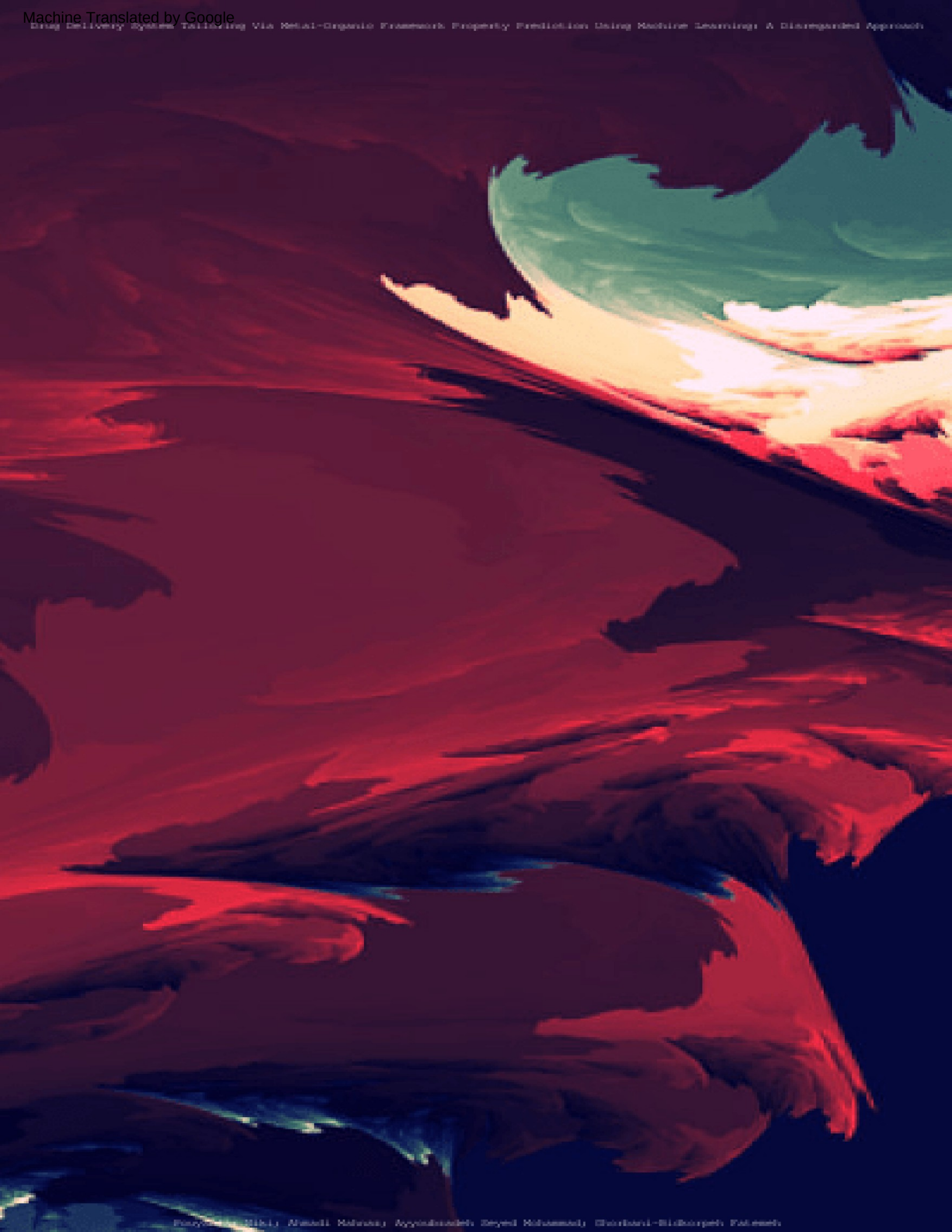


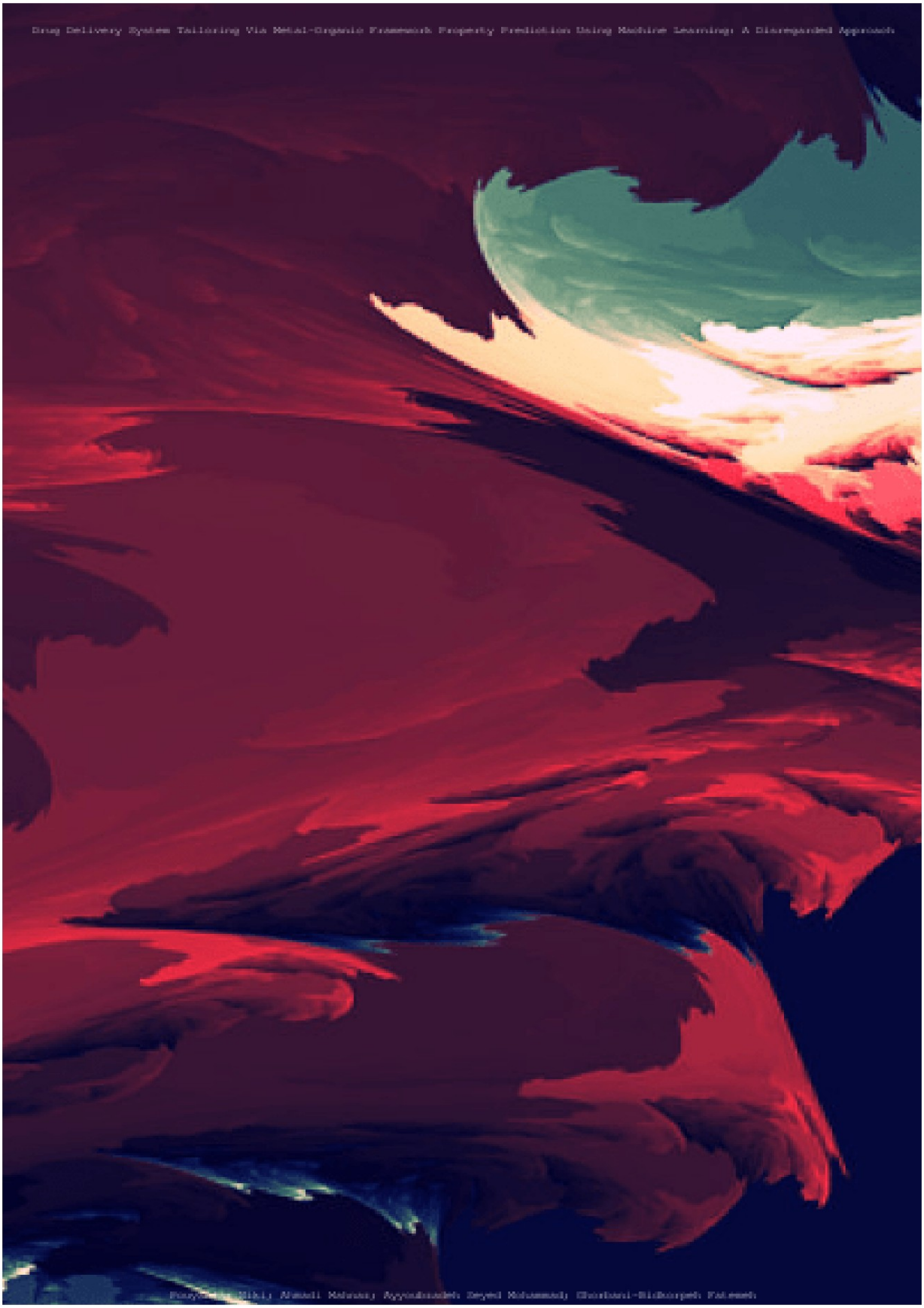


Table of Contents

Drug delivery system tailoring via metal-organic framework property prediction using machine learning: A disregarded approach

Metal-organic frameworks (MOFs) are a class of crystalline and porous substances that have found widespread applications in various fields, including nanomedicine, catalysis, gas storage, and sensing technology. These materials are composed of metal ions/clusters and organic binders, which provide a large surface area for interaction. The characteristics of MOFs depend on the specific components, the length of the organic binders, the presence of functional groups in the structure, as well as the availability of free metal centers and other factors [1-4]. Understanding the characteristics of MOFs is a key factor in selecting the most suitable MOF for a specific application. For example, nanoscale MOFs with a high surface area, magnetic resonance imaging (MRI) capability, and small pore diameters may be useful for theranostics, while for hydrogen storage, MOFs with a high void ratio, large pore volume, and low density may be useful. Predicting these properties is not always easy. Research in the laboratory can be tedious and unproductive in studying MOFs. Taking this into account, the use of the trial and error method in the synthesis and study of MOFs does not seem interesting, since it is time-consuming, gives low yields and is expensive [9]. The Cambridge Structural Database (CSD), managed by the Cambridge Crystallographic Data Center (CCDC), contains an extensive collection of crystal structures demonstrating the features of various metal-organic frameworks (MOFs) and their properties. Currently, the OIF CSD database contains more than 100,000 structures, of which more than 12,000 are available for free use without

restrictions [10]. Another MOF database is CoRE MOF, which contains MOF structures derived from computer calculations. CoRE MOFs and CSD refer to databases containing already synthesized (real) MOFs. However, many databases contain only hypothetical MOFs (hMOFs) [5]. MOF databases can be used to perform molecular dynamics calculations and Monte Carlo (MC) simulations. (MD) predict new MOF structures using computer technology, statistics and probability theory. The properties of MOFs can be predicted using such models after studying characteristics such as the ability to store gases, separate and adsorb on the surface [11-13]. Software packages such as topology-based crystal designers (ToBaCCo) provide another set of tools for creating MOFs with a high degree of order. These programs connect the building blocks of the MOF using data obtained from extensive databases [14]. A conventional method for screening various data to find MOFs with desired properties could be high-throughput computer screening

(HTCS). The screening process often employs methods such as density functional theory (DFT) or grand cantilever molecular dynamics (GCMC) to identify MOF candidates with promising properties. Even with such methods, screening for HTCS is time-consuming, expensive, and ineffective. Artificial intelligence (AI) and, in particular, machine learning (ML) can become the fastest way to predict the properties of MOFs with high accuracy. ML is the application of algorithms or technologies that allow a system to forecast trends and make decisions. ML can be used to classify data in genetics, biological medicine, etc. In the process of synthesizing various materials, including MOFs, patterns can be identified using ML, since in this case the prediction of the properties of materials is faster than when using methods of high-temperature catalytic synthesis. ML also helps improve the properties of materials by linking molecular chemistry, structure and properties [18-20].

The input data for AI algorithms can be material data, including geometric, chemical and spatial characteristics. This data is used to train and form a model that predicts a variety of MOF functions. The functions can include properties such as conductivity, mechanical characteristics, adsorption on the surface and absorption of gases [21, 22]. To achieve this goal, various machine learning algorithms can be used, which can be divided into four main groups: 1. Supervised learning algorithms, including support vector machines (SVMs) and nearest neighbors (KNN); 2. Unsupervised algorithms such as probability distribution and compressed autoencoder; 3. Algorithms with semi-supervised learning, including semi-supervised SVM and the graph theory method; 4. Algorithms with reinforcing, in particular, the Q-learning method [23-25]. One of the recently developing applications of MOFs is drug delivery due to the biocompatibility and biodegradability of these nanostructures and their high drug loading capacity.
Predicting the critical characteristics of MOFs, including pore size, surface density, and polarity, can provide valuable insights into the functionality of these compounds as drug carriers. It can predict their drug incorporation capacity, cellular uptake capacity, and safety profiles.
However, to our knowledge, there is currently no available literature that describes the application of this approach. Therefore, the relevance of this area of research seems very significant.
[27-30]. Current DDSs have several disadvantages, such as low biodegradability in most carriers, risk of dose release, rapid clearance from the body by immune cells and organs such as the liver and kidneys, and unpredictability of pharmacokinetic parameters [31, 32]. This article reviews machine learning methods used to predict various drug delivery-related properties of MOFs and discusses their capabilities and relationships with DDS. First, machine learning methods are reviewed, then the properties of MOFs are described, and then the feasibility of them is discussed.

predictions using these methods and their relationship with DDS. Various machine learning algorithms can be used to predict different parameters. These fall into four categories: supervised learning algorithms (e.g. machine support vector methods, k-dimensional nearest neighbors), unsupervised learning algorithms (e.g. mixed distribution method, compressed autoencoder), semi-supervised learning algorithms (e.g. semi-supervised support vector method, graph theory method) and algorithms with reinforcing (for example, Q-learning method) [23-25]. Supervised learning is a category of machine learning methods that train models using input and target features extracted from data. The model determines the relationship or correspondence between input and target features. The trained model can then use this relationship to classify unknown samples (classification) or to predict numerical values of a target feature (regression). The most well-known supervised learning models include neural networks (ANN), Bayes, support vector machine (SVM), tree problem solving, random forests (RF), and nearest neighbors (KNN). [33, 34]. In this category of methods, data is transferred to the model without reference to labels. The model should group data based on record similarities. In these methods, records with a high degree of similarity will be assigned to the same cluster. This category includes methods of clustering, identifying association rules, applying the principle of principal components (PCA) and self-learning algorithms. In this category of methods, the input data contains both previously known and unknown values. In addition, models in this category try to extract as much information as possible from the data containing this mixture in order to identify patterns. This category combines the two categories described above [33,34]. This category of methods refers to those who try to learn through trial and error, seeking to maximize the result obtained and minimize the signal of punishment. Q-learning and temporal differential learning methods fall into this category ([33], [34]). In the previous section, machine learning algorithms

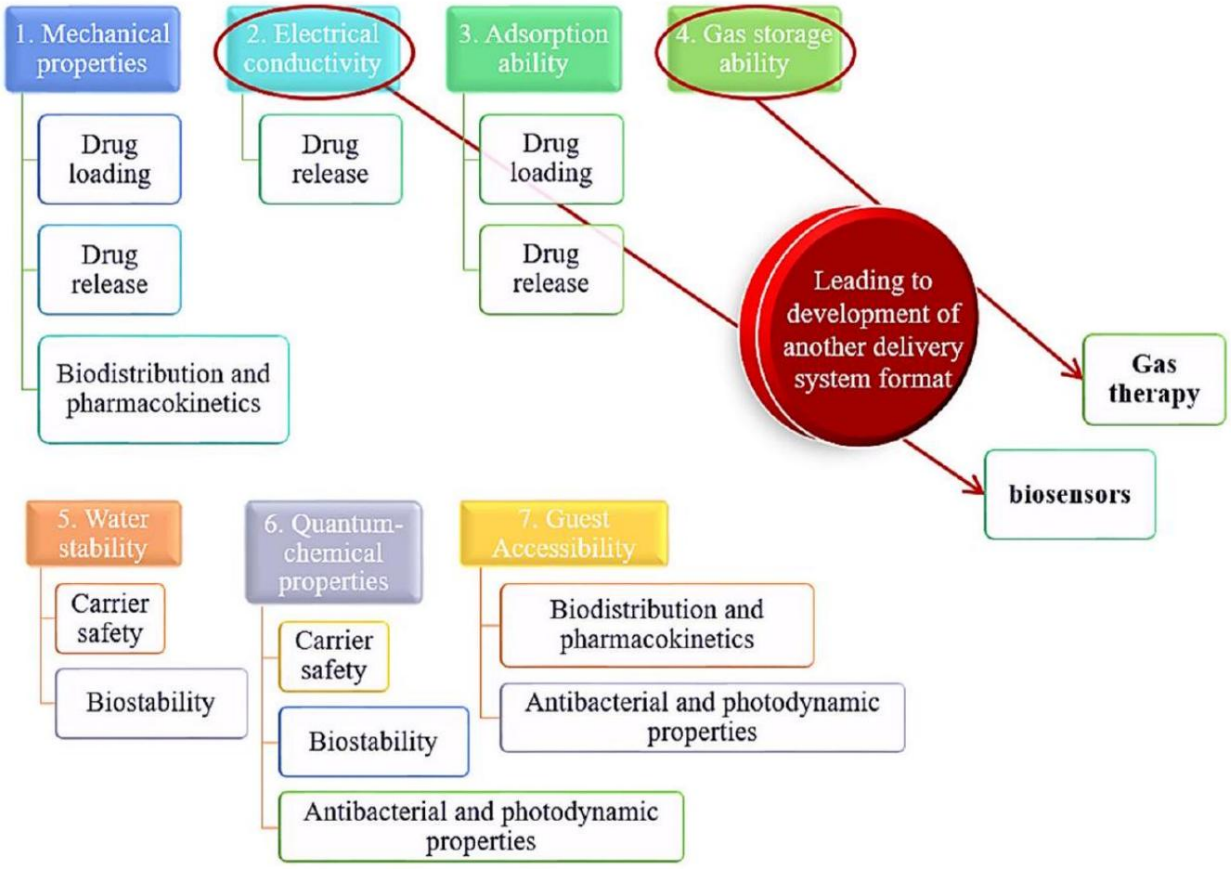
classified based on their ability to solve various problems, including classification, regression, and clustering [35]. When classifying the input information, it is assigned a category that has been previously defined as a target, for example, it is determined whether a given MOF material is electrically conductive or not. Regression uses input information to predict a numerical value, for example, to determine the gas absorption capacity of an MOF based on its characteristics. Clustering groups objects based on their input characteristics without considering their properties. For example, MOFs can be grouped by their characteristics, assuming that materials in the same group have similar properties. There are several aspects involved in the task of training machine algorithms. One of them is the data size, which plays an important role. The more data, the higher its quality. Typically, large data sets include samples derived from a wide range of input parameter combinations. Another aspect to consider is the data type. It can take many forms, including recordings, signals, images, or combinations thereof [36]. Feature selection is a valuable approach for efficient data processing, especially when working with high-dimensional data sets, and can solve various machine learning problems. This process selects the most important features from the original dataset to achieve the following goals: simplifying models, improving data processing performance, and providing clean and understandable data [37]. Various metrics can be used to select the optimal machine learning model based on its performance. Classification problems often use metrics derived from the confusion matrix, such as accuracy, classification error, and Cohen's d. For regression problems, metrics such as mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), and squared relative error (RRSE) are often used. These metrics provide valuable information about the model's performance and compare it with other machine learning models [38]. Like other materials, MOFs have various properties, including physicochemical

characteristics and differences in their properties depending on the application. Sometimes determining their properties and predicting their effectiveness in various applications is a complex, time-consuming and expensive task, so the use of AI methods can be useful and promising for predicting the properties of MOFs. Some properties of MOFs have been predicted using machine learning methods, which are described in the following sections [39]. MOFs have porosity and high surface activity, which makes them excellent candidates for applications in various fields, such as catalysis or drug delivery [40]. Traditional drug delivery faces various challenges. Drug molecules can spread throughout the body, causing unwanted side effects. Nanoscale delivery could be useful in providing more efficient drug distribution and fewer side effects in the treatment of diseases ranging from cancer to inflammatory diseases. Such diseases show increased vascular permeability in the affected areas, which promotes increased permeability and retention of nanocarriers (EPR) and application of targeted binding molecules on their surface for precise drug delivery and hence more effective treatment [41, 42]. Over the years, various substances have been used as drug carriers, such as liposomes, dendrimers, polymeric micelles, etc. [43, 44]. MOFs can be produced at the nanoscale and have high drug entrapment and loading efficiency. They can also be tailored with various metal clusters and organic ligands using various synthesis methods, making them promising for medical applications such as diagnostics and theranostatics [27]. Machine learning algorithms can determine MOF structures with fairly high quality and simple synthesis methods [45]. MOF nanoparticles have great potential for medical applications such as drug delivery and imaging [46]. However, in the laboratory, a lot of time and effort must be spent on inefficient syntheses of MOF carriers, which can only be discovered after testing in cell cultures or under conditions

living organism. Another source of unproductive work in the laboratory is the low efficiency of drug entrapment in MOF nanoparticles or complex procedures that result in micronized MOF molecules that cannot be used as drug carriers due to their rapid elimination from the body through the reticuloendothelial vascular system and other parts of the immune system [47-50]. Instead of testing every possible formula in the laboratory, it is advisable to use machine learning algorithms to improve accuracy when designing DDSs and new nanotransport systems such as MOFs [39,51]. A possible solution to such problems in the laboratory is the use of machine learning algorithms to predict such properties in silicone models prior to the experimental phase [52]. The process of using ML to predict the properties of MOFs begins with the collection of data, which can be obtained either from the literature or in the laboratory. For MOFs, the data (input values) represent the basic characteristics of the MOF, such as ligand, metal cluster, size, surface tension, etc. In the case of supervisory learning, the output (label) is also given for each sample as an aid to training, and in this case it will be the characteristics of the MOF that need to be predicted. In the case of unsupervised learning, the algorithm is able to find a connection between input data (information) without knowing the values of the labels. One way to collect data for laboratory experiments in ML training is through experimental design techniques. Once the data is collected, it needs to be pre-processed as it may not be sufficient to train the algorithm, requiring techniques such as data augmentation or outlier removal. Data normalization is also an important step, especially for algorithms that respond to distance, such as logistic regression and k-nearest neighbors. In the case of algorithms that do not depend on distance, such as problem solving using trees or decision forests, the data does not require normalization [53]. The data is then split into training, test and validation sets, which are then

models are presented. The model's accuracy and sensitivity (as well as other metrics used to evaluate its performance, such as f1-score or cross-validation results) are then tested to determine the optimal hyperparameter settings for the model. Using a properly trained algorithm, it is possible to predict the values of features corresponding to new input data [54]. According to the authors, there is still little information on this issue. Therefore, proposals regarding the use of machine learning techniques can be useful to usher in a new era in the field of molecular property prediction based on biomedical MOF nanoparticle data. Using data on the properties of biomedical MOF nanoparticles as input for training machine learning models can be a good start in this direction [55]. Many aspects can be beneficial in the synthesis of MOF nanocarriers: (a) drug loading, (b) drug release, (c) safety of nanocarriers, (d) biostability, (e) biodistribution and pharmacokinetics, (f) antibacterial and photodynamic properties. Figure 1 shows how the properties of MOF molecules are related to the characteristics of drugs, and demonstrates that predicting each property of a molecule provides insight into the properties of a drug. (a) The dimensions of the surface adsorption layer and the pore sizes in MOFs (see Fig. 1) [51] determine the amount of drug that can be loaded into them. The pore sizes must be larger than the dimensions of the drug molecule for the MOF to effectively load the drug. However, if synthesis methods are planned to be used under conditions close to natural conditions, large pore sizes are not required to load the drug and ensure a high degree of its confinement within the MOF. It is most likely that MOF nanoparticles with high surface density (as measured by the Brunauer, Emmett, and Teller (BET) test) will exhibit high loading efficiency [56–58]. A recent study collected 40 different MOFs from the literature and used them to predict ibuprofen loading capacity using support vector analysis (SVR), categorical gain (CatBoost), adaptive gain

(AdaBoost) and neural network algorithms with 5 input parameters (surface, functional groups, ligand and metal type, pore size and volume), which made it possible to predict the loading capacity of the MOF. CatBoost was found to be the most effective prediction method, showing a correlation coefficient of 0.76 [59]. This work, although the only one devoted to drug delivery into MOFs using machine learning methods, does not fully address the topic: MOF nanoparticles differ in the type of ligands and ion clusters that bind them, so predicting loading capacity using different MOFs can lead to inaccuracies. In addition, 40 samples are not enough to fully capture the diversity described in the paper. The most effective approach would be to use one type of MOF molecule, such as ZIF-8, and obtain a variety of laboratory data instead of using only literature sources. (b) Large pores also provide rapid drug release. Therefore, for cargoes that need to be released at a controlled or slow rate, it is recommended to use MOF molecules with smaller pores. On site, they can be loaded with transport nanoparticles, which are most effective in the treatment of infectious or oncological diseases, for example, in response to changes in pH. It is also possible to predict the presence of functional groups on the surface of MOFs so that the most effective therapeutic ligands can be attached to them. In 2022, predicting the release of drug ligands. polymer nanoparticles using a neural network using a release model showed good results. The polymers were erosive, and 22 different input parameters were synthesized in the laboratory to produce 22 release schedules. The input parameters were related to the drug, polymer used, excipient content and pH. High values of the Pearson coefficient demonstrated successful learning and modeling [62].



Rice. 1. Molecular characteristics of MOFs that contribute to the development of new drug delivery systems or predict the properties of MOFs.

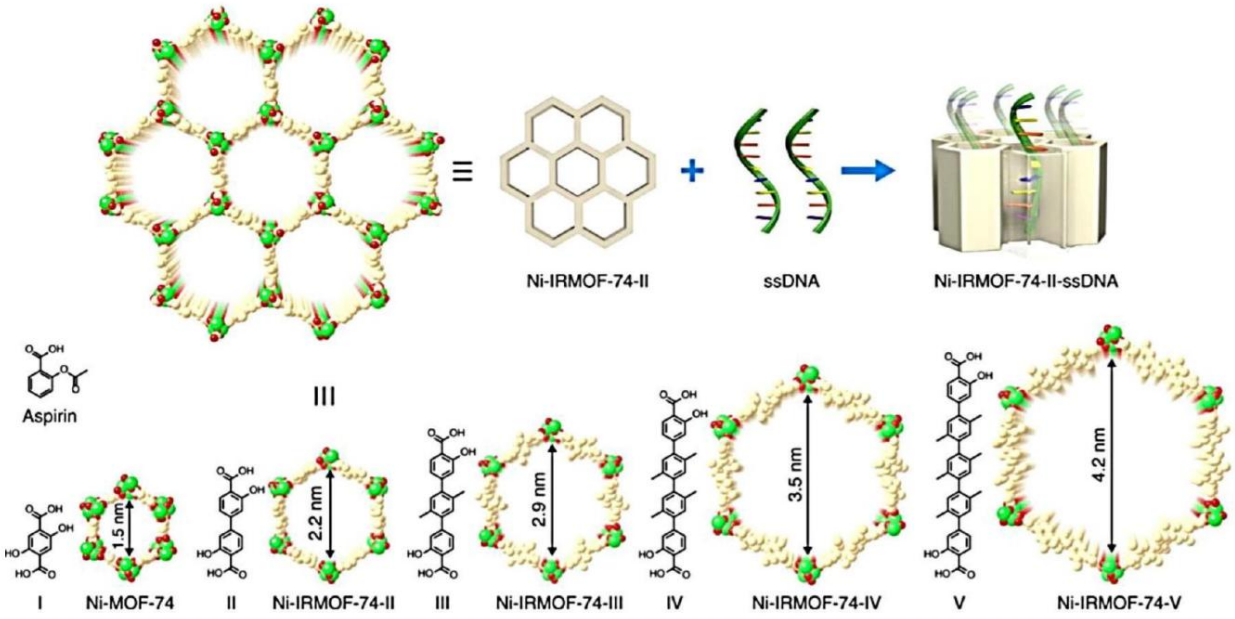


Figure 1. Pore size optimization for better DNA loading into

MOF ligands. Adopted from [88]. Research may also be useful in identifying different molecular release mechanisms in MOF nanoparticles to realize the compound's potential in responsible drug delivery. © Studying the properties of the atoms of MOF molecules and determining toxicity in cell cultures in vitro makes it possible to predict the toxicity of MOF before conducting complex studies on the toxicity of cells and animals. The polarity and size of the nanocarrier may also determine how the nanocarrier is likely to be distributed in the body and which organs are at greatest risk of toxicity. An example is the toxicity assessment of quantum dots and metal oxide nanoparticles using ML. The properties of the nanomaterial, the characteristics of the biological environment, and the experimental parameters were used as input data, and indicators such as the concentration at which complete inhibitory activity is observed (IC50) were used as output data. The effectiveness of the algorithm was tested using the coefficient of determination and the correlation coefficient R2. This study demonstrates the possibility of using ML methods to assess the toxicity of other nanoparticles, including MOFs [66]. (e) The biostability of MOF nanoparticles depends on the metal clusters and organic ligands used and is significantly related to how these parts combine to form MOF structures [67, 68]. It is common knowledge that ZIF-8 MOF nanoparticles are acid labile, making them suitable for on-demand and activation-assisted drug delivery in diseases such as cancer or some inflammatory diseases such as rheumatoid arthritis [69, 70]. Another issue that raises concerns when using ZIF-8, one of the most famous and easily synthesized structures

MOF is the ability to react with a solution containing phosphate ions (PBS) and form zinc phosphate crystals [71, 72]. This question may be problematic for several reasons, as this solution is often used in biological research because it is isosmolar and contains ions in a concentration similar to the concentration of ions in the human body, which forces researchers to use other nanoparticles to coat them.

materials such as folic acid or hyaluronic acid to improve results [73, 74]. Predicting this problem with the biological stability of ZIF-8 nanoparticles at the very beginning of research could avoid lengthy synthesis and testing processes, and would also allow the use of other coating materials or changes in the structure of the MOF before further research begins. The Bond Graph Convolution Neural Network (BE-CGCNN) based on the principles of Bond-type crystallography effectively demonstrates the electrochemical stability of metal nanoparticles. This technique has been shown to be more effective than simple CGCNN or density function of state (DFT) calculation-based methods for predicting the stability of metals [75]. Similar approaches can also be used to predict the stability of MOFs. It is also possible to predict the interaction of nanomaterials with biological environments, which can help predict the biological stability of nanomolecules, such as organometallic compounds. In 2020, neural networks with convolutional layers were used to convert images of 147 nanoparticles into information about physicochemical properties and about uptake and adsorption on their surface by cellular structures [76]. (f) MOF structures differ in their biodistribution and pharmacokinetic properties depending on the different metal and organic segments, which may have different surface potentials, sizes and functional groups [27,77]. MIL-100 (Fe) is a MOF composed of iron and benzotricarboxylic acid clusters. In 2019, this nanocluster was loaded with hemicitabine, resulting in nanoparticles with approximately equal hydrodynamic diameters of approximately and electrical potentials of approximately . Pharmacokinetic studies in rats showed that when hemicitabine was loaded into MIL-100(Fe) nanoparticles, the half-life increased and the biodistribution showed that in the free form the drug was distributed in large quantities in the animals' bodies, while when it was loaded into nanoparticles, it was mainly concentrated in the blood and less distributed in various organs. The drug was discovered using the method

high performance liquid chromatography (HPLC). It was mainly found in the liver after passing through the circulatory system [78]. Another study used ZIF-8 nanoparticles radiolabeled with technetium-99m to evaluate biodistribution and pharmacokinetics, which mainly detected ZIF-8 nanoparticles in the lung. Coupling of pertectinate with ZIF-8 MOFs increased the area under the plasma concentration-time curve (AUC), residence time (MRT), and half-life compared with the radiolabel in the free state, while reducing its elimination and volume of distribution [79].]. Since large particles typically end up in the lungs, these results indicate the possibility of nanoparticle aggregation. If it is necessary to avoid the accumulation of MOF in the lungs, the use of a polyethylene glycol (PEG) coating around the ZIF-8 MOF can be considered [80]. The application of machine learning methods and other predictive devices can be crucial to determine the distribution results of biomolecules and provide recommendations for the use of more preferable materials for the synthesis of MOFs or their coating. A physiology-based pharmacokinetics model was applied in conjunction with 5 models to predict nanoparticle delivery to tumors in vivo. To train the model, the nano-tumor database was used, which collected 200 studies. Twenty-one variables were used as parameters to train the model, creating a new tool for early screening of anticancer drugs that could ultimately lead to a reduction in the number of animals used in experiments. The problem with the database was that the nanomaterials used belonged to different groups, and some groups were more numerous than others (for example, the frequency of use of gold nanoparticles compared to nanocrystals), which could bias the algorithm and make it more biased towards certain nanomaterials. It would be optimal if all the nanomaterials presented in this paper were MOFs, but since this is not the case, this tool could be useful for other MOFs and for the treatment of diseases other than cancer. (f) Determination of antibacterial properties of MOFs

has a significant impact on delivery systems used in the treatment of wounds or infections [83]. Photodynamic MOFs such as ZIF-8 and porphyrin-containing particles are capable of producing reactive oxygen species (ROS), which can kill nearby cells (bacterial or cancerous). At the same time, MOFs based on copper and silver have innate antibacterial properties without requiring exposure to light [84, 85]. These effects could have been predicted in the early stages of development of the wound healing agent to determine the need for the addition of an additional antibacterial or photodynamic agent or the self-sufficiency of the synthesized MOF in this regard. After training 5 ML models using input data from 60 papers, RF was selected as the most effective algorithm for predicting the antibacterial properties of nanostructures. The key parameters were the type of bacteria, the size of the nanoparticles and the dose. A high correlation coefficient indicated good performance of the model [86]. Similarly, a 2023 study showed how normal and tumor tissues respond to photodynamic therapy coupled with the delivery of gold nanoparticles. The use of machine learning methods made it possible to evaluate such reactions and demonstrate the advantages of the nanoparticles used over tumor cells, as well as their toxicity, which was minimal for normal tissues [87]. Such capabilities can also be used for MOFs to predict antibacterial or photodynamic effects. As mentioned earlier, experimental analysis of material properties can take a long time, from several hours to several months, depending on the conditions, the availability of materials and the methods used. Therefore, various AI methods have been developed to predict various properties of materials. With the help of AI, experimental analysis time can be reduced from days to minutes or hours. In addition, these methods allow large-scale monitoring and screening of materials in large volumes, such as MOFs. Among the various AI methods, the most effective are machine learning methods, which allow you to speed up the analysis of materials and expand the search area using

smaller data set to a larger space of MOFs [89]. Selecting and preparing input and output data for ML models is essential for property prediction. In the case of MOFs, researchers typically use input data that includes geometric, chemical, and energetic properties that can be calculated using computer tools, as well as output data that includes large arrays of calculated values that must contain a sufficient number of data points [90]. The following sections will describe studies in which ML has been used to predict the properties of MOFs, as well as the benefits of these properties in the design of various DDSs. Mechanical properties, like physical properties, can influence the performance of MOFs in various applications [4]. There are studies that have examined the behavior of MOFs under mechanical load. The mechanical stability and flexibility of MOFs are important parameters to consider in a specific application. At larger scales, some MOFs, such as MIL-53 and MIL-88, can be flexible, allowing them to be used in a variety of applications. Therefore, predicting and controlling the mechanical properties of MOFs is an important step for their effective application [95]. Mechanical properties such as pore volume or surface area play an important role in the application of MOFs in the field of drug delivery, especially in drug loading and release [67, 96]. In addition, the flexibility of such nanoparticles can be useful in the development of various carriers. Compared to rigid structures, flexible nanoparticles can provide better blood circulation and endocytosis [96]. In addition to experimental analysis, computer methods can also help in understanding the mechanism of deformation under mechanical loads [97]. However, computer analysis of the mechanical properties of MOFs on large scales is not economically feasible. Therefore, machine learning methods can be a promising alternative for assessing mechanical properties. In this context, Moghadam et al. [98] analyzed the mechanical properties of MOFs based on 3385 hypothetical MOFs with different topologies using a multiscale modeling method. In particular, to identify the most suitable

MOFs, a high-throughput screening technique based on classical molecular simulations was used to determine the values of the G and K parameters for MOFs from the database. As a result, interesting patterns were revealed, in particular, the influence of binder length, pore volume, surface and maximum coordination number (MCN) on the elastic properties of materials. The elastic properties of MOFs were predicted using a neural network built on the basis of 13 MOFs selected taking into account the optimality criteria. Two different MOF models were constructed, one containing topology information and the other not, and these models were used as descriptive characterizations to identify the influence of topology on the mechanical properties of MOFs. The differences in the prediction accuracy of these models were quite noticeable. Model with descriptive characteristics

showed the value . while the other model (without descriptive characteristics) showed the value . These results indicate a significant influence of topology on the mechanical properties of MOFs. After deeper analysis, it was found that the model with descriptive characteristics provided high prediction accuracy for all values of the elastic modulus, while the model without descriptive characteristics gave a worse result for MOFs with a large elastic modulus. After analysis using machine learning methods, a molecular simulation was carried out using force field parameters obtained by a priori calculation. The molecular simulation results confirmed previously identified patterns obtained using ANN modeling and large-scale screening. In particular, a gradual decrease in the elastic modulus was confirmed with increasing equilibrium cell length (approximately corresponding to the length of the liquid crystal display (LCD) or interface), indicating a stable nature of this process. In addition, molecular dynamics results have shown that at a certain pressure a lack of crystallinity can be detected, allowing a direct relationship to be established between this pressure and the elastic modulus. In general, machine learning-based models can be useful for assessing the mechanical properties of MOFs. However

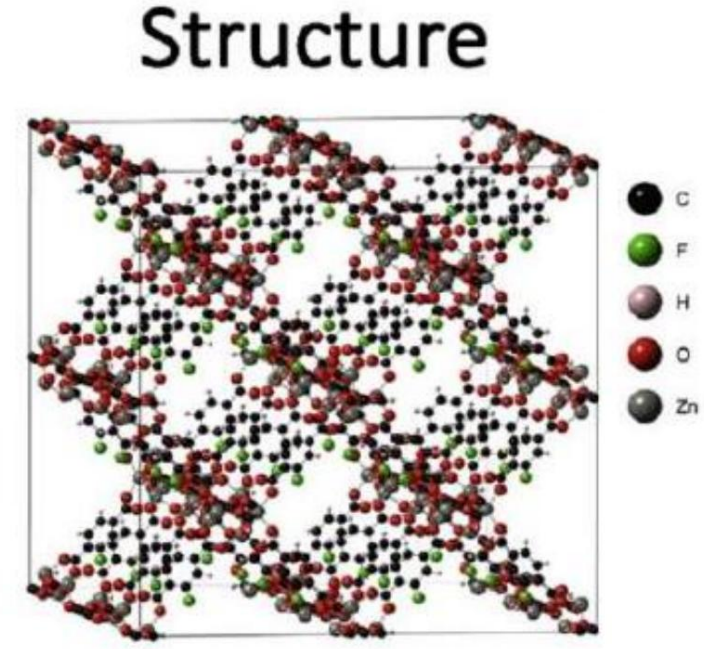
More research is needed to determine their impact on the drug delivery process. The large pores and electrically inactive regions, such as the carboxylate linkages in the structure of MOFs, make them electrical insulators. However, scientists were able to obtain electrically conductive structures while maintaining porosity by rationally designing MOFs. Dinka et al. showed that MOFs can become electrically conductive with rational design and retain their porosity forever [99]. Subsequently, other studies were carried out in which electrically conductive MOFs were obtained, which experimentally showed that their conductivity can reach [100-103]. These MOFs can be used in various applications, such as transistors [104], sensors [105], and supercapacitors [106]. The application of electrical conductivity in DDS has a variety of applications. Conductive MOFs can be used to create biosensors to evaluate drug delivery efficiency or be used to activate cargo release using electrical waves ([107,108]). By increasing the contact between the molecules of interest and the conductive channels due to the internal porosity of the MOFs, conductive MOFs are obtained. Since little is known about such MOFs, the use of AI technologies, in particular machine learning (ML) methods, makes it possible to discover new conductive MOFs and predict their conductivity. For the first time, ML methods were applied to predict the conductivity of MOFs by He and co-workers. For this purpose, data from different but related classes of inorganic materials from the Open Quantum Materials Database were used. The logistic regression method consists of building a model that determines the probability of a particular outcome based on the input parameter. The Support Vector Classifier method is used to classify input data processed as a single set. The trained models were then used to predict conductivity using ML methods. The performance of these models was evaluated, and accuracy rates ranged from for the logistic regression model to for the support vector model. Before Predicting Conductivity

The t-distributed stochastic neighbor embedding (t-SNE) method of MOFs was used to compare the distribution of MOFs with the distribution of inorganic materials included in the database. The revealed significant overlap of distributions indicates that models developed for inorganic materials can be effectively used to predict the conductivity of MOFs. In addition, for the cross-training learning method, a multi-class classification scheme was applied and results were obtained for all four models. The CoRE MOF database containing 2932 organometallic oxide compounds was analyzed, and organometallic oxides with electrical properties were discovered using ML models to be used as triggers for drug delivery or to create biosensors [111]. If it is important to obtain organometallic oxide compounds with electrical properties for activated drug delivery or for the creation of biosensors, then their availability can be assessed using artificial intelligence models before synthesis. One of the abilities of MOF structures is adsorption. Numerous papers [112–116] have reported the use of MOFs for adsorption, but recently, researchers have increasingly turned to using ML modeling to study this property of MOFs. Although this approach has not been tested in the development of drug transport systems, the adsorption of drug molecules provides high loading capacity and high encapsulation efficiency, resulting in high yield and cost-effectiveness when using MOF molecules [117]. Acetylene is one of the most important alkynes and is difficult to store due to the risk of explosion at high pressure [118]. A recent study presented a method for predicting adsorption on MOFs in less than one minute [119]. Copper open metal center MOFs (Cu-OMS) were found to be most effective in acetylene adsorption compared to other MOFs. 7855 Cu-OMS MOFs were synthesized using the ToBaCCo program based on 877 Cu-OMS MOFs obtained from CSD. Of the synthesized MOFs, only those whose hole diameter was smaller than the kinetic diameter of acetylene were selected (8681 structures). Decision Tree (DT) Model

showed that characteristics such as open metal sites (Cu-OMS), pore volume (PV), and surface mass density (GSA) are sufficient to predict high MOF adsorption. The Gradient Boosting Decision Tree (GBDT) algorithm was found to be the most effective in determining the relationship between acetylene adsorption and the characteristics of MOFs. The prediction accuracy of GBDT was high, with a correlation coefficient of 0.93, indicating the high accuracy and speed of the method. Activated carbon materials (OCMs) are used to adsorb heavy metals from natural media such as water, which is necessary for solving environmental problems and wastewater treatment [120]. In 2022, the neural network method was used to simulate the structure of UiO-66-(Zr) MOF grown on layered hydroxide sheets for the adsorption of Pb(II) and Cd(II) from water. As a result, the adsorption capacity increased and the concentration of metals in the aquatic environment decreased, which was associated with a change in the rate of mass exchange and the force leading to the adsorption process. Due to the strong bond between MOF and Cd(II), Cd(II) was dominant over Pb(II) during the adsorption process. A correlation coefficient R2 greater than 0.99 indicates that the training and testing algorithm is highly accurate. Another example of metal uptake from aqueous environments is a 2022 study on arsenate uptake [122]. It identified 13 different types of absorption. Figure 2. Copyright © 2018, American Chemical Society. The knowledge transfer method was used to predict the electrical conductivity of MOFs "Revised with permission from [111]". MOF structures have been studied under a wide range of experimental conditions, including temperature, pH, and adsorbent surface activity. The study was conducted using several machine learning algorithms, in particular Random Forest (RF), Extreme Gradient Boosting (XGBoost) [123], GBDT and Light Gradient Boosting Machine (LightGBM) algorithms. LightGBM performed best among these models, exhibiting the smallest average absolute relative error (AAPRE) of 2.88% and the smallest standard error deviation (STD) of 0.0628. These examples can be of great importance for applying the prediction of this property when designing DDS in

future. Gases are an important source of energy, and gas storage is one of the most effective ways to provide clean energy. It also prevents toxic gases from entering the atmosphere, thereby controlling the greenhouse effect. Gas therapy is considered as a new approach to the treatment of various diseases, from cancer to infectious and cardiovascular problems [126]. The difficulty of delivering most gases forces researchers to look for new ways to store and transport them [68]. Also, co-delivery of gases and drugs makes it possible to obtain a synergistic and cumulative effect from their combined use [127]. One of the best ways to store gases is through the use of MOF nanostructures due to their porous structure, large surface area, adjustable openings and easy functionalization [128–130]. Hydrogen ( ) can be used as a fuel, but its storage in fuel cells is difficult, but possible with the help of MOFs [131, 132]. Prediction of optimal structures of MOFs for hydrogen storage is possible using machine learning methods [133, 134]. In 2021, a model method based on Extreme Random Tree (ERT) was used for hydrogen uptake (Fig. 3) [5]. The algorithm was trained on 24,674 MOFs, which made it possible to make accurate predictions based on 820,039 structures in seven crystallographic aspects, among which pore volume and void volume fraction were the most effective for determining volumetric and mass capacities, respectively. The algorithm made it possible to determine suitable MOFs for use in the compression-expansion process and optimal structures for use under changes in temperature and pressure. The high correlation and accuracy of the modeling made it possible to create a website (HyMarc), which uses the model used in the article to predict the structures of MOFs. In 2022, the Gradient Improvement Regression (GBR) method was used to predict five properties of metal-organic compounds (MOCs), temperature and pressure-temperature changes that were applied in the liquid natural gas (LNG) gasification process combined with the absorbed natural gas charging method. gas (ANG). LNG-ANG can be used for storing and transporting methane. Received

the data were used to synthesize two MOC compounds (one with a cobalt core, the other with a copper core) for storing methane in an LNG-ANG mixture. Metal-centered cobalt MOC was found to be more stable after repeated methane adsorption and desorption process, which was predicted by molecular dynamics data. Predicting the gas storage capacity of MOF structures could lead to the design of more effective nanoparticles for gas therapy and delivery of gases, especially oxygen and nitrogen.



Speech Machine Learning

Figure 3. Prediction of hydrogen storage in MOF structures using a machine learning method, adapted with permission from [5]. in particular, such as nitrite acid, hydrogen sulfide and carbon dioxide, which are considered as medicines [136]. Water purification ensures the necessary balance in the environment and safe drinking water [137]. MOFs can be useful in water purification by removing contaminants from it. Therefore, it is necessary to test the water resistance in these MOFs [138, 139]. In addition, some other applications of MOFs (e.g., adsorption) can be carried out under high humidity conditions, which requires MOFs that are water-stable [140]. The ML method can classify MOFs into groups based on their water resistance or predict their water resistance based on structure [141]. MOFs, which are easily destroyed in water, cannot be used for drug delivery due to the rapid

destruction in biological fluids and drug release in environments other than the target organ. Therefore, to transport drugs, the MOFs used must be stable in water throughout the entire period of their use as a carrier [140]. However, to release the drug at the site of action (e.g., tumor), MOFs need to be degraded, so their structure must contain properties such as low sensitivity to acidity, which requires a certain finesse in their formation [142]. Hopefully, machine learning algorithms will be able to predict such features in the structure of molecules and prevent the use of factors that can change the surface properties and size of nanoparticles (for example, the presence of multiple layers or coatings), which could lead to undesirable consequences [143]. The model, trained on approximately 200 MOFs, used metal sites, organic binders, and metal/ligand ratios to predict stability in aqueous environments. Models with two and three classes (to distinguish between stable and unstable MOFs), as well as SVM, RF and GB algorithms were used. Data from 10 already known MOFs were used to test the accuracy of the model, and it was also used to analyze the stability of new MOFs. There were previous discussions in the gas storage section where it was found that MOF materials could be effective in this process. However, there is a serious problem associated with the adsorption of carbon dioxide in the presence of water vapor. This process is complicated by the need for efficient absorption of carbon dioxide in the presence of water vapor [145, 146]. To address this issue, the researchers used Monte Carlo tree and recurrent neural networks (MCTS and RNN) methods to evaluate the ability of MOFs to absorb carbon dioxide under humid conditions. (Fig. 4) [147]. Low Henry coefficient and hydrophobicity are properties that provide resistance to water vapor. Synthesizability and high ability to absorb carbon dioxide are properties that provide better adsorption. The structure-property relationship showed that the optimal value for pore size is due to the fact that specificity is a priority

. and the optimal surface

adsorption at lower values of this parameter, and at larger values - a greater ability to absorb carbon dioxide. In this study, promising MOFs were developed for the adsorption of carbon dioxide in a press-and-switch adsorption process using the methods applied in this study. Algorithms for predicting water stability can also help predict stability in other environments, such as the blood or metabolic organs such as the liver. They can even facilitate the production of oral MOF formulas by recommending structures that are stable in gastrointestinal fluids. It is also possible to evaluate various quantum chemical characteristics of MOFs and use them as classification criteria. These characteristics include charge, spin and density of states, partial charges and optimized geometry. Computer models can be used to predict some properties of hMOFs. The models can be divided into two groups: (A) quantum chemical electron structural methods and (B) classical methods (e.g. molecular dynamics and molecular simulations). The Schrödinger functional equation method (WFT) and the Dirac functional equation method (DFT) are two ways to solve the Schrödinger equation, which is the main tool for describing various quantum chemical properties. DDS are used for various purposes. As is known, properties characteristic of quantum chemistry can show that chemical reactions occur in molecules. Therefore, predicting such properties in drug carriers may reveal whether this enhances the solubility of the loaded substance (as is the case with ABRAXANE) or whether the carrier can accommodate multiple drugs without fear of drug interactions [152–155]. At present, such a view may seem unrealistic, but such predictions have already been made for MOFs, although not yet for drugs, given all the subtle interactions between the body, drug, vehicle and food [156]. The Quantum MOF database has been introduced, containing the quantum chemical characteristics of approximately 14,000 MOFs (Fig. 5) [152, 157]. The work used a combination of DFT methods and various machine learning methods. Among them

The best results were shown by the crystallographic convolution neural layer (CGCNN) [158] and the smooth overlap of atomic positions (SOAP) method [159]. The mean absolute deviation, R2 correlation coefficient, and Spearman correlation coefficient for CGCNN were 0.274 ± 0.008, 0.876 ± 0.011, and 0.932 ± 0.005, respectively. For SOAP, these values were 0.357±0.008, 0.822±0.010, and 0.910±0.003. This model predicts various quantum chemical properties, in particular, it predicts the band gap to determine the conductive structures of MOFs based on low band gap materials. Although the prediction accuracy of SOAP is slightly lower than that of CGCNN, it can be used in conjunction with non-supervisory dimensionality reduction techniques to identify patterns between structure and properties in the database generated by the study. Machine learning methods tested using structural features included Stoichiometric-45 and Stoichiometric-120. The 2021 paper resulted in new data on MOFs, raising hopes that such predictions could be used in the development of drug formulas.

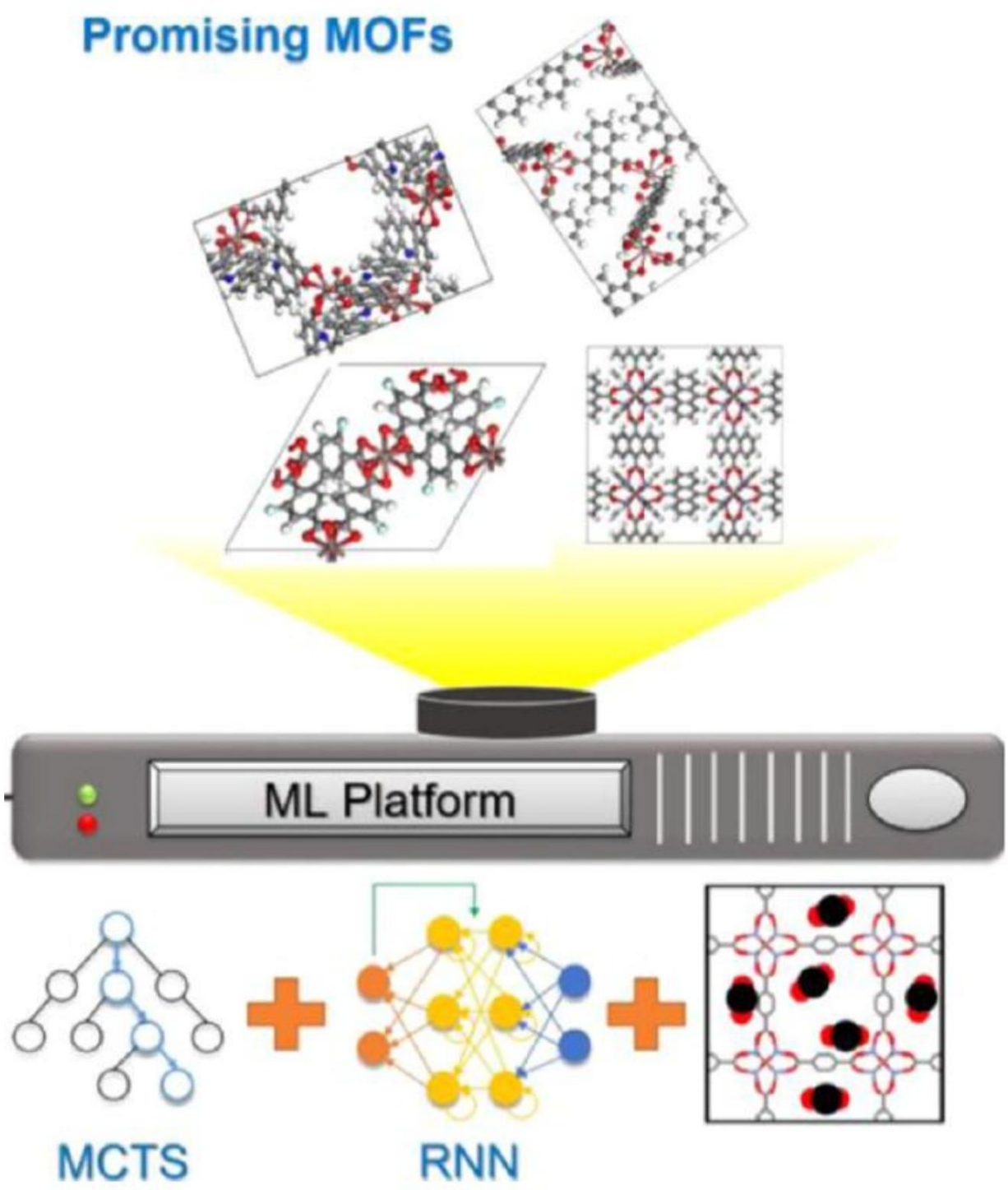
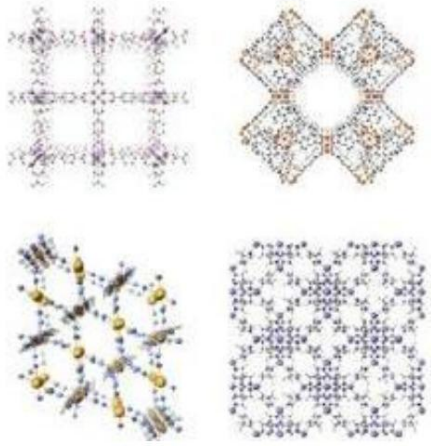
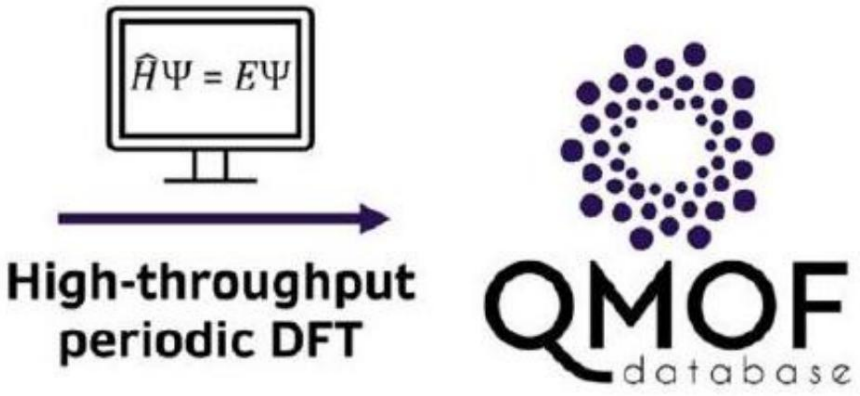


Figure 4. Simulation of the CO2 absorption process under high humidity conditions using ML algorithms. Note: Adapted with permission from [147].


Experimentally synthesized MOFs



15,713 structures Figure 5. Schematic of the quantum MOF database. Translated with permission from [152,157]. The use of MOF materials as adsorbents, for the separation of substances, and in other areas depends on the accessibility of their structure. The accessibility of the MOF structure to guest molecules determines the ability of the material to perform its functions. ML can use various algorithms to determine the relationship between the structural or chemical properties of MOFs and the accessibility of guest molecules [160–163]. The functional flexibility of the MOF structure and its ability to interact with cells or their receptors can be predicted by properties that determine the accessibility of guest molecules, which can be of great importance for the development of drug delivery system (DDS). Many carriers bind to ligands for internal transport in tissues affected by inflammation or cancer through cell receptors [166]. The binding potential of MOFs to such ligants (from antibodies to small molecules) can be predicted using molecular linear models, which avoids pointless laboratory experiments and unnecessary consumption of materials [167]. IN

A 2021 paper examined the chemical properties of metals and organic ligands to predict guest availability in MOFs based on these factors using machine learning algorithms (see Figure 6) [164]. A database of 3D MOFs from CSD was collected, and the RF algorithm was selected as the most efficient for correlating the metal-ligand ratio with the corresponding guest accessibility, defined as pore size greater than . 14296 MOFs were selected as the 1M1L3D set, each containing one metal and one ligand. To train the algorithm, various metals, ligands, and guest accessibility values were used depending on the pore size. Atomic mass, number, radius, polarizability, electronegativity and electron affinity were taken as metallic characteristics. A second algorithm was then developed that takes into account whether the ports of the MOF are larger than . Figure 6. MOF model for determining guest availability for interaction using a machine learning method, adapted with permission from [164]. The model determined whether the pores were small (2.4 to 4.4 Å), medium or large (greater than 4.4 Å), and a third model determined whether the pores were medium (4.4 to 5.9 Å) or large (more than 5.9 Å). The accuracy of these models was 80.5%, 76.4% and 68.5%. The results of the study can be used in the field of absorption, separation and catalysis. Knowledge of such important molecular properties would improve drug development and synthesis, preventing problems during development and during in vitro and in vitro testing. Machine learning (ML) has emerged as a new tool to predict properties of materials that are difficult to determine using traditional methods. In this review, various ML methods for predicting the properties and capabilities of MOFs in drug delivery systems (DDS) were presented. As the studies presented in this review have shown, it is clear that ML models have a significant impact on MOF research. ML technologies have been successfully used to predict mechanical properties, electrical conductivity, adsorption properties, gas storage capacity, gas separation capacity, carbon dioxide absorption capacity, water resistance and accessibility to

guest particles. However, there is a lack of research focusing on predicting the properties of MOFs in biological applications, especially in the field of drug delivery. Although significant advances have been made in this field, the application of ML technologies in MOF research is at an early stage and more research is required. Predicting the properties of MOFs will become more relevant in the near future as work is currently underway to synthesize MOFs for drug delivery purposes. Combining ML technologies with laboratory research on the synthesis of MOFs for drug delivery purposes may be the key to faster development of drug carriers. The research reported in this article was supported by Tehran Shahid Behesti University of Medical Sciences. Ahmadi Mahnaz: Visualization, writing - original version. Pouyanfar Niki: Conceptualization, revision, writing - original draft. Gorbanbidkorpekh Fatemeh: Obtaining funding, project organization, supervision, writing - editing and additions. Ayubzadeh Syed Mohammad: Writing - original version.