлена динамика изменения функции потерь и точности в процессе обучения модели на архитектуре ResNet.

На рисунке 11 представлена динамика изменения функции потерь и точности в процессе обучения модели на архитектуре DenseNet121. Итоговая точность для архитектуры ResNet составила 90.33%, в то время как DenseNet121 показала более высокое значение – 93.75%.

В таблице 2 приведены итоговые показатели точности и значения функции потерь для каждой из исследованных архитектур нейронных сетей на валидационном наборе данных. Поскольку модели не обучались на валидационной выборке, данная оценка является объективной, так как разделение набора данных осуществлялось в пропорции 70:30.

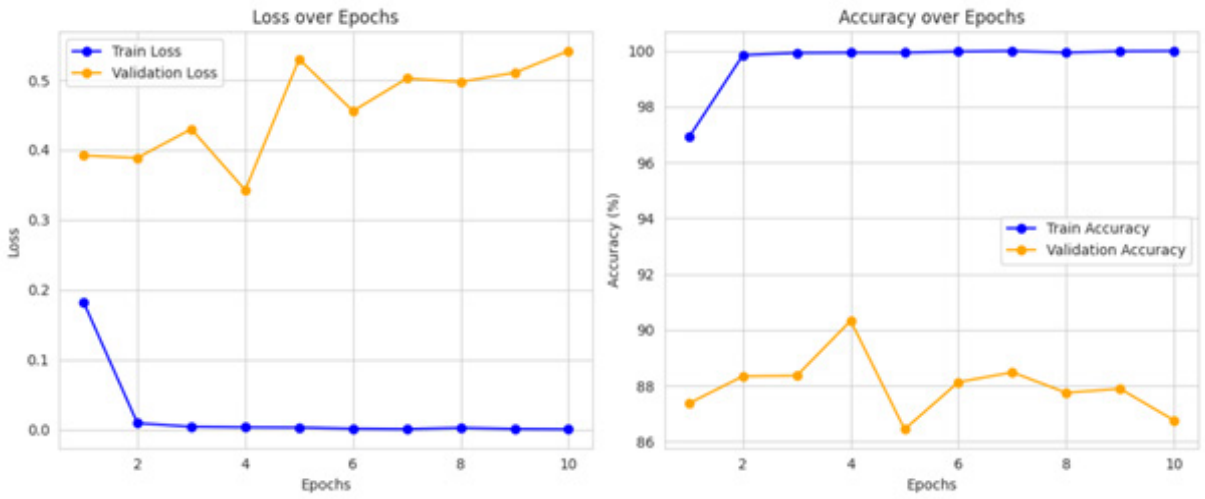

Рисунок 10 – Изменение точности и функции потерь при обучении модели ResNet (средняя точность 90.26%, низкая для Leaf Smut – 58.96%)

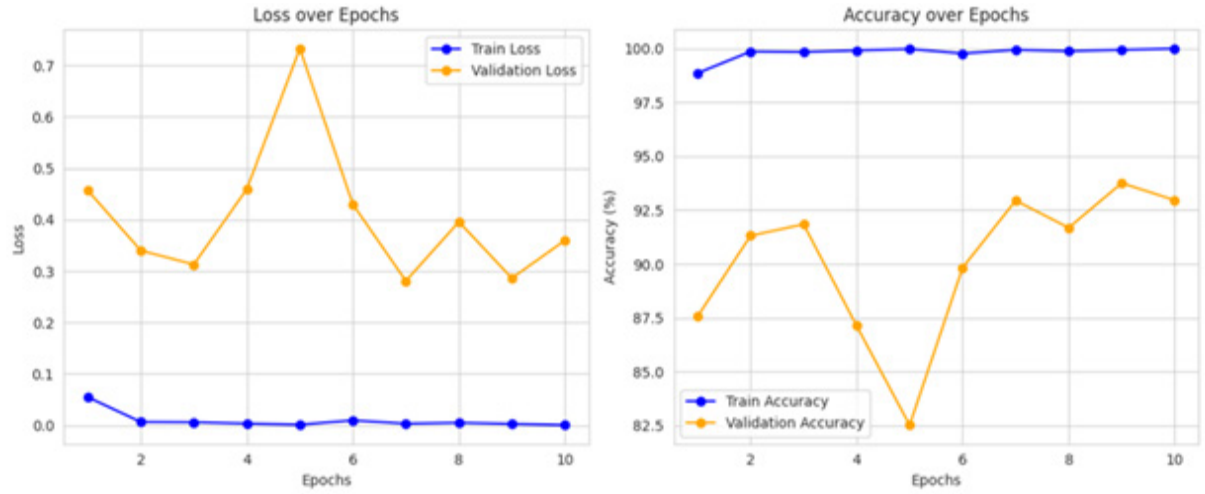

Рисунок 11 – Изменение точности и функции потерь при обучении модели DenseNet121 (средняя точность 93.75%).

Таблица 2 – Диаграмма изменения точности и потерь для каждой архитектуры

	ResNet	DenseNet121	ViT	MaxViT
Precision	90.26	94.09	91.69	93.97
Recall	87	93.74	91.29	93.83
F1-score	88.93	93.50	90.49	93.75
Accuracy	90.33	93.75	91.21	94.10

Заключение

Авторами были проведены эксперименты по обучению и сравнению четырех различных архитектур нейронных сетей: DenseNet121, ResNet, ViT (Vision Transformer) и MaxViT. Результаты показали, что модели MaxViT достигли наибольшей точности (94.10%), благодаря своей способности учитывать как локальные, так и глобальные признаки изображений. Трансформеры демонстрируют преимущества в точности благодаря возможности учитывать глобальные зависимости, а сверточные сети обладают высокой скоростью вычислений и стабильностью в условиях ограниченных данных. Внедрение подобных технологий в сельское хозяйство может значительно повысить эффективность диагностики болезней риса, ускорить принятие решений и снизить потери урожая, делая мониторинг более точным и автоматизированным.

ЛИТЕРАТУРА

1 Ahad, M.T., Li, Y., Song, B., Bhuiyan, T. Comparison of CNN-based deep learning architectures for rice diseases classification. Artificial Intelligence in Agriculture, 9, 22–35 (2023). https://doi.org/10.1016/j.aiia.2023.07.001.

2 Sethy, P.K., Barpanda, N.K., Rath, A.K., Behera, S.K. Deep feature based rice leaf disease identification using support vector machine. Computers and Electronics in Agriculture, 175, 105527 (2020). https://doi.org/10.1016/j.compag.2020.105527.

3 Simhadri, C.G., Kondaveeti, H.K., Vatsavayi, V.K., Mitra, A., Ananthachari, P. Deep learning for rice leaf disease detection: A systematic literature review on emerging trends, methodologies and techniques. Information Processing in Agriculture (2024, May). https://doi.org/10.1016/j.inpa.2024.04.006.

4 Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T. et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint (2020). URL: https://arxiv.org/abs/2010.11929.

5 Haque, M.A., Pastor-Escuredo, D., Brigui, I., Kesswani, N., Bordoloi, S., Ray, A. K. Rice disease identification using Vision Transformer (ViT) based network. The Future of Artificial Intelligence and Robotics (Cham: Springer, 2024), pp. 732–741. https://doi.org/10.1007/978-3-031-60935-0_63.

6 Mhaned, A., Salma, M., El Haji, M., Jamal, B. Plant disease detection using vision transformers. International Journal of Electrical and Computer Engineering, 15 (2), 2334–2344 (2025). https://doi.org/10.11591/ijece.v15i2.

7 Zhang, H. Attention-based feature enhancement for rice leaf disease recognition. Proc. 2nd Int. Conf. Artificial Intelligence and Automation in High-Performance Computing (AIAHPC 2022), pp. 41–49. https://doi.org/10.1117/12.2641832.

8 Tharani, P., Baranidharan, B. A hybrid ViT-CNN model premeditated for rice leaf disease identification. International Journal of Computational Methods and Experimental Measurements, 12 (1), 35–43 (2024). https://doi.org/10.18280/ijcmem.120104.

9 Mustofa, S., Munna, M., Emon, Y., Rabbany, G., Ahad, M. A comprehensive review on plant leaf disease detection using deep learning // arXiv preprint (2023). URL: https://arxiv.org/abs/2308.14087.

10 Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. IEEE Transactions on Neural Networks and Learning Systems (2021), pp. 1–21. https://doi.org/10.1109/TNNLS.2021.3084827.

11 Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., Li, Y. MaxViT: Multi-axis vision transformer. arXiv preprint (2022). URL: https://arxiv.org/abs/2204.01697.

12 Ahad, Md & Li, Yan & Song, Bo & Bhuiyan, Touhid. Comparison of CNN-based deep learning architectures for rice diseases classification. Artificial Intelligence in Agriculture, 9, 22–35 (2023). https://doi.org/ 10.1016/j.aiia.2023.07.001.

[1]Джураев Д.Б.,
бакалавр, ORCID ID 0009-0008-5800-9960, e-mail: doninadirov@mail.ru
[1]*Уалиева И.М.,
ф.-м.ғ.к., қауымдастырылған профессор, ORCID ID: 0000-0003-3853-8896,
*e-mail: i.ualiyeva@gmail.com
[2]Акжалова А.Ж.,
PhD, профессор, ORCID ID: 0000-0002-1141-7595, e-mail: a.akzhalova@kbtu.kz

[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан
[2]Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан

КҮРІШ ЖАПЫРАҚТАРЫНЫҢ АУРУЛАРЫН АВТОМАТТЫ ЖІКТЕУ ҮШІН ТРАНСФОРМЕР ЖӘНЕ СВЕРТКІШ НЕЙРОЖЕЛІ АРХИТЕКТУРАЛАРЫНЫҢ ӘСЕРЛІЛІГІН САЛЫСТЫРА ТАЛДАУ

Андатпа

Бұл мақалада заманауи нейрондық желі архитектуралары – конволюциялық нейрондық желілер (CNN) және трансформерлер – күріш жапырақтарының ауруларын автоматты түрде диагностикалаудағы салыстырмалы талдауы ұсынылады. Эксперименттер барысында DenseNet121, ResNet, Vision Transformer (ViT) және MaxViT модельдері оқытылып, сыналды, содан кейін олардың дәлдігі мен есептеу тиімділігі тұрғысынан бағаланды. Зерттеу шынайы сау және ауру күріш жапырақтарының бейнелерінен құралған ірі ауқымды деректер жиыны негізінде жүргізілді, бұл нәтижелерді ауыл шаруашылығы ғылымы мен тәжірибесі үшін аса өзекті етеді. Эксперименттер гиперпараметрлерді оңтайландыруды, деректерді аугментациялау әдістерін қолдануды, сондай-ақ модельдердің жалпылау қабілетін арттыру мақсатында шығын функциялары мен регуляризация әдістерін пайдалануды қамтыды. Бағалау метрикаларына классификация дәлдігі, F1-мера, сондай-ақ есептеу тиімділігінің көрсеткіштері – болжам жасау уақыты мен ресурстарды тұтыну кірді. Нәтижелер көрсеткендей, трансформерлер негізіндегі модельдер, әсіресе MaxViT, 94,10%-ға дейін дәлдікке қол жеткізді. Бұл олардың назар механизмдері мен терең контекстуализация арқылы кескіндердің жергілікті және ғаламдық белгілерін тиімді модельдей алу қабілетіне байланысты. Сонымен қатар, DenseNet121 және ResNet сияқты CNN архитектуралары шектеулі есептеу ресурстары жағдайында жоғары жылдамдық пен тұрақтылық көрсетті.

Тірек сөздер: терең оқыту, сверткіш нейрондық желілер, Vision Transformer, MaxViT, өсімдіктерді диагностикалау, күріш, кескіндерді жіктеу.

[1]Jurayev D.B.,
Bachelor student, ORCID ID 0009-0008-5800-9960, e-mail: doninadirov@mail.ru
[1,2]*Ualiyeva I.M.,
Cand.Phys.-Math.Sc., Associate Professor, ORCID ID 0000-0003-3853-8896,
*e-mail: i.ualiyeva@gmail.com
[2]Akzhalova A.Zh,
PhD, professor, ORCID ID 0000-0002-1141-7595, e-mail: a.akzhalova@kbtu.kz

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]Kazakh-British Technical University, Almaty, Kazakhstan

COMPARATIVE ANALYSIS OF THE EFFECTIVENESS OF TRANSFORMER AND CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES FOR AUTOMATIC CLASSIFICATION OF RICE LEAF DISEASES

Abstract

This article presents a comparative analysis of modern neural network architectures, convolutional neural networks (CNNs) and transformers, for the automatic diagnosis of rice leaf diseases. In the experiments,

DenseNet121, ResNet, Vision Transformer (ViT), and MaxViT models were trained and tested, followed by their evaluation in terms of accuracy and computational efficiency. The study was conducted on a large-scale dataset containing real images of healthy and diseased rice leaves, which makes the results highly relevant for agricultural science and practice. The experiments included hyperparameter optimization, application of data augmentation techniques, and the use of loss functions and regularization methods to improve the generalization ability of the models. The evaluation metrics comprised classification accuracy, F1-score, as well as computational efficiency indicators such as prediction time and resource consumption. The results showed that transformer-based models, particularly MaxViT, achieve accuracy of up to 94.10%. This is attributed to their ability to effectively capture both local and global image features through attention mechanisms and deep contextualization. At the same time, CNN architectures such as DenseNet121 and ResNet demonstrate high processing speed and robustness under limited computational resources.

Keywords: deep learning, convolutional neural networks, Vision Transformer, MaxViT, plant disease diagnosis, rice, image classification.

Дата поступления статьи в редакцию: 29.08.2025