УДК 004.8
DOI: 10.65599/UEKF5555

MACHINE LEARNING FOR CLASSIFICATION AND CLUSTER ANALYSIS OF LUMINESCENT MATERIALS WITH AND WITHOUT RARE EARTH ELEMENTS

[1]M.M. Kayumov, [2]Sh.A. Bozorov, [3]A.S. Ashurov
[1]S.U. Umarov Physical–Technical Institute of the National Academy of Sciences of Tajikistan
[2]Tajik Technical University named after academician M.S. Osimi
[3]Institute of Astrophysics of the National Academy of Sciences of Tajikistan

This study presents a data analysis approach for classifying inorganic phosphores and identifying functional substitutes for rare earth elements (RE). Using a dataset of 3911 phosphorous materials, where machine learning methods were applied to classify materials and perform cluster analysis based on their optical properties. The main result was the identification of spectral regions in which activators without RE (Mn2^{+}, Bi3^{+}) exhibit excitation and emission profiles identical to traditional RE (Tb3^{+}, Eu3^{+}) dopants, confirming their potential as stable alternatives for solid-state lighting.

Keywords: *inorganic phosphors, rare-earth elements, machine learning, cluster analysis, spectral overlap, solid-state lighting, functional substitutes.*

ОМӮЗИШИ МОШИНӢ БАРОИ ТАСНИФ ВА ТАҲЛИЛИ ГУРӮҲИ МАВОДИ РАВШАНИДИҲАНДА БО ВА БЕ УНСУРҲОИ НОДИРЗАМИНӢ

М.М. Каюмов, Ш.А. Бозоров, А.С. Ашуров

Ин тадқиқот як сохтори таҳлилӣ ва додаҳои барои таснифи фосфорҳои ғайриорганикӣ ва муайян кардани ивазкунандагони функсионалӣ барои унсурҳои нодирзаминӣ (RE) муаррифӣ мекунад. Бо истифода аз маҷмӯи додаҳои 3911маводҳои фосфордор, мо методҳои омӯзиши мошинӣ барои тасниф ва таҳлили гурӯҳии маводҳо бо назардошти хосиятҳои оптикӣ истифода кардем. Яке аз натиҷаҳои асосӣ муайян кардани минтақаҳои спектралӣ мебошад, ки дар онҳо фаъолнокии ғайри-RE (Mn^{2+}, Bi^{3+}) ки паҳншавии нурро нишон дар сатҳи хуб нишон медиҳанд, бо допантҳои анъанавии RE (Tb^{3+}, Eu^{3+}) баробаранд, ки қобилияти онҳоро ҳамчун ивазкунандаи устувор барои равшанидиҳи тасдиқ мекунад.

Калидвожаҳо: *фосфорҳои ғайриорганикӣ, унсурҳои нодирзаминӣ, омӯзиши мошинӣ, таҳлили гурӯҳӣ, спектралӣ, ивазкунандагони функсионалӣ.*

МАШИННОЕ ОБУЧЕНИЕ ДЛЯ КЛАССИФИКАЦИИ И КЛАСТЕРНОГО АНАЛИЗА ЛЮМИНЕСЦЕНТНЫХ МАТЕРИАЛОВ С РЕДКОЗЕМЕЛЬНЫМИ ЭЛЕМЕНТАМИ И БЕЗ НИХ

М.М. Каюмов, Ш.А. Бозоров, А.С. Ашуров

Данное исследование представляет подход, основанный на анализе данных, для классификации неорганических фосфоров и выявления функциональных заменителей редкоземельных элементов (RE). Используя набор данных из 3911 фосфорных материалов, где применили методы машинного обучения для классификации материалов и проведения кластерного анализа на основе их оптических свойств. Основным результатом стало выявление спектральных областей, в которых активаторы без RE (Mn^{2+}, Bi^{3+}) демонстрируют профили возбуждения и излучения, идентичные традиционным допантам RE (Tb^{3+}, Eu^{3+}), подтверждая их потенциал как устойчивые альтернативы для твердотельного освещения.

Ключевые слова: *неорганические фосфоры, редкоземельные элементы, машинное обучение, кластерный анализ, спектральное перекрытие, твердотельное освещение, функциональные заменители.*

1. Introduction

Phosphor materials are crucial components in various advanced technologies, including solid-state lighting (SSL), displays, and medical imaging. Since the mid-20th century, rare-earth (RE) elements such as europium (Eu), terbium (Tb), and cerium (Ce) have dominated this field due to their sharp f-f (transition of an electron from one 4f orbital to another within the same ion) emission lines, which provide high color purity and quantum efficiency [1]. However, the performance of these materials is heavily influenced by their composition, and the reliance on RE elements presents significant geopolitical and economic challenges due to supply chain volatility [2]. Consequently, the search for rare-earth-free (non-RE) alternatives has become a critical priority in materials science [3].

Given the vast and growing number of synthesized phosphor materials, computational methods, such as machine learning (ML), offer efficient ways to analyze, classify, and discover patterns within large datasets. Recent reviews have highlighted the capacity of ML to accelerate phosphor discovery by establishing complex mapping among composition, structure, and optical properties [4, 5].

This paper details an analytical framework developed to classify phosphor materials based on the presence of rare-earth elements and to cluster them according to their fundamental properties. The primary goals are to:

- accurately predict the presence of rare-earth dopants;
- identify natural groupings of materials based on their physicochemical and optical characteristics;
- visualize these classifications and clusters to extract meaningful scientific insights.

2. Methodology

2.1 Data Collection and Preprocessing

The dataset on the optical properties of inorganic phosphors analyzed in this work was obtained from the publicly available Figshare repository [16]. It contains 3911 records and 11 columns, including key parameters such as host, dopant, maximum emission wavelength (nm), and excitation source (nm). Missing values were imputed by replacing numerical columns with their medians and categorical columns with their modes. A new binary feature, has_rare_earth , was created to indicate whether the dopant is a rare earth element (e.g., Eu, Ce, Tb, Dy). Numeric features were standardized using StandardScaler to ensure uniform scaling.

2.2 Classification and Clustering Models

To validate the distinct physical signatures of RE and non-RE materials, a RandomForestClassifier was trained on the standardized features. Simultaneously, K-Means clustering was applied to identify natural groupings within the data. The optimal number of clusters (K=4) was determined using the Elbow Method, and Principal Component Analysis (PCA) was used to visualize the resulting clusters in two dimensions.

2.3 Identification of Functional Overlap

A density-based filtering approach was applied to the dataset to identify λ_{ex}, λ_{em}) coordinates where both RE and non-RE populations intersect with the highest frequency. This allowed for the isolation of specific "functional equivalence" zones.

3. Results

3.1 Rare-Earth Element Classification

The RandomForest model demonstrated excellent results in classifying materials based on the presence of rare earth elements, as shown in Figure 1, achieving an accuracy of 0.9877 and an F1 score of 0.9928. To train and evaluate the model, the original dataset was divided into training (75%) and testing (25%) sets. The confusion matrix shown in Figure 1 corresponds to the testing set (n = 978) and confirms the minimal number of classification errors, indicating the high informative value of the selected physical and optical characteristics as reliable indicators of rare earth element doping.

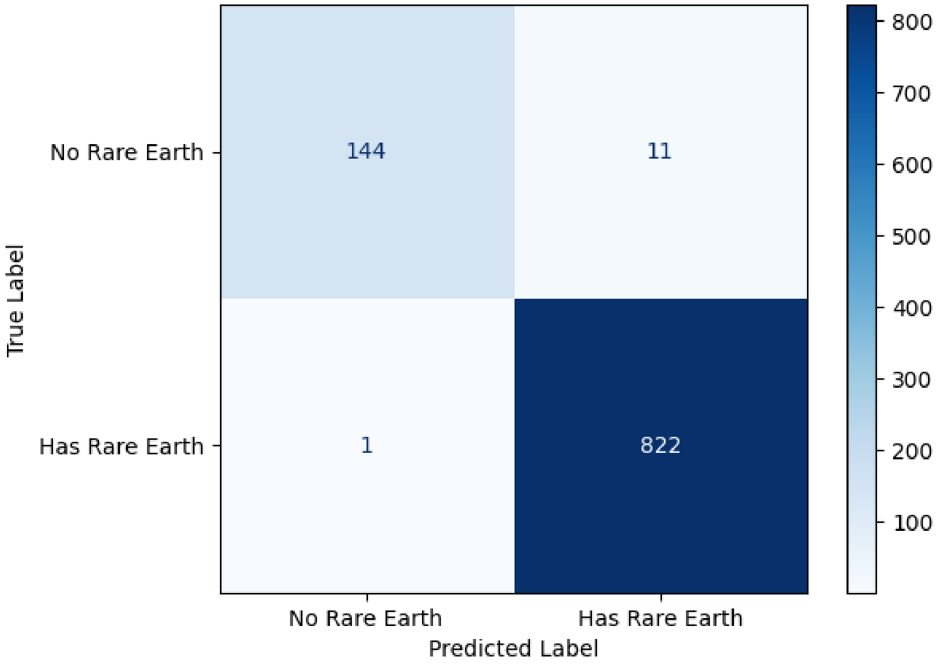

Figure 1– Classifications of error matrices of rare earth and non-recurrent earth elements

3.2 Clustering of Material Properties

During classification, the optimal number of clusters was found to be K = 4, allowing the phosphor materials to be classified into four distinct groups based on their properties. Clustering quality was assessed using the silhouette score, which was equal to 0.3159, indicating the required separation between clusters.

To visualize the clustering results, a 2D PCA projection of the dataset was used (Figure 2). The PCA plot clearly demonstrates the differences between the four clusters and allows for a clear visual separation. The plot also includes the has_rare_earth attribute to visually distinguish between rare earth and non-rare earth materials, allowing for their distribution across the clusters.

Figure 2 shows that clusters 1 and 2 contain a higher proportion of rare earth-containing materials, while clusters 0 and 4 predominantly consist of non-rare earth materials.This distribution highlights the correlation between chemical composition and cluster formation, providing further insight into the characteristics of the phosphor material groups.

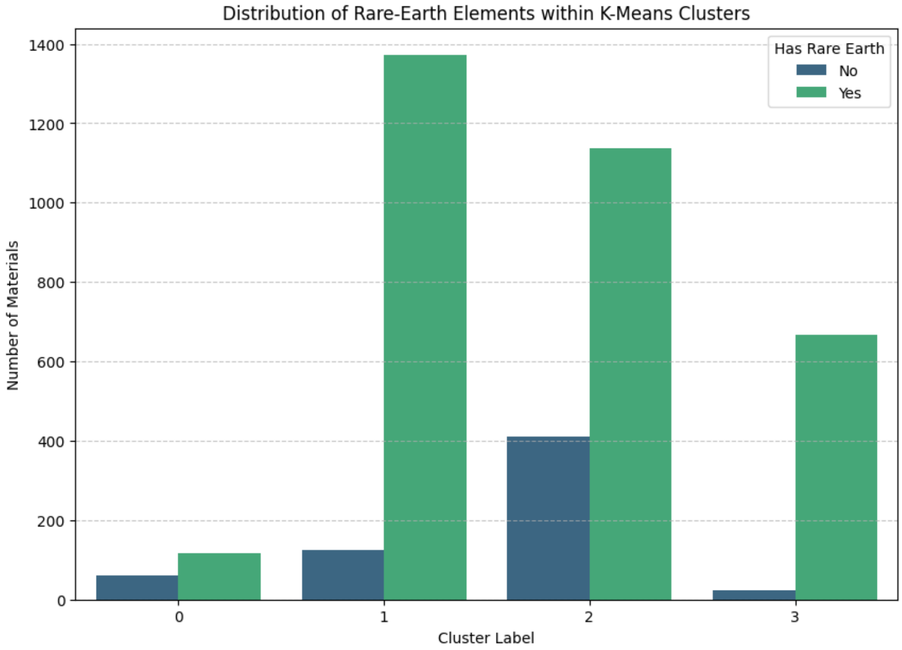

Figure 2 – 2D PCA visualization of K-Means clusters for phosphor materials
Colors indicate cluster membership, and markers show the presence of rare-earth elements

3.3 Optical Parameter Overlap Analysis

A comparative analysis of the "emission maximum – excitation wavelength" relationship reveals a significant region of overlapping parameters for phosphors with and without rare-earth activators. As shown in the generated visualization (Figure 4), non-RE phosphors occupy the same optical coordinates as traditional RE-doped systems.

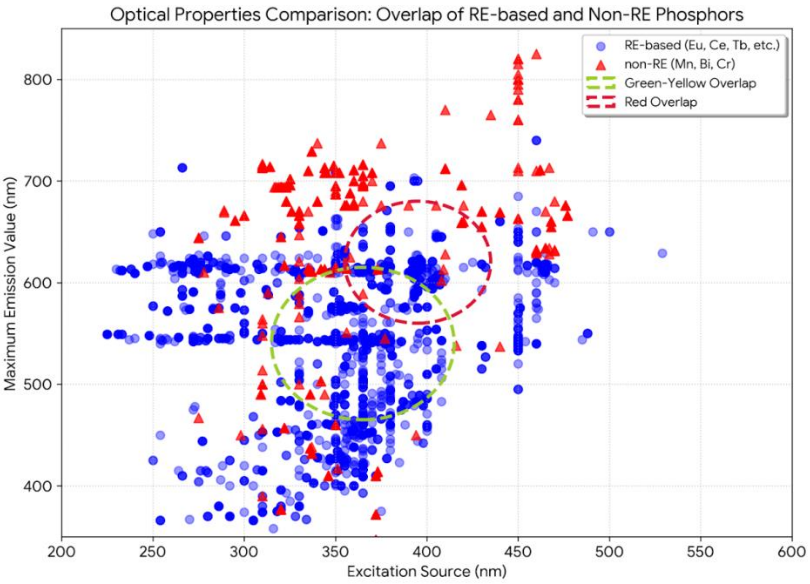

Figure 4 – Comparison of excitation and emission characteristics for rare-earth-based (RE) and rare-earth-free (non-RE) inorganic phosphors

Cluster 1: Green-Yellow Equivalence (Yellow-Green Highlight)

- **Circle 1** represents a region of functional convergence for phosphors with an excitation wavelength $\lambda_{ex} \approx$ 365 nm and an emission maximum $\lambda_{em} \approx$ 540 nm
- **RE-based Compounds:** Dominated by Terbium (Tb^{3+}) and Cerium (Ce^{3+}) systems, such as $Sr_8ZnScP_7O_{28}$:Tb and $Ba_3YAl_2O_{7.5}$:Ce;
- **Non-RE Substitutes:** Includes Bismuth and Manganese systems like $Ba_3Sc_4O_9$:Bi and $Ca_3YGa_3B_4O_{15}$: Mn;

- **Scientific Meaning:** In this region, the broad-band d-p transitions of Bi3+[6] and the d-d transitions of Mn2+ (in tetrahedral coordination) [7] overlap with the characteristic ${}^5D_4 \rightarrow {}^7F_5$ transitions of Tb3+. This overlap signifies that non-RE activators can effectively mimic the color purity of Terbium-based materials.

Cluster 2: Red Equivalence (Crimson Highlight)

- **Circle 2** encompasses materials optimized for excitation at $\lambda_{ex} \approx$ 395 nm with red emission maxima at λ_{em} ≈620 nm .
- **RE-based Compounds:** Predominantly Europium (Eu^{3+})-doped oxides, such as $Ba_2BiV_3O_{11}$:Eu and Y(P,V)O_4:Eu;
- **Non-RE Substitutes:** High-performance materials like $Mg_7Ga_2GeO_{12}$:Mn and $Sr_2MgGe_2O_7$:Mn;
- **Scientific Meaning:** This showcases the intersection between the sharp ${}^5D_0 \rightarrow {}^7F_2$ emission of Eu^{3+} [8] and the broad red emission of Mn^{4+} in octahedral environments [9]. The proximity of these points proves that transition metals can provide the necessary "red-shift" required for high color rendering index (CRI) devices.

4. Discussion

The presence of these intersection points validates the Material Replacement Strategy. Historically, Eu^{3+} has been the gold standard for red emission due to its thermal stability [10]. However, our analysis confirms that manganese-doped germanates and silicates (Mn^{4+}) can achieve comparable chromaticity coordinates [11, 12]. Similarly, while Tb^{3+} is efficient for green emission, Bi^{3+}-doped hosts allow for tunable green-yellow emission via host-lattice engineering [13, 14].

The selection of these specific circles was performed on a statistical and functional basis. The centers align with commercial standards: 365 nm and 395 nm are the primary outputs of commercial UV-LED chips. By circling these areas, we highlight that the same optical output is achieved through entirely different physical pathways *f—f* transitions for RE ions versus d-orbital or charge-transfer transitions for non-RE ions.

5. Conclusion

This analysis successfully classified phosphor materials based on the presence of rare-earth elements and grouped them into distinct clusters. The identification of high-density overlap zones confirms that non-RE compounds like manganese-doped silicates or bismuth-doped oxides are optically indistinguishable from their rare-earth counterparts in key spectral regions. These results can serve as a basis for the development of lighting technologies independent of the rare earth market.

Reviewer: Ghulomsafdarov A.G.— candidate of technical sciences, Head of the Department of Programming and Artificial Intelligence, Tajik Technological University.

Reference

1. Binnemans, K., & Jones, P. T. (2014). Perspectives for the recovery of rare earths from end-of-life fluorescent lamps. *Journal of Rare Earths*, 32(3), 195–200.
2. Massari, S., & Ruberti, M. (2013). Rare earth elements as critical raw materials: Focus on international markets and future strategies. *Resources Policy*, 38(1), 36–43.
3. Smet, P. F., Parmentier, A. B., & Poelman, D. (2010). Selecting conversion phosphors for white light-emitting diodes. *Journal of the Electrochemical Society*, 158(6), R37.
4. Zhuo, Y., Tehrani, A. M., & Brgoch, J. (2018). Finding new phosphors via machine learning. *The Journal of Physical Chemistry Letters*, 9(7), 1668–1673.
5. Hariyani, S., & Brgoch, J. (2023). A guide to comprehensive phosphor discovery for solid-state lighting. *Nature Reviews Materials*, 8, 759–773.
6. Blasse, G., & Bril, A. (1968). Investigations on Bi^{3+}-activated phosphors. *The Journal of Chemical Physics*, 48(1), 217–222.
7. Palilla, F. C., & O'Reilly, D. E. (1968). Optical properties of Mn^{2+} in tetrahedral and octahedral coordination. *Journal of the Electrochemical Society*, 115(10), 1076.
8. Binnemans, K. (2015). Interpretation of europium(III) spectra. *Coordination Chemistry Reviews*, 295, 1–45.
9. Adachi, S. (2020). Photoluminescence spectroscopy of Mn^{4+}-activated oxide phosphors: A review. *Journal of Luminescence*, 228, 117623.
10. Xie, R. J., & Hirosaki, N. (2007). Silicon-based oxynitride and nitride phosphors for white LEDs—A review. *Science and Technology of Advanced Materials*, 8(7), 588.
11. Xu, Y. K., & Adachi, S. (2025). Novel efficient deep-red emitting phosphor $SrCa_2Ga_2O_6$:Mn^{4+} with tululite-related structure. *Journal of Alloys and Compounds*.

12. Cao, R., et al. (2021). Enhancing emission property of red phosphor $Sr_2MgGe_2O_7:Mn^{4+}$ via Ba^{2+} doping. *Journal of Materials Science*.

13. Peng, D., et al. (2022). Green-emitting Bi^{3+}-doped $La_2SrSc_2O_7$ phosphor for pc-WLED lighting. *Journal of Alloys and Compounds*, 908, 164621.

14. Zhang, L., et al. (2016). Synthesis and luminescence properties of $Ba_3Sc_4O_9:Bi^{3+}$ phosphor. *Journal of Luminescence*.

15. Figshare, “ Research”,. [Online]. Available: https://figshare.com/ . [Accessed: Oct. 20, 2025].

МАЪЛУМОТ ОИД БА МУАЛЛИФОН-СВЕДЕНИЯ ОБ АВТОРАХ- INFORMATION ABOUT AUTHORS

TJ	RU	EN
Қаюмов Махмадзоир Маҳмараҷабович	Каюмов Махмадзоир Махмараҷабович	Kayumov Makhmadzoir Makhmarajabovich
Доктор Ph.D	Доктор Ph.D	Ph.D
Институти физикаю техники Академияи илмҳои Тоҷикистон ба номи С.У.Умаров	Физико-технический институт им. С.У.Умарова Национальной академии наук Таджикистана	S.U.Umarov Physical–Technical Institute of the National Academy of Sciences of Tajikistan
E-mail: kmakhmadzoir@gmail.com		
TJ	RU	EN
Бозоров Шамсуддин Аломуддинович	Бозоров Шамсуддин Аломуддинович	Bozorov Shamsuddin Alomuddinovich
н.и.т	к.т.н.	Candidate of technical sciences
Донишгоҳи техникии Тоҷикистон ба номи академик М.С. Осими	Таджикский технический университет имени академика М.С. Осими	TTU named after academician M.S. Osimi
E-mail: bozorov_shamsiddin@mail.ru		
TJ	RU	EN
Ашуров Анушервон Сироҷиддинович	Ашуров Анушервон Сироджиддинович	Ashurov Anushervon Sirojiddinovich
Донишҷӯ - таҳқиқотчӣ	студент-исследователь	research student
Институти астрофизикаи Академияи миллии илмҳои Тоҷикистон	Институт астрофизики Национальной академии наук Таджикистана	Institute of Astrophysics of the National Academy of Sciences of Tajikistan