Predicting Emission Wavelengths and Quantum Yields of Diverse Bis-cyclometalated Iridium(III) Complexes Using Machine Learning

Sergei V. Tatarin,[+,][a] Lev V. Krasnov,[+,][a] Ekaterina V. Nykhrikova,[a] Maxim M. Minin,[a,b] Daniil E. Smirnov,[a] Andrei V. Churakov,[a] Stanislav I. Bezzubov*[a]

Cyclometalated iridium(III) complexes are excellent emitters for phosphorescent organic light-emitting diodes, but the design of such complexes require substantial cost and experimental efforts. In turn, the predictive power of density functional theory calculations is seldom enough for reliable prediction of the excited state properties of iridium(III) complexes. In this work, we aimed at data-driven prediction of the emission energies and photoluminescence quantum yields of such complexes. To this end, we created a database (IrLumDB) that contains experimentally measured luminescence properties for over 1200 literature *bis*-cyclometalated iridium(III) complexes. Based on this database, we developed machine learning models that are capable of predicting the energy of emission maxima and photoluminescence quantum yields for the iridium phosphors with mean absolute errors of 18.26 nm and 0.129, respectively, requiring only SMILES of ligands. Furthermore, we validated the model for emission wavelength prediction on the set of 33 experimentally obtained luminescence spectra for newly synthesized and characterized iridium(III) complexes. Our data-driven methodology will complement quantum chemical calculations as an efficient alternative approach for the prediction of the excited-state properties of large sets of bis-cyclometalated iridium(III) complexes, facilitating computational discovery of efficient emitters. The emission properties prediction and the dataset exploration are available at (https://irlumdb.streamlit.app/).

Introduction

Over the past decades, octahedral cyclometalated iridium(III) complexes have become commonly utilized emitters in phosphorescent organic light-emitting diodes (PhOLEDs)[1,2], photosensitizers in dye-sensitized solar cells[3,4], photocatalysts[5–7], in cellular imaging[8] and photodynamic therapy agents[9,10]. The key advantage of these complexes is their bright luminescence enabled by effective intersystem crossing (ISC) of the photo- or electroexcited singlet state to the triplet excited state without population of nonemissive metal-centered (MC) excited states[11].

Many iridium-based PhOLEDs with high external quantum efficiency (EQE) in green and yellow spectral regions have been reported[12–14], although shifting the emission color toward both spectral ends (blue and red) is accompanied by substantial drop of the emission efficiency. In the blue region, the thermal population of nonemissive MC states becomes significant, causing a weakening of the metal-ligand bonds in the excited state and, hence, a decrease in the long-term stability of emitters[15]. Nevertheless, a long-lasting PhOLED was manufactured exploiting pyrimidine modification of the FIrPic emitter with sky-blue λ_{max}=475 nm[16], whereas effective and relatively stable blue PhOLEDs were created using tris-cyclometalated iridium(III) complexes with N-heterocyclic carbene ligands[17,18]. On the other hand, red and NIR iridium-based emitters suffer from low quantum yields owing to nonradiative relaxation of the excited state due to the energy gap law[19]. Increasing the rigidity of ligands[20], appropriate choice of substituents and their positions[21] and incorporation of peripheral fragments with large steric hindrance to eliminate unfavorable stacking interactions[22] may be effective to enhance the intensity of low-energy emission but all these design strategies have limitations.

Overall, the development of efficient iridium phosphors requires significant cost and experimental efforts, while the predictive power of state-of-the-art density functional theory (DFT) calculations for heavy metal complexes still needs to be substantially improved to enable rational design of iridium emitters with targeted emission characteristics. In turn, there is a very large body of experimental photophysical data that can be used for training machine learning (ML) algorithms to develop a fast yet relatively accurate tool for predicting the photophysical properties of iridium complexes based solely on their molecular formula.

Application of the ML methods in predicting various molecular properties has become extremely popular in recent years[23,24]. However, the space of individual molecules is usually limited strictly to organic compounds, as their representation in machine-readable format is quiet straightforward, with SMILES[25] being the most common. In a number of works is was shown that ML can be utilized to predict UV-Vis absorption spectra of organic and BODIPY dyes[26–28]. Moreover, a machine learning model based on a database of solvated organic fluorescent dyes was capable of predicting both emission wavelengths and photoluminescence quantum yields (PLQY) for such molecules with mean absolute errors

[a] S.V. Tatarin, L.V. Krasnov, E.V. Nykhrikova, D.E. Smirnov, M.M. Minin, Dr. A.V. Churakov, Dr. S.I. Bezzubov*
N.S. Kurnakov Institute of General and Inorganic Chemistry, Russian Academy of Sciences, Leninskii pr. 31, Moscow, 119991, Russia
E-mail: bezzubov@igic.ras.ru

[b] M.M. Minin
Department of Chemistry, Lomonosov Moscow State University, Lenin's hills 1/3, Moscow 119991, Russia

[+] These authors contributed equally.

(MAE) of 0.080 eV for emission energy and 0.13 for PLQY, respectively, becoming thus comparable to time-dependent DFT (TD DFT) calculations[29]. Given that PLQY is one of the most important measured quantities defining applicability of emitters in OLEDs[30], its prediction could be especially useful. In turn, coordination compounds are scarcely involved in such approaches because of difficulties in the interpretation of coordination bonds[31]. Nevertheless, a number of works has emerged in this field considering the optical properties of iridium(III) complexes. In one of the pioneering studies, Kulik *et.al.* utilized the experimental properties of high-throughput-synthesized 1440 distinct heteroleptic $[Ir(CN)_2(NN)]^+$ complexes obtained in[32] to train an artificial neural network[33]. Although outperforming TD-DFT in terms of predicting the emission energies on the test set, the class of molecules in the model is limited by cationic *bis*-cyclometalated complexes, predominantly with conventional ligand scaffolds. Most other works aiming at application of machine learning to predict photophysical properties of transition metal complexes tend to rely on parameters predicted by means of quantum chemistry, being limited with the abovementioned high-throughput-dataset obtained by Bernhard et.al.[34], by other relatively small datasets[35,36] or working with the entirely computational data[37]. Thus, a systematic exploration of experimental luminescence properties of the cyclometalated iridium(III) phosphors and construction of a big diverse dataset is needed to enable the creation of ML models directly from molecular formula of the corresponding complexes.

In this work we have created a dataset uniting 1454 entries of experimental data on luminescent properties of 1287 iridium(III) complexes from 340 peer-reviewed articles. We developed a data-driven algorithm that requires only SMILES of the ligands as an input and enables predicting the luminescence wavelength of *bis*-cyclometalated iridium(III) complexes with MAE = 18.3 nm on 10-fold cross-validation (Figure 1). Furthermore, we performed synthesis of 12 new complexes and measured 33 experimental luminescence spectra to implement validation of the model on our data (MAE = 16.9 nm). Moreover, we developed regressor ML models to predict photoluminescence quantum yields of the complexes, as well as classifier ML models to identify highly-emissive iridium phosphors. This study will help to evaluate photophysics of cyclometalated iridium(III) complexes from scratch thus enabling fast computational screening of thousands of new potential iridium phosphors in order to accelerate the development of new effective emitters.

Results and Discussion

Database collection

The algorithm for the database collection was adapted from[38] (Figure 2). The peer-reviewed articles containing words "iridium" and "complex(es)" in the title were found using Cobalt search engine[39]. For each complex the luminescence data (maximal peak wavelength (λ_{max}), quantum yield and excited state lifetime) and structural information were manually extracted along with the digital object identifier (DOI) of the source publication, creating an entity of record. Speaking about the complexes representation, a couple of notes should be pointed out. We have narrowed the dataset to the space of octahedral *bis*-cyclometalated iridium (III) complexes with three bidentate ligands sur-

rounding the Ir(III) ion, which represent the most popular (but not the only) type of such compounds. Taking into account that vast majority of these complexes possess similar geometry with trans-N,N' alignment of both cyclometalated ligands and with coordinated atoms of ancillary ligand arranged against the metalated carbon atoms[40], we assumed that a complex molecule can be unambiguously defined by listing the ligands in a certain fashion. Such representation can accurately describe both cationic and neutral *bis*-cyclometalated complexes, which is taken into consideration by formal deprotonation of corresponding atoms in the ancillary ligand. The cyclometalation site itself is specified in a similar way formally deprotonating the carbon atom in the corresponding SMILES string. In contrary, describing *tris*-cyclometalated complexes is not that straightforward, due to possibility of *fac/mer* isomers formation and strong dependence of photophysical properties on the isomer geometry[41,42]. Moreover, it is important to note that luminescence of such complexes might arise from solvatochromic charge-transfer excited states[43], thus their luminescence energy is strongly dependent on the solvent polarity, so the solvent was included in the dataset for every entity. In addition, conditions for PLQY measurements were also extracted. For the majority of the complexes inert atmosphere was stated in the source articles, and all the deviations from such (as well as other specific considerations) were stated in an additional column. The total database consists of 1454 individual luminescence measurements for 1287 iridium complexes from 340 papers, which provides a valuable resource for ML application in discovery of effective phosphors. Still, the reported dataset is by no means comprehensive, and we will continue to expand this database in order to support its ML applications. The database itself can be directly accessed at Zenodo[44].

During the data analysis we have depicted several trends. Firstly, it should be noted that the complexes with all the possible quantum yield values (from near zero to near unity) are observed with emission wavelengths from 450 to c.a. 650 nm, indicating a complete coverage of the luminescence properties space in this region (Figure 3). However, in the low-energetic region a considerable bias towards low values of PLQY is observed. This fact can be attributed to acceleration of non-radiative processes as governed by the energy gap law[19]. It is explicitly shown that maximum number of compounds emit light in the middle of visible spectra (500-650 nm, 1023 out of 1454 values), whereas in the both spectral ends number of phosphors significantly decreases (Figure 4a). This is explained by the fact that the archetypal compounds display emission in the greenish-yellow range and significant structural deviations are needed to increase/decrease the energy of triplet excited state considerably. The quantum yield distribution is significantly biased towards low values (Figure 4b) with maximum number of compounds being extremely low emissive (PLQY<0.025) what indicates that non-radiative relaxation is still a considerable issue while designing new complexes. The most common solvent employed is by a wide margin CH_2Cl_2 (815 from 1455 rows of data, 59%), which is easy to operate with and ensures good solubility for most organoiridium compounds (Figure 4c). The most popular ligands are easily defined in both cyclometalated and ancillary role (Figure 4d,e). In the former group ligands based on a 2-phenylpyridine core are the most popular with parent ppy⁻ being by a wide margin the leader, followed by dfppy⁻, and π-extended pq⁻

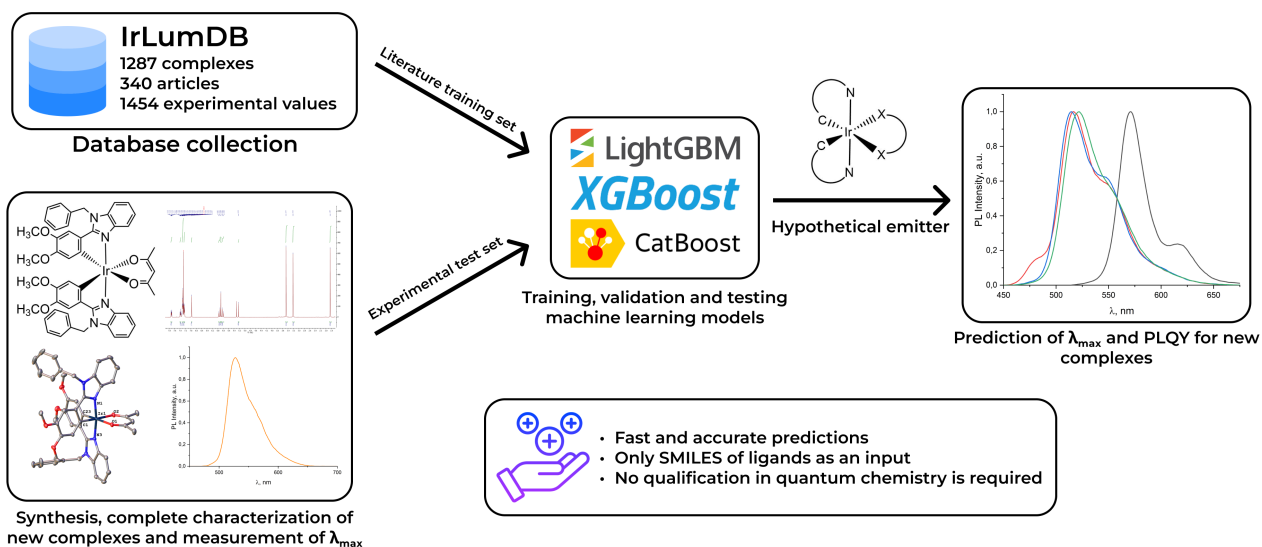**Figure 1.** The general strategy of luminescence wavelength prediction for *bis*-cyclometalated iridium(III) complexes.
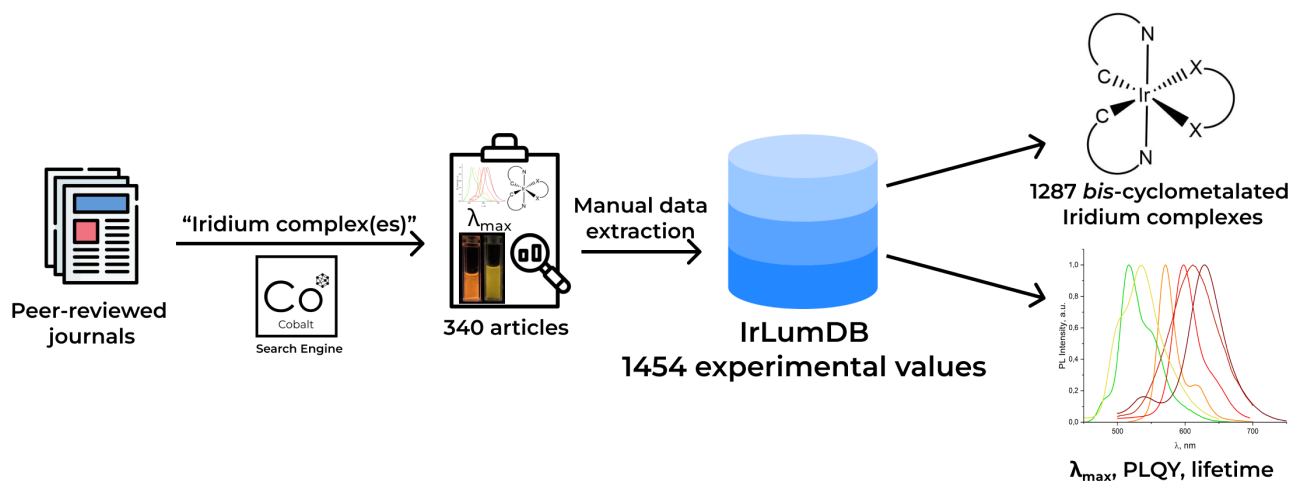

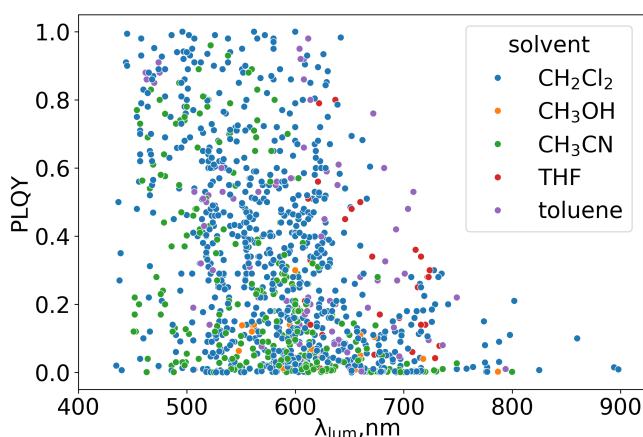**Figure 2.** The algorithm of database collection

Figure 3. Luminescence properties distribution of bis-cyclometalated iridium(III) complexes across the IrLumDB database. Emission parameters for top-5 most frequently observed solvents are shown.

and piq$^-$, whereas in the ancillary role acac$^-$ is by far the most popular, followed by anionic picolinate and neutral 2,2'-bipyridine. This in well in line with the general strategy of fundamental research, when one type of ligands is kept standard and simple while the second type is subjected to substantial variation. There are 638 rows of data for cationic $[Ir(CN)_2(XX)]^+$ complexes and 801 rows of data for neutral $[Ir(CN)_2(LX)]$ complexes, respectively, which shows profound interest in development of both kinds of compounds (Figure 4f).

Construction of experimental test set

In order to further verify the ML model predictions, we aimed to collect a set of our experimental luminescence measurements and to test the trained model on this data. The latter included: a) previously measured luminescence spectra in CH$_2$Cl$_2$ solvent[4,45]; b) luminescence spectra for previously investigated compounds, for which luminescence studies in CH$_2$Cl$_2$ have not been performed yet[46,47]; c) luminescence spectra for newly synthesized series of complexes. A new series was constructed exploiting 1-benzyl-2-aryl-benzimidazole ligand framework for the cyclometalated part and a commonly used acac$^-$ ligand as well as a number of 2,2'-bipyridine derivatives as an ancillary part.

The aryl-benzimidazole scaffold is well-known for ease of incorporation of various substituents in different parts of the molecule. Such substitution was shown to have profound impact on the luminescence spectra of the corresponding complexes, which is dependent on both the type and position of substitution[4,48,49]. To enrich variation of the ancillary ligand, three complexes containing 1,2-diphenyl-phenanthroimidazole cyclometalated part and various bpy-derived ligands were also synthesized. The structure and composition of the new complexes were reliably determined by SCXRD (9 out of 12 complexes)[50], PXRD, NMR and HRMS (see SI for complete characterization details). Data for the abovementioned complexes was complemented by spectra for the complexes published in[46,47] measured in CH$_2$Cl$_2$ and by data from[4,45]. In these works extensive modifications of cyclometalated 1-phenyl-2-arylbenzimidazole or 1,2-diphenyl-phenanthroimidazole and ancillary aromatic

1,3-diketonates were performed. Summing up, we have ended with total of 33 complexes (12 new) with luminescence maxima varying from 509 nm to 629 nm covering a significant part of the visible spectral range (several representatives with their luminescence spectra are presented in Figure 5). The obtained complexes demonstrated expected absorption profile with high extinction in the ultraviolet region and moderate extinction in the visible region (400 - 450/500 nm) (Figures S43 - S46). All the complexes expectedly possessed mainly reversible oxidation with E$_{ox}$= 0.13 - 0.92 V vs Fc$^+$/Fc, with additional irreversible processes arising when extra methoxy/chlorine substituents are presented (Figures S53 - S64). The luminescence of the complexes falls into the classification of excited states presented by Bernhard et.al.[32], thus the complexes being within the framework of typical iridium(III) complexes and possessing ^{3}MLLCT/^{3}LC mixed excited states (Figures S47 - S52).

Machine learning models development

Since SMILES representation is not generally applicable to the coordination compounds, it is an obstacle in developing the ML models with transition metals complexes. Given that the general geometrical arrangement of the ligands around the Ir(III) ion is very similar for all the molecules in our dataset[40], we supposed that the complexes representation can be adapted from[31] and generalized to describe far more diverse space of *bis*-cyclometalated iridium(III) complexes. Namely, for each ligand its SMILES representation was canonized using RDKit[51] and converted to the widely applied extended-connectivity fingerprints (ECFP4[52]) with 2048-bits length via molfeat[53]. The fingerprint describing desired complex is then created as a linear combination of the fingerprints of its ligands (Figure 6).

Given that the solvent affects photophysical properties of the complexes[43], only data in CH$_2$Cl$_2$ (as the most popular solvent) was used for training ML models. Ligands containing boron, bromine and silicon atoms were also removed from the training dataset due to rare occurrence of these atoms in the database (8, 13 and 26 molecules, respectively) and molecules with small L3 ligand (1 molecule). The final training dataset contained 818 rows. For prediction of PLQY only the rows, for which deaerated conditions of measurements are clearly stated in the source articles, were picked, as the phosphorescence quantum yields are highly dependent on the presence of oxygen in the atmosphere[54], which results in the training dataset of 748 rows.

Five ML algorithms: kNN[55], SVR[56], LightGBM[57], CatBoost[58], XGBoost[59], which commonly show good results, were chosen to training. The hyperparameter optimization was performed for all models using the Optuna[60] package. Hyperparameter values are presented in the Table S5. To evaluate the model precision on the train set, we used the 10-fold cross-validation (10CV) using Scikit-learn[61]. MAE, root-mean-square error (RMSE) and R^2 metrics were used to evaluate the quality of the ML models. Metrics for 10CV and for test set are presented in Tables 1,2.

Considering the emission wavelength prediction, it is explicitly shown that modern boosting algorithms (LightGBM, CatBoost and XGBoost) demonstrate superior to kNN and SVR performance with MAE not exceeding 18.78 nm on the 10CV for the former group of algorithms (Table 1). The performance on the test set maintains the same correlations, with modern boosting algorithms performing the best,
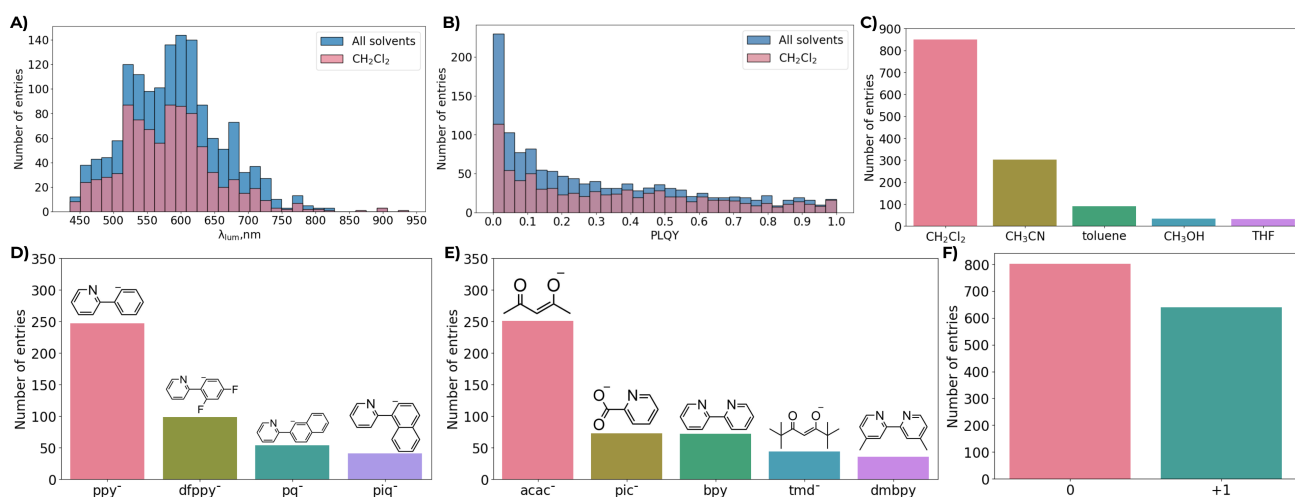
Figure 4. Statistics in the IrLumDB. A) Distribution of the emission wavelength values. B) Distribution of the photoluminescence quantum yield values. C) Distribution of the solvent media. D) Most common cyclometalated ligands. E) Most common ancillary ligands. F) Distribution of the complex total charge values.
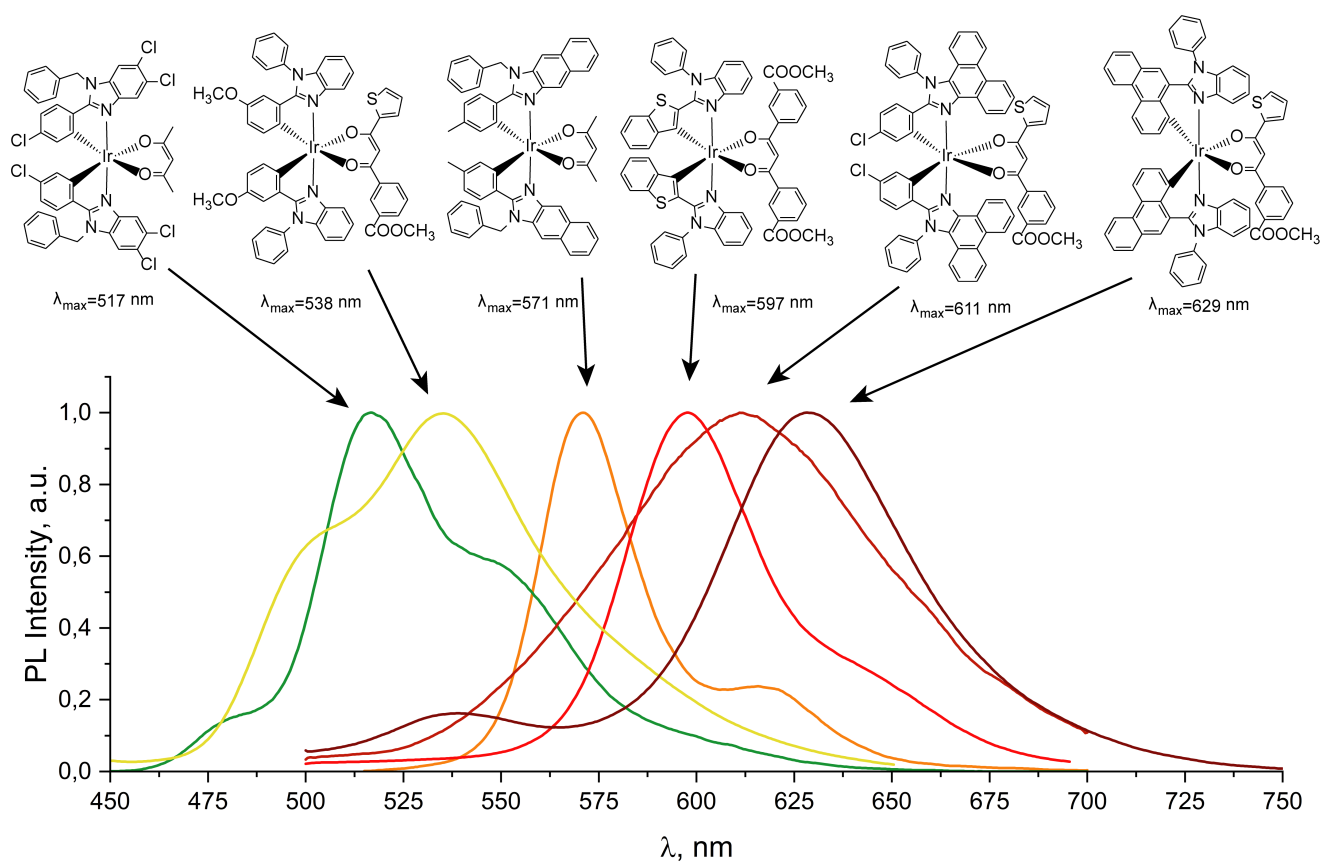

Figure 5. Examples of luminescence spectra experimentally measured in the test dataset with corresponding structural formulas and maximum emission wavelengths.
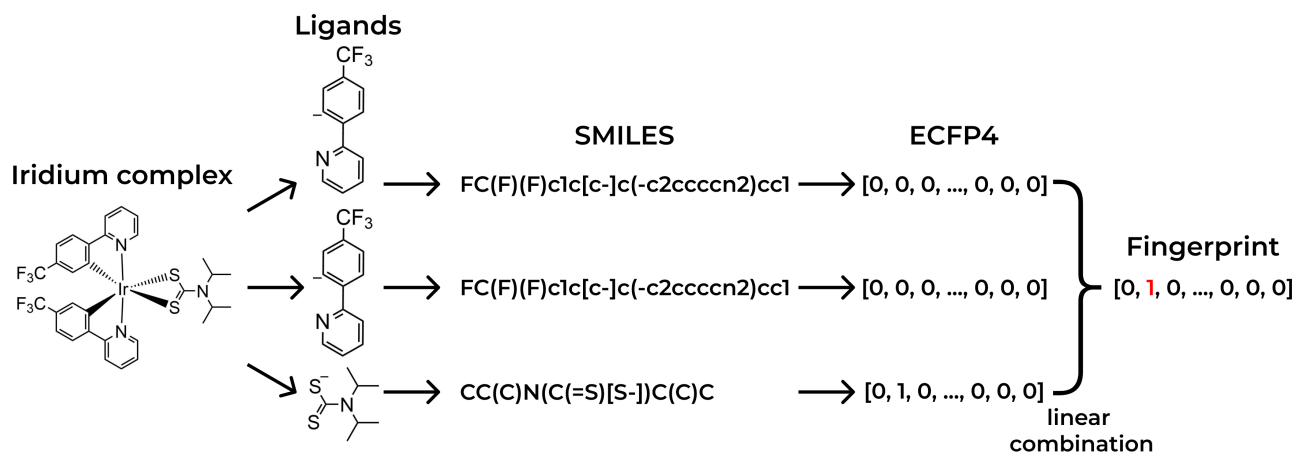
Figure 6. Schematic illustration of fingerprints generation. [31]

Table 1. Performance of the ML models on emission wavelength prediction.

Algorithm	10-fold cross-validation			Test		
	MAE	RMSE	R^2	MAE	RMSE	R^2
kNN	25.31 ± 2.05	39.17 ± 4.83	0.69 ± 0.06	29.3	36.64	0.18
SVR	21.45 ± 1.45	31.54 ± 2.27	0.8 ± 0.04	21.01	25.81	0.59
LightGBM	18.26 ± 1.95	27.32 ± 3.34	0.84 ± 0.06	18.53	21.55	0.72
CatBoost	18.78 ± 1.47	27.37 ± 2.41	0.84 ± 0.05	16.85	20.26	0.75
XGBoost	18.36 ± 1.38	26.79 ± 1.76	0.85 ± 0.04	19.56	22.54	0.69

followed by a wide margin by SVR and kNN algorithms. The MAEs on the test set is of the similar magnitude comparing to MAEs on the 10CV, indicating that models are capable of prediction of properties for the molecules they have not seen before, and thus possibility of exploration of new compounds. We have analyzed the molecules, for which maximal error is observed (Figure 8). Two of these molecules contain electron-donating alcoxy-fragments, but in different position. Although it is known that position of such mesomeric donors impact the resulting luminescence energies significantly [4,48,62], it is hard for the algorithm to recognize such dependence, as the ligands seem to be very similar. Thus, both hypsochromic effect of the meta-positioned -OMe fragment and bathochromic effect of the para-positioned -OMe fragment become underestimated, resulting in overestimating the former emission energy and underestimating the latter. Another pattern to highlight is the overestimation of luminescence energies for complexes with cyclometalated phenanthro[9,10-d]-imidazole, especially with the plain 2,2'-bipyridine ligand. This fact can be attributed to existence of compounds with cyclometalated phenanthroim-idazole fragments in the literature dataset [63,64], which possess blueshifted luminescence spectra compared to the complexes investigated in our work. The resulting RMSE metric for the best algorithm (26.8) is comparable to such for prediction of emission wavelengths of organic compounds (24.8) [29]. Thus, the developed system is reliable and precise in terms of energetic predictions.

Prediction of photoluminescence quantum yield still remains a challenging task, predominantly due lo large experimental errors (reaching 10 % or even more for relative measurements) and variation of experimental conditions [65]. Development of quantum chemical approaches is also limited due to complexity of photophysical processes upon molecule excitation. The situation becomes ever more complicated for iridium(III) emitters as they involve triplet excited states, and several calculations are needed to predict PLQY of one molecule [66]. In our hands, however, use of LightGBM, CatBoost and XGBoost algorithms provided reasonable accuracy with MAE's approx. 0.13 on the 10CV, what we can consider a satisfactory result at least for semi-quantitative screening. The kNN algorithm, as expected, performed worse, but the SVR algorithm, to our surprise, provided us with comparable to modern boosting algorithms metrics (Table 2). Moreover, such MAE and RMSE values (0.13 and 0.18) are comparable to those reported for prediction of PLQY of organic fluorophores (0.11 and 0.16, respectively) [29], indicating good performance of our models even on considerably more complicated bis-cyclometalated iridium complexes.

Still, for technological application of cyclometalated iridium complexes the exact value of PLQY itself is not so important. Namely, for purposes where emitting light is the key feature (emitters in OLEDs, bioimaging and sensing) the highest efficiency possible is desired. In contrary, for other tasks which require realization of non-luminescent relaxation pathways (generation of singlet oxygen or heat) low PLQY values may be needed. Thus, in the hope of achieving higher reliability, we turned to solving the problem of classification rather than regression. We grouped the

Table 2. Performance of the ML regressor models on photoluminescence quantum yield prediction.

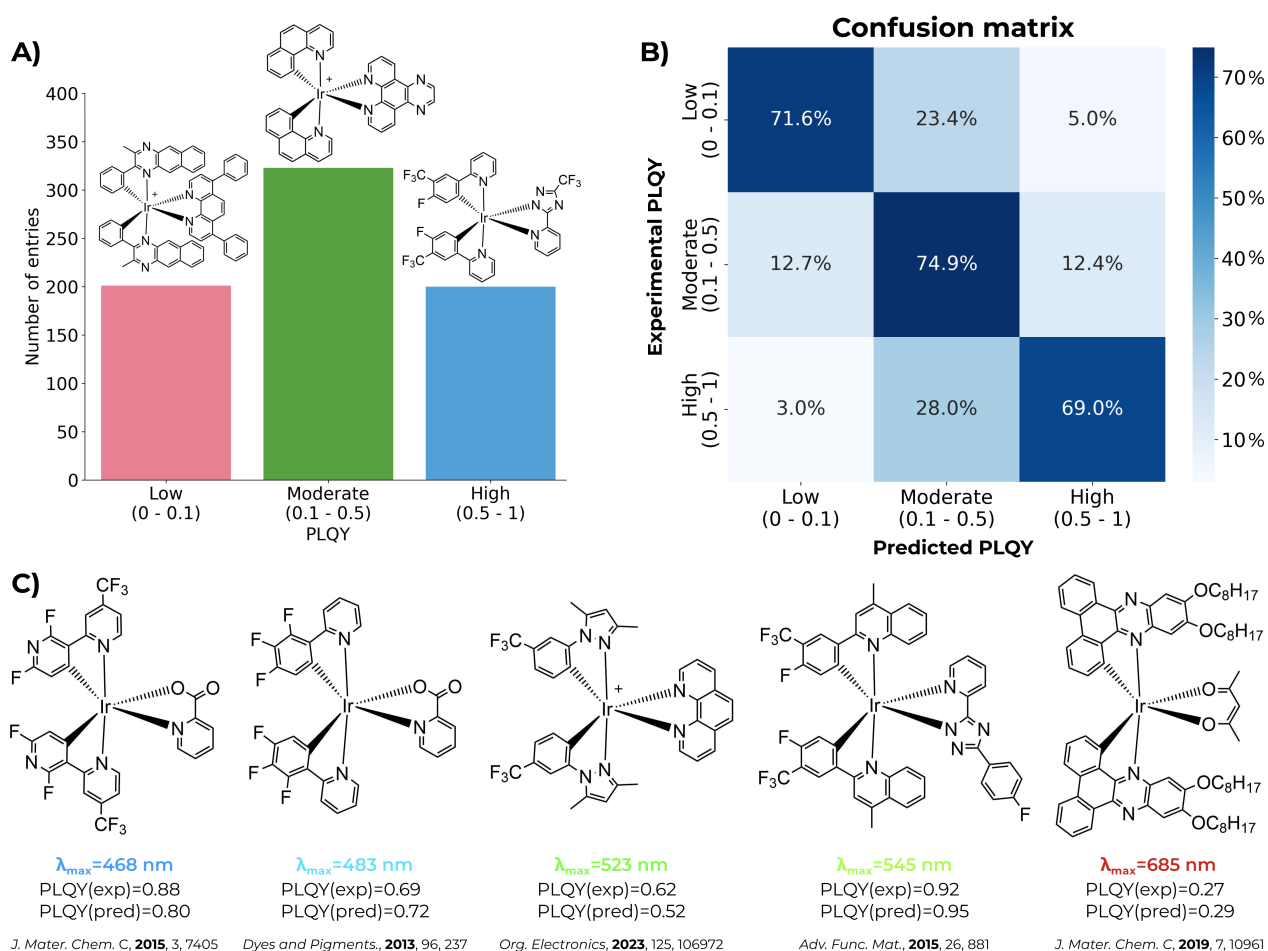| Algorithm	10-fold cross-validation		
	MAE	RMSE	R^2
kNN	0.142 ± 0.009	0.207 ± 0.014	0.441 ± 0.099
SVR	0.128 ± 0.010	0.181 ± 0.015	0.577 ± 0.074
LightGBM	0.133 ± 0.009	0.183 ± 0.013	0.571 ± 0.049
CatBoost	0.129 ± 0.009	0.180 ± 0.015	0.583 ± 0.043
XGBoost	0.129 ± 0.010	0.185 ± 0.013	0.561 ± 0.035

Figure 7. A) Distribution of the complexes among the PLQY classes in the IrLumDB dataset (numbers only for measurements in CH_2Cl_2 are shown) and structural formulas of examples from the each class. B) Confusion matrix showing performance of the CatBoost ternary classifier on the 10-fold cross-validation. C) Several representatives of highly emitting complexes from the literature [67–71] (not included in the original IrLumDB) and the predicted PLQY values by XGBoost regressor model.

Table 3. Performance of the ML classifier models on photoluminescence quantum yield prediction.

Algorithm	10-fold cross-validation Accuracy, %
kNN	68.1 ± 4.1
SVC	70.3 ± 3.2
LightGBM	71.4 ± 2.3
CatBoost	72.4 ± 4.1
XGBoost	70.2 ± 3.3

complexes in three classes - low (0 - 0.1), moderate (0.1 - 0.5) and high (0.5 - 1) PLQY, respectively. Such demarcation seems reasonable for both distribution of data from IrLumDB across these three classes (201, 323 and 200 values in CH_2Cl_2, respectively, Figure 7A) as well as for potential suitability in realistic applications. The performance of the classifier is described by the confusion matrix in Figure 7B. The accuracies of the resulting CatBoost classifier for identification of low-emitting, moderately emitting and high-emitting iridium(III) complexes are 71.6, 74.9 and 69 %, respectively, giving rise to a satisfactory overall accuracy of 72.4 %. Moreover, it should be noted that most mistakes are prediction of the class which is adjacent to the correct one. It means that percentage of completely incorrect predictions (when highly emissive complex is predicted to be low-emissive and vice versa) is negligible (5.0 and 3.0 %, respectively), what further coincides the usability of our models for semi-quantitative screening of large spaces of molecules. In addition, we tested our PLQY regressor on several highly-emitting complexes from the literature, which were not included in the original IrLumDB dataset (Figure 7C). Successful prediction of effective emission for them supports the applicability of the classifier for identification of molecules possessing high PLQY from space of previously unknown iridium phosphors.

Quantum chemical calculations

Furthermore, we have performed quantum chemical exploration of the experimentally investigated compounds. The theoretical emission wavelength was computed as the $\Delta_S CF$ value corresponding to the difference between the triplet and singlet single point energy calculations at the optimized triplet minima. This approach was recently found to be superior to TDDFT upon examination of Pt(II) and Ir(III) complexes[72,73]. It can be noticed that there is a systematical overestimation of emission wavelength predicted by quantum chemical calculations, though the general correlation is maintained (Figure 8). Such systematic errors are commonly observed upon computational examination of transition metal complexes[33,36,72,74], even with spin-orbit coupled TDDFT predictions[75]. The origin of this fact might be explained while examining the structure of emission bands. In particular, some of the spectra possess several emission bands, which originate from numerous triplet sublevels. The ΔSCF calculation refers to the lowest triplet excited state, which does not always contribute to the most intensive emission maxima. In contrary, ML approach uses only the peak emission wavelengths as an input, what explains far better correlation with the experimental values. Thus, the developed predictive model can effectively com-

plement and in some cases even replace DFT approaches, when fast and precise prediction of emission wavelength of the Ir(III) chromophores is needed without any additional insights in their electronic properties.
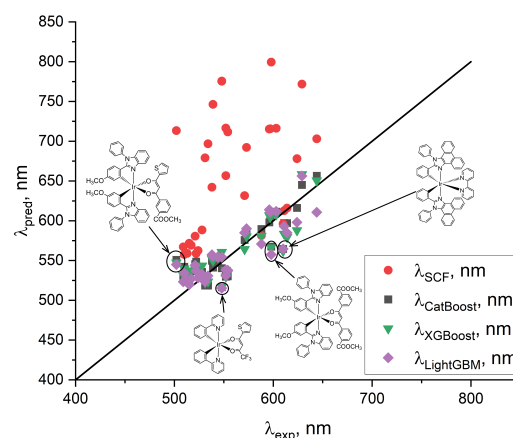

Figure 8. Comparison of ML and ΔSCF λ_{max} predictions to experimental (in nm) across 33 test set iridium complexes from the experimental dataset. The black line is included as a reference and corresponds to perfect agreement between prediction and experiment. The complexes with maximum prediction error (averaged for three modern boosting algorithms) are highlighted with their molecular formulas depicted.

Conclusion

We have collected a standardized database containing experimental luminescence characteristics for 1287 *bis*-cyclometalated iridium complexes. We demonstrated that a simple representation of heteroleptic complex as a linear combination of its ligands is a prominent way to describe such diverse coordination compounds for data-driven analysis. Thus, successful training of ML models was performed in order to predict luminescence energies and photoluminescence quantum yields of the complexes. The use of modern boosting algorithms allowed us to achieve accuracy comparable to state-of-the-art ML models for organic fluorophores and surpassing that of density functional theory calculations, reaching MAE 18.3 nm for λ_{max} and 0.13 for PLQY, respectively. A ternary PLQY classifier, which allows for identifying highly efficient and low efficient phosphors with 72.4 % accuracy, was also developed. In addition, the model developed requires only SMILES representation of the desired ligands as an input, making computational screening of the potential emitters extremely easy and enabling potential automatization of the workflow. We believe that both the database and the developed algorithms will provide a promising pathway for high-throughput screening and the acceleration of chemical discovery, as they can be used for quick evaluation of optical properties of thousands of hypothetical iridium complexes, highlighting the promising candidates for subsequent synthesis.

Author contributions

L.K., S.T., E.N., S.B. prepared, processed and evaluated the data, L.K. wrote the code for the ML. S.T., D.S., M.M. performed the synthesis, characterization and spectroscopic study of the complexes. S.T. and L.K. composed the manuscript. S.B. and A.C. revised the manuscript. S.B. supervised the project. All authors discussed, commented and approved the final manuscript.

Notes

The authors have cited additional references in the supporting information [76–83].

Acknowledgments

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

All code used for training is available on the GitHub repository upon https://github.com/levakrasnovs/irlumdb/. The IrLumDB dataset is available on Zenodo [44].

Keywords: iridium • metallacycles • database • machine learning • luminescence wavelength prediction • ligands

References

[1] J. Jayabharathi, V. Thanikachalam, S. Thilagavathy, *Coordination Chemistry Reviews* **2023**, *483*, 215100.

[2] T. Y. Li, W. Jing, Z. G. Wu, Y. Zheng, J.-L. Zuo, Y. Pan, *Coordination Chemistry Reviews* **2018**, *374*, 55.

[3] M. V. Bobo, A. Paul, A. J. Robb, A. M. Arcidiacono, K. M. Hanson, A. K. Vannucci, M. Smith, *Inorganic Chemistry* **2020**, *59*, 6351.

[4] S. V. Tatarin, E. A. Meshcheriakova, S. A. Kozyukhin, V. V. Emets, S. I. Bezzubov, *Dalton Transactions* **2023**, *52*, 16261.

[5] M. C. Carson, C. R. Liu, M. C. Kozlowski, *Journal of Organic Chemistry* **2024**, *89*, 3419.

[6] S. D. Kreijger, O. Schott, E. Cauët, L. Troian-Gautier, G. S. Hanan, B. Elias, *Inorganic Chemistry* **2022**, *61*, 5245.

[7] Y. Wang, Y. Huang, S. Chen, J. GAO, Y. Zhang, Y. C. Duan, P. Deng, *Inorganic Chemistry* **2023**, *62*, 7212.

[8] J. Zhou, J. Li, K. Y.-J. Zhang, S. Liu, X. Hu, *Coordination Chemistry Reviews* **2022**, *453*, 214334.

[9] P. C. Ho, C. H. Ho, W. yeung Wong, *Coordination Chemistry Reviews* **2020**, *413*, 213267.

[10] A. Galindo, A. Gandioso, M. Svitelova, G. Vigueras, S. Nonell, M. C. Santana, V. Brabec, V. Marchán, V. Novohradsky, A. Rovira, C. Hally, R. Bresoli-Obach, H. Kostrhunova, L. Markova, J. Kasparkova, *Angewandte Chemie - International Edition* **2019**, *58*, 6311.

[11] E. Zysman-Colman, *Iridium (III) in optoelectronic and photonics applications*, John Wiley & Sons **2017**.

[12] X. J. Liao, Z. Jin-jun, Z. Yan, Z. L. Tu, M. X. Mao, J. J. Lu, W. Zhang, Y. Zheng, Y. Li, *Materials Chemistry Frontiers* **2021**, *5*, 6951.

[13] N. Su, S.-B. LI, K. Yang, F. Zhou, J. Song, L. Zhou, J. Qu, *Dyes and Pigments* **2021**, *195*, 109733.

[14] Z. Yan, Z. WANG, X. Zhuang, Z. Li, C. Kai, X. Song, J. Liang, H. Bi, Y. Wang, *Advanced Optical Materials* **2024**.

[15] C. Wu, K. Shi, S. Li, J. Yan, Z.-Q. Feng, K.-N. Tong, S. Zhang, Y. Zhang, D. Zhang, L.-S. Liao, Y. Chi, G. Wei, F. S. Kang, *EnergyChem* **2024**, *6*, 100120.

[16] M. Sarma, W. L. Tsai, W.-K. Lee, Y. Chi, C.-C. Wu, S.-W. Liu, P.-T. Chou, K.-T. Wong, *Chem* **2017**, *3*, 461.

[17] K. J. Kim, H. Lee, S. Kang, T. Kim, *Chemical Engineering Journal* **2022**, *448*, 137671.

[18] J. Yan, D. Y. Zhou, L.-S. Liao, M. Kuhn, X. Zhou, S.-M. Yiu, Y. Chi, *Nature Communications* **2023**, *14*.

[19] Y. Zhang, J. Qiao, *iScience* **2021**, *24*, 102858.

[20] Z. Yan, M. Mao, Q.-M. Liu, Y. Li, X.-F. Luo, X. Liao, W. Cai, Y. Zheng, *Advanced Functional Materials* **2024**.

[21] Z. Chen, H. Zhang, D. Wen, W. Wu, Q. Zeng, W. Wong, S. Chen, *Chemical Science* **2020**, *11*, 2342.

[22] C. Ou, Y.-C. Qiu, C. Cao, H. Zhang, J. Qin, Z. L. Tu, J. Shi, Z. G. Wu, *Inorganic Chemistry Frontiers* **2023**.

[23] B. Dou, Z. Zhu, E. Merkurjev, K. Lu, L. Chen, J. Jiang, Y. Zhu, J. Liu, B. Zhang, G.-W. Wei, *Chemical Reviews* **2023**, *123*, 8736.

[24] E. Heid, K. Greenman, Y. Chung, S.-C. Li, D. E. Graff, F. H. Vermeire, H. Wu, W. Green, C. J. Mcgill, *Journal of Chemical Information and Modeling* **2023**, *64*, 9.

[25] D. Weininger, *Journal of Chemical Information and Modeling* **1988**, *28*, 31.

[26] K. Greenman, W. Green, R. Gómez-Bombarelli, *Chemical Science* **2022**, *13*, 1152.

[27] A. McNaughton, R. P. Joshi, C. R. Knutson, A. Fnu, K. J. Luebke, J. P. Malerich, P. B. Madrid, N. Kumar, *Journal of Chemical Information and Modeling* **2023**, *63*, 1462.

[28] A. A. Ksenofontov, M. M. Lukanov, P. S. Bocharov, M. B. Berezin, I. V. Tetko, *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy* **2022**, *267*, 120577.

[29] H. Bai, R. Liu, C.-R. JU, B. Li, *Journal of Chemical Information and Modeling* **2021**, *61*, 1053.

[30] H. Shi, W. Jing, L. Wu, Z. LI, B. Qiao, S. Zhao, Z. Xu, D. Song, Y. Li, *ACS Omega* **2022**, *7*, 7893.

[31] M. Orsi, B. S. Loh, C. Weng, W. H. Ang, A. Frei, *Angewandte Chemie - International Edition* **2024**.

[32] S. Diluzio, V. Mdluli, V. Vanbenschoten, J. Lewis, T. U. Connell, S. Bernhard, *Journal of the American Chemical Society* **2021**, *143*, 1179.

[33] G. G. Terrones, C. Duan, A. Nandy, H. J. Kulik, *Chemical Science* **2023**, *14*, 1419.

[34] T. A. Fiala, W. B. Swords, J. Schmidt, Y. Pal, T. P. Yoon, J. R. Schmidt, *ChemPhysChem* **2024**.

[35] M. HATANAKA, H. Kato, M. Sakai, K. Kariya, N. Shunsuke, T. Yoshimura, T. Inagaki, *Journal of Physical Chemistry A* **2023**, *127*, 7630.

[36] S. Wang, C. Yam, shuguang Chen, L. Hu, L. Li, F. F. Hung, J. FAN, C.-M. Che, G. CHEN, *Journal of Computational Chemistry* **2023**, *45*, 321.

[37] C. Zheng, J. Liu, T. JIANG, M. Chen, F. Dai, Z. Gao, G. Ke, Z. Zhao, Q. Ou, *Advanced Optical Materials* **2023**, *11*.

[38] L. Krasnov, S. Tatarin, D. Smirnov, S. Bezzubov, *Scientific data* **2024**, *11*.

[39] Cobalt search engine, can be found under https://cobalt.colab.ws.

[40] J. C. Babón, P.-L. T. Boudreault, M. A. Esteruelas, M. A. Gaona, S. Izquierdo, M. Oliván, E. Oñate, J.-Y. Tsai, A. Vélez, *Inorganic Chemistry* **2023**, *62*, 19821.

[41] T. Batagoda, O. Abimbola, M. Idris, S. C. Kapper, A. C. Tadle, P. I. Djurovich, C. Coburn, S. R. Forrest, M. E. Thompson, D. S. M. Ravinson, J. Kim, *Advanced Optical Materials* **2021**, *9*, 2001994.

[42] B. S. Yun, S. Y. Kim, J. H. Kim, H. J. Son, S.-W. Kang, *Journal of Materials Chemistry C* **2021**, *9*, 4062.

[43] A. P. Wilde, R. J. Watts, *The Journal of Physical Chemistry* **1991**, *95*, 622.

[44] S. Tatarin, L. Krasnov, E. Nykhrikova, D. Smirnov, M. Minin, A. Churakov, S. Bezzubov, *Zenodo* **2024**, 10.5281/zenodo.13987455.

[45] M. A. Kiseleva, A. V. Churakov, I. V. Taydakov, M. T. Metlin, S. A. Kozyukhin, S. I. Bezzubov, *Dalton Transactions* **2023**, *52*, 17861.

[46] S. V. Tatarin, P. Kalle, I. V. Taydakov, E. A. Varaksina, V. Korshunov, S. I. Bezzubov, *Dalton Transactions* **2021**, *50*, 6889.

[47] S. V. Tatarin, D. E. Smirnov, I. V. Taydakov, M. T. Metlin, V. V. Emets, S. I. Bezzubov, *Dalton Transactions* **2023**, *52*, 6435.

[48] Y. Jiao, M. Li, N. Wang, T. Lu, L. Zhou, Y. Huang, Z. Lu, D. Luo, X. Pu, *J. Mater. Chem. C* **2016**, *4*, 4269.

[49] C. Ciambrone, W. L. D. Hauret, C. Latouche, F. Loiseau, E. M. Vollbert, P. H. Lanoë, *Inorganic Chemistry* **2022**, *61*, 3033.

[50] Deposition Number(s) 2393803 (for **[Ir(mnim)2(acac)]**), 2393804 (for **[Ir(bim)2(acac)]**), 2393805 (for **[Ir(cbim)2(acac)]**), 2393806 (for **[Ir(tbim)2(acac)]**), 2393807 (for **[Ir(mbim)2(acac)]**), 2393808 (for **[Ir(tdcbim)2(acac)]**), 2393809 (for **[Ir(dmbim)2(acac)]**), 2393810 (for **[Ir(phi)2(dmbpy)]**) and 2393811 (for **[Ir(phi)2(bpy)]**) contain(s) the supplementary crystallographic data for this paper. These data are provided free of charge by the joint Cambridge Crystallographic Data Centre and Fachinformationszentrum Karlsruhe Access Structures service www.ccdc.cam.ac.uk/structures.

[51] RDKit: Open-Source Cheminformatics Software, can be found under https://www.rdkit.org.

[52] D. ROGERS, M. Hahn, *Journal of Chemical Information and Modeling* **2010**, *50*, 742.

[53] E. Noutahi, C. Wognum, H. Mary, H. Hounwanou, K. M. Kovary, D. Gilmour, thibaultvarin r, J. Burns, J. St-Laurent, t, DomInvivo, S. Maheshkar, rbyrne momatx, datamol-io/molfeat: 0.9.4 **2023**.

[54] M. Samandarsangari, D. O. Kozina, V. V. Sokolov, A. D. Komarova, M. V. Shirmanova, I. S. Kritchenkov, S. P. Tunik, *Biosensors* **2023**, *13*, 680.

[55] G. Guo, H. Wang, D. Bell, Y. Bi, K. GREER, KNN Model-Based Approach in Classification, pages 986–996, Springer Nature **2003**.

[56] M. A. Hearst, S. T. Dumais, E. Osuna, J. PLATT, B. Scholkopf, M. A. HEARST, J. R. Platt, B. Schölkopf, *IEEE Intelligent Systems and their Applications* **1998**, *13*, 18.

[57] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, *Advances in neural information processing systems* **2017**, *30*.

[58] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, A. Gulin, *Advances in neural information processing systems* **2018**, *31*.

[59] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* **2016** pages 785–794.

[60] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework, in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* **2019** pages 2623–2631.

[61] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., *the Journal of machine Learning research* **2011**, *12*, 2825.

[62] L.-L. Wen, C. X. Zang, Y. Gao, G. F. Li, K.-Z. Shao, G. Shan, X. Wang, W. Xie, Z. Su, *Inorganic Chemistry* **2022**, *61*, 3736.

[63] J. Jayabharathi, R. Sathishkumar, V. Thanikachalam, *Journal of Physical Organic Chemistry* **2014**, *27*, 504.

[64] A. I. Solomatina, K. M. Kuznetsov, V. V. Pavlovskiy, R. Evarestov, S. P. Tunik, V. V. Gurzhiy, V. Porsev, *Dalton Transactions* **2020**, *49*, 6751.

[65] H. Ishida, S. Tobita, Y. Hasegawa, R. Katoh, K. NOZAKI, *Coordination Chemistry Reviews* **2010**, *254*, 2449.

[66] D. Escudero, *Chemical Science* **2016**, *7*, 1262.

[67] A. Kumar, M. Reddy, K. S. Bejoymohandas, S. Varughese, E. Varathan, V. R. Subramanian, *Journal of Materials Chemistry C* **2015**, *3*, 7405.

[68] L. Kang, Y. Gao, X. Song, H. Bi, J. Liang, *Organic Electronics* **2024**, *125*, 106972.

[69] W.-L. Li, S.-J. Liu, Q. Yu, P. Tao, J. Zhang, S. Guo, X. Hu, H. Wang, B. Wei, X. hui Zhou, B. Xu, W. Huang, *Advanced Functional Materials* **2015**, *26*, 881.

[70] C. You, D. Liu, Y. Wang, J. Yu, S. Su, W. Zhu, F. Meng, S. Wang, *Journal of Materials Chemistry C* **2019**, *7*, 10961.

[71] C. Wu, S. Tao, M. Chen, F. L. Wong, Q. Yang, H. W. Mo, W. ming Zhao, C. Lee, *Dyes and Pigments* **2013**, *96*, 237.

[72] P. Kumar, D. Escudero, *Inorganic Chemistry* **2021**, *60*, 17230.

[73] P. Kumar, M. Pérez-Escribano, D. M. E. V. Raamsdonk, D. Escudero, *Journal of Physical Chemistry A*

2023, *127*, 7241.

[74] X. Li, B. F. Minaev, H. Ågren, H. Tian, *Journal of Physical Chemistry C* **2011**, *115*, 20724.

[75] K. D. Dobbs, J. M. Younker, *Journal of Physical Chemistry C* **2013**, *117*, 25714.

[76] D. E. Smirnov, S. V. Tatarin, M. Kiseleva, I. V. Taydakov, M. T. Metlin, S. I. Bezzubov, *Russian Journal of Inorganic Chemistry* **2023**, *68*, 1178.

[77] A. A. Bilyalova, V. D. Dolzhenko, P. Kalle, S. I. Bezzubov, S. V. Tatarin, *Chemistry - A European Journal* **2018**, *24*, 12779.

[78] A. Becke, *Physical Review A* **1988**, *38*, 3098.

[79] C. Lee, W. Yang, R. G. Parr, *Physical Review B* **1988**, *37*, 785.

[80] W. KOHN, A. D. Becke, R. Parr, *The Journal of Physical Chemistry* **1996**, *100*, 12974.

[81] M. Cossi, N. Rega, G. Scalmani, V. Barone, *Journal of Computational Chemistry* **2003**, *24*, 669.

[82] R. Ahlrichs, F. Weigend, *Physical Chemistry Chemical Physics* **2005**, *7*, 3297.

[83] F. Neese, F. Wennmohs, U. Becker, C. Riplinger, *Journal of Chemical Physics* **2020**, *152*, 224108.

Entry for the Table of Contents
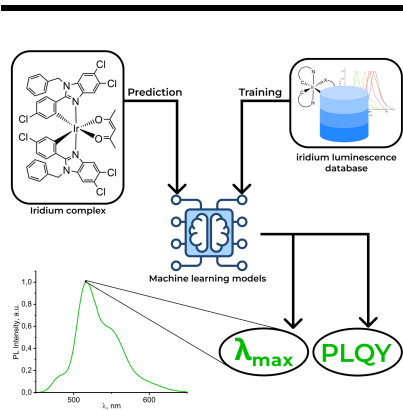
A database of optical properties of 1287 Iridium(III) complexes was created to provide the data basis for the computational exploration of iridium-based emitters. Machine learning models capable of predicting luminescence energy and quantum yield of bis-cyclometalated iridium(III) complexes were created and validated on the experimental data.