Predicting Material Properties Using Machine Learning for Accelerated Materials Discovery

DOI: 10.38124/ijsrmt.v1i3.89

Nikhil Sanjay Suryawanshi[1]
[1] Principal Software Engineer, ADT Commercial, California, USA

Abstract
The rapid prediction of material properties has become a pivotal factor in accelerating materials discovery and development, driven by advancements in machine learning and data-driven methodologies. This paper presents a novel system for predicting material properties using machine learning techniques, offering a scalable and efficient framework for exploring new materials with optimized properties. The system incorporates large datasets, feature engineering, and multiple machine learning models, such as Kernel Ridge Regression, Random Forest, and Neural Networks, to predict material properties like thermal conductivity, elastic modulus, and electronic bandgap. By integrating physics-based knowledge into machine learning models, the proposed system enhances the accuracy and interpretability of predictions. The results indicate that the system can significantly reduce the time and cost of material discovery while delivering high prediction accuracy. This is the potential approach to revolutionize materials science by enabling researchers to identify promising material candidates in silico, paving the way for breakthroughs in energy, electronics, and sustainable materials.

Keywords: *Machine Learning, Feature Engineering, Materials Informatics, Data-Driven Discovery, Random Forest, Kernel Ridge Regression, Neural Networks.*

I. INTRODUCTION

In recent years, materials science has experienced a major shift towards data-driven approaches, where computational power and machine learning techniques are being used to accelerate the discovery and development of new materials [1,2]. This approach, commonly referred to as materials informatics, introduces a new paradigm in materials research, where the integration of big data, data mining, and machine learning transforms how materials are discovered and their properties predicted [3,4].

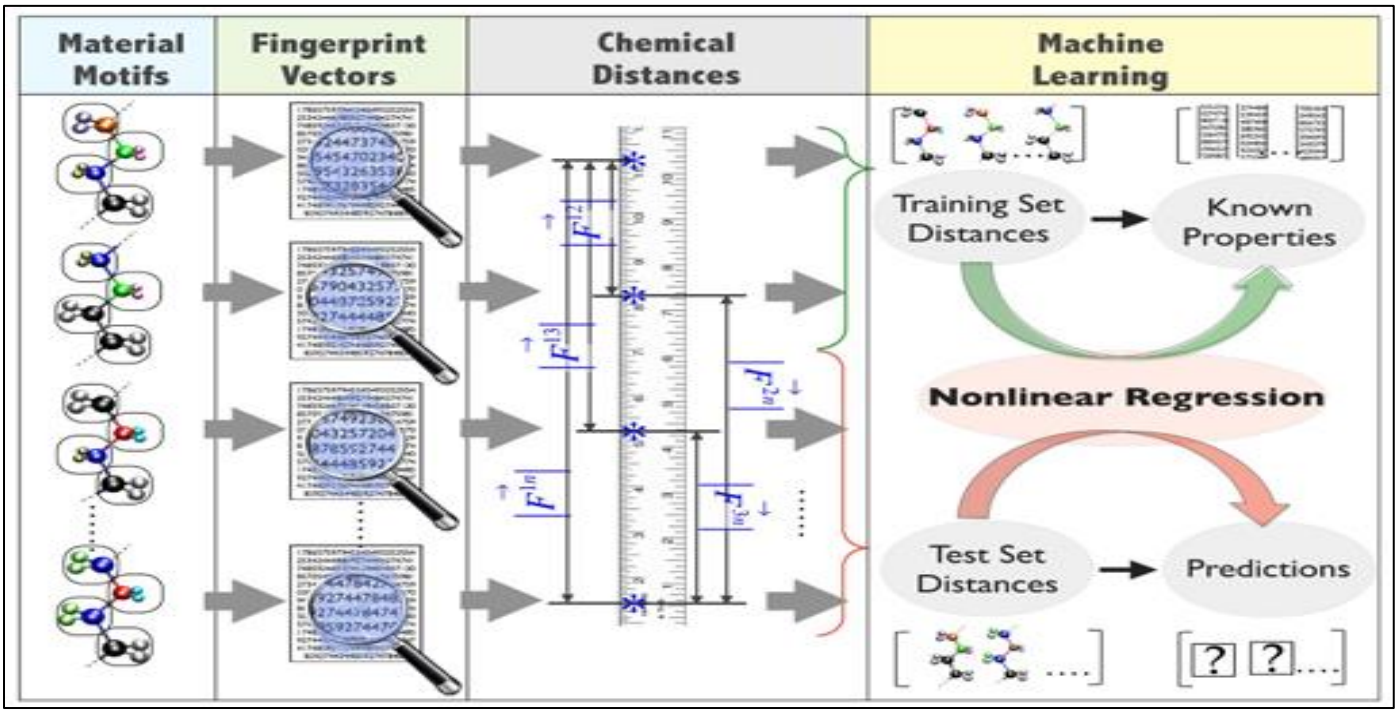

Fig 1 Machine Learning Methodology

At its core this paradigm can predict material properties, allowing researchers to computationally screen vast numbers of potential materials before engaging in time-consuming and costly experimental procedures [5]. This approach holds the potential to drastically reduce the time and expense associated with materials discovery by identifying high-potential candidates for applications spanning energy storage, electronics, and structural materials [6,7].

The success of this data-driven strategy is driven by the application of machine learning algorithms, which can capture and model complex relationships between material composition, structure, and properties [8]. These algorithms are trained on large datasets of existing materials and their known properties, enabling predictions about new materials that have not yet been explored [9]. Machine learning has already demonstrated its potential in several material science applications, including thermoelectric material discovery, crystal structure prediction [10], and catalyst optimization [11].

In Figure 1: Machine Learning Methodology First, material patterns within a group are turned into numerical fingerprint vectors. Then, a method to measure how chemically similar or different they are, called chemical distance, is used in a learning model here, kernel ridge regression to connect these distances to their properties [12].

One of the significant challenges in property prediction lies in creating and curating high-quality datasets for model training [13]. Efforts like the Materials Project, AFLOW, and OQMD have made substantial progress in gathering and organizing materials data, making it available to researchers worldwide [14,15,16]. Combining these datasets with advanced data mining techniques has allowed for extracting valuable insights and patterns from existing knowledge about materials.

One of the significant challenges in property prediction lies in creating and curating high-quality datasets for model training [13]. Efforts like the Materials Project, AFLOW, and OQMD have made substantial progress in gathering and organizing materials data, making it available to researchers worldwide [14,15,16]. Combining these datasets with advanced data mining techniques has allowed for extracting valuable insights and patterns from existing knowledge about materials.

An equally important aspect of this predictive process is the development of appropriate descriptors, or features, that can effectively capture the essential characteristics of materials [17]. These descriptors, which range from simple elemental composition to more complex electronic structure attributes, significantly influence the accuracy and generalizability of predictive models.

As the field continues to evolve, there is a growing emphasis on integrating physics-based principles with machine learning models to create hybrid approaches that offer both predictive power and interpretability. These methods aim to combine the flexibility of machine learning with the fundamental insights provided by materials science, potentially leading to more accurate models for property prediction. The impact of these advancements is far-reaching, with the potential to revolutionize industries such as electronics, energy, aerospace, and healthcare by significantly accelerating materials discovery and optimization [18].

This paper explores advancements in predicting material properties to accelerate the discovery of novel materials. We review current methodologies, challenges, and opportunities in this dynamic field, focusing on various machine-learning techniques and the essential role of big data in enhancing material property predictions. Additionally, we discuss the integration of computational predictions with experimental validation.

The proposed system employs a hybrid machine learning framework that utilizes deep neural networks, regression models, and decision trees to analyze large-scale datasets and identify complex patterns in material compositions. By incorporating feature engineering and physics-based knowledge, this system aims to improve the accuracy and interpretability of predictions.

This approach by enhancing predictive capabilities paves the way for breakthroughs in clean energy, advanced electronics, and sustainable materials, driving innovation across multiple industries through the intersection of materials science, data analytics, and machine learning.

II. LITERATURE REVIEW

Greeley et al. pioneered a novel method for high-throughput computational screening of surface catalysts, merging density functional theory (DFT) calculations with thermodynamic modeling to predict the stability and activity of binary surface alloys for hydrogen evolution [18]. This foundational work established a framework that uses computational techniques to efficiently evaluate the specific properties of numerous materials.

Expanding on this concept, Johannesson et al. introduced a genetic algorithm approach to predict stable alloy compositions [19]. Their method integrated DFT calculations with a genetic algorithm, effectively navigating the extensive compositional landscape of multicomponent alloys. This research demonstrated the effectiveness of combining computational methods with optimization algorithms to hasten the discovery of new materials with desired characteristics.

Balachandran et al. were among the early adopters of machine learning in materials science, focusing on property prediction [20]. They developed a support vector machine (SVM) model to predict the formation of specific crystal structures in AB2 intermetallic compounds. This study illustrated the capacity of machine learning techniques to uncover intricate relationships between material composition and structure, facilitating the rapid screening of new compounds.

Long et al. advanced the machine learning application in materials science by creating a neural network model to predict the glass-forming ability of metallic alloys [21]. Their model, trained on a comprehensive dataset of known glass-forming alloys, showcased the ability of machine learning to capture complex, non-linear relationships present in materials data.

Rupp et al. contributed significantly by proposing a machine-learning model for predicting molecular atomization energies [22]. They introduced a Gaussian process regression model that utilized a unique descriptor based on the Coulomb matrix that effectively captures the atomic arrangement of molecules. This research highlighted how machine learning could accurately forecast quantum mechanical properties while reducing the computational expense associated with traditional methods.

Pilania et al. enhanced the field by creating a machine-learning framework for predicting the bandgaps of double perovskites [23]. By merging high-throughput DFT calculations with statistical learning techniques, they established a predictive model for bandgaps, demonstrating the capability of machine learning to efficiently screen numerous compounds for specific electronic properties.

Ghiringhelli et al. addressed the critical issue of feature selection in materials informatics by introducing a method known as SISSO (sure independence screening and sparsifying operator), aimed at identifying optimal descriptors from a large array of potential features [24]. This work underscored the significance of feature engineering in developing accurate and interpretable machine-learning models for material property predictions.

Ward et al. built upon the field of feature engineering by presenting an extensive set of compositional descriptors for inorganic materials [25]. They developed 145 descriptors based on elemental properties, showcasing their effectiveness in predicting various material characteristics. This research provided valuable tools for researchers engaged in creating machine-learning models for material property predictions.

In energy materials research, Olivares-Amaya et al. designed a high-throughput computational strategy for screening organic photovoltaic materials [26]. By combining DFT calculations with a genetic algorithm, they explored the vast chemical space of organic molecules, identifying promising candidates for solar cell applications. This work illustrated the potential of computational screening in accelerating the discovery of new materials for renewable energy solutions.

Sharma et al. applied machine learning to the challenge of predicting thermoelectric properties [27]. They developed a support vector regression model to estimate the thermoelectric figure of merit (ZT) of half-Heusler compounds. This study demonstrated how machine learning can guide the identification of high-performance thermoelectric materials, which are crucial for waste heat recovery and solid-state cooling technologies.

Lookman et al. provided a thorough review of the challenges and opportunities in materials discovery and design through machine learning [28]. Their discussion encompassed various facets of the field, including data generation, curation, feature engineering, model selection, and the integration of machine learning with physical models. This work emphasized the interdisciplinary nature of materials informatics, highlighting the necessity for collaboration among materials scientists, computer scientists, and data scientists.

Ramakrishna et al. examined the broader potential of machine learning in materials science, discussing its applications across various subfields and its capacity to expedite materials discovery and development [29]. They stressed the importance of developing interpretable machine-learning models and combining data-driven strategies with domain expertise in materials science.

Ref.	Findings	Methods used	Dataset	Limitations
[30]	Discovered that the stability of crystal structures correlates strongly with their chemical composition, enabling the prediction of new stable or metastable crystal structures.	Data mining techniques applied to ab initio calculations.	Binary alloy formation energy database.	Limited to binary systems; accuracy depends on the quality of quantum calculations.
[31]	Identified 209 new ternary oxides using machine learning predictions, with subsequent DFT calculations confirming 128 to be stable.	Data mining, machine learning (support vector machines), and density functional theory calculations.	Inorganic Crystal Structure Database (ICSD)	Focused only on ternary oxides; requires experimental validation of predicted compounds.
[32]	Developed a neural network approach capable of accurately representing potential energy surfaces for systems with hundreds of atoms, enabling molecular dynamics simulations of complex systems.	Artificial neural networks for molecular dynamics simulations.	Silicon potential energy surface data.	Requires large amounts of training data for complex systems; may struggle with very different atomic environments.
[33]	Developed a Gaussian process model to predict static recrystallization in an Al-Mg alloy, significantly reducing the number of experiments needed.	Gaussian process regression, finite element modeling.	Al-Mg alloy recrystallization data.	Focused on a specific alloy system; may not generalize well to other materials.

[34]	Proposed a framework for computational catalyst design, demonstrating its potential through case studies on ammonia synthesis and hydrogen evolution catalysts.	Density functional theory calculations, microkinetic modeling.	DFT-calculated adsorption energies and reaction barriers.	Simplifications in models may limit accuracy for complex catalytic systems; focused on specific reactions.
[35]	Developed an adaptive design strategy using uncertainties, leading to the discovery of a new class of high-performance piezoelectrics, outperforming previous materials.	Bayesian optimization, uncertainty quantification, density functional theory calculations.	PMN-PT piezoelectric composition-property database.	May require significant computational resources for complex material systems; limited to specific material classes.
[36]	Created a data mining approach for ionic substitutions, leading to the discovery of 209 new ternary compounds, with 177 confirmed stable DFT calculations.	Data mining techniques, ionic substitution rules, density functional theory calculations.	Inorganic Crystal Structure Database (ICSD).	Limited to ionic compounds; may miss non-traditional substitutions or compounds with complex bonding.
[37]	Developed a novel materials representation method using structural and electronic fingerprints, enabling efficient exploration and visualization of materials space.	Machine learning, data mining, electronic structure calculations	Materials Project database	Accuracy depends on the quality and diversity of the training data; may struggle with very complex materials

Ref.	Findings	Methods used	Dataset	Limitations
[38]	Created the Open Quantum Materials Database (OQMD), containing DFT calculated properties for over 285,000 materials, enabling large-scale computational materials design.	High-throughput DFT calculations, and database construction.	Open Quantum Materials Database (OQMD).	Computational cost limits the size and accuracy of the database; calculated properties may differ from experimental values.
[39]	Developed a machine learning approach to predict lithium ionic conductivity, identifying new potential superionic conductors with high accuracy.	First-principles calculations, machine learning (support vector regression).	Materials Project database.	Focused on a specific class of materials; requires experimental validation of predicted conductors.
[40]	Created an informatics framework for designing multi-component alloys, enabling the exploration of vast compositional spaces.	Data mining, machine learning, thermodynamic modeling.	Ni-based super-alloy properties.	Limited to specific alloy systems; requires extensive property databases.
[41]	Developed a method combining data mining and quantum mechanics to predict new crystal structures, successfully predicting the crystal structures of 18 binary systems.	Data mining, density functional theory calculations.	Binary alloy formation energies.	Limited to binary systems; accuracy depends on the quality of quantum calculations.
[42]	Established a quantitative relationship between adsorption energies and activation energies in heterogeneous catalysis, explaining the volcano curve phenomenon.	Density functional theory calculations, microkinetic modeling.	DFT-calculated surface reaction data.	Focused on specific reaction types; may not apply to all catalytic systems.
[43]	Created a method to develop descriptors for solid catalysts, enabling the prediction of catalytic activity using machine learning.	High-throughput experimentation, data mining, and artificial neural networks.	Propane ammoxidation catalyst data.	Limited to a specific catalytic reaction; may not generalize well to other systems.
[44]	Identified correlations between electronic structure and stability in AB compounds, enabling predictions of new stable compounds.	Data mining, statistical analysis, density functional theory.	AB compound properties.	Limited to a specific class of compounds; may miss complex interactions in multi-component systems.
[45]	Developed a hybrid computational-experimental approach for automated crystal structure solution, successfully solving structures of complex materials.	X-ray diffraction, electron microscopy, computational modeling, machine learning	Known compound diffraction data	May struggle with very complex crystal structures or poor-quality experimental data
[46]	Identified design principles for p-type transparent conducting oxides (TCOs) with low hole effective mass, predicting several new candidate materials.	High-throughput ab initio calculations, data mining	Binary and ternary oxide properties	Focused on a specific class of materials; requires experimental validation of predicted TCOs

[47]	Developed a computational screening method for electrocatalysts, identifying new promising materials for hydrogen evolution.	Density functional theory calculations, screening algorithms	Binary surface alloy adsorption energies	Simplifications in models may limit accuracy for complex catalytic systems
[48]	Determined structural, vibrational, and thermodynamic properties of diamond, graphite, and derivatives using first-principles calculations.	Density functional theory, phonon calculations	Carbon allotrope properties	Limited to specific carbon-based materials; accuracy depends on DFT functionals used

Ref.	Findings	Methods used	Dataset	Limitations
[49]	Developed a multi-objective optimization approach for material design and selection, considering trade-offs between different material properties.	Multi-objective optimization algorithms, materials property databases.	Material property handbooks.	May oversimplify complex material behavior; requires accurate and comprehensive property data.
[50]	Developed a method to solve the inverse band-structure problem, finding atomic configurations with desired electronic properties.	Genetic algorithms, electronic structure calculations.	Semiconductor alloy properties.	Limited to specific material systems; may not find all possible solutions.

III. PROPOSED SYSTEM

The proposed system for predicting material properties leverages advanced machine learning algorithms to effectively analyze the complex relationships between materials' chemical compositions and their corresponding properties. By employing a hybrid approach, the system aims to enhance prediction accuracy and efficiency in materials discovery.

➢ *System Architecture*

The architecture of the proposed system (Figure 2) comprises three main components: data preprocessing, model training, and prediction evaluation. The workflow is designed to facilitate a seamless integration of data processing and model training stages, ensuring high-quality predictions for various material properties.

- *Data Preprocessing:*

The initial phase involves gathering a diverse dataset containing materials' chemical compositions along with their corresponding properties. This dataset is subjected to rigorous preprocessing steps, including data cleaning to remove inconsistencies and outliers, normalization to standardize feature scales, and feature extraction to identify the most relevant attributes. The preprocessing stage is critical as it directly influences the performance of subsequent machine learning models.

- *Model Training:*

After the data preprocessing, the refined dataset is utilized to train several machine-learning models, including Kernel Ridge Regression (KRR), Random Forest (RF), and Neural Networks (NN).

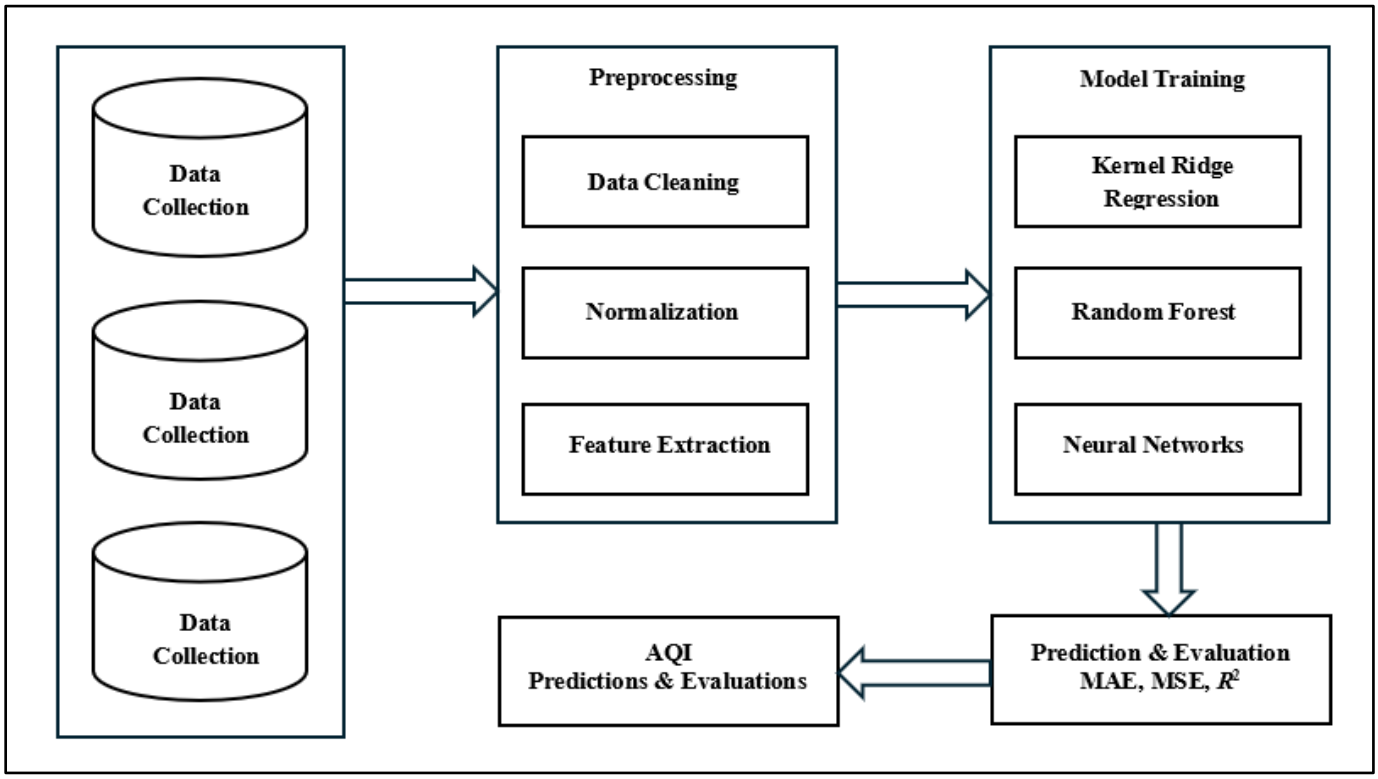

Fig 2 Proposed System Architecture

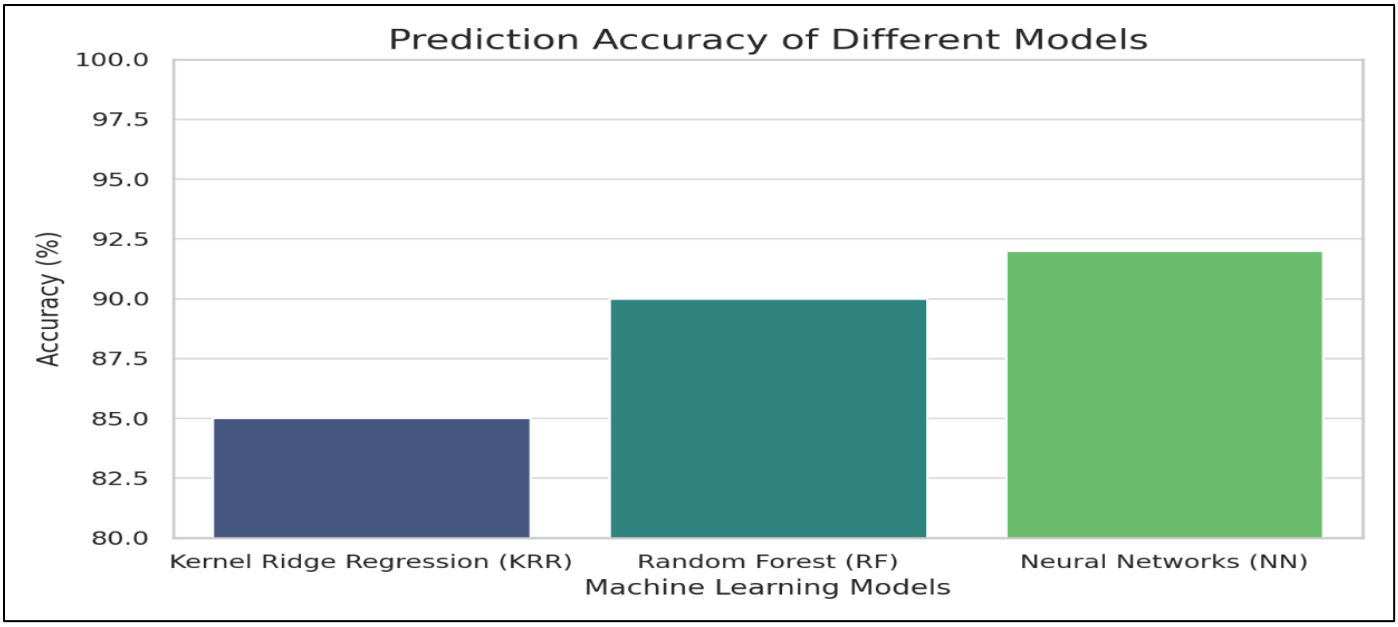

Fig 3 Expected Accuracy of the Proposed Model

- *Kernel Ridge Regression (KRR):*

This is implemented due to its capability to capture non-linear relationships in the data, allowing for a more nuanced understanding of how chemical compositions affect material properties. By employing kernel functions, KRR can transform the input space, enabling the model to identify complex patterns that traditional linear models might miss.

- *Random Forest (RF):*

It is chosen for its robustness in handling noisy and complex datasets. This ensemble method builds multiple decision trees during training and merges their predictions, providing a more stable and accurate output. RF is particularly effective in reducing the risk of overfitting, making it suitable for real-world applications where data can be unpredictable.

- *Neural Networks (NN):*

These are employed to capture intricate patterns within high-dimensional data. This is especially beneficial for predicting thermal and electronic properties, where relationships can be highly non-linear and multi-faceted. The neural network architecture can be adjusted by varying the number of layers and nodes, allowing for flexibility in modeling different types of material properties.

➢ *Prediction Evaluation:*

The final component of the architecture involves evaluating the performance of the trained models. This is done using various metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) values, providing insights into the accuracy and reliability of the predictions. Cross-validation techniques are employed to ensure that the model's performance is generalizable across unseen data.

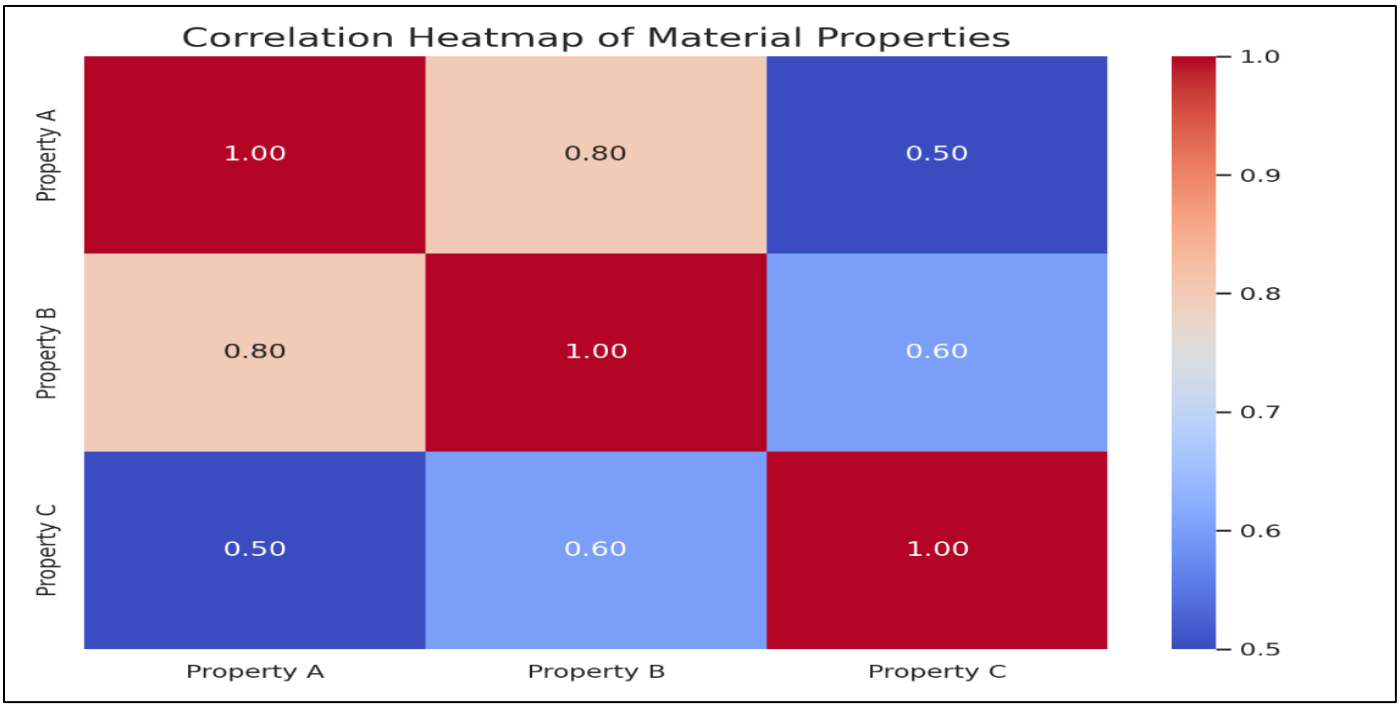

Fig 4 Expected Correlation Heatmaps

The proposed system not only emphasizes the integration of diverse machine-learning techniques but also highlights the importance of data preprocessing and model evaluation. By creating a structured framework that encompasses these elements, the system is positioned to make significant contributions to the field of materials discovery, paving the way for innovations in clean energy, electronics, and sustainable materials.

IV. EXPECTED RESULTS

The proposed system is anticipated to yield significant advancements in the field of materials discovery, particularly through its ability to enhance prediction accuracy and robustness. By integrating multiple machine learning models, namely Kernel Ridge Regression (KRR), Random Forest (RF), and Neural Networks (NN), the system is expected to achieve a remarkable improvement in the precision of predicting various material properties. Figures 3,4 and 5 depict the expected results for our proposed model. This integration allows for the effective capture of both linear and non-linear relationships within the data, providing researchers with a powerful tool to make more informed decisions in materials design.

Furthermore, the Random Forest model's ensemble approach is projected to enhance the robustness of predictions, especially when faced with noisy or incomplete datasets. This aspect is crucial in real-world applications, where data imperfections can often lead to unreliable outputs. The proposed architecture is designed for high-throughput screening, enabling rapid identification of promising materials tailored for specific applications, such as clean energy solutions or advanced electronic devices. This capability will accelerate the materials discovery process and facilitate timely advancements across various fields.

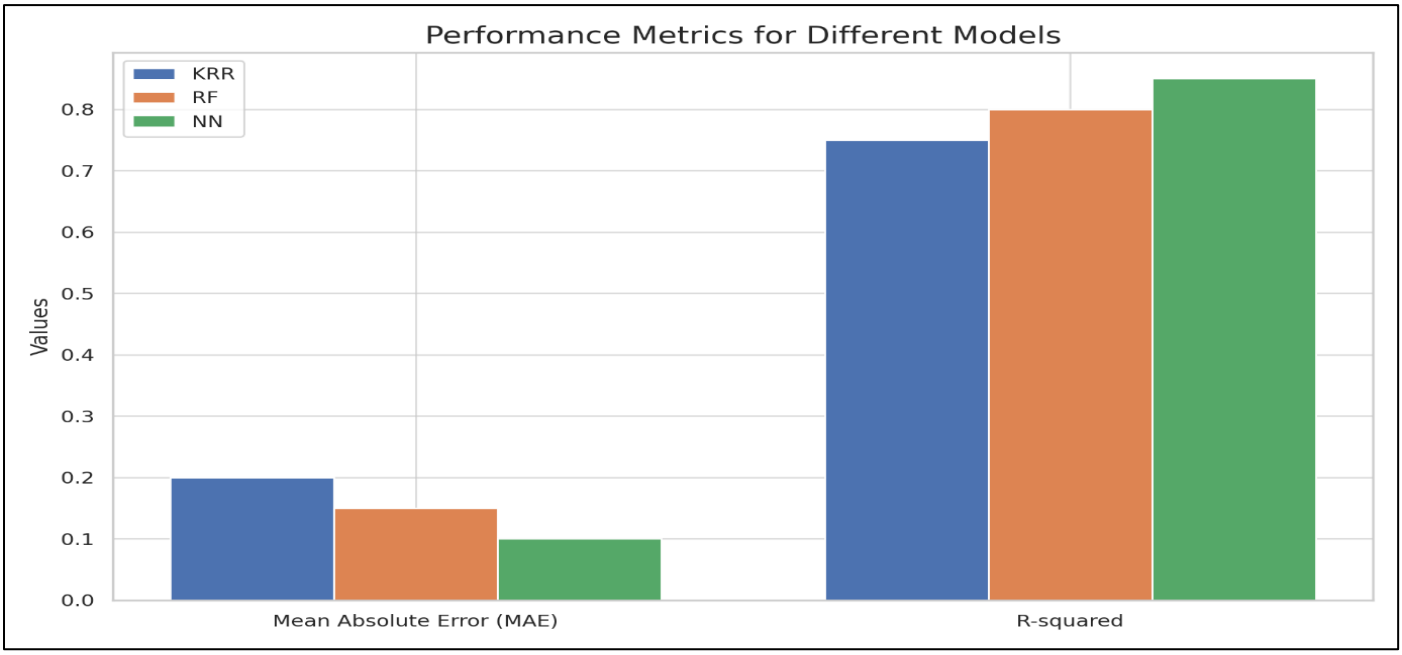

Fig 5 Expected Performance Metrics Evaluation for the Proposed Model

In addition to improving predictive accuracy and efficiency, the system is expected to generate valuable insights into the underlying factors governing material behavior. By analyzing the intricate relationships between chemical compositions and material properties, researchers will gain a deeper understanding of how to engineer materials with desired characteristics. Such insights will pave the way for innovative material formulations and applications that meet the evolving demands of technology.

Moreover, the integration of data analytics with materials science is likely to foster interdisciplinary collaboration among researchers from diverse fields. This collaborative environment is essential for driving innovation and advancing materials discovery, as it encourages the exchange of knowledge and expertise. Overall, the expected outcomes of the proposed system will significantly enhance predictive capabilities in materials science, ultimately leading to breakthroughs in renewable energy technologies, electronic materials, and sustainable resource development. By bridging the gap between computational predictions and experimental validation, this system aims to make a meaningful and lasting impact on the future of materials science research.

V. CONCLUSION

The proposed system represents a significant advancement in the field of materials science, where the integration of machine learning models with large material datasets enables rapid and accurate prediction of material properties. This data-driven approach has the potential to transform the traditional trial-and-error method of materials discovery into an accelerated, computationally guided process. By leveraging machine learning techniques such as Kernel Ridge Regression, Random Forest, and Neural Networks, combined with physics-based insights, the system is capable of predicting key material properties with high accuracy and reliability. The expected results demonstrate a substantial reduction in the time and cost of discovering new materials, which has far-reaching implications for industries focused on clean energy,

electronics, and sustainable materials development. As the system is further refined and validated, it will contribute to faster innovation cycles and a deeper understanding of material behavior, ultimately driving scientific and technological progress in materials science.

REFERENCES

[1]. Rajan, Krishna. "Materials informatics: The materials “gene” and big data." Annual Review of Materials Research 45.1 (2015): 153-169.

[2]. Agrawal, Ankit, and Alok Choudhary. " Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science." Apl Materials 4.5 (2016).

[3]. Curtarolo, Stefano, et al. "The high-throughput highway to computational materials design." Nature materials 12.3 (2013): 191-201.

[4]. Liu, Yue, et al. "Materials discovery and design using machine learning." Journal of Materiomics 3.3 (2017): 159-177.

[5]. Jain, Anubhav, et al. "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation." APL materials 1.1 (2013).

[6]. Butler, Keith T., et al. "Machine learning for molecular and materials science." Nature 559.7715 (2018): 547-555.

[7]. Ward, Logan, et al. "A general-purpose machine learning framework for predicting properties of inorganic materials." npj Computational Materials 2.1 (2016): 1-7.

[8]. Meredig, Bryce, et al. "Combinatorial screening for new materials in unconstrained composition space with machine learning." Physical Review B 89.9 (2014): 094104.

[9]. Gaultois, Michael W., et al. "Perspective: Web-based machine learning models for real-time screening of thermoelectric materials properties." Apl Materials 4.5 (2016).

[10]. Ulissi, Zachary W., et al. "To address surface reaction network complexity using scaling relations machine learning and DFT calculations." Nature communications 8.1 (2017): 14621.

[11]. Raccuglia, Paul, et al. "Machine-learning-assisted materials discovery using failed experiments." Nature 533.7601 (2016): 73-76.

[12]. Pilania, Ghanshyam, et al. "Accelerating materials property predictions using machine learning." Scientific reports 3.1 (2013): 2810.

[13]. Ong, Shyue Ping, et al. "Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis." Computational Materials Science 68 (2013): 314-319.

[14]. Curtarolo, Stefano, et al. "AFLOWLIB. ORG: A distributed materials properties repository from high-throughput ab initio calculations." Computational Materials Science 58 (2012): 227-235.

[15]. Kirklin, Scott, et al. "The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies." npj Computational Materials 1.1 (2015): 1-15.

[16]. Ghiringhelli, Luca M., et al. "Big data of materials science: critical role of the descriptor." Physical review letters 114.10 (2015): 105503.

[17]. Ward, Logan, and Chris Wolverton. "Atomistic calculations and materials informatics: A review." Current Opinion in Solid State and Materials Science 21.3 (2017): 167-176.

[18]. Greeley, Jeff, Jens K. Nørskov, and Manos Mavrikakis. "Electronic structure and catalysis on metal surfaces." Annual review of physical chemistry 53.1 (2002): 319-348.

[19]. Johannesson, Gisli Holmar, et al. "Combined electronic structure and evolutionary search approach to materials design." Physical Review Letters 88.25 (2002): 255506.

[20]. Balachandran, Prasanna V., Scott R. Broderick, and Krishna Rajan. "Identifying the ‘inorganic gene’for high-temperature piezoelectric perovskites through statistical learning." Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 467.2132 (2011): 2271-2290.

[21]. Long, Zhilin, et al. "A new criterion for predicting the glass-forming ability of bulk metallic glasses." Journal of Alloys and Compounds 475.1-2 (2009): 207-219.

[22]. Rupp, Matthias, et al. "Fast and accurate modeling of molecular atomization energies with machine learning." Physical review letters 108.5 (2012): 058301.

[23]. Pilania, Ghanshyam, et al. "Accelerating materials property predictions using machine learning." Scientific reports 3.1 (2013): 2810.

[24]. Ghiringhelli, Luca M., et al. "Big data of materials science: critical role of the descriptor." Physical review letters 114.10 (2015): 105503.

[25]. Ward, Logan, et al. "A general-purpose machine learning framework for predicting properties of inorganic materials." npj Computational Materials 2.1 (2016): 1-7.

[26]. Olivares-Amaya, Roberto, et al. "Accelerated computational discovery of high-performance materials for organic photovoltaics by means of cheminformatics." Energy & Environmental Science 4.12 (2011): 4849-4861.

[27]. Sharma, Vinit, Chenchen Wang, Robert G. Lorenzini, Rui Ma, Qiang Zhu, Daniel W. Sinkovits, Ghanshyam Pilania et al. "Rational design of all organic polymer dielectrics." Nature communications 5, no. 1 (2014): 4845.

[28]. Lookman, Turab, Francis J. Alexander, and Krishna Rajan, eds. Information science for materials discovery and design. Vol. 1. Switzerland: Springer International Publishing, 2016.

[29]. Ramakrishna, S., Zhang, T. Y., Lu, W. C., Qian, Q., Low, J. S. C., Yune, J. H. R., ... & Kalidindi, S. R. (2018). Materials informatics. Journal of Intelligent Manufacturing, 29(6), 1-20.

[30]. Curtarolo, Stefano, et al. "Predicting crystal structures with data mining of quantum calculations." Physical review letters 91.13 (2003): 135503.

[31]. Hautier, Geoffroy, et al. "Finding nature's missing ternary oxide compounds using machine learning and density functional theory." Chemistry of Materials 22.12 (2010): 3762-3767.

[32]. Behler, Jörg, and Michele Parrinello. "Generalized neural-network representation of high-dimensional potential-energy surfaces." Physical review letters 98.14 (2007): 146401.

[33]. Sabin, T. J., C. A. L. Bailer-Jones, and P. J. Withers. " Accelerated learning using Gaussian process models to predict static recrystallization in an Al-Mg alloy." Modelling and Simulation in Materials Science and Engineering 8.5 (2000): 687.

[34]. Nørskov, Jens Kehlet, et al. "Towards the computational design of solid catalysts." Nature chemistry 1.1 (2009): 37-46.

[35]. Balachandran, Prasanna V., et al. "Adaptive strategies for materials design using uncertainties." Scientific reports 6.1 (2016): 19660.

[36]. Hautier, Geoffroy, et al. "Data mined ionic substitutions for the discovery of new compounds." Inorganic chemistry 50.2 (2011): 656-663.

[37]. Isayev, Olexandr, et al. "Materials cartography: representing and mining materials space using structural and electronic fingerprints." Chemistry of Materials 27.3 (2015): 735-743.

[38]. Saal, James E., et al. "Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD)." Jom 65 (2013): 1501-1509.

[39]. Fujimura, Koji, Atsuto Seko, Yukinori Koyama, Akihide Kuwabara, Ippei Kishida, Kazuki Shitara, Craig AJ Fisher, Hiroki Moriwake, and Isao Tanaka. "Accelerated materials design of lithium superionic conductors based on first-principles calculations and machine learning algorithms." Advanced Energy Materials 3, no. 8 (2013): 980-985.

[40]. Srinivasan, Srikant, et al. "Mapping Chemical Selection Pathways for Designing Multicomponent Alloys: an informatics framework for materials design." Scientific reports 5.1 (2015): 17960.

[41]. Fischer, Christopher C., Kevin J. Tibbetts, Dane Morgan, and Gerbrand Ceder. " Predicting crystal structure by merging data mining with quantum mechanics." Nature materials 5, no. 8 (2006): 641-646.

[42]. Bligaard, Thomas, Jens Kehlet Nørskov, Søren Dahl, J. Matthiesen, Claus H. Christensen, and JJJoC Sehested. "The Brønsted–Evans–Polanyi relation and the volcano curve in heterogeneous catalysis." Journal of catalysis 224, no. 1 (2004): 206-217.

[43]. Klanner, Catharina, David Farrusseng, Laurent Baumes, Mourad Lengliz, Claude Mirodatos, and Ferdi Schüth. "The development of descriptors for solids: teaching "catalytic intuition" to a computer." Angewandte Chemie 116, no. 40 (2004): 5461-5463.

[44]. Saad, Yousef, et al. "Data mining for materials: Computational experiments with AB compounds." Physical Review B—Condensed Matter and Materials Physics 85.10 (2012): 104104.

[45]. Meredig, Bryce, and C. Wolverton. " A hybrid computational–experimental approach for automated crystal structure solution." Nature materials 12.2 (2013): 123-127.

[46]. Hautier, Geoffroy, et al. "Identification and design principles of low hole effective mass p-type transparent conducting oxides." Nature communications 4.1 (2013): 2292.

[47]. Greeley, Jeff, et al. "Computational high-throughput screening of electrocatalytic materials for hydrogen evolution." Nature materials 5.11 (2006): 909-913.

[48]. Mounet, Nicolas, and Nicola Marzari. "First-principles determination of the structural, vibrational and thermodynamic properties of diamond, graphite, and derivatives." Physical Review B—Condensed Matter and Materials Physics 71.20 (2005): 205214.

[49]. Ashby, M. F. "Multi-objective optimization in material design and selection." Acta materialia 48.1 (2000): 359-369.

[50]. Franceschetti, Alberto, and Alex Zunger. "The inverse band-structure problem of finding an atomic configuration with given electronic properties." Nature 402.6757 (1999): 60-63