

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ / ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

DOI: <https://doi.org/10.60797/IRJ.2024.143.121>

ИЗУЧЕНИЕ ВЛИЯНИЯ ПАРАМЕТРОВ ПЕРОВСКИТНЫХ СОЛНЕЧНЫХ БАТАРЕЙ НА ИХ ЭФФЕКТИВНОСТЬ ПРИ ПОМОЩИ МАШИННОГО ОБУЧЕНИЯ

Научная статья

Новоселов И.Э.¹, Смирнов А.А.^{2,*}, Жидков И.С.³

¹ ORCID : 0009-0005-7078-2663;

² ORCID : 0000-0002-7312-7219;

³ ORCID : 0000-0001-8727-4730;

^{1,2} Уральский Федеральный Университет имени первого Президента России Б. Н. Ельцина, Екатеринбург, Российская Федерация

³ Институт физики металлов УрО РАН, Екатеринбург, Российская Федерация

* Корреспондирующий автор (a.a.smirnov[at]urfu.ru)

Аннотация

В работе представлен подход к оптимизации перовскитных солнечных элементов (PSC) с использованием машинного обучения. Цель исследования – поиск наилучших параметров PSC, обеспечивающих высокую эффективность преобразования энергии (PCE). Для обучения алгоритмов был создан набор данных, содержащий информацию о признаках PSC, таких как толщина перовскитного слоя, тип электронно-транспортного слоя и т.д., а также целевые переменные. Были применены различные алгоритмы машинного обучения, результаты показали, что низкую ошибку на обоих этапах обучения (регрессия, где целевая переменная PCE, и многоцелевая регрессия, где четыре переменных: PCE, Voc, Jsc и FF) показали XGBoosting, CatBoost и Random Forest. Наибольшую важность для PCE имеют такие параметры, как Voc, Jsc и FF, меньшими по важности: Pero th, ETL, Cs, MA, FA, I и HTL.

Ключевые слова: перовскитные солнечные элементы, машинное обучение, оптимизация, эффективность преобразования энергии, PCE.

THE STUDY OF THE PEROVSKITE SOLAR CELL PARAMETER INFLUENCE ON THEIR EFFICIENCY BY USING MACHINE LEARNING

Research article

Novoselov I.E.¹, Smirnov A.A.^{2,*}, Zhidkov I.S.³

¹ ORCID : 0009-0005-7078-2663;

² ORCID : 0000-0002-7312-7219;

³ ORCID : 0000-0001-8727-4730;

^{1,2} Ural Federal University named after the first President of Russia B. N. Yeltsin, Ekaterinburg, Russian Federation

³ Institute of Metal Physics, Ural Branch of the Russian Academy of Sciences, Ekaterinburg, Russian Federation

* Corresponding author (a.a.smirnov[at]urfu.ru)

Abstract

This paper presents an approach to optimize perovskite solar cells (PSC) by using machine learning algorithms. The aim of the study is to find the best PSC parameters that provide high power conversion efficiency (PCE). To train the algorithms, the dataset with PSC feature information (the perovskite layer thickness, the electron transport layer type, the target variables, etc.) was created. Various machine learning algorithms were applied, the XGBoosting, CatBoost and Random Forest showed the low error in both learning stages (regression where the target variable is PCE, multi-target regression where the four variables are PCE, Voc, Jsc and FF as well). The most important variables for PCE are Voc, Jsc and FF, the less important ones – Pero th, ETL, Cs, MA, FA, I and HTL.

Keywords: perovskite solar cells, machine learning, optimization, power conversion efficiency, PCE.

Введение

Машинное обучение становится все более популярным инструментом для исследований в области материаловедения. При помощи методов машинного обучения можно оптимизировать поиск исходных структур соединений, что способствует созданию новых материалов с определёнными свойствами, их улучшению, ускорению процесса разработки новых инновационных решений, сокращению затрат ресурсов и времени на исследование [1], [2], [3], [4], [5]. Перовскиты – класс материалов с широким спектром потенциальных применений, включая солнечные батареи, светодиоды, детекторы и др. [6], [7], [8], [9], [10]. В данной работе рассматривается применение алгоритмов машинного обучения для решения задачи поиска наилучших параметров для получения соединения с высокой эффективностью преобразования энергии (PCE).

Методы и принципы исследования

Набор данных для обучения алгоритмов машинного обучения для решения задачи поиска наилучших параметров для получения соединения с высокой эффективностью (PCE) был получен при помощи моделирования PSC различной архитектуры (зарядовые слои, перовскиты, толщины перовскитного слоя) в программном пакете SCAPS [11].

Перед непосредственной настройкой алгоритмов и их обучением требуется предварительная обработка набора данных, поскольку полученный набор имеет пропуски в данных. На рисунке 1 представлен фрагмент датасета после обработки.

	Pero th, nm	ETL	Cs	MA	FA	I	Br	HTL	FTO	Voc	Jsc	FF	PCE
0	400	2	0.0	1.00	0.00	0.00	3.00	11	0	1.768197	6.541885	73.5625	8.5092
1	500	5	0.0	0.05	0.95	0.95	0.05	11	0	1.129240	24.014978	57.0812	15.4797
2	800	2	0.0	1.00	0.00	3.00	0.00	7	1	1.146535	25.711566	83.7178	24.6794
3	300	3	0.0	0.50	0.50	3.00	0.00	2	0	1.190994	23.257929	83.4099	23.1046
4	400	2	0.0	0.15	0.85	0.85	0.15	3	0	1.331036	21.053159	85.9299	24.0797

Рисунок 1 - Фрагмент датасета
DOI: <https://doi.org/10.60797/IRJ.2024.143.121.1>

Примечание: Примечание: Pero th, nm – толщина перовскитного слоя, nm; ETL – тип электронно-транспортного слоя; Cs – количество Cs в A-положении; MA – количество MA в A-положении, FA – количество FA в A-положении, I – количество йода в X-положении, Br – количество брома в X-положении, HTL – тип дырочно-транспортного слоя, FTO – присутствие слоя FTO, Voc – напряжение холостого хода, Jsc – ток короткого замыкания, FF – фактор заполнения, PCE – эффективность преобразования энергии

Тип зарядово-транспортных слоёв задавался цифрами, кодировка представлена в таблице 1.

Таблица 1 - Кодировка типов зарядово-транспортных слоёв

DOI: <https://doi.org/10.60797/IRJ.2024.143.121.2>

Тип ETL	Номер в датасете	Тип HTL	Номер в датасете
PCBM	1	C ₆₀	1
TiO ₂	2	Cu ₂ O	2
ZnO	3	CuI	3
CdS	4	Spiro-OMeTAD	4
WO ₃	5	CuSCN	5
WS ₂	6	NiO	6
IGZO	7	PЗНТ	7
SnO ₂	8	CuSbS ₂	8
–	–	PEDOT:PSS	9
–	–	CuO	10
–	–	MoO ₃	11

Обучение было разбито на два этапа:

1. Решение задачи регрессии (несколько признаков и одна целевая переменная).
2. Решение задачи многоцелевой регрессии (несколько признаков и несколько целевых переменных).

Всего получилось 7182 объекта, которые будут анализироваться по 12 признакам ('Pero th', 'ETL', 'Cs', 'MA', 'FA', 'I', 'Br', 'HTL', 'FTO', 'Voc', 'Jsc', 'FF') и 1 целевой переменной ('PCE') для первого этапа, а также по 9 признакам ('Pero th', 'ETL', 'Cs', 'MA', 'FA', 'I', 'Br', 'HTL', 'FTO') и 4 целевым переменным ('Voc', 'Jsc', 'FF', 'PCE'), значение которых и будут предсказываться на втором этапе.

Перед загрузкой данных в алгоритм была построена тепловая карта взаимокорреляций Пирсона для признаков (см. рис. 2). Корреляция – важнейший фактор, лежащий в основе анализа данных. Он сообщает, как переменные в наборе данных связаны друг с другом.

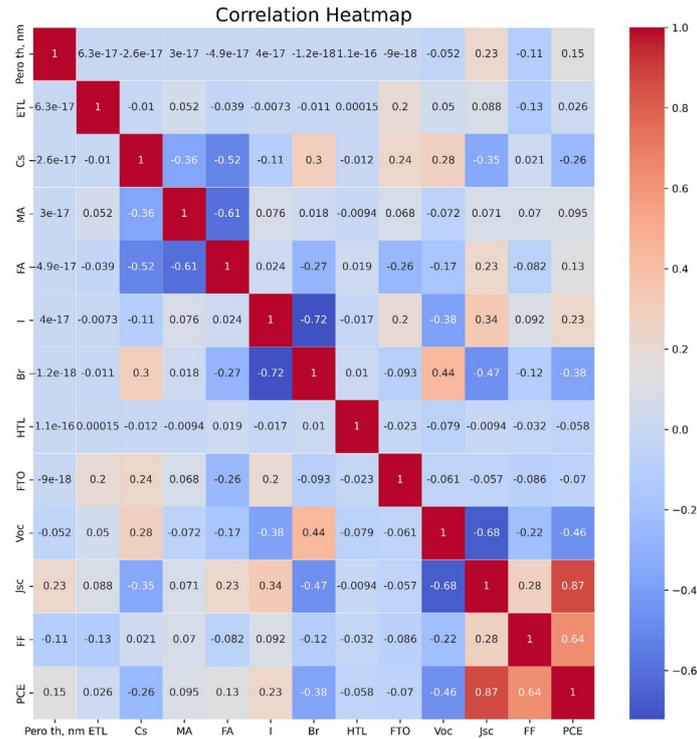


Рисунок 2 - Тепловая карта матрицы корреляции признаков и целевых переменных

DOI: <https://doi.org/10.60797/IRJ.2024.143.121.3>

Поскольку коэффициент корреляции Пирсона обладает несколькими недостатками: работает только с непрерывными переменными, учитывает только линейную связь между ними, чувствителен к выбросам, использовалась PhiK (φк) корреляция, лишенная данных недостатков (см. рис. 3).

Согласно рисунку 3 сильная корреляция (больше 0,6) со значениями целевых переменных наблюдается у следующих признаков: HTL, Cs, MA, FA, I, Br.

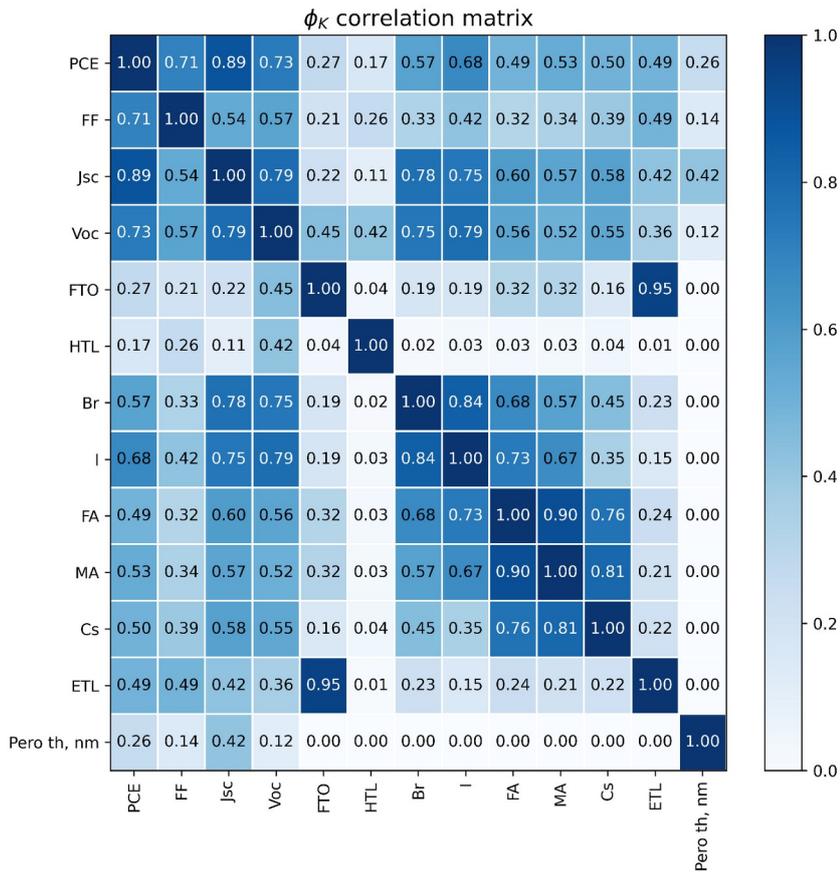


Рисунок 3 - Матрица коэффициентов корреляции ϕ_k
 DOI: <https://doi.org/10.60797/IRJ.2024.143.121.4>

Однако при оценке корреляций необходимо также исследовать статистическую значимость данной корреляции.

Для определения значимости, необходимо учесть значение р-критерия. Значение р менее 0,05 обычно считается статистически значимым, и в этом случае нулевую гипотезу следует отклонить на взятом уровне значимости, то есть связь между двумя наблюдаемыми событиями существует. Значение р больше 0,05 означает, что нулевая гипотеза не отвергается, следовательно, не существует связи между двумя наблюдаемыми событиями.

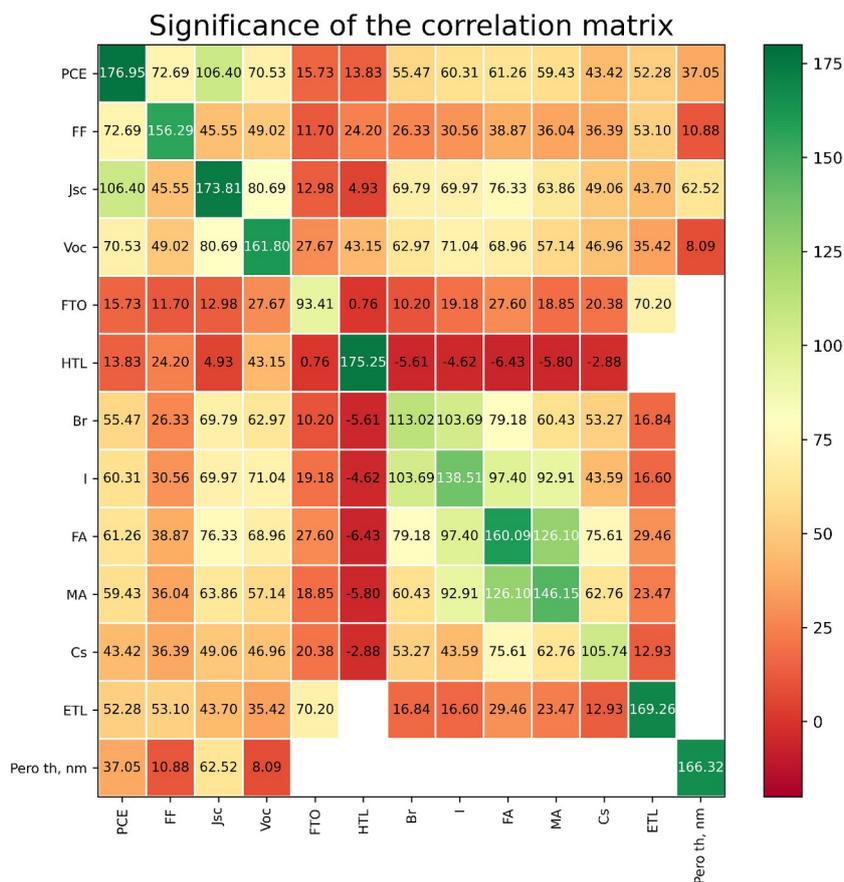


Рисунок 4 - Значимость корреляции
DOI: <https://doi.org/10.60797/IRJ.2024.143.121.5>

Согласно рисунку 4 почти все коэффициенты корреляции признаков (кроме 'Pero th, nm', 'HTL', 'FTO', 'Cs', 'ETL') с целевыми переменными статистически значимы при выбранном уровне значимости. Также стоит отметить, что в матрице отсутствуют некоторые значения коэффициентов значимости корреляции, поскольку нет зависимости между данными переменными.

В качестве алгоритмов машинного обучения применялись не только базовые алгоритмы, такие как линейная регрессия (англ. Linear Regression), K-ближайших соседей (англ. K-Nearest Neighbors), метод опорных векторов (англ. Support vector Machine), дерево решений (англ. Decision Tree), но и современные ансамблевые методы, такие как случайный лес (англ. Random Forest) и методы основанные на бустинге (англ. Extreme Gradient Boosting (XGBoosting), Light Gradient Boosting Machine (LightGBM), Gradient Boosting Regressor). Для тонкой настройки алгоритмов, а также для борьбы с переобучением необходимо использовать разбиение набора данных на обучающую и тестовую выборки, кросс-валидацию, различные регуляризации, масштабирование признаков, подбор гиперпараметров при помощи метода GridSearchCV и другие приемы для увеличения обобщающей способности алгоритмов. В результате на основе выбранных алгоритмов можно будет оценить важность выбранных признаков, а также оценить точность алгоритмов по метрикам средняя квадратичная ошибка (англ. mean squared error, MSE) и средняя абсолютная ошибка (англ. mean absolute error, MAE).

Далее представлены результаты обучения алгоритмов, а также важности признаков для первого этапа.

Таблица 2 - Результаты обучения алгоритмов для одноцелевой регрессии

DOI: <https://doi.org/10.60797/IRJ.2024.143.121.6>

Алгоритм	MSE	MAE
Support Vector Machine	1,63	0,93
Linear Regression	1,56	0,92
Ridge	1,88	1,05
Lasso	3,86	1,54
K-nearest neighbors	0,75	0,54
Decision Tree	0,16	0,22
Random Forest	0,05	0,14
XGBoosting	0,04	0,15

LightGBM	0,05	0,16
CatBoost	0,02	0,10
Gradient Boosting	0,15	0,30

Исходя из данных таблицы 2, низкую ошибку показали ансамблевые методы (Random Forest и методы, основанные на бустинге).

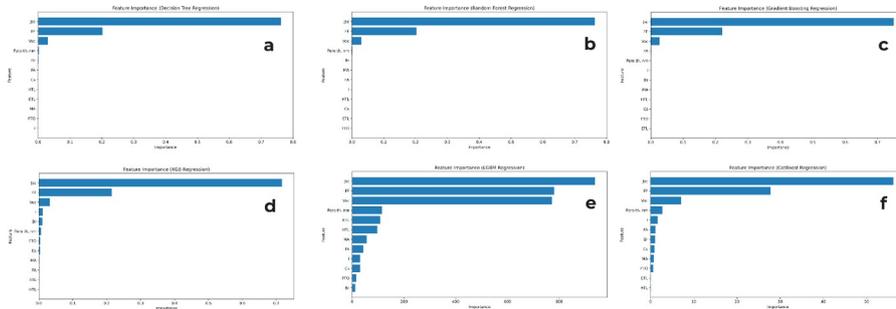


Рисунок 5 - Важность признаков в алгоритмах
DOI: <https://doi.org/10.60797/IRJ.2024.143.121.7>

Примечание: Примечание: a – decision tree regression, b – random forest regression, c – gradient boosting regression, d – xgboosting regression, e – lightgbm regression, f – catboost regression

Почти все алгоритмы показали примерно одинаковые значения важности признаков (см. рис. 5). Из рисунков видно, что наибольшую важность вносят следующие признаки: 'Voc', 'Jsc', 'FF'. Методы, основанные на бустинге LightGBM и CatBoost показали, что помимо вышеупомянутых признаков, также важны, но в меньшей степени, такие признаки как 'Pero th', 'ETL', 'Cs', 'MA', 'FA', 'T', 'HTL'.

Далее представлены результаты обучения алгоритмов для второго этапа.

Таблица 3 - Результаты обучения алгоритмов для многоцелевой регрессии

DOI: <https://doi.org/10.60797/IRJ.2024.143.121.8>

Алгоритм	MSE				MAE			
	Voc	Jsc	FF	PCE	Voc	Jsc	FF	PCE
SVM	0,05	62,07	271,67	58,65	0,15	5,61	10,83	6,52
Linear Regression	0,04	39,34	205,79	48,39	0,14	4,65	11,72	5,75
Ridge	0,04	39,34	205,79	48,40	0,14	4,65	11,72	5,74
Lasso	0,05	43,55	212,35	50,66	0,17	5,27	12,10	6,05
KNN	0,03	22,52	127,06	22,75	0,11	3,29	8,45	3,72
Decision Tree	0,02	8,02	135,05	18,12	0,09	2,24	8,53	3,37
Random Forest	0,00	0,11	4,96	0,48	0,01	0,18	1,19	0,43
XGBoosting	0,00	0,05	4,96	0,21	0,01	0,12	1,45	0,32
LightGBM	0,00	0,15	11,84	0,67	0,02	0,22	2,21	0,58
CatBoost	0,00	0,06	6,29	0,31	0,01	0,13	1,64	0,37
Gradient Boosting	0,00	0,98	43,85	4,37	0,05	0,68	5,20	1,59

Исходя из данных табл. 3, самые низкие значения ошибки показали следующие алгоритмы: XGBoosting, CatBoost и Random Forest.

Заключение

В работе были применены методы машинного обучения для поиска наилучших параметров PSC с высоким PCE. Для обучения алгоритмов был создан набор данных, содержащий информацию о PSC, таких как толщина перовскитного слоя, тип электронно-транспортного слоя и т.д., а также целевые переменные: Voc, Jsc, FF и PCE. Были применены различные алгоритмы машинного обучения, включая линейную регрессию, K-ближайших соседей, метод опорных векторов, дерево решений, случайный лес и методы, основанные на бустинге. Результаты показали, что:

1) низкую ошибку на обоих этапах обучения показали XGBoosting, CatBoost и Random Forest;

2) наибольшую важность для PCE имеют такие параметры, как Voc, Jsc и FF;

3) дополнительными важными параметрами, в меньшей степени влияющими на PCE, являются Pero th, ETL, Cs, MA, FA, I и HTL.

Данная работа демонстрирует потенциал применения машинного обучения для оптимизации PSC и может быть использована для ускорения разработки новых материалов с высокой PCE.

Финансирование

Исследование выполнено при поддержке РФФ (грант 22-61-00047).

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Funding

The study was supported by the Russian Science Foundation (grant 22-61-00047).

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы на английском языке / References in English

1. Morgan D. Opportunities and challenges for machine learning in materials science / D. Morgan, R. Jacobs // *Annual Review of Materials Research*. — 2020. — Vol. 50. — P. 71–103.
2. Cai J. Machine learning-driven new material discovery / J. Cai, X. Chu, K. Xu [et al.] // *Nanoscale Advances*. — 2020. — Vol. 2. — No. 8. — P. 3115–3130.
3. Tao Q. Machine learning for perovskite materials design and discovery / Q. Tao, P. Xu, M. Li [et al.] // *npj Computational Materials*. — 2021. — Vol. 7. — No. 1. — P. 23.
4. Carrasquilla J. Machine learning for quantum matter / J. Carrasquilla // *Advances in Physics: X*. — 2020. — Vol. 5. — No. 1. — P. 1797528.
5. Hart G. L. W. Machine learning for alloys / G. L. W. Hart, T. Mueller, C. Toher [et al.] // *Nature Reviews Materials*. — 2021. — Vol. 6. — No. 8. — P. 730–755.
6. Tu Y. Perovskite solar cells for space applications: progress and challenges / Y. Tu, J. Wu, G. Xu [et al.] // *Advanced Materials*. — 2021. — Vol. 33. — No. 21. — P. 2006545.
7. Zhang L. Advances in the application of perovskite materials / L. Zhang, L. Mei, K. Wang [et al.] // *Nano-Micro Letters*. — 2023. — Vol. 15. — No. 1. — P. 177.
8. Li J. Perovskite single crystals: synthesis, optoelectronic properties, and application / J. Li, Z. Han, Y. Gu [et al.] // *Advanced Functional Materials*. — 2021. — Vol. 31. — No. 11. — P. 2008684.
9. Gladkikh V. Machine learning for predicting the band gaps of ABX₃ perovskites from elemental properties / V. Gladkikh, D. Y. Kim, A. Hajibabaei [et al.] // *The Journal of Physical Chemistry C*. — 2020. — Vol. 124. — No. 16. — P. 8905–8918.
10. Lei L. Metal halide perovskites for laser applications / L. Lei, Q. Dong, K. Gundogdu [et al.] // *Advanced Functional Materials*. — 2021. — Vol. 31. — No. 16. — P. 2010144.
11. Burgelman M. Modelling polycrystalline semiconductor solar cells / M. Burgelman, P. Nollet, S. Degraeve // *Thin solid films*. — 2000. — Vol. 361. — P. 527–532.