Au nanoclusters. Accuracy rates of 0.78 (R2) and 0.17 (Pearson's r) were achieved in a prediction of the adsorption energy using a two-step procedure for feature engineering and feature selection approach. Hulva et al. [7] observed that the most influential factor in CO adsorption sites was the Ag atoms distribution in Au after deciphering the most significant nodes of random forest. The ML model is readily generalizable to various Au-based nanoclusters. We anticipate the model's primary use will be as a filter to identify promising sources for in-depth study.

Predicting the Property of New Materials

Jiang et al. [8] aim to optimize properties such as electrolyte conductivity, thermodynamic Seebeck coefficients, and the efficiency of power conversion of organic-inorganic perovskite. Even after extensive trial and error, results from numerical simulations or the perception of chemical experts are sometimes disappointing. Applications of ML models, which can forecast the characteristics and structures of materials with a reasonable degree of accuracy prior to synthesis, may therefore be of great use. The ML model built in MATLAB was utilized by Benea and Benea [19] to locate rare solid electrolytes amongst over 12,000 different substances. Next, they looked through the research to find 40 crystalline solids with known atomic structures that matched a popular set of electrolytes utilized in exercise. Despite the relatively small quantity of the dataset, the "smart" characteristic based on past physical knowledge is required for an accurate representation of the data.

In order to characterize the local organization of atoms and crystalline chemical features, the author uses the atomic structure downloaded from ICSD as inputs and computes approximately 20 features according to their atomic geography, electro-negativity, mass and atomic radii of the structural element, such as lithium bond iconicity, atomic volume, lithium neighbouring components, and the minimum anion-to-anion separation extent.

The values of the experiment with the conductivity of lithium ions are applied as outputs, while a list of 40 known materials serves as inputs to a machine learning method. By adjusting the model's parameters in a steady fashion, it is possible to screen and categorize solid electrolytes, and eventually 317 compounds were indicated as promising. The updated MATLAB model proved to be two times as effective as Stanford graduate students working in a similar area, and three times as effective as random guessing, when it came to finding prospective new materials. The F1 score is almost 50% lower than the DFT values..

High-throughput simulations are just as useful as practical testing for ML training data. The thermodynamic stability of double perovskite halides was investigated by Dimopoulos et al. [10] utilizing high-throughput computation and ML. They began by creating a high-throughput DFT-based breakdown energy database, which was then linked to the thermodynamic constancy of 352 potential perovskite materials. They used this data to teach a machine learning model. Experiments on perovskite formability of 246 A_2B (I) B (III) X_6 elements (F1 95% score) further corroborated its prediction performances. The results of this study demonstrate that the model of prediction in ML is both cost-effective and efficient compared to experimental approaches.

Many important materials, including organic-inorganic lead-free hybrid perovskite, LEDs, OLEDs, LEDs, and monoatomic catalysts have been designed using similar approaches. The experimental validity of the latter two approaches has also been shown. In its current form, material science still relies somewhat on tried-and-true methods. There are still some hypotheses being employed to cut down on the amount of needed tests, and this trend is only expected to increase in importance. Instead, the regression model may be used to narrow down a pool of potential materials to only one that shows promise, cutting down on wasted time and effort spent on trial-and-error procedures.

Synthetic Route Planning

As organic synthesis follows a tried-and-true protocol, Haggin [11] may easily develop software to solve complex synthesis-related challenges. In the eyes of computer experts, chemical reactions are just a bunch of numbers that represent the link between two substances. Data structures like graphs and networks may be used to represent this presence. This structural information might then be used by AI to direct the synthesis process. Using a feedback loop and realtime spectrum analysis, Yang, Hansen, and Baldi [12] showed a robot that can carry out six separate organic synthesis experiments concurrently.

A chemically-assembled pressure pump and a raw-materials storage tank are its core components. The six reaction bottles that run in parallel are fed reactants by these pumps. In addition, the robot employs NMR and IR spectroscopy for real-time evaluation of reactions and SVM algorithm to routinely classify the reactions mixtures into both non-reactive and reactive combination. This approach can anticipate the reaction of reagent combinations and is quicker than manual experimentation. Moreover, the robot found four additional responses and had a prediction accuracy of over 80% after only gathering data from about 10% of the experimental dataset.

Experimental Parameter Optimization

In conventional material development, many factors in synthesis, processing, and device assembly must be examined and changed by hand. The efficiency is quite poor, and it may not even be possible to determine the best settings. To pinpoint the optimal value in the vast parameter space, ML makes use of its potent nonlinear regression capabilities. The concept has been implemented in welding procedures. Recently developed friction stir welding (FSW) has found widespread use in sectors as diverse as aerospace, shipbuilding, automotive manufacturing, and more. The impacts of the traditional

parameters of welding, such as the maximum shear strain on the tool spin, strain rates, torque, temperature, and possible causal agents on void formations, were investigated by Habibizadeh, Honarpisheh, and Golabi [13], who collected about 107 independent data for experiments from a wide-range authoritative literatures for the purpose of training different model of ML, including decision trees, and the neural networks. The findings demonstrate that both algorithms are capable of making reliable predictions about the emergence of errors, with a maximum prediction precision of 96%. The model permits for a complete optimization of the parameters of the welding process, preventing adverse phenomena like void development in FSW from Occurring.

Examples like this have been used in 3D printing. Microelectronic devices on flexible substrates are often fabricated using AJP (aerosol jet printing), a non-contact three-dimensional printing method. While it has the capacity to deposit unique designs, the printing quality will be heavily influenced by the intricate interplay between the primary process factors. To find the optimal AJP process window across various design domains, Caldana et al. [14] suggested a novel hybrid ML approach. Data clustering, knowledge transfer, classification, and experimental sampling are all staples of traditional ML approaches. The technique employs the experimental design known as Latin hypercube sampling to exhaustively probe the two-dimensional design spacing at a specific printing velocity. Then, using the K-means clustering technique, we studied how the SHGFR and CGFR affected the printing line quality, and we used a support vector machine to zero in on the sweet spot for running the presses (see **Fig 4**). The transfer learning technique exploits the association between several operation process windows in order to accurately detect more of these windows at varying printing rates. As a result, the improved printing speed allows for a much smaller sample size of rows to determine the new operating process window.

Lastly, an incremental categorization approach is employed to generate a 3D operating process windows that strike a balance between CGFR, printing speed, and SHGFR. Instead than relying on trial-and-error experimentation, as is the case with traditional approaches to quality improvement in 3D printing, this new approach is grounded on the principles of data mining and knowledge discovery. As a result, all of the information needed to maximize printing line quality may be gleaned from the many design spaces that exist.

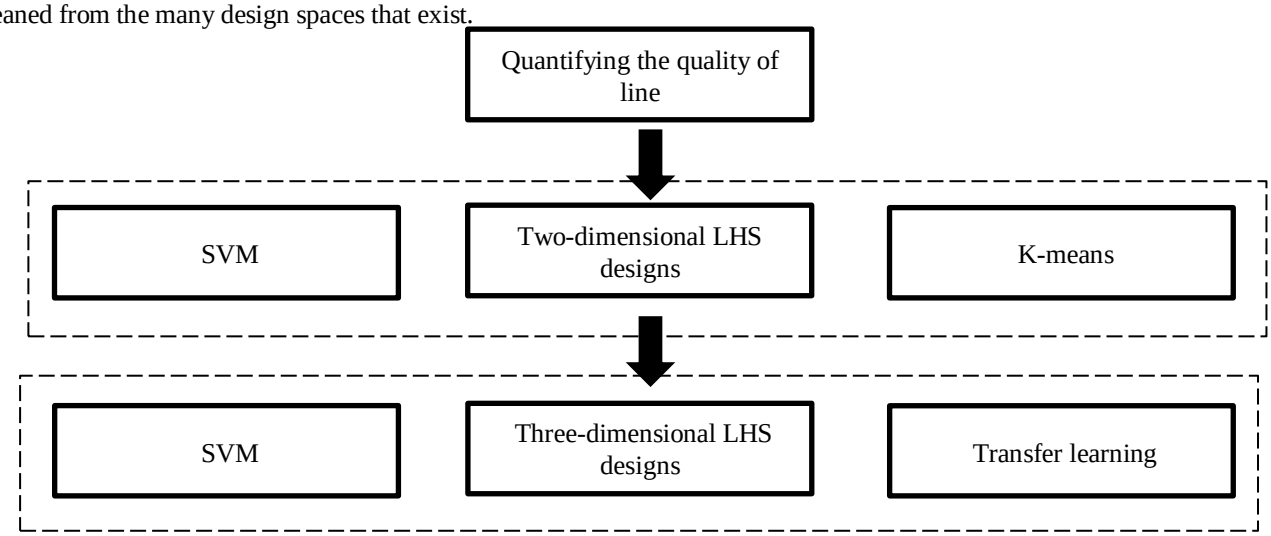

Fig 4. High-level Overview of the Hybrid ML Approach to Printing Parameter Optimization

As material synthesis moves toward complete automation, it will be included into industrial production 4.0 alongside tools like polymers' computable high-throughput analysis platforms. To begin this kind of high-throughput analysis successfully, ML must first analyse parameter spaces to identify the raw material's optimal ratio and the catalyst delivery rate for synthesizing organic molecules with the desired molecular weight, distribution, and side reaction profile.

Upgrading of Characterization Methods

Due to improvements in representational technologies, scientists have been able to examine structures and motions on the atomic scale, leading to the discovery of new rules governing materials. High-throughput analysis and preparation of materials based on the application of AI will be a given once the Material Genome Project matures. Due to deep learning's effective use of convolutional neural networks, significant development has been witnessed in the field of picture recognition.

This skill and knowledge in pattern recognition may be simply applied to the characterisation of micromaterials in images. Material science relies heavily on electron microscopy and defect analysis because they provide invaluable information about the microscopic structures and behaviour a wide range of materials systems and bio-materials. Establishing an adaptable and robust framework for automatic fault detection and categorization in electron microscopy would allow for faster analysis both post-recording and during picture capture. Manual identification is still the norm, despite being time-consuming and prone to error. This is because a higher number of images is needed to obtain statistical insights. As of late, Chakroun, Bouhamed, Kallel, Solaiman, and Derbel [15] used ML, image analysis, and computer

vision methods to learn about the magnitude and nature of faults (see **Fig 5**). The current performance of the application agrees with the quality analysis performed manually. If the software is upgraded, it will soon be possible to analyze massive data volumes in real time.

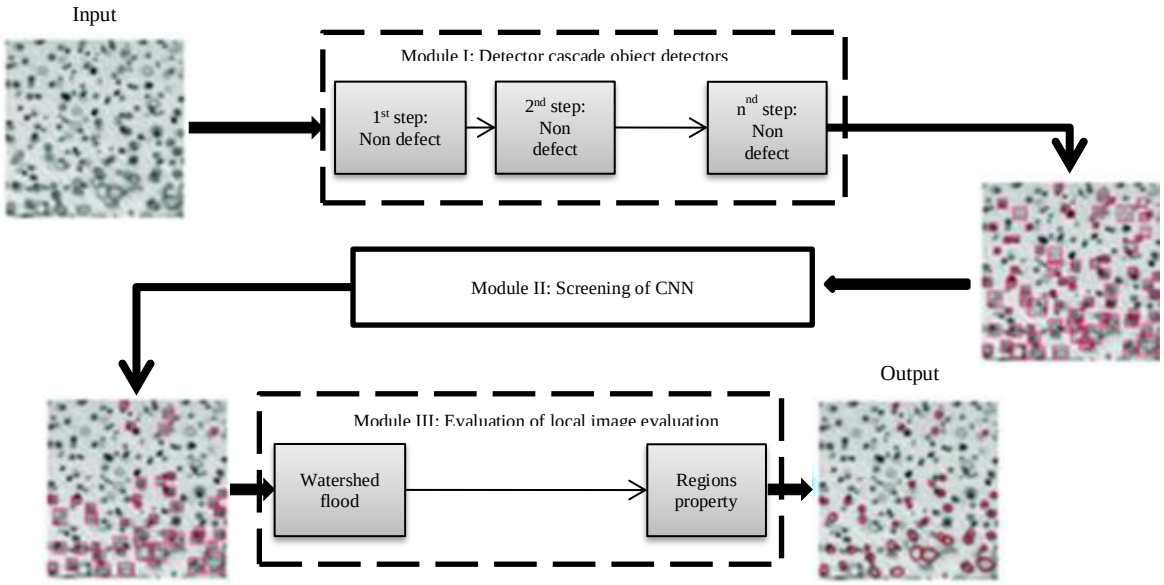

Fig 5. Proposed Automated Detection Method

Fig 5 is a high-level diagram depicting the suggested automated detection method. The first step in processing an incoming micrographic picture is the Cascade Object Detector, followed by the Convolutional Neural Network (CNN) detection and lastly the LIA (Local Image Analysis) module. When module I is completed, module II is then used to refine the bounding boxes and loop location, hence reducing false positives. Artificial intelligence (AI) may also be used to evaluate data from X-ray diffraction (XRD). When dealing with massive amounts of high-throughput characterization data, performing individual analyses to extract representative samples is a laborious and time-consuming process. Scientists may use ML to increase the speed of their analyses and unearth previously unknown patterns in their data.

Min [16] deposited ternary FeGaPd compound films on a single silicon wafer, resulting in 535 samples with varying ternary FeGaPd compositions, each measuring 1.75 mm^2 in area. X-ray diffraction (XRD) characterisation data was collected for 273 samples. Then, ML is used to perform unsupervised clustering on 273 XRD sample data using a hierarchical clustering technique, with the goal of combining as many single-phase samples into a single cluster as feasible. The analysis is far more efficient since just a small subset of the data in each cluster is examined. Based on these findings, it is clear that ML's dimensionality reduction and clustering technique may be used to effectively evaluate high-throughput XRD dataset, identify the stage distribution pattern and the multi-phase junction, and speed up the discovery process.

Increasing the number of times you cycle a lithium-ion battery will reduce its capacity. Researchers into battery technology have always been interested in improving cycle life. A new, data-driven, large-scale model has been constructed by Sandie et al. [17]. The capacity to employ ANNs to assess the laws of high-dimensional dataset allows for the prediction of the whole lifecycle of graphite batteries, or commercial lithium iron phosphate using just the charging and discharging dataset of the first couple of cycles, without any analysis of the process of battery deterioration. The first 100 cycles are used in the regression setup, and the prediction error is reduced to 9.1%. Using data from the first five cycles, the author constructs a classification scheme that yields a prediction error of just 4.9%. As a result, there will be fresh chances to improve battery manufacturing, cascade use, and optimization. Manufacturers of rechargeable batteries may, for instance, shorten the time it takes to create new battery types, verify the viability of proposed production methods, and categorize batteries in terms of their expected lifespan. In a similar vein, the expected lifespan of batteries in consumer electronics may be estimated. As a whole, the work places emphasis on the integration of data collection and data-driven modeling, an approach with promising implications for the study and improvement of complex systems like lithium-ion batteries.

The ambiguity associated with analyzing impedance data may also be cleared out with the use of ML. When it comes to studying and diagnosing electro-chemical battery and possible electro-chemical storage systems for energy distribution in the future, electro-chemical impedance spectroscopy (EIS) is a significantly effective technique. On the other hand, analyzing a huge quantity of EIS data is notoriously challenging. Many optimization techniques have gaps in their functionality. To achieve the right convergence of the fitting, Researchers must carefully create the equivalent circuit (EC) model, choose acceptable beginning values for the parameters of each component, and continually validate the output during the process. For this reason, Altay and Yildiz [18] recommended a ML-inverse approach that condensed a million

separate fitting optimization issues into a solitary one. Using many perspectives from the ML literature resulted in an error rate of less than 1% for addressing a single optimization task. The physical model parameter could be effectively fitted to the gathered set of data if an open-source model is constructed for FIS testing, and the system can be simply adaptable to other impedance spectra. There is no need for human intervention and the results are consistent every time. You may get a copy of the code used in this study by visiting -https://github.com/samuel-buteau/eisfitting. This self-deprecation of material science research as "stir-fried meals" is a current problem. It uses trial and error to find new materials, like as salt and water, to add to existing ones. High-throughput computers and machine learning allow material scientists to reduce the time and effort spent on trial and error.

Free open-source software that integrates AI data processing capabilities with a user-friendly interface may be necessary for the development of tangible AI in the future. Artificial intelligence has the potential to monitor all areas of scientific inquiry and provide other analyses to address representational issues. For the benefit of other researchers trying to address the same problems, researchers may post their own experimental procedure and findings. The use of AI will not eliminate the need for synthetic chemists, though. Indeed, AI will definitely become much significant to chemists to enable them to determine synthetic pathways better and promptly, but synthetic chemists will continue to discover novel reactions through actual scientific evaluation and diversify the theoretical framework of Chemistry. Using ML technology in conjunction with existing experimental data and theoretical foundations, AI-assisted materials design, application, characteristics, and synthesis research will significantly enhance the research effectiveness of scientists in the materials field and aid in rapid advancement of materials science.

IV. CONCLUSION AND FUTURE RESEARCH

This article provides a summary of the current state of the art in materials AI research, focusing on its most salient applications and benefits over more traditional approaches. High-throughput experimentation, high-throughput characterization, and high-throughput simulation computations are all necessary for the further growth of material informatics. In this section, we shall discuss the software and hardware future. Data analysis (statistical approach) is at the heart of ML, and the necessary data strives for breadth, depth, and impartiality. The calculated parameters of previous material informatics investigations lacked sufficient precision, limiting the scope of the research. The difference will be substantial if the datasets are made of more precise experimental findings. However, due to the overwhelming concentration of attention-grabbing research hubs, the existing experimental samples are insufficient. Generative adversarial networks, active learning, transfer learning, and auto-encoders are just a few examples of models that function well with little data. To prevent the "Black Box" problem, ML frameworks have to be changed into physical visuals or real-time knowledge. It may be possible to get some insight by averaging the responses of the neurons to the various descriptions. Materials informatics might be advanced by the use of other explanatory frameworks (for example decision trees) that can show the importance of different elements through the relative weight of their branches and nodes.

In most cases, a large amount of data is required for successful training of ML models. There are several potential sources for this kind of information, including web databases, published articles, and high-throughput experimental equipment. The use of online databases like ImageNet is becoming more common in the implementation of deep learning. Similar infrastructure is required for the advancement of material informatics. To compile data on electrolytes such their ionic conductivity, transference number, and chemical stability, Researchers created a database. In addition to a wealth of information on materials, published papers are a great source. Once these publications are organized by defined article formats, Researchers may conduct targeted information searches with ease using natural-language-processing technologies. High-throughput synthesis and characterisation tools may be upgraded by adding more sensors and software. The data acquired by these instruments is sent back into AI models for use in fine-tuning trial conditions. The settings may then be tweaked to generate samples with the desired characteristics. By these efforts, the "composition-structure-property-processing-application" link will be mapped in materials informatics.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding was received to assist with the preparation of this manuscript.

Ethics Approval and Consent to Participate

The research has consent for Ethical Approval and Consent to participate.

Competing Interests

There are no competing interests.

References

[1]. W. Sha et al., "Artificial intelligence to power the future of materials science and engineering," Adv. Intell. Syst., vol. 2, no. 4, p. 1900143, 2020.

[2]. B. G. Hyde and S. Andersson, Inorganic Crystal Structure. Nashville, TN: John Wiley & Sons, 1989.

[3]. M. Pentz, Handling Experimental Data. Buckingham, England: Open University Press, 1988.

[4]. J. X.-Y. Lim, D. Leow, Q.-C. Pham, and C.-H. Tan, "Development of a robotic system for automatic organic chemistry synthesis," IEEE Trans. Autom. Sci. Eng., vol. 18, no. 4, pp. 2185–2190, 2021.

[5]. P. Krishnamurthi, Y. Raju, Y. Khambhaty, and P. T. Manoharan, "Zinc Oxide-Supported Copper Clusters with High Biocidal Efficacy for Escherichia coli and Bacillus cereus," ACS Omega, vol. 2, no. 6, pp. 2524–2535, 2017.

[6]. S. Magnussen and D. Burgess, "Stochastic resampling techniques for quantifying error propagations in forest field experiments," Can. J. For. Res., vol. 27, no. 5, pp. 630–637, 1997.

[7]. J. Hulva et al., "Unraveling CO adsorption on model single-atom catalysts," Science, vol. 371, no. 6527, pp. 375–379, 2021.

[8]. Z. Jiang et al., "Ag/Br dual-doped Li6PS5Br electrolyte with superior conductivity for all-solid-state batteries," Scr. Mater., vol. 227, no. 115303, p. 115303, 2023.

[9]. M. L. Benea and O. D. Benea, "Mathematical modelling in Matlab of the experimental results shows the electrochemical potential difference - temperature of the WC coatings immersed in a NaCl solution," IOP Conf. Ser. Mater. Sci. Eng., vol. 106, p. 012025, 2016.

[10]. E. A. Dimopoulos et al., "HAYSTAC: A Bayesian framework for robust and rapid species identification in high-throughput sequencing data," PLoS Comput. Biol., vol. 18, no. 9, p. e1010493, 2022.

[11]. J. Haggin, "Easier-To-Use Computer Systems Available to Organic Chemists: New SYNLIB and Chemist's Personal Software Series systems are usable with personal computers, aim at synthesis chemists," Chem. Eng. News Archive, vol. 63, no. 34, pp. 13–16, 1985.

[12]. Y. Yang, L. Hansen, and A. Baldi, "Suppression of simultaneous Fmoc-his(trt)-OH racemization and Nα-DIC-endcapping in solid-phase peptide synthesis through design of experiments and its implication for an amino acid activation strategy in peptide synthesis," Org. Process Res. Dev., vol. 26, no. 8, pp. 2464–2474, 2022.

[13]. A. Habibizadeh, M. Honarpisheh, and S. Golabi, "Effect of friction stir spot welding parameters on the microstructure and properties of joints between aluminium and copper," Weld. World, vol. 66, no. 9, pp. 1757–1774, 2022.

[14]. A. C. F. Caldana, J. H. P. P. Eustachio, B. Lespinasse Sampaio, M. L. Gianotto, A. C. Talarico, and A. C. da S. Batalhão, "A hybrid approach to sustainable development competencies: the role of formal, informal and non-formal learning experiences," Int. J. Sustainability Higher Educ., vol. 24, no. 2, pp. 235–258, 2023.

[15]. M. Chakroun, S. A. Bouhamed, I. K. Kallel, B. Solaiman, and H. Derbel, "Feature selection based on discriminative power under uncertainty for computer vision applications," ELCVIA Electron. Lett. Comput. Vis. Image Anal., vol. 21, no. 1, pp. 111–120, 2022.

[16]. H. S. Min, "A review on the Penternary compound thin films," Researchgate.net. [Online]. Available: https://www.researchgate.net/publication/317637819_A_review_on_the_Penternary_compound_thin_films. [Accessed: 03-Mar-2023].

[17]. A. B. Sandie et al., "Observed versus estimated actual trend of COVID-19 case numbers in Cameroon: A data-driven modelling," Infect. Dis. Model., vol. 8, no. 1, pp. 228–239, 2023.

[18]. A. Altay and A. B. Yildiz, "Complete electrical equivalent circuit based modeling and analysis of permanent magnet direct current (DC) motors," WSEAS Trans. Circuits And Syst., vol. 21, pp. 182–187, 2022.