Accelerating 2D Device Discovery via Machine Learning

Dr. Ritesh Kumar
Assistant Professor, Department of Physics, Maulana Azad National Urdu University, Polytechnic, Bangalore – 560072

***Abstract*—Two-dimensional materials are about to change the world of next-generation electronics by making it possible to make smaller, more energy-efficient devices with new features. Because they are so thin and have great electrical, mechanical, and optical properties, they are perfect for making ultra-scaled transistors, as well as neuromorphic and quantum technologies. But getting these materials ready for production is hard because there are so many different factors that go into making and combining them. This problem can be solved very well with machine learning. ML speeds up the development cycle by finding hidden patterns in large experimental datasets and automating high-throughput testing. This paper examines that pivotal intersection, elucidating how machine learning improves material characterisation, refines growth processes to regulate morphology, and optimises fabrication parameters for enhanced device performance.**

#Machine #Learning #2D #ML #Quantum #Technology

I. INTRODUCTION

The primary objective of this research is to provide a comprehensive analysis of the expanding synergy between machine learning and two-dimensional materials. This study will show the most recent achievements, the problems that will still be there, and the huge chances that are still to come in the field of device engineering. Machine learning (ML) has grown from being a niche academic field to a game-changing technology that is changing everything from advanced speech and image recognition to medical diagnostics [1–8].

Machine learning has made it much easier to make sense of complicated experimental data in the field of materials science. This success has naturally carried over to the field. Researchers can find small patterns and connections that aren't straight lines in huge datasets by using these algorithms. These are insights that are often hard to see with just standard analysis. This processing power, along with the current lab setup, makes things even more effective. The rise of automated, high-throughput experimental platforms has made it much cheaper and faster to run each experiment, which has made it possible to collect and analyse experimental data much more quickly [9–12]. Machine learning is now seen as not only a useful tool, but also a key factor in speeding up the creation of new 2D material technologies. All of these accomplishments point to a shift in the way things are done, which puts machine learning as.

II. ML ALGORITHMS

Choosing an algorithm is not a routine box-ticking step in a machine learning workflow; it is a deliberate decision that must reflect the character of the dataset how large it is, how many features it has, how sparse the data may be, and whether the outputs are continuous values or discrete categories. Because these conditions vary widely from one problem to another, no single method can be applied universally. Instead, researchers select among three broad families of approaches supervised, unsupervised, and reinforcement learning depending on the nature of the task (Figure 1).

Within 2D material informatics, supervised learning has become the mainstay [13]. Using labelled datasets, models are trained to capture relationships between input descriptors and target properties, either through regression when the target is a continuous quantity or classification when it falls into distinct classes. In automated laboratory environments, methods such as support vector machines (SVMs) and tree-based algorithms are routinely employed to predict material properties and guide experiments. More recently, deep neural networks (DNNs), including large language models (LLMs), have been introduced to cope with increasingly complex and high-dimensional data structures [14,15].

When labels are not available, unsupervised learning offers an alternative route. Techniques like k-means clustering, principal component analysis (PCA), and t-distributed stochastic neighbour embedding (t-SNE) are used to reduce dimensionality and uncover patterns in raw data [16,17]. These methods help reveal underlying structure, group similar samples, and make high-dimensional material spaces more accessible and interpretable. Reinforcement learning provides yet another perspective: here, a learning agent interacts with its environment and gradually improves its decisions by maximising a cumulative reward. Algorithms such as Q-learning and deep deterministic policy gradients (DDPG) are particularly suited to developing strategies for long-term experimental planning and optimisation [18–20]. Together, these approaches, summarised schematically in Figure 1, form a versatile toolkit for exploring, designing, and optimising 2D materials.

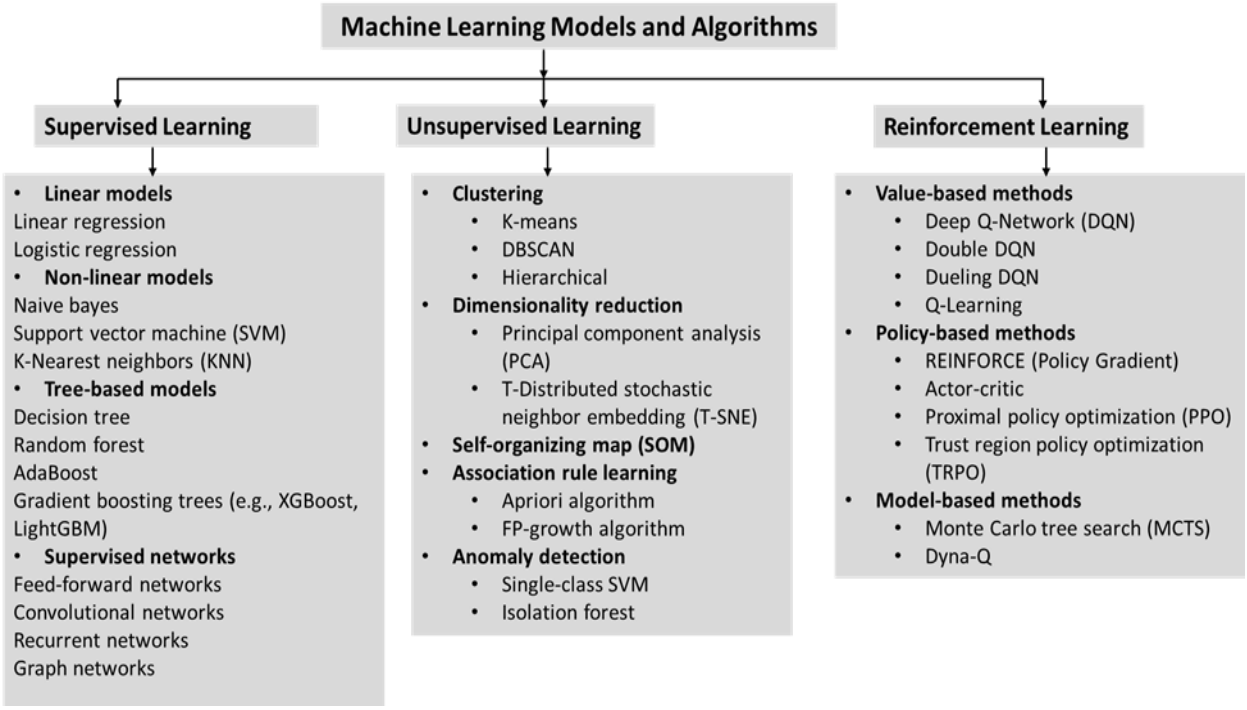

Figure 1. Various Machine Learning Models and Algorithms

III. TRAINING, VALIDATION, AND TESTING

Any researcher who uses machine learning knows that training, validation, and testing are not separate procedures; they are three parts of a whole that affect how dependable a model will be in the end [21]. What makes a concept that appears good on paper work in the real world is getting this pipeline right. The model is given a batch of labelled examples during training, and it steadily changes its internal weights to find the patterns in the data. The problem is that a model might get too excellent at fitting the training data, which means it remembers examples instead of learning the rules that would let it apply what it learnt to new circumstances.

Validation is what this is all about. Researchers can honestly judge how well the model is doing as they change its settings like how complicated it should be, how much regularisation to use, or what learning rate to employ by setting aside a distinct set of data that the model never trains on. If you don't do this step, you won't be able to tell if the model is really learning or just overfitting.

The model's final test is the test set, which it takes once it has been fully trained and all the settings are established. This dataset has examples that the model has never seen before, thus it gives the most accurate picture of how it will work in the real world. Training educates the model, validation picks the best one, and testing tells you if it's ready for the real world.

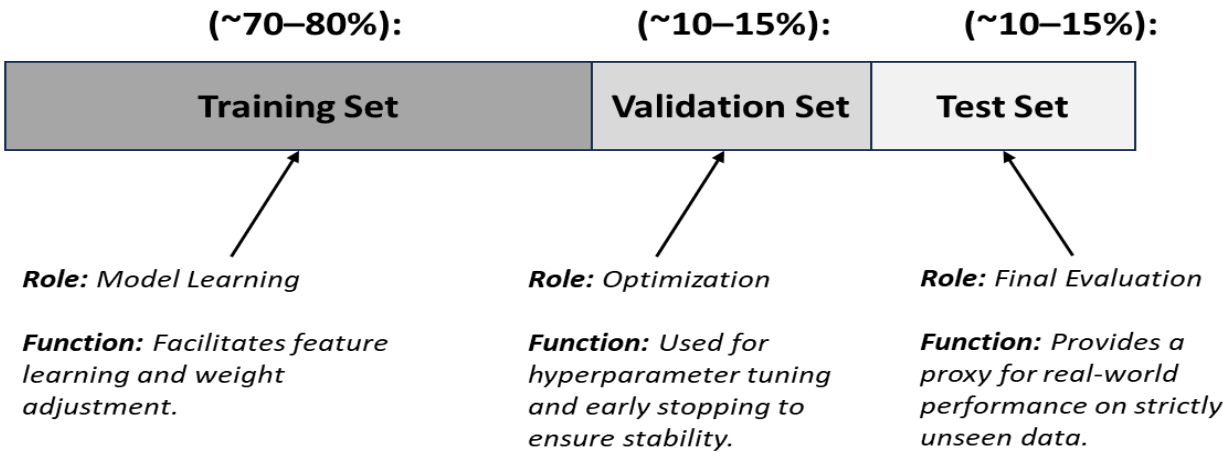

Figure 2. Training, validation, and test set

We can make a good machine learning model in one go. It happens in three stages: training, validation, and testing. These stages are all very connected, and each one affects the nextTraining is where it all begins. The model is given a set of labelled examples and starts to change its internal weights so that it can find patterns in the data. The model reads through examples, makes mistakes, corrects itself, and slowly gets better at finding the important features. It's a bit like learning from a textbook. But there's a problem. If the model pays too much attention to the training examples, it might just memorise them instead of learning general rules. When that happens, it has a hard time with anything new. That is where validation comes in. Researchers can see how well the model is really learning by keeping back a separate piece of data that the model never trains on. You can change settings like how deep the network should be or how aggressively to regularise at this stage. You can also catch overfitting before it ruins the model. If you skip this step, it's like giving a student the same questions they practiced on; the results don't tell you much about how well they understand.

The test is the last step and the final decision. After all the tuning is done and the model is set in stone, it has to deal with data it has never seen before. This new test shows how the model will work in the real world. In short, training gives the model its knowledge, validation checks that knowledge, and testing shows whether the model is really ready to be used in real life.

IV. RECENT PROGRESS IN ML

Recent improvements in machine learning have changed materials science from trial-and-error to data-driven discovery. This lets scientists quickly search through huge chemical spaces for new candidates. Machine learning and physics-informed models can now accurately predict how different structures affect properties and find the best synthesis parameters, which saves a lot of time and money on experiments. Combining generative algorithms with high-throughput experimentation is making the goal of inverse design a reality. This lets researchers design materials with specific, pre-defined functions on their own. Because of ML's recent amazing progress in material synthesis, it is now faster and easier to find new materials. Merchant et al. [22] introduced GNoME, an advanced AI tool that predicts stable material structures, to speed up the synthesis and discovery of materials. Szymanski et al. [23] built on GNoME's features by using its predictions in a separate lab called A-Lab. This innovative approach led to the successful synthesis of 41 new materials in just 17 days. This shows that machine learning can help material science by automating synthesis and discovery.

V. ML FOR 2D MATERIAL GROWTH PROCESSES

Through high-throughput computationally exfoliating of empirically recognised substances, a large number of 2D layered materials have been found as building blocks of 2D-material–based nanodevices [24]. Furthermore, density functional theory predicts a large

number of new 2D materials. ML-based structure–function relations [25], hopping methods [26,27], evolutionary algorithms [28-30], simulated annealing [31,32], metaheuristic–ML hybrid methods [33], crystal structure classification [34], generative models[35], and chemical exfoliation screening based on large-scale computation [36,37] are just a few of the varied methods for material prediction that have recently surfaced with the rise of AI.

Recent progress in making 2D materials has led to the creation of self-driving platforms that use active learning to get around the limits of traditional chemical vapour deposition scalability. These platforms can quickly navigate complex parameter spaces to stabilise hard-to-find metastable phases. This evolution is even stronger thanks to the combination of deep learning with in-situ characterisation. This makes real-time feedback loops possible, which help with precise defect engineering and grain boundary control in wafer-scale van der Waals heterostructures. As a result, the creation of a closed-loop synergy between computer-based prediction and robotic experimentation is making it possible to quickly reverse-engineer functional materials, which greatly shortens the time between theoretical discovery and device use.

Table 1 Research advances in 2D material synthesis techniques combined with experimental methods and ML.

Synthesis Technology	Target	ML Algorithms	Features	Year	Ref
CVD	Controlled growth of 2D materials	XGBoost	parameters for the experimental growth ofWTe2	2021	[38]
flash Jouleheating	Predicted graphene yield	XGBoost	flash Joule heating reaction parameters	2022	[39]
CVD	Projected growth area	Gaussian processregression	h-BN growth parameters	2023	[40]
laser Induction	Parameter optimization	Bayesian optimization	laser power, scanning speed, reactionchamber pressure and gas type	2024	[41]
laser Induction	Predict conductivity,morphology, and sheet resistance	Gaussian processregression	laser speed, power, duty cycle	2024	[42]

VI. RESEARCH ADVANCES IN 2D DEVICE DESIGN USING ML

Researchers now look at microscopy and spectroscopy data in a different way thanks to machine learning in the characterisation lab. Things that used to take hours of manual interpretation can now be done in seconds. Deep learning models that have been trained on methods like Raman and photoluminescence spectroscopy can almost instantly show layer thickness, strain distributions, and defect densities. This combination of computer power and experimental tools makes it possible to do high-throughput quality control and find small structural details that the human eye would probably miss.

Machine learning is changing how 2D electronics are made on the device side. The old way of doing things running experiment after experiment, changing parameters one at a time, and hoping for the best is being replaced by something much better. Researchers can now use algorithms to predict important performance numbers like contact resistance, bandgap, and carrier mobility. This lets them optimise heterostructure layouts before making a single device. Generative models and Bayesian optimisation are changing the design process in ways that go beyond just making predictions. Researchers can now ask, "What structure will give me these properties?" instead of "What properties will this structure have?" and get an answer. This inverse design strategy is very helpful for speeding up the creation of new logic circuits and optoelectronic devices.

Table 2 Research advances in 2D material characterization techniques combined with experimental methods and ML.

CharacterizationTechnique	Target	ML Algorithms	Year	Ref
optical microscope	identification of monolayer MoS2and 3D characterization	SVM, KNN and randomforest	2021	[43]
PL spectroscopy	identify perturbation components	PCA and k-meansclustering	2021	[44]
Raman spectroscopy	determine the range of torsionangles	random forest, SVM	2022	[45]
AFM	predict crystal coverage	CNN	2024	[46]
AFM	identify films at differenttemperatures	CNN	2024	[47]
optical microscope	identify thickness and predictingtwist angle	semantic segmentationCNN and ResNet	2024	[48]
optical microscope	detection, classification 2Dmaterials	Gaussian mixturemodel	2024	[49]

VII. CONCLUSION

In a nutshell, machine learning is causing a shift in the manufacturing process of two-dimensional material devices by removing experimental barriers and empowering designers to engage in independent efforts. Because it makes the entire process, from synthesis to fabrication, more efficient, it reduces costs and speeds up the innovation process. It is a well-known fact that the region continues to face considerable challenges, including a lack of sufficient data, standards that are not always consistent, and the inability to duplicate outcomes. The path forward, however, is crystal clear: the amalgamation of these computer techniques with hands-on testing is the most effective method for accelerating the discovery and enhancement of electronics of the future generation.

REFERENCES

[1] M.M. Waldrop, Nature 530 (2016) 144–148.

[2] R.S. Jacobsen, K.N. Andersen, P.I. Borel, J. Fage-Pedersen, L.H. Frandsen, O. Hansen, M. Kristensen, A.V. Lavrinenko, G. Moulin, H. Ou, C. Peucheret, B. Zsigri, A. Bjarklev, Nature 441 (2006) 199–202.

[3] G.P. Lansbergen, R. Rahman, C.J. Wellard, I. Woo, J. Caro, N. Collaert, S. Biesemans, G. Klimeck, L.C.L. Hollenberg, S. Rogge, Nat. Phys. 4 (2008) 656–661.

[4] B. Radisavljevic, M.B. Whitwick, A. Kis, ACS Nano 5 (2011) 9934–9938.

[5] H. Li, Z. Yin, Q. He, H. Li, X. Huang, G. Lu, D.W.H. Fam, A.I.Y. Tok, Q. Zhang, H. Zhang, Small 8 (2012) 63–67.

[6] H. Qiao, J. Yuan, Z. Xu, C. Chen, S. Lin, Y. Wang, J. Song, Y. Liu, Q. Khan, H.Y. Hoh, C.-X. Pan, S. Li, Q. Bao, ACS Nano 9 (2015) 1886–1894.

[7] N.R. Pradhan, D. Rhodes, Y. Xin, S. Memaran, L. Bhaskaran, M. Siddiq, S. Hill, P.M. Ajayan, L. Balicas, ACS Nano 8 (2014) 7923–7929.

[8] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, F. Wang, Nano Lett. 10 (2010) 1271–1275.

[9] L. Li, Y. Yu, G.J. Ye, Q. Ge, X. Ou, H. Wu, D. Feng, X.H. Chen, Y. Zhang, Nat. Nanotechnol. 9 (2014) 372–377.

[10] D. Rodan-Legrain, Y. Cao, J.M. Park, S.C. de la Barrera, M.T. Randeria, K. Watanabe, T. Taniguchi, P. Jarillo-Herrero, Nat. Nanotechnol. 16 (2021) 769–775.

[11] K. Roy, M. Padmanabhan, S. Goswami, T.P. Sai, G. Ramalingam, S. Raghavan, A. Ghosh, Nat. Nanotechnol. 8 (2013) 826–830.

[12] M.S. Choi, G.-H. Lee, Y.-J. Yu, D.-Y. Lee, S.H. Lee, P. Kim, J. Hone, W.J. Yoo, Nat. Commun. 4 (2013) 1624.

[13] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Npj Comput. Mater. 3 (2017) 1–13.

[14] A. Ghods, D.J. Cook, Data Min. Knowl. Discov. 35 (2021) 46–87.

[15] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu, Z. Wu, L. Zhao, D. Zhu, X. Li, N. Qiang, D. Shen, T. Liu, B. Ge, Meta-Radiol 1 (2023) 100017.

[16] G. Ivosev, L. Burton, R. Bonner, Anal. Chem. 80 (2008) 4933–4944.

[17] M. Zhong, K. Tran, Y. Min, C. Wang, Z. Wang, C.-T. Dinh, P. De Luna, Z. Yu, A.S. Rasouli, P. Brodersen, S. Sun, O. Voznyy, C.-S. Tan, M. Askerka, F. Che, M. Liu, A. Seifitokaldani, Y. Pang, S.-C. Lo, A. Ip, Z. Ulissi, E.H. Sargent, Nature 581 (2020) 178–183.

[18] H. Zhang, S. Li, Y. Zheng, Ind. Eng. Chem. Res. 59 (2020) 17987–17999.

[19] Y. Bao, Y. Zhu, F. Qian, Ind. Eng. Chem. Res. 60 (2021) 5504–5515.

[20] C. Tan, C. Wang, J. Tian, H. Niu, Q. Wei, X. Zhang, ACS Omega 8 (2023) 23739–23753.

[21] E. Lopez, J. Etxebarria-Elezgarai, J.M. Amigo, A. Seifert, Anal. Chim. Acta 1275 (2023) 341532.

[22] A. Merchant, S. Batzner, S.S. Schoenholz, M. Aykol, G. Cheon, E.D. Cubuk, Nature (2023) 1–6.

[23] N.J. Szymanski, B. Rendy, Y. Fei, R.E. Kumar, T. He, D. Milsted, M.J. McDermott, M. Gallant, E.D. Cubuk, A. Merchant, H. Kim, A. Jain, C.J. Bartel, K. Persson, Y. Zeng, G. Ceder, Nature (2023) 1–6.

[24] N. Mounet, M. Gibertini, P. Schwaller, D. Campi, A. Merkys, A. Marrazzo, T. Sohier, I.E. Castelli, A. Cepellotti, G. Pizzi, N. Marzari, Nat. Nanotechnol. 13 (2018) 246–252.

[25] Q. Zhou, S. Lu, Y. Wu, J. Wang, J. Phys. Chem. Lett. 11 (2020) 3920–3927.

[26] D.J. Wales, J.P.K. Doye, J. Phys. Chem. A 101 (1997) 5111–5116.

[27] S. Goedecker, J. Chem. Phys. 120 (2004) 9911–9917.

[28] A.R. Oganov, C.W. Glass, J. Chem. Phys. 124 (2006) 244704.

[29] G. Trimarchi, A. Zunger, Phys. Rev. B 75 (2007) 104113.

[30] B. Meredig, C. Wolverton, Nat. Mater. 12 (2013) 123–127.

[31] J. Pannetier, J. Bassas-Alsina, J. Rodriguez-Carvajal, V. Caignaert, Nature 346 (1990) 343–345.

[32] J.C. Schön, M. Jansen, Angew. Chem. Int. Ed. Engl. 35 (1996) 1286–1304.

[33] Q. Tong, L. Xue, J. Lv, Y. Wang, Y. Ma, Faraday Discuss. 211 (2018) 31–43.

[34] K. Ryan, J. Lengyel, M. Shatruk, J. Am. Chem. Soc. 140 (2018) 10158–10168.

[35] D.M. Anstine, O. Isayev, J. Am. Chem. Soc. 145 (2023) 8736–8750.

[36] J. Björk, J. Zhou, P.O.Å. Persson, J. Rosen, Science 383 (2024) 1210–1215.

[37] A. Thakur, B. Anasori, Science 383 (2024) 1182–1183.

[38] M. Xu, B. Tang, Y. Lu, C. Zhu, Q. Lu, C. Zhu, L. Zheng, J. Zhang, N. Han, W. Fang, Y. Guo, J. Di, P. Song, Y. He, L. Kang, Z. Zhang, W. Zhao, C. Guan, X. Wang, Z. Liu, J. Am. Chem. Soc. 143 (2021) 18103–18113.

[39] J.L. Beckham, K.M. Wyss, Y. Xie, E.A. McHugh, J.T. Li, P.A. Advincula, W. Chen, J. Lin, J.M. Tour, Adv. Mater. 34 (2022) 2106506.

[40] J.-H. Park, A.-Y. Lu, M.M. Tavakoli, N.Y. Kim, M.-H. Chiu, H. Liu, T. Zhang, Z. Wang, J. Wang, L.G.P. Martins, Z. Luo, M. Chi, J. Miao, J. Kong, Nano Lett. 23 (2023) 4741–4748.

[41] H. Wahab, J. Heil, A.S. Tyrrell, T. Muller, J. Ackerman, L. Kotthoff, P.A. Johnson, Ceram. Int. 50 (2024) 9114–9124.

[42] J.C. Santos-Ceballos, F. Salehnia, A. Romero, X. Vilanova, Sci. Rep. 14 (2024) 10363.

[43] Y. He, Y. Ju, Q. Wang, Appl. Surf. Sci. 565 (2021) 150530.

[44] P.V. Kolesnichenko, Q. Zhang, C. Zheng, M.S. Fuhrer, J.A. Davis, Mach. Learn. Sci. Technol. 2 (2021) 025021.

[45] P. Solís-Fernández, H. Ago, ACS Appl. Nano Mater. 5 (2022) 1356–1366.

[46] I.A. Moses, C. Wu, W.F. Reinhart, Mater. Today Adv. 22 (2024) 100483.

[47] I.A. Moses, W.F. Reinhart, Mater. Charact. 209 (2024) 113701.

[48] H. Yang, R. Hu, H. Wu, X. He, Y. Zhou, Y. Xue, K. He, W. Hu, H. Chen, M. Gong, X. Zhang, P.-H. Tan, E.R. Hernández, Y. Xie, Nano Lett. (2024).

[49] J.-L. Uslu, T. Ouaj, D. Tebbe, A. Nekrasov, J.H. Bertram, M. Schütte, K. Watanabe, T. Taniguchi, B. Beschoten, L. Waldecker, C. Stampfer, Mach. Learn. Sci. Technol. 5 (2024) 015027.