

## Распознавание бренда одежды по изображению с использованием методов машинного обучения

*Д.А. Якупова, Д.М. Бартнев, Д.А. Болгов, А.А. Гуреева, А.Р. Донская*  
*Волгоградский государственный технический университет, Волгоград*

**Аннотация:** В статье рассматривается разработанная модель для распознавания бренда одежды по изображению. Модель не только прогнозирует тип и бренд одежды, но также может определять их схожесть. На первоначальном этапе был собран набор данных, содержащий изображения одежд различных брендов общим объемом 9000 изображений. В данной работе использовалась архитектура нейронной сети ViT (Vision Transformer) - модель для работы с изображениями, которая была представлена специалистами из Google Brain. В качестве представителя архитектуры трансформеров выступила модель vit-base-patch16-224. Перед обучением все изображения были преобразованы в черно-белый вид, а также была использована аугментация данных: поворот изображения на случайный угол, зеркальное преобразование. Все фотографии были нормализованы – координаты пикселей приведены к промежутку [0,1].

**Ключевые слова:** нейронная сеть, модель, машинное обучение, Vision Transformer, индустрия моды, предсказание бренда одежды, предсказание типа одежды, определение схожести брендов.

### Актуальность

Ежегодно McKinsey [1] анализируют индустрию моды, согласно их отчетам на данную индустрию приходится 2% мирового ВВП. Так, например, во Франции в 2015 году сектор моды обошел по показателю ВВП такие важные отрасли, как аэрокосмическую и автомобильную.

За последние 20 лет спрос на одежду очень вырос. Потребители модной одежды покупают более 80 миллиардов новых предметов одежды ежегодно, что на 400% больше, чем всего два десятилетия назад. В среднем в 2021 году люди потратили 220,7 долларов США на 16,7 миллиардов единиц одежды и обуви во всем мире. А с развитием информационных технологий темпы расширения индустрии стремительно увеличиваются и появляется необходимость быстрого производства, разработки и создания новых идеи, концепций. Этот рост способствует также необходимости анализа огромных объемов данных, быстрого приспособления к изменениям на рынке, конкурентной способности создавать товары и услуги, которые заинтересуют

клиентов. Все это стало возможным благодаря развитию машинного обучения и компьютерного зрения, которые, в свою очередь, на данный момент задают новые тенденции в индустрии моды. Современные алгоритмы позволяют предсказывать новые тренды, оказывать помощь в создании образов, искать продукт по фотографии, воспользоваться виртуальной примеркой и многое другое [2, 3, 4]. Эти возможности искусственного интеллекта не только привлекают внимание клиентов, упрощают их взаимодействие с продуктом, но и способствуют созданию продуктов, в которые вложены меньше ресурсов, то есть, технологии помогают как развитию компаний, так и улучшению пользовательского опыта [5, 6].

### **Предлагаемое решение**

В данной работе предлагается рассмотреть модель, которая определяет бренд одежды по фотографии. Работа была разделена на три основных блока:

1. Анализ предметной области, индустрии мод и ее связи с искусственным интеллектом;
2. Сбор данных для обучения модели;
3. Разработка модели для распознавания бренда одежды по фотографии.

Целью данной работы является определение бренда одежды по фотографии.

Для реализации цели работы необходимо выполнить следующие задачи:

1. Провести обзор предметной области (взаимосвязь и влияние искусственного интеллекта на индустрию моды);
  2. Собрать набор данных для обучения;
  3. Разметить данные;
  4. Разработать модель для классификации типа одежды;
  5. Разработать модель классификации бренда.
-

## Сбор данных

Для дальнейшего обучения модели было решено собрать набор данных, который содержит изображения одежды различных брендов. На рисунке 1 изображены гистограммы, которые описывают количество изображений для каждого анализируемого бренда и каждого выбранного типа одежды соответственно. Было решено отобрать по 1000 изображений для каждого бренда и по 800 изображений для каждого типа одежды. В работе были отобраны следующие бренды: Louis Vuitton (Франция), Burberry (Великобритания), Chanel (Франция), Diesel (Италия), Dolce&Gabbana (Италия), Gucci (Италия), Guess (США) и Prada (Италия). Все типы одежды, которые были использованы в обучении, изображены на рисунке 1. Общий объем выборки составил 9000 изображений.

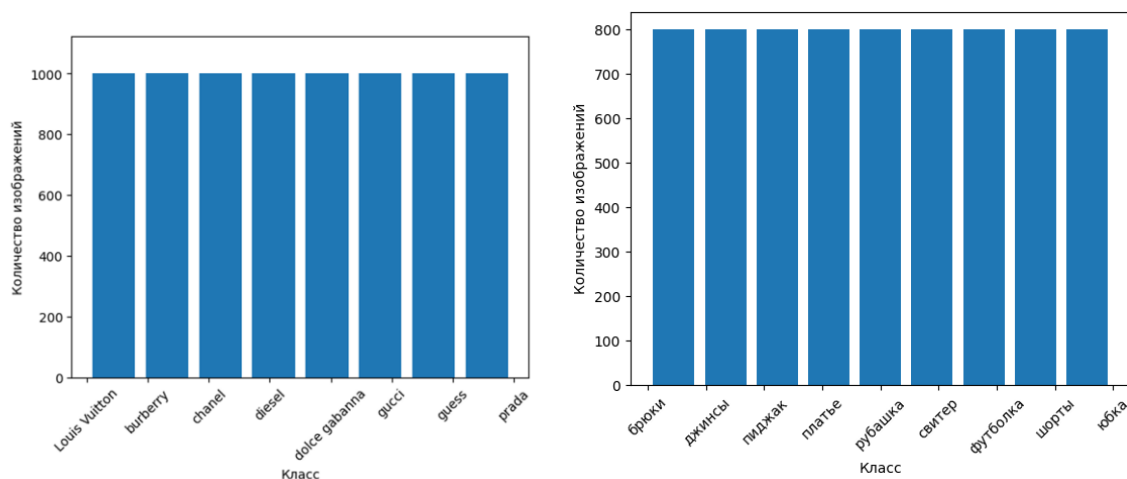


Рис. 1. – Гистограмма распределения количества изображений относительно брендов и типов одежд

К изображениям были предъявлены следующие требования:

Требования к иерархии каталога с изображениями:

1. Директории соответствуют следующей иерархии: <страна>-><бренд>-><тип одежды>;
2. Внутри последней директории (<тип одежды>) располагаются сами изображения;

3. Строго соответствовать иерархии (то есть не перепутать типы одежды, принадлежность одежды бренду/стране);

4. Директории не дублируются;

Требования к изображению:

1. На фото только 1 человек;

2. Выгружать из официальных источников (например: официальный сайт бренда, сайт lamoda, цума и т. п.);

3. Человек на фото должен быть расположен в профиль;

4. Название изображения должно соответствовать следующему шаблону: <бренд одежды> \_<тип одежды>\_<порядковый номер изображения> (отчет начинаем с 1);

5. На каждый тип одежды нужно не менее 150 изображений.

Для работы было собрано более 10 000 изображений. Примеры изображений можно видеть на рисунках 2, 3.



Рис. 2. – Пример фото для набора данных. На фото модель изображен в пиджаке от бренда Guess [7]



Рис. 3. – Пример фото для набора данных. На фото модель изображен в свитере от бренда Burberry[8]

## Модель

В данной работе было решено использовать архитектуру нейронной сети ViT (Vision Transformer) - модель для работы с изображениями, которая была представлена специалистами из Google Brain [9, 10]. В случае с изображениями ViT работает не с отдельными пикселями, а с квадратиками, на которые делит изображение и которые содержат группы пикселей, — это делает обработку более быстрой и при этом не снижает точности. В качестве представителя архитектуры трансформеров была выбрана модель vit-base-patch16-224 от Google, предобученная на 15 миллионах изображений, которые включают в себя около 23 тысяч классов. Благодаря системе многоголового внимания, он не создает иерархию сверточных фильтров, выделяющих признаки, а разбивает изображение на прямоугольные сегменты и воспринимает их одновременно, вследствие чего редуцируется один из недостатков сверточной модели – проблема с выявлением зависимостей между далеко расположенными пикселями. В случае с классификацией брендов: на фотографиях, где модели демонстрируют брюки или джинсы, может также присутствовать обычная футболка или рубашка, однако благодаря механизмам внимания, трансформер может выделить фрагменты изображения с брендовыми вещами и строить предсказания, не фокусируясь на побочных деталях [11, 12].

Исходный датасет был разделен на две части: только бренды и только типы. Далее каждая часть была разбита еще на три части: тренировочная, валидационная и проверочная в соотношении 7:1, 5:1,5. Все фотографии подверглись уменьшению до 50 176 пикселей (224x224) в целях более быстрого обучения, а также из-за ограничений оперативной и видеопамяти.

Перед обучением модели, схема которой изображена на рисунке 4, все изображения из соответствующего датасета были преобразованы в черно-белый вид, чтобы на распознавание не влиял цвет одежды, также была

---

использована аугментация данных: поворот изображения на случайный угол, зеркальное преобразование. Помимо этого, все фотографии были нормализованы – координаты пикселей приведены к промежутку  $[0, 1]$ .

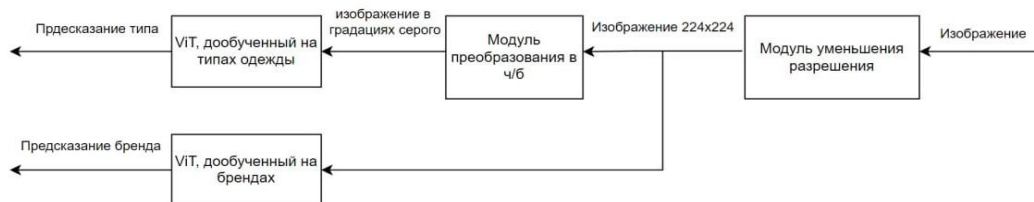


Рис. 4. – Схема модели определения бренда одежды

### Интерпретация результата

В результате обучения модели удалось получить наибольшую точность в 85% для модели, классифицирующей бренд одежды и 74% для модели, классифицирующей тип одежды, показатели производительности приведены в таблицах 1 и 2 соответственно.

Таблица № 1

Показатели производительности классификации бренда одежды

№ п/п	Метрика	Значение
1	precision	0.8493864483392404
2	recall	0.8495867768595041
3	accuracy	0.8495867768595041
4	f1	0.8492554164674724
5	macro average precision	0.8495406491342242
6	macro average recall	0.85078125
7	macro average f1	0.8492554164674724
8	weighted average precision	0.8493864483392404
9	weighted average recall	0.8495867768595041
1	weighted average f1	0.8485468056992578

0		
---	--	--

Таблица № 2

Показатели производительности классификации бренда одежды

№ п/п	Метрика	Значение
1	precision	0.7397420189591004
2	recall	0.737037037037037
3	accuracy	0.737037037037037
4	f1	0.7353050474426417
5	macro average precision	0.7397420189591004
6	macro average recall	0.737037037037037
7	macro average f1	0.7353050474426417
8	weighted average precision	0.7397420189591004
9	weighted average recall	0.737037037037037
1 0	weighted average f1	0.7353050474426417

На рисунках 5, 6 приведены матрицы ошибок для модели определения бренда одежды и типа одежды соответственно. Изучая матрицу ошибок распознавания брендов, мы можем отметить, что бренд Burberry имеет больше схожих черт с Dolce&Gabbana, нежели с Diesel или Louis Vuitton. Также можно увидеть, что модель отмечает схожесть британского дома моды Burberry и парижского Chanel, что объясняется утонченностью концепций и любовью этих брендов к винтажным мотивам. Burberry, как и Chanel, относятся к модным домам класса люкс, которые часто выпускают продукции, вдохновленные предыдущими (классическими) коллекциями.

Еще один интересный момент, который можно заметить, изучая матрицу - это схожесть бренда Guess и Dolce&Gabbana, хотя бренды

относятся к совершенно разным ценовым категориям. Изучая предметы одежды брендов, можно предположить, что их схожесть определяется использованием ярких узоров, неким винтажным духом, сочетанием многих оттенков цвета на одном предмете.

Американский бренд Guess известен своими джинсовыми атрибутами одежды, традиционно выпускают разнообразные джинсы, куртки, топы. Итальянский дом хоть и представляет премиум-сегмент, но также создает джинсовые вещи, которые пользуются большой популярностью. Dolce&Gabbana имеет концептуальную и классическую коллекции джинсов, последняя очень схожа с концепцией Guess. Концептуальная коллекция итальянского дома выражается через яркие узоры, стразы, интересное сочетание различных видов ткани на одном изделии. В коллекции Guess обычно представлена джинса в классическом варианте, но также выпускаются изделия из ярких цветов (розовый, зеленый), со стразами или вышивками.

Также матрица схожести позволяет проанализировать схожесть типов одежды. Так, например, модель определяет схожесть брюк и джинсов или же платья и юбки. При этом есть предсказания, которые вызывают вопросы. А именно, модель указывает на схожесть таких типов, как: платье и джинсы, пиджак и брюки.

Можно предположить, что схожесть платья и юбки объясняется одинаковым фасоном низа (за основу низа платья берется край юбки). По рисунку 6 видно, что алгоритм принимает юбку за платье чаще, чем платье за юбку.

Объяснить идентификацию пиджака в качестве брюк можно схожим силуэтом этих вещей. Скорее всего, модель распознает штанину брюк/джинсов как рукав пиджака/рубашки и тем самым происходит ошибочное распознавание типов одежды. Кроме того, по рисунку 6 мы

---



видим, что количество ошибок выше при распознавании джинсов и брюк, а не при распознавании пиджаков и рубашек.

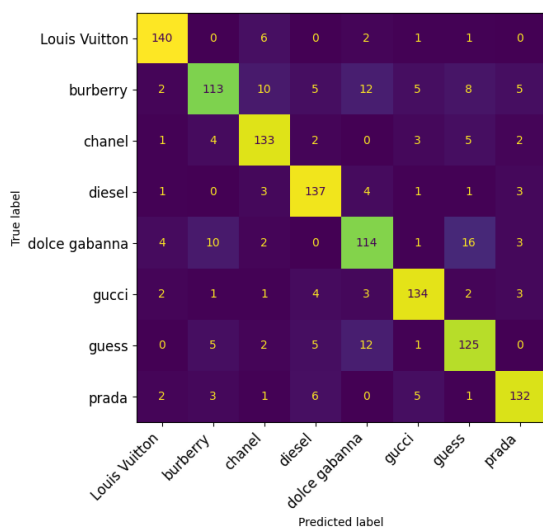


Рис. 5. – матрица ошибок модели определения бренда одежды

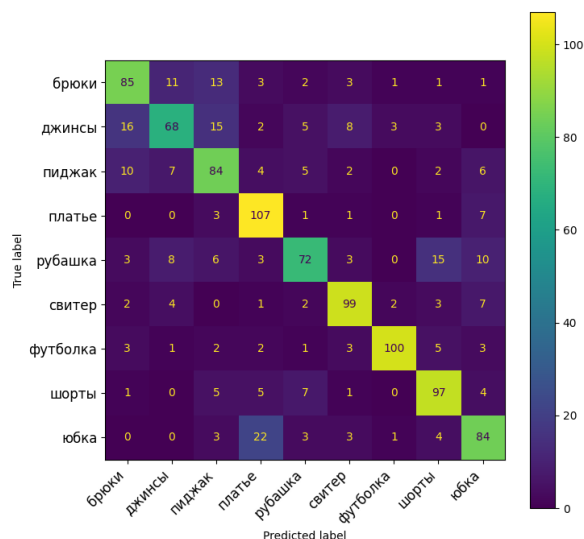


Рис. 6. – матрица ошибок модели определения типа одежды

### Заключение

Индустрия моды развивается очень быстрыми темпами, поэтому необходимость в передовых идеях и технологиях с каждым годом возрастает. Речь идет не только о новых инструментах создания, проектирования продуктов, но и о предсказываниях тенденций, взаимодействии с конечным покупателем.

На данный момент искусственный интеллект представляет возможность виртуальной примерки, позволяя подобрать размер, образ. Образ подбирается, основываясь на множестве критериев: цветовая гамма, сочетаемость типов, подбор аксессуаров, тренды.

Набирает огромную популярность digital fashion - это пространство моды будущего, целью которого является создание виртуальных атрибутов одежды. Это течение позволяет творить интересные и концептуальные вещи, примерять эти вещи, а также при необходимости воплотить задумку вживую.

Возможность автоматического предсказания бренда по фотографии является актуальной задачей, которая даст возможность подбирать похожие товары, создавать с помощью современных технологий новые и уникальные продукты, ускорять процесс разработки и дизайна. В данной работе была предложена модель, которая использует архитектуру нейронной сети ViT (vit-base-patch16-224 от Google). Для обучения модели датасет был разбит на две части: бренды и типы. В свою очередь, каждая из этих частей была разбита еще на три части: тренировочная, валидационная и проверочная.

В результате проделанной работы удалось добиться поставленной цели и выполнить все задачи, которые были обозначены. В итоге, были получены следующие показатели: точность модели, равная 85% для модели, классифицирующей бренд одежды и 74% для модели, классифицирующей тип одежды. Таким образом, разработанный алгоритм и программа позволяют предсказывать тип одежды и бренд по фотографии.

### Литература

1. Generative AI: Unlocking the future of fashion // mckinsey.com: официальный сайт. URL: [mckinsey.com/industries/retail/our-insights/generative-ai-unlocking-the-future-of-fashion](https://mckinsey.com/industries/retail/our-insights/generative-ai-unlocking-the-future-of-fashion).

2. Artificial Intelligence in Fashion: Reshaping the Entire Industry // 3dlook.ai: официальный сайт. URL: [3dlook.ai/content-hub/artificial-intelligence-in-fashion/](https://3dlook.ai/content-hub/artificial-intelligence-in-fashion/).

3. Liu K., Chen T., Chen C. MVC: A Dataset for View-Invariant Clothing Retrieval and Attribute Prediction // ACM on International Conference on Multimedia Retrieval. 2016. URL: [researchgate.net/publication/303901989\\_MVC\\_A\\_Dataset\\_for\\_View-Invariant\\_Clothing\\_Retrieval\\_and\\_Attribute\\_Prediction](https://researchgate.net/publication/303901989_MVC_A_Dataset_for_View-Invariant_Clothing_Retrieval_and_Attribute_Prediction).

4. Al-Halah Z., Stiefelhagen R., Grauman K. Fashion Forward: Forecasting Visual Style in Fashion // Conference: IEEE International Conference on

---

Computer Vision. 2023. URL:  
[researchgate.net/publication/317010098\\_Fashion\\_Forward\\_Forecasting\\_Visual\\_Style\\_in\\_Fashion/](https://researchgate.net/publication/317010098_Fashion_Forward_Forecasting_Visual_Style_in_Fashion/).

5. Renaningtyas L., Dwitasari P., Ramadhani N. Implementing The Use of AI for Analysis and Prediction in the Fashion Industry // The Sustainability in Creative Industries and the Urban City. 2023. Vol. 7. №1. URL:  
[researchgate.net/publication/367555308\\_Implementing\\_The\\_Use\\_of\\_AI\\_for\\_Analysis\\_and\\_Prediction\\_in\\_the\\_Fashion\\_Industry/](https://researchgate.net/publication/367555308_Implementing_The_Use_of_AI_for_Analysis_and_Prediction_in_the_Fashion_Industry/).

6. Kenan E. A., Lim J. H., Tham J. Y., Kassim A. Multi-Attribute Similarity Learning Towards Attribute-based Fashion Search // Winter Conference on Applications of Computer Vision. 2018. URL:  
[researchgate.net/publication/324728522\\_Efficient\\_Multi-Attribute\\_Similarity\\_Learning\\_Towards\\_Attribute-based\\_Fashion\\_Search](https://researchgate.net/publication/324728522_Efficient_Multi-Attribute_Similarity_Learning_Towards_Attribute-based_Fashion_Search).

7. Men's coats and jackets // Guess: официальный сайт. URL: [guess.eu/en-sk/guess/men/clothing/coats-and-jackets/front-patch-blazer-blue/M3RN24WF4A0-G7V2.html](https://guess.eu/en-sk/guess/men/clothing/coats-and-jackets/front-patch-blazer-blue/M3RN24WF4A0-G7V2.html).

8. Men's Designer Hoodies & Sweatshirts // In-cc: официальный сайт. URL: [in-cc.com/en-gb/men/clothing/sweatshirts/burberry-logo-hooded-sweatshirt-in-black-bur0140001blk.html](https://in-cc.com/en-gb/men/clothing/sweatshirts/burberry-logo-hooded-sweatshirt-in-black-bur0140001blk.html).

9. Vision Transformers vs. Convolutional Neural Networks // medium.com: официальный сайт. URL: [medium.com/?faheemrustamy/vision-transformars-vs-convolutional-neural-networks-5fe8f9e18efc](https://medium.com/?faheemrustamy/vision-transformars-vs-convolutional-neural-networks-5fe8f9e18efc).

10. ViT — на кухне фаворит // Хабр: официальный сайт. URL: [habr.com/ru/articles/599677/](https://habr.com/ru/articles/599677/).

11. Белякова А. Ю., Беляков Ю. Д., Замятин П. С. Решение задачи распознавания объектов и инцидентов на фотоматериалах, полученных с беспилотных летательных аппаратов с использованием методов глубокого



обучения // Инженерный вестник Дона. 2021. №5. URL:  
ivdon.ru/ru/magazine/archive/n5y2021/6985.

12. Пучков Е.В. Сравнительный анализ алгоритмов обучения искусственной нейронной сети // Инженерный вестник Дона. 2013. №4. URL:  
ivdon.ru/ru/magazine/archive/n4y2013/2135.

### References

1. Generative AI: Unlocking the future of fashion. URL:  
mckinsey.com/industries/retail/our-insights/generative-ai-unlocking-the-future-of-fashion.

2. Artificial Intelligence in Fashion: Reshaping the Entire Industry. URL:  
3dlook.ai/content-hub/artificial-intelligence-in-fashion/.

3. Liu K., Chen T., Chen C. MVC: A Dataset for View-Invariant Clothing Retrieval and Attribute Prediction. ACM on International Conference on Multimedia Retrieval. 2016. URL:  
researchgate.net/publication/303901989\_MVC\_A\_Dataset\_for\_View-Invariant\_Clothing\_Retrieval\_and\_Attribute\_Prediction.

4. Al-Halah Z., Stiefelhagen R., Grauman K. Fashion Forward: Forecasting Visual Style in Fashion. Conference: IEEE International Conference on Computer Vision. 2023. URL:  
researchgate.net/publication/317010098\_Fashion\_Forward\_Forecasting\_Visual\_Style\_in\_Fashion/.

5. Renaningtyas L., Dwitasari P., Ramadhani N. Implementing The Use of AI for Analysis and Prediction in the Fashion Industry. The Sustainability in Creative Industries and the Urban City. 2023. Vol. 7. №1. URL:  
researchgate.net/publication/367555308\_Implementing\_The\_Use\_of\_AI\_for\_Analysis\_and\_Prediction\_in\_the\_Fashion\_Industry/.

6. Kenan E. A., Lim J. H., Tham J. Y., Kassim A. Multi-Attribute Similarity Learning Towards Attribute-based Fashion Search. Winter Conference on

---



Applications of Computer Vision. 2018. URL:  
[researchgate.net/publication/324728522\\_Efficient\\_Multi-Attribute\\_Similarity\\_Learning\\_Towards\\_Attribute-based\\_Fashion\\_Search](https://researchgate.net/publication/324728522_Efficient_Multi-Attribute_Similarity_Learning_Towards_Attribute-based_Fashion_Search).

7. Men's coats and jackets // Guess: официальный сайт. URL : [guess.eu/en-sk/guess/men/clothing/coats-and-jackets/front-patch-blazer-blue/M3RN24WF4A0-G7V2.html](https://guess.eu/en-sk/guess/men/clothing/coats-and-jackets/front-patch-blazer-blue/M3RN24WF4A0-G7V2.html).

8. Men's Designer Hoodies & Sweatshirts // In-cc: официальный сайт. URL: [in-cc.com/en-gb/men/clothing/sweatshirts/burberry-logo-hooded-sweatshirt-in-black-bur0140001blk.html](https://in-cc.com/en-gb/men/clothing/sweatshirts/burberry-logo-hooded-sweatshirt-in-black-bur0140001blk.html).

9. Vision Transformers vs. Convolutional Neural Networks. URL: [medium.com/?faheemrustamy/vision-transformers-vs-convolutional-neural-networks-5fe8f9e18efc](https://medium.com/?faheemrustamy/vision-transformers-vs-convolutional-neural-networks-5fe8f9e18efc).

10. ViT — на кухне favorit [ViT in the kitchen favorite]. URL: [habr.com/ru/articles/599677/](https://habr.com/ru/articles/599677/).

11. Belyakova A. Yu., Belyakov Yu. D., Zamyatin P. S. Inzhenernyj vestnik Dona. 2021, №5. URL: [ivdon.ru/ru/magazine/archive/n5y2021/6985](https://ivdon.ru/ru/magazine/archive/n5y2021/6985).

12. Puchkov E.V. Inzhenernyj vestnik Dona. 2013. №4. URL: [ivdon.ru/ru/magazine/archive/n4y2013/2135](https://ivdon.ru/ru/magazine/archive/n4y2013/2135).

**Дата поступления: 13.12.2023**

**Дата публикации: 25.01.2024**