A Review on AI Integration With Green Chemistry

Krishna sawale[1], Dr. Priya Joshi[2], Sukhraj Girme[3], Rudra Khandelwa[4], Devesh Bhosle[5]
[1, 2, 3, 4, 5]Dept of Applied Sciences and Humanities
[1, 2, 3, 4, 5] Pimpri Chinchwad college of engineering, , Sector 26, Pradhikaran, Pune, Maharashtra, India

Abstract- *This review aims to demonstrate the combined power of Artificial Intelligence (AI) and Machine Learning (ML) with Green Chemistry, an active process for designing chemicals that seeks to suppress or eliminate deleterious substances right from the source. With the gradual shift in the chemical industry and pharmaceutical sectors towards the concept of the circular economy, the manual process required for the optimization of chemical synthesis routes and analysis of environmental effects has become inadequate in dealing with complex regions within chemical compounds. This review evaluates various computational technologies, particularly the use of ChemPager for Process Mass Intensity Prediction, mUCT-dc-V algorithm for optimal retrosynthetic analysis, and accuracy in carbon footprint predictions offered by FineChem2. In addition, it also weighs the use of open-source platforms in the measurement of molecular complexity and the "Five Pillars" framework for predictive toxicity. This review brings together different use cases of "artificial intelligence" and combines them in an innovative manner by highlighting the ways in which "green by design" intelligence derived from these use cases can be utilized to optimize chemical structures by removing waste, making them less toxic, as well as making them energy-efficient. It also discusses some of the key issues that exist in "black box" intelligence and the energy needs of "computational intelligence" training.*

Keywords: Artificial Intelligence, Green Chemistry, Sustainable Synthesis, Retrosynthetic Analysis, Carbon Footprint Prediction, Predictive Toxicity

I. INTRODUCTION

Green Chemistry, also known as sustainable chemistry, is an innovative approach in the field of chemical engineering and research pertaining to product and process design. This approach aims at limiting or avoiding the use and production of hazardous substances. Another difference is that while the conventional approach is reactive in nature, pertaining to treatment after the waste is produced (remediation oriented), Green Chemistry is proactive, redesigning, at its core, the scientific process for preventing pollution at its source with the aim of making chemistry safer.

Origins and Core Philosophy The idea was formalized, in the 1990s mainly through the chemists Paul Anastas and John Warner, as a set of 12 Green Chemistry Principles. These provide a guideline set for scientists working on methods of a sustainable nature. They call for a moratorium on toxic solvents and non-renewable resources in favor of renewable materials, energy-efficient processes (reactions conducted at room temperature) and "atom economy"-the maximization of every atom from the reactant being incorporated into the product molecule itself.

Key Differences and Impact

Green Chemistry must be differentiated on the basis of the special concern it holds compared to another type of chemistry, namely the impact it could have on nature due to the adverse effects of various harmful chemicals that can be found under the domain of a specific type of chemistry called Environmental Chemistry. While the former conducts research on the harmful effects of various chemicals on nature, the latter seeks a halt to the production of the said harmful chemicals that end up in nature. This will save industries money on disposing of the waste and will save the Earth as well.

Green Chemistry is the backbone of the greener future of the industrial age of the current era. It is applied while manufacturing biodegradable plastics and safer medicines.

Feature	**Green Chemistry**	**Traditional/Environmental Chemistry**
Focus	Prevention (Design phase)	Remediation (Cleanup phase)
Goal	Eliminate hazard generation	Monitor and clean existing pollution
Key Metric	Atom Economy & Efficiency	Parts per million (pollution levels)
Philosophy	"Benign by design"	"End-of-pipe" treatment

1.1 Chem Pager: Now expanded for ever greener chemistry

The challenge of sustainability in the pharmaceutical sector has led to a need to find a new way of implementing "Green and Sustainable Chemistry" to ensure that the minimization of ecological footprints will not be opposed by the requirements of cost-effectiveness. An important aspect of the sustainability challenge has been the Process Mass Intensity (PMI), proposed by the ACS Green Chemistry Institute Pharmaceutical Roundtable (ACS GCIPR), that emphatically focuses on the material use volume in terms of the volume of the product. An important issue continues to remain in the assessment of the synthesis process, because it can greatly differ based on the stage of maturity to which that particular process belongs. For example, the comparison of a commercial process to that of a synthesis process that has been drafted on paper.

In this review, we examine the 2019 update of ChemPager, which was created by researchers at F. Hoffmann-La Roche as a solution for this problem. Originally based on the TIBCO Spotfire platform with the intention of displaying critical performance indicators (KPIs) such as production cost and solvent consumption, ChemPager was updated to incorporate the "PMI Predictor" web application. This enables chemists to calculate the cumulative process mass intensity (cPMI) value for proposed synthetic routes relative to a database of large-scale reactions irrespective of experimental values.

Being able to compare "real" data with "predicted" probability distributions, this tool can act as a decision-support tool to show where green chemistry enhancements might be possible at quite early stages in the scouting stage. In this review, the functional development of ChemPager will be discussed and assessed for methodology in predicting environmental impact, and in this regard, applied to a specific case study involving RG7834—a proposed treatment for Hepatitis B—that successfully pinpointed which green pathways existed for a chiral pool synthesis. In this manner, this review illustrates how data analysis is important for proactively developing greener chemical processes.

1.2 Towards Efficient Discovery of Green Synthetic Pathways with Monte Carlo Tree Search and Reinforcement Learning

There has been an increased demand lately in efficient retrosynthetic planning methods due to the trend of automation in both the pharmaceutical and fine chemical sectors. While deep learning significantly improved the prediction of chemical reactions, the search algorithms being used in combination with chemical reactions, such as MCTS, have faced efficiency issues.

Traditional algorithms like the Predictor + Upper Confidence Bound applied to Trees, popularized through the AlphaGo system, were heavily dependent on previous probabilities yielded through the use of policy networks. In the case of retrosynthesis, these were never the case and thus always suboptimal. Consequently, the PUCt algorithms tend to be "greedy," starting from very probable reaction paths early on and then neglecting the not-so-obvious but potentially better reaction paths. Apart from this, traditional CASP tools had the "validity" as the only parameter while ignoring the important process parameters like the toxicity of the solvent and the number of steps.

The Wang et al. paper offers a solution to both problems by using a revised search strategy, which represents a balance between search and PUCT, and a scoring function where non-green solvents are penalized.

It comprises three technical novelties, which in turn make up the mUCT-dc-V algorithm.

One major step forward is the shift from static or policy-weighted exploration to the concept of Dynamic Exploration “C”

In standard UCB formulae, C itself remains a fixed exploration parameter. In this modified version, parameter ‘c’ changes dynamically as a function of search progress:

C = max(Q)/2

where 'Q' is the quality value of the node.

The mechanism provided forces the algorithm to re-explore “underrated” retrosynthetic moves, which are actions that may be rated as a low probability by the policy network, but can open much shorter or more efficient paths. The mechanism ensures that the system is robust to a poor policy network that is only used to focus the search beam, rather than decide the exploration-exploitation ratio.

In order to further accelerate search speed, a "value network trained with RL self-play (bootstrapping)" is introduced. The network predicts "the synthesis of easiness" of an "intermediate" molecule, acting as a "look-ahead heuristic" that assists tree search in discovering commercially available precursors faster than random rollouts.

This article expresses qualitative guidelines of green chemistry in quantity form.
Reaction Solvent Score: This is a neural network that provides reaction solvent suggestions. These suggestions receive scores: +1 for Green, 0 for Mediocre, and -1 for Non-Green. This RSS will be the weighted average of the scores.

Compound Solvent Score (CSS) The objective function for searching, which calculates the Cumulative Reaction Solvent Penalty (RSP). A lesser RSS introduces a small penalty, for instance, -0.1; Green solvents The toxic solvents possess lower RSS values and are penalized heavily, i.e., -1.0. Optimization aims at the least toxic solvents, that is, the minimum number of penalties or the shortest route.

1.3 Chemical data intelligence for sustainable chemistry

There is an urgent need for a change in the chemical industry for a sustainable future through a paradigm shift away from their current reliance on fossil fuels to the integration of a circular economy. With the chemical industry attempting to exploit renewable feed stocks such as biomass and waste materials, the level of complexity in the problem of identifying efficient and sustainable reaction pathways increases exponentially. Currently, research work in the field of identifying efficient and sustainable alternatives itself is a manual task in itself with a considerable emphasis on chemical intuition, which is inadequate for exploration of the vast chemical space.

However, the review will mainly emphasize the importance of Weber et al. (2021), which introduces the concept of "Chemical Data Intelligence" and how this can be done to tackle these issues. It has been based on the rapid digitization procedure that has been completed for the chemical data, and this technique provides an opportunity for a systematic approach towards automation processes. The following sections will shed some light on the three important aspects which have been identified as significant for this shift to take place successfully. These aspects have been (i) Data, (ii) Assessment Metrics, and (iii) Decision Making.

1.4Enhanced Deep-Learning Model for Carbon Footprints of Chemicals:

With an increasing need to make the chemical industry a sustainable one, being able to assess the environmental footprint of molecules at an initial design phase is now more necessary than ever before. Unfortunately, there exists a large gap: there is no PCF data available for hundreds of millions of substances. Traditional Life Cycle Assessment processes have always been quite resource-intensive; this is particularly true for novel substances that lack industrial experience.

Prior machine learning approaches like FineChem 1, or rapid-ANN/ANN-DP tried to bridge this divide with a concentration on algorithms for molecular structures. Though these programs represented a revolution in chemistry calculations, they faced a number of key difficulties:

"Limited Accuracy: High errors in predicting complex biological structures."
Narrow Applicability Domain (AD): Inability to generalize well across different molecular scaffolds and other chemical substructure classes.

Lack of Interpretability: Behaving like "black boxes" without providing any explanation for why a particular molecule has a large carbon footprint.

For addressing these issues, we introduce FineChem 2, which is a deep learning-based model designed on the basis of an atom-bond transformer architecture. By using message passing neural networks and self-attention components, FineChem 2 is able to capture the perception of molecular structure and its impact on the environment in a way that is superior to the existing model.

FineChem 2 is unique for the coexistence of the following two major innovations:

Architecture

The architecture of the model comprises of the atom-bond transformer network. The model uses molecular feature vectors, calculated Descriptors, as well as inter-atomic matrices. Such constructs help the model detect the complex relationships that exist between the atoms.

Data Quality: FineChem 2 is built with high-quality training dataset, unlike the previous versions. The dataset is derived from three high-quality sources:

The Ecoinvent Database

The IDEA Database “First-hand Industry Data There”

These developments allow FineChem 2 to better predict the effects of these molecules in the environment.

1.5 Evaluating Molecular Complexity with Open-Source Machine Learning Techniques to Predict Process Mass Intensity

Towards the realization of Green Chemistry, the Process Mass Intensity (PMI), which is essentially the sum of the masses of raw materials (solvents, reagents, reactants) divided by the mass of the final product, is the most optimal method of measurement.

Nonetheless, a purely raw PMI value has to be set in relation to some kind of baseline. Obviously, a complex molecule will require a greater PMI than a simple one. In the past, there was a "SMART-PMI" model (Sheridan et al.) that correlated "crowdsourced complexity" to PMI values. That older scheme was dependent on commercial software (MOE) and used 207 descriptors to effectively put sustainability metrics behind a "paywall."

The problem is addressed by the reviewed article by creating an open-source solution that not only preserves the same level of accuracy as the contemporary method but also reduces the computational structure by a huge extent.

1.6 Review of "Machine Learning Methods for the Forecasting of EnvironmentalImpacts in Early-Stage Process Design" • This paper presents an approach that

In the chemical industry, there is a paradox concerning sustainable engineering. This sustainable engineering field is significantly influenced by the decisions that are made at the early stages of design. This is paradoxical because it’s at the early stages of design that little sustainability information is available. This paper highlights the issue raised by Aboagye et al. concerning sustainable engineering by applying ML.

"Greenby-design" decisions refer to decisions that involve design that takes into account, or is informed by, green principles. In other words, decisions that could only be realized after resource commitment through advances of science, and yet are informed through green principles, Design decisions informed by green principles, but implemented only after resource commitment through advances of science, meaning through science that advances only after resource commitment.

"Greenby-design" decisions refer to decisions that involve design that takes into account, or is informed by, green principles. In other words, decisions that could only be realized after resource commitment through advances of science, and yet are informed through green principles.

Inthis The authors’ data set comprised a reliable collection of 350 common solvents for which the authors were able to extract impact data available on the $\text{SimaPro}^{\circledR}$ data base. This approach can further be subdivided on three steps:

Feature Engineering; The project involved two groups of input variables.

Thermodynamic Descriptors-15: The critical temperature, heat capacity, and standard Gibbs free energy of the compounds are taken from Python libraries.

Molecular Descriptors 200 structural descriptors, such as molecular weight, carbon atom numbers, and functional groups, derived from the SMILES strings by $.RDKit$

Data Preparation: Missing label data in the training data was dealt with by the authors using $\text{k-Nearest Neighbors (kNN)}$ as the missing data

Model Architecture Choice

The engine that will be used is the algorithm XGBoost, which is an ensemble model and extremely efficient as well as accurate.

They used a method called Sequential Backward Feature Selection ($\\text{SBFS}$) to reduce the complexity of models. As a result of this process, they achieved extremely efficient models that could predict a set of targets with only 10 features, 5 thermodynamic and 5 molecular features. It reduced computational costs drastically

1.7 Artificial Intelligence in Green Organic Chemistry: Pathway to Sustainable and Eco-Friendly Chemistry

There is an imperative emerging in the chemical industry for a changeover to a more sustainable path due to increasing levels of environmental degradation and resource consumption. Green organic chemistry responds to the pressing concern for a more sustainable process in the chemical industry by ensuring that the twelve principles are followed to the letter to emphasize the elements of the avoidance and reduction of waste and toxicity, as well as the preservation of energy during chemical procedures.

In spite of the enormous progress achieved in such areas as biocatalysis or photoredox catalysis, the task of optimization of synthesis routes is extremely complex. The huge amount of experimental data, as well as the complicated non-linear dependencies between the variables of the reactions, make it extremely problematic to find the most eco-friendly synthesis route.

Artificial Intelligence and Machine Learning are currently transforming this field by providing high-end solutions for the purpose of sustainability enhancement. AI finds a vital role in this context as it helps in the optimization of chemical reactions and reduces their sustainability. AI can classify a large number of sets on the basis of which it can optimize chemical reactions and provide a different synthetic pathway that is more sustainable.

It aims at evaluating the revolutionizing influence of AI and ML, specifically in terms of their uses in the critical fields of chemical synthesis:

• Reaction optimization
• Artificial intelligence solvent selection
•Catalyst and Reagent Design

However, going ahead with focusing on the widespread use of AI poses numerous challenges, and among them, the absence of quality and carefully gathered data related to chemicals and the challenges posed because of the “black box problem” because of the lack of transparency of AI is one of the most relevant ones. After a rigorous analysis of these situations, this paper aims at making the reader comprehend the crucial role that AI and ML can assume for purposes of attaining a sustainable and ecologically friendly future related to organic synthesis.

1.8 Machine Learning for Toxicity Prediction Using Chemical Structures: Pillars for Success in the Real World

The pharmaceutical industry is at a critical stage at the moment.The escalating cost of developing new medications is pitted against a high rate of failures, where safety risks and toxicities have continued to dominate. Although conventional in vivo studies have remained the ultimate for safety testing for many years, these are resource-intensive and pose a range of ethical concerns and have often proven inefficient in simulating human-specific toxicities. Hence, there has been a paradigm shift to the implementation of New Approach Methodologies (NAMs), including Machine Learning (ML)-oriented solutions.

Machine learning (ML) based Quantitative Structure-Activity/Property Relationship (QSAR/QSPR) modeling can potentially integrate into drug research to predict toxicity of a molecule before it is actually synthesized. Yet, there are challenges associated with using machine learning algorithms. Biasing related to

That is, non-representative training data, the use of inappropriate algorithms, or wrong procedures regarding validation may produce non-accurate predictions or non-optimal decisions. In response to these issues and in order to fill the gap that exists between modeling and applied reality, a framework called "Five Pillars for Success" is proposed by Seal et al. The following are the most essential points that this review will cover: (1) The strict criteria needed for selecting data sets with relevance to human toxicology. (2) The representation of chemical structures into valid forms. (3) The use of valid algorithms. (4) The imperative for prospective validations in modeling, and (5) The use and influence of modeling predictions within the drug discovery process. These pillars are crucial to strictly follow for increasing the predictive validities of models using machine learning and for accelerating applied research to develop safe and effective drugs.

II. DISSCUSION

2.1 Chem Pager: Expanded for Ever Greener Chemistry

Hugo Loureiro1, et al. describe the extension of the ChemPager resource that assists in dealing with the gap existing between the theoretical "paper" routes and optimization routes through the application of the Monte Carlo algorithm for calculating the predicted cPMI distributions. The resource uses the ACS GCIPR PMI Predictor and assists in performing an objective and early-stage evaluation of sustainability, similar to that observed in the optimization of the RG7834 candidate Hepatitis B, where it demonstrated that the "green" optimization of the "chiral pool" involving valine was correct over the inefficient existing route. In addition, the resource assists in furnishing details with regard to step-by-step information through the aid of median comparisons within industries in terms of yields of reactions and the illustration of the destinies of solvents through visual images, such that correctly identifying that the incinerated solvents, such that solvents like dichloromethane (DCM), should actually be replaced.

2.2 Towards Efficient Discovery of Green Synthetic Pathways with Monte Carlo Tree Search and Reinforcement Learning
In recent years, combinatorial problems have gained

In their research paper on Chemical Science, Xiaoxue Wang[2] and the team from the Klaus Jensen Lab demonstrated the failure of traditional search algorithms in achieving a balance between search speed and the search of high-quality chemical pathways. The team evaluated their proposed mUCT-dc-V algorithm against the traditional UCT and AlphaGo variants of the PUCT algorithm and found that greedy algorithms are prone to getting stuck in local minima. The PUCT algorithm favors exploration along the most probable pathways. The pathways that can result in valid but inefficient pathways according to the neural networks.

The model has addressed this problem by implementing the dynamic adjustment of the exploration parameter (c) by the model, thus addressing the issue. This method allows the search process to resort to the low-probability branches, which might end up being neglected by the model. This resulted in a positive impact, which gave the success rate of mUCT-dc-V a 19.7% improvement over the success rate of the Classic UCT model, along with PUCT, which recorded a relative superiority of 16% over the success rate of the PUCT model. This model does not allow the search process to settle on the first solution that is available.

Anyway, the greatest success was in the quality of the “green” chemistry optimization. The dc-V optimization demonstrated an improvement in 71.4% of cases in the length and greenness of the paths compared to the PUCT method. Once again, this proves that the search should be dynamic in order to successfully extract the “hidden gems” – those rare paths that have a smaller length and are more green compared to the general paths suggested by the policy networks.

2.3 Chemical Data Intelligence for Sustainable Chemistry

Starting from here, the commentary will move on to discuss the integration of data, metrics, and decision framework as conceptualized by Weber[3] et al. (2021) in this section. Overall, the point being made in this section is that while each of these areas of research has made strides in their own right, it will take each one being able to contribute in a way that supports an overall “Chemical Data Intelligence” approach in order for these new pathways in the realm of chemicals production, namely ones that are more sustainable, to be produced.

The greatest challenge that has so far hindered the automation of the process of sustainability assessment, and the most significant challenge in general, has to do with the status of the data in the subject of chemistry. Although there has been some acceleration of initiatives regarding the digital representation of the data of chemistry, the process tends to be unorganized in most instances, and even then, it would probably fall below the reasoning power of the computer.

Incompleteness and Standardization: One of the most critical aspects deals with the “data completion” problem. The mass balance calculations, a prerequisite in the computation of, for example, the E-factor, among other sustainability factors, can easily never be completed due to the certainty of the absence of stoichiometric coefficients and the absence of multi-step reaction description in the reaction entries. Besides, critical process information, such as the weight of the solvents, the reaction temperature and pressure, and the reaction yield, can neither be stated nor pointed out.

Semantic Web Solution - A paradigm shift from the existing issue of mere digitization of information can occur with a new approach called 'semantic web'. This will be accomplished with the help of ontology and knowledge graphs. These will enable the use of URIs to link various information sources such as properties, market information, as well as toxicology.

The Need for Inference in Data: Due to the presence of numerous gaps in the reported data, the need for the application of "data inference" techniques has come into the limelight. The need for recommendatory systems, like the application of NLP techniques, has become paramount in the prediction of missing reaction conditions and yields, thereby acting as gap-fillers in the chemical space.

A paradigm shift is necessary in the transition from “green chemistry” to the concept of “sustainable chemistry” involved in the process of reaction routes. Present methods, including the idea of atom economy, fail because they generally lack independence.

Resource-Based Assessment: The vision of the authors regarding the importance of considering the parameters of thermodynamics, especially that of exergy, which represents the loss of quality of energy and material resources, has paramount importance in the sense that “exergy has the unique status of being the only DPR that spans the two domains of the environmental impact space, resource depletion and economic cost in terms of the quality of energy.

System Boundary in a Circular Economy In a circular economy, the “waste” in the system boundary becomes one of the issues that need consideration. That is, something that can be a by-product of a particular chemical reaction can also serve as a raw material for a different chemical reaction, wherein the definition of waste is based solely on the boundary set within the system.

Al The final area of the concern of this problem statement involves figuring out the most optimal paths through the massive “reaction network.” This extends beyond simple routing, and actually involves optimization. Network Topology: The chemical reaction networks are "forward/backward" branched trees and not the road map corresponding to the optimal control problem, due to the optimization at each step involving multiple coreactants and coproducts. The present network needs algorithms for tackling the interplay between the whole. Model Fidelity and Feasibility The degree of superstructure optimization concerning details like MINLP models is highly accurate in terms of intense computational costs regarding process units as well as heat transfer. On the contrary, Reaction Network Flux Analysis in decision-making during early stages enables screening through huge databases based on simplified linear prediction models. Weber et al. think that heuristics and machine learning will emerge as trends in decision-making in the coming future.

2.4 Deep-Learning Model Enhancement: Carbon Footprints of Chemicals, FineChem 2

FineChem 2, proposed in the 2024 paper published in ACS Sustainable Chemistry & Engineering by Dachuan Zhang[4] and et all (2024)is a deep learning model that outperforms previous pre-LCA predictions by 30%-55% in the RMSE, or accuracy of predictions, of carbon footprint predictions on test datasets. What is perhaps the most fascinating aspect of FineChem 2 is that it has Applicability Domain greatly expanded, which means that the model is able to make predictions on the entire newly discovered class of molecule structures. The value added to the scientific community of drug and material researchers, who venture into the uncharted territories of the chemical space that is well beyond the database limits, is immense.

The important innovation in the FineChem 2 model lies in the fact that instead of the “black box” approach, it uses the Explainable AI (XAI) approach. By making use of the “attention” part of the model, it manages to identify the “hotspot” substructures, which are the atom arrangements in the molecular structure contributing maximally to the molecular ‘Product Carbon Footprint’ (PCF). The authors observed that the ‘attention’ weights are strongly analogous to the computed PCF contributions via the traditional LCA method. The possibility for chemists to conduct ‘hotspot analysis’ in terms of modifying the ‘hotspot’ of the molecule in order to reduce carbon footprint prior to actually synthesizing the compound in the laboratory has finally been realized.

Despite this progress, Zhang and Hellweg acknowledge the fact that certain limitations exist in the current implementation of this model. The FineChem 2 has developed in terms of purely organic chemicals, which has shown mediocre capabilities in terms of working with polymers and inorganic compounds. In fact, this model may create certain problems in the pharmaceutical sector of specialty chemicals. In this regard, the carbon footprint would be energy-oriented in terms of purification rather than chemical substances in this matter. Despite this fact, FineChem 2 has made revolutionary progress with respect to incorporating sustainability in the initial process of chemical engineering.

Table 1: Comparison of Applicability Domain Coverage

Metric	**Previous Tools / LCA Databases**	**FineChem 2**	**Improvement**
Overall AD Expansion	Baseline	**+ sim>75%**	Significant
OECD Organic HPV Chemicals	sim10% covered	**> 80% covered**	8x Increase
Sectors Covered	Limited	Food additives, Plastics, Daily chemicals	High Diversity

2.5 Assessing Molecular Complexity Using Open-Source Machine Learning Techniques to Predict Process Mass Intensity

This is a masterclass on "Explainable AI" as applied to the chemical sciences. Nicole Tin and Remus Osan [5] as a team have boldly walked away from the dark woods of obfuscated 'black box' systems and demystified a molecular complexity as one of a problem set in real-space tethered to properties such as chirality and torsion as opposed to purely abstract problem-solving. Their choice to examine the Random Forest algorithm, integrated with SHAP and MDI analysis, resulted in a 1,500 descriptor high-dimensional dataset being translated into a 'glass box' model using only four features. The more extreme dimensionality reduction benefits predictive accuracy; the GS-04 model shows superior values of \$R^2\$ within the class and lower RMSE. It also allows chemists to operationalize 'Green by Design' within the domain of Process Mass Intensity (PMI); and it gives real-time actionable intelligence. Moreover, Tin and Osan have facilitated the extension of the concept of democratic unsustainability by making the tool available as an open-source python package and web application, thus enabling even the most petite research laboratories to breakeven against the imperative of industrial green compliance and embargoed expensive commercial solutions.

2.6 Machine Learning Methods for the Forecasting of Environmental Impacts in Early-stage Process Design

The new framework, for Life Cycle Assessment (LCA) in this research, by Emmanuel A. Aboagye[6] and K.

Yenkie shows that the chemical structure of a molecule is the factor that decides the chemical structure of a molecule's outcome. The researchers found a Human Health Impact (HHI) correlation of 0.997. The Human Health Impact (HHI) correlation is very high. I think this result matters. The result shows that perfect predictions of toxicity can be made with models that use data. The models can use descriptors such, as critical temperature and XLogP. In my view Aboagye and Yenkie used the model to do a Cradle-to-Cradle analysis of the solvent NMP. Their work added knowledge to the industry. They found that the production phase (C2G) has a environmental impact than the energy used in the recovery phase (G2G). . Such an extent of focus shift clearly reflects that the most important element, in order to achieve sustainability in the process, is the target molecule. The R^2 of the other parameters has some instability issues that require further enhancement. The improvement achieved in basic properties such as branching (HallKierAlpha) in estimating complex parameters such as Global Warming Potential has potential implications in terms of the broad scope of achieving the use of fundamental principles and chemical properties to estimate the impact. It gives chemists the ability to choose solvents for laboratory work in such a way that it is "Green by Design" but at the same time maintains a screening level.

2.7. AI in green organic chemistry: The way forward in sustainable and eco-friendly chemistry

Gurinderdeep Singh[7] et al., is a prime example of a paradigm shift in the world of science that has taken place in the recent past, from being a "trial and error" and intuition-based "chemistry," which was a popular practice in the early days of chemistry, to a "glass box," data-based solution. The association of the abilities of AI with the 12 Principles of Green Chemistry makes it abundantly clear in Gurinderdeep Singh's research that accuracy of prediction is the best way of avoiding wastage.He obtains an R^2 value > 0.99 in ionic liquid properties prediction tasks, pinpointing particular "catalyst genes" for non-toxic activation. This model clearly establishes that the aim of being "green" can no longer remain an objective or a necessity in itself, as it becomes possible to measure it. Notably, Gurinderdeep Singh et al. tackle the "Green AI Paradox," analyzing the massive carbon impact of a particular deep learning approach. Through their advocacy of "Explainable AI (XAI)" practices, along with energy-efficient algorithms, Gurinderdeep Singh et al. make sure that this time, there aren't higher costs associated with discovery as opposed to the direct, positive "green" impact of the particular optimizations in the realm of chemistry.

2.8 Machine Learning for Toxicity Prediction Using Chemical Structures: Pillars for Success in the Real World

The framework laid by Srijit Seal [8] and co-authors gives a vital guide to the important area of predictive toxicology. It goes all the way from algorithm performance to real regulatory use. By defining the "Five Pillars for Success," the authors remind one that a toxicological model is only as good as its weakest link. These range from balancing high-quality data sets and chemical representations against the strict demands of prospective validation. The present review truly bridges the "translation gap." It underlines the requirement that machine learning should, when intended to replace animal testing, function within a well-defined applicability domain. In addition, it calls for transparent, explainable AI-driven decision support instead of simple yes/no answers. Seal and his team rightfully foresee in this field a future of integration: one in which structural data will be integrated with new approach methodologies such as organ-on-a-chip technology toward multi-modal risk assessments. Finally, this work advances a culture of open science and standardized validation. It ensures that computationally impressive AI-driven safety assessments will also be legally and scientifically sound enough to protect human health in drug development.

III. CONCLUSION

The appearance of Green Chemistry, together with the application of Artificial Intelligence, proves to be a paradigm shift in reaction to the trial-and-error stage previously exercised in the application, leaning instead towards an entirely informed method, not lacking in specificity either. Based on the explanation provided in the context of the discussion, it is clear that, among the current capabilities of AI-driven tools with regard to predicting the result of a reaction, identifying the most optimal option for solvent selection, or even performing an analysis with regard to strategy on an environmental level at an early reaction stage, specificity is much more pronounced compared to the preceding standards. The "glass box" movement, along with the adoption of Explainable AI (XAI), clearly dispels the lack of transparency previously occurring, hampering the application of these tools/technologies in the environment of chemistry, as identified.

However, there remain a number of challenges that have to be overcome in arriving at a fully sustainable chemical future. First of all, there is "the data bottleneck" or an inefficient condition characterized by partially and not fully developed chemical data. Then, ironically, there is the issue of energy use in deep learning algorithm trainings that, in some fashion, may somehow offset the gains made in chemical processes through optimization. In this regard, in order to advance toward a fully efficient chemical future, a full merger between chemical efficiency and computation of energy employed in computations must be realized through fully open strategies with regard to data sharing.

REFRENCES

[1] Loureiro, H.; Prem, M.; Wuitschik, G. ChemPager: Now Expanded for Even Greener Chemistry.*CHIMIA*2019, *73* (9), 721–726. https://doi.org/10.2533/chimia.2019.721

[2] Wang, X.; Schwaller, P.; Jensen, K. F. Towards Efficient Discovery of Green Synthetic Pathways with Monte Carlo Tree Search and Reinforcement Learning.*Chem. Sci.*2020, *11*, 10959–10972. https://doi.org/10.1039/D0SC03544H

[3] Weber, J. M.; Guo, Z.; Zhang, C.; Schweidtmann, A. M.; Lapkin, A. A. Chemical Data Intelligence for Sustainable Chemistry.*Chem. Soc. Rev.*2021, *50*, 12013–12038. https://doi.org/10.1039/D1CS00471A

[4] Zhang, D.; Wang, Z.; Oberschelp, C.; Bradford, E.; Hellweg, S. Enhanced Deep-Learning Model for Carbon Footprints of Chemicals.*ACS Sustainable Chem. Eng.*2024, *12*, 2700–2708. https://doi.org/10.1021/acssuschemeng.3c07256

[5] Tin, N.; Chauhan, M.; Agwamba, K.; Sun, Y.; Parsons, A.; Payne, P.; Osan, R. Evaluating Molecular Complexity with Open-Source Machine Learning Approaches to Predict Process Mass Intensity.*ACS Omega*2024, *9*, 28476–28484. https://doi.org/10.1021/acsomega.4c01234

[6] Aboagye, E. A.; Yenkie, K. Systems-Level Roadmap for Solvent Recovery and Reuse in Industries.*Sustainable Production and Consumption*2021, *24*, 101–114. https://doi.org/10.1016/j.spc.2020.08.005

[7] Singh, G.; Rajasekhar, E. S. K.; Mounika, K.; Tulasi, K. R. S. K.; Dondapati, T.; Himasaila, M.; Pulipati, S. Artificial Intelligence in Green Organic Chemistry: Pathway to Sustainable and Eco-Friendly Chemistry.*Asian J. Chem.*2024, *36* (12), 2731–2743.

[8] Seal, S.; *et al.*Machine Learning for Toxicity Prediction Using Chemical Structures: Pillars for Success in the Real World.*Chem. Res. Toxicol.*2025, *38*, 759–807. https://doi.org/10.1021/acs.chemrestox.4c00321