

Development of a Hybrid Classical-Quantum SVM Model for Predicting Perovskite Formation Energy

Fauzan Arif Prayogi^{1*}, Muhamad Akrom^{2**}

* Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia

** Research Center for Quantum Computing and Materials Informatics, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia

fauzanarifp@gmail.com¹, m.akrom@dsn.dinus.ac.id²

Article Info

Article history:

Received 2026-01-12

Revised 2026-02-22

Accepted 2026-04-08

Keyword:

Explainable AI,
Formation Energy,
Hybrid SVM,
Perovskite,
Quantum Machine Learning.

ABSTRACT

The global energy crisis and the threat of climate change are driving the acceleration of the transition towards new and renewable energy (NRE), with perovskite solar cells emerging as a leading candidate due to their high efficiency (>25%) and low production costs. A major challenge in the development of perovskite materials is their structural stability, which can be assessed through Formation Energy (FE). However, FE calculations using Density Functional Theory (DFT) are computationally expensive and not scalable for screening thousands of material candidates. This research develops a Quantum Machine Learning (QML) model to predict the Formation Energy of ABX_3 perovskite materials as an efficient alternative to conventional DFT approaches. Three variants of Support Vector Machine algorithms are compared comprehensively: the classical SVM with an RBF kernel as a baseline, Quantum SVM (QSVM) utilizing a quantum kernel with ZZFeatureMap to represent data in a high-dimensional quantum feature space, and Hybrid SVM (HSVM), which combines the expressiveness of the quantum feature map with the flexibility of classical parameter optimization. Evaluation results show that Hybrid SVM outperforms others with $R^2 = 0.7221$, RMSE = 0.4203, and MAE = 0.3228 on test data, improving by 4-8% compared to classical SVM ($R^2 = 0.6674$) and surpassing QSVM ($R^2 = 0.6966$). Interpretability analysis using Explainable AI with permutation feature importance reveals that the bandgap (HSE gap indirect) and effective hole mass are the most crucial predictors, confirming that electronic properties dominate the thermodynamic stability of perovskites. This study validates the potential of QML, particularly the hybrid approach, as an intelligent solution to accelerate the screening and design of renewable energy materials, supporting Indonesia's target of a 23% energy mix by 2025 and the Vision of Indonesia 2045.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Krisis energi global, fluktuasi harga minyak dunia, dan ancaman perubahan iklim akibat emisi karbon mendorong percepatan transisi menuju energi baru dan terbarukan (EBT). Indonesia berkomitmen mencapai target bauran energi terbarukan sebesar 23% pada tahun 2025 dan 31% pada tahun 2030 guna mendukung Sustainable Development Goals (SDGs), khususnya akses energi bersih dan penanganan perubahan iklim [1]. Dalam upaya ini, sel surya perovskite dengan struktur ABX_3 muncul sebagai teknologi yang

menjanjikan, menawarkan efisiensi konversi energi tinggi (>25%) dengan biaya produksi yang lebih rendah dibandingkan silikon konvensional [2]. Namun, tantangan utama komersialisasi material ini terletak pada stabilitas strukturalnya terhadap faktor lingkungan. Parameter termodinamika krusial untuk menilai stabilitas ini adalah *Formation Energy* (FE), di mana nilai FE yang lebih negatif menandakan kestabilan yang lebih baik terhadap degradasi [3], [4].

Metode konvensional untuk menghitung FE, seperti *Density Functional Theory* (DFT), terbukti akurat namun memiliki kendala skalabilitas komputasi yang masif. Kompleksitas algoritma DFT standar umumnya berskala kubik, atau $O(N^3)$, terhadap jumlah atom dan elektron di dalam sistem. Akibatnya, simulasi relaksasi struktur dan kalkulasi energi untuk satu kandidat material perovskite (terutama yang menggunakan *supercell* kompleks) dapat memakan waktu puluhan jam hingga berhari-hari pemrosesan secara kontinu pada klaster *High-Performance Computing* (HPC). Beban komputasi dan waktu proses yang sangat mahal ini menjadikannya tidak pragmatis dan tidak efisien untuk melakukan *high-throughput screening* pada ribuan ruang desain kandidat material [5], [6]. Sebagai solusi, pendekatan Machine Learning (ML) berkembang pesat dalam *materials informatics* untuk memprediksi sifat material secara efisien [7], [8]. Beberapa studi terbaru menunjukkan keberhasilan ML klasik; misalnya, penelitian [9] dan [10] menggunakan metode ensemble untuk memprediksi sifat mekanik dan band gap perovskite dengan akurasi tinggi. Sementara itu, penelitian [11] dan [12] memanfaatkan ML untuk klasifikasi stabilitas struktural dan desain material berskala besar. Meskipun demikian, metode ML klasik sering kali memiliki keterbatasan dalam menangkap kompleksitas sistem kuantum material secara efisien [13], [14].

Seiring kemajuan teknologi, *Quantum Machine Learning* (QML) hadir sebagai frontier baru yang memanfaatkan prinsip superposisi dan *entanglement* untuk memproses data berdimensi tinggi [15]. Adaptasi algoritma kuantum, seperti Quantum Support Vector Machine (QSVM), menawarkan potensi representasi fitur yang lebih kaya melalui quantum kernel dibandingkan metode klasik [16]. Namun, penelitian yang secara spesifik menerapkan QML untuk prediksi FE material perovskite masih sangat terbatas, dan evaluasi komparatif antara model kuantum dengan model klasik maupun hybrid belum banyak dieksplorasi.

Penelitian ini bertujuan untuk mengembangkan model prediksi FE material perovskite ABX_3 menggunakan pendekatan QML dan membandingkan performanya terhadap model Support Vector Machine (SVM) klasik, Quantum SVM, dan Hybrid SVM. Kebaruan penelitian ini terletak pada pengembangan arsitektur Hybrid SVM yang mengintegrasikan quantum kernel computation melalui ZZFeatureMap dengan *classical parameter optimization*, komparasi sistematis ketiga pendekatan dengan hyperparameter konsisten untuk *fairness* evaluasi, serta integrasi *Explainable AI* untuk meningkatkan interpretabilitas model quantum. Penelitian ini diharapkan dapat mempercepat

penemuan material energi terbarukan dan berkontribusi pada target energi nasional Indonesia.

II. METODE

Penelitian ini mengembangkan pendekatan QML untuk memprediksi FE perovskit ABX_3 dan memvalidasinya melalui analisis komparatif dengan model Support Vector Machine (SVM) klasik, Quantum SVM (QSVM), dan Hybrid SVM. Model SVM klasik dipilih sebagai baseline karena ketangguhannya (*robustness*) dalam menangani data berdimensi tinggi dan efektivitasnya yang telah teruji dalam berbagai studi *materials informatics* [17]. Model QSVM diterapkan untuk memanfaatkan quantum kernel method yang mampu memetakan data ke dalam Hilbert space berdimensi tinggi, guna menangkap korelasi non-linear yang sulit dijangkau oleh kernel klasik. Selanjutnya, model Hybrid SVM dikembangkan untuk menggabungkan keunggulan representasi fitur kuantum dengan efisiensi optimasi klasik. Pendekatan QML ini diusulkan untuk mengeksplorasi potensi representasi data yang lebih efisien dan akurat dalam menyelesaikan permasalahan prediksi FE yang kompleks.

A. Waktu dan Tempat Penelitian

Penelitian berlangsung selama 7 bulan terhitung sejak bulan Juni sampai Desember 2025 di Laboratorium Quantum Computing and Material Informatics, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro.

B. Alat Penelitian

Aktivitas utama penelitian adalah pengembangan model komputasional melalui layanan Google Colab dan IBM Quantum Experience untuk eksperimen QML, dengan memanfaatkan library Scikit-Learn, NumPy, Pandas, Matplotlib, Seaborn, PyTorch, TensorFlow, dan Qiskit berbasis Python. Eksperimen dilakukan menggunakan StatevectorSampler simulator dari Qiskit yang menyediakan simulasi eksak *quantum circuits* tanpa *noise* untuk evaluasi algoritma QSVM dan Hybrid SVM.

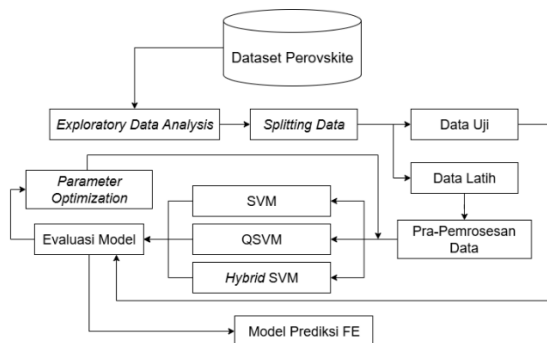
C. Dataset

Dataset yang digunakan merupakan data sekunder yang berasal dari studi [18] dengan jumlah 127 baris data poin, 8 fitur (F1-F8) meliputi E-hull, miu/miuB, PBE_gap direct, PBE_gap indirect, HSE_gap direct, HSE_gap indirect, m*h, P(LDA) sebagai variabel independen, dan 1 target (Y) yaitu nilai FE sebagai variabel dependen. Sampel dataset 5 baris teratas terdapat pada tabel I.

TABEL I SAMPEL DATASET PEROVSKITE

No	F1	F2	F3	F4	F5	F6	F7	F8	Y
1	0,00	0,00	1,85	2,91	3,15	3,95	0,69	66	-2.748
2	0,00	0,00	0,00	0,00	0,09	0,31	0,00	1	-3.188
3	0,00	0,00	1,00	1,00	1,57	1,57	0,14	1	-1.807
4	0,00	0,00	2,14	2,64	3,58	4,12	10,57	31	-3.506
5	0,00	0,00	2,12	2,79	3,39	4,10	2,31	0	-3.098

D. Tahapan Penelitian



Gambar 1. Diagram Alir Penelitian

Penelitian diawali dengan pengumpulan dataset material Perovskite. Tahap selanjutnya adalah *Exploratory Data Analysis* (EDA) yang bertujuan untuk memahami distribusi, korelasi, dan karakteristik data secara menyeluruh. Setelah analisis awal, dilakukan proses *splitting* data untuk memisahkan dataset menjadi dua bagian, yaitu data latih dan data uji. Pada data latih, dilakukan tahap pra-pemrosesan data (*pre-processing*) guna membersihkan dan mentransformasi fitur agar siap digunakan dalam proses pelatihan algoritma.

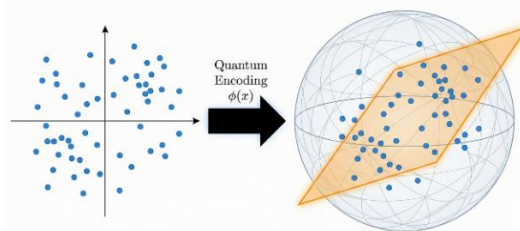
Pemodelan inti dilakukan melalui tiga pendekatan komparatif, yaitu Support Vector Machine (SVM) sebagai representasi metode klasik, serta Quantum Support Vector Machine (QSVM) dan Hybrid SVM sebagai pendekatan berbasis kuantum. Ketiga model ini menjalani proses *parameter optimization* secara iteratif untuk mendapatkan konfigurasi *hyperparameter* yang paling optimal. Kinerja model yang telah terbentuk kemudian dievaluasi menggunakan data uji untuk mengukur tingkat akurasi. Tahapan ini bermuara pada terbentuknya Model Prediksi FE final guna memvalidasi potensi penerapan algoritma kuantum dalam *materials informatics* dibandingkan dengan metode klasik.

E. Quantum Machine Learning

Machine Learning (ML) merupakan cabang dari *Artificial Intelligence* (AI) yang berperan penting dalam pemecahan masalah kompleks di berbagai domain, termasuk *materials science*, dengan kemampuan untuk belajar dari data tanpa pemrograman eksplisit. Secara umum, ML terbagi menjadi tiga paradigma utama: *supervised learning*, yang memanfaatkan data berlabel untuk mempelajari pemetaan *input-output* dan digunakan secara luas untuk prediksi properti material seperti Formation Energy, bandgap, dan stabilitas struktural; *unsupervised learning*, yang mengeksplorasi data tak berlabel untuk mengungkap struktur atau pola laten seperti *clustering* material dengan sifat serupa atau *dimensionality reduction* untuk visualisasi ruang kimia; dan *reinforcement learning*, yang melatih agen melalui proses *trial-and-error* dengan umpan balik penguatan untuk mengoptimalkan *reward* kumulatif, relevan untuk optimasi komposisi material atau parameter sintesis. Meskipun ML klasik telah memberikan kontribusi signifikan dalam

materials informatics, keterbatasan dalam menangani kompleksitas sistem kuantum material dan eksplorasi ruang fitur berdimensi tinggi mendorong perkembangan pendekatan baru yang memanfaatkan komputasi kuantum. [19], [20]

Quantum Machine Learning (QML) merupakan paradigma komputasi yang menggabungkan prinsip-prinsip mekanika kuantum dengan algoritma *machine learning* untuk mengatasi keterbatasan pendekatan klasik dalam memproses informasi kompleks. QML memanfaatkan fenomena kuantum fundamental seperti superposisi (kemampuan qubit untuk berada dalam kombinasi linear dari beberapa state secara simultan, memungkinkan paralelisme kuantum dalam eksplorasi ruang solusi) dan *entanglement* (korelasi kuantum non-lokal antar qubit yang tidak dapat dijelaskan oleh teori klasik, menciptakan representasi data yang lebih ekspresif) untuk mencapai keunggulan komputasi dalam tugas-tugas tertentu. Dalam konteks *materials science*, QML menawarkan potensi *breakthrough* untuk memprediksi sifat material dengan akurasi lebih tinggi dan efisiensi komputasi lebih baik dibandingkan ML klasik, terutama karena sistem material itu sendiri bersifat kuantum sehingga dapat direpresentasikan secara lebih natural dalam *quantum feature space*. Pendekatan QML dapat diimplementasikan dalam berbagai bentuk: *fully quantum approaches* yang menjalankan seluruh algoritma pada quantum hardware, *hybrid quantum-classical approaches* yang mengkombinasikan *quantum circuits* untuk *feature extraction* dengan *classical optimizer* untuk *parameter optimization* (relevan untuk era NISQ saat ini), dan *quantum-inspired classical algorithms* yang mengadopsi prinsip kuantum dalam arsitektur klasik. Keunggulan teoritis QML meliputi quantum speedup dalam algoritma pencarian dan optimasi, eksponen lebih tinggi dalam kapasitas representasi melalui *quantum feature maps* yang memetakan data ke Hilbert space berdimensi tinggi, serta kemampuan untuk menangkap korelasi non-linear kompleks yang sulit diakses oleh kernel klasik, karakteristik yang sangat relevan untuk memprediksi Formation Energy material perovskite di mana interaksi many-body elektron dan efek kuantum memainkan peran krusial dalam menentukan stabilitas termodinamika. [15]



Gambar 2. Ilustrasi Quantum Feature Space

Secara konseptual, *quantum feature space* merupakan sebuah pendekatan matematis untuk mengatasi kompleksitas sebaran data non-linear pada ruang klasik berdimensi rendah. Melalui mekanisme *quantum encoding* (seperti penggunaan ZZFeatureMap), titik-titik data pengamatan

ditransformasikan dan dipetakan ke dalam ruang Hilbert berdimensi eksponensial. Transformasi spasial ini memungkinkan distribusi data yang sebelumnya saling tumpang tindih dan sulit dimodelkan menjadi terurai secara lebih terstruktur di dalam quantum state. Sebagai implikasinya, algoritma ML dapat dengan lebih efisien mengonstruksi sebuah hyperplane linier untuk mengekstraksi pola tersembunyi, yang pada akhirnya mengoptimalkan akurasi dan presisi prediksi model. [21], [22]

F. Model Machine Learning

1) *Support Vector Machine (SVM) Klasik*: Support Vector Machine adalah algoritma *supervised learning* yang bekerja dengan menemukan *hyperplane* optimal dalam ruang fitur berdimensi tinggi untuk memisahkan atau memprediksi data dengan memaksimalkan margin antara support vectors. Dalam konteks regresi (SVR), SVM meminimalkan *error* prediksi menggunakan fungsi kernel seperti Radial Basis Function (RBF) atau polynomial untuk memetakan data non-linear ke ruang fitur yang lebih tinggi. SVM klasik dipilih sebagai *baseline* model karena terbukti efektif dalam menangani dataset berukuran kecil hingga menengah dengan dimensi tinggi, *robust* terhadap *outlier* melalui parameter epsilon-tube, dan mampu menangkap hubungan non-linear kompleks antara fitur fisikokimia dengan FE melalui kernel trick tanpa memerlukan perhitungan eksplisit di ruang fitur berdimensi tinggi [23].

2) *Quantum Support Vector Machine (QSVM)*: Quantum SVM merupakan adaptasi kuantum dari SVM klasik yang menggantikan *classical kernel* dengan quantum kernel yang dihitung menggunakan *quantum circuit* untuk mengukur *similarity* dalam *quantum feature space*. Algoritma ini memanfaatkan prinsip superposisi dan *entanglement* untuk menciptakan *feature map* kuantum yang dapat merepresentasikan data dalam Hilbert space berdimensi eksponensial, memungkinkan identifikasi pola kompleks yang sulit diakses oleh kernel klasik. QSVM berpotensi memberikan *quantum advantage* dalam prediksi FE material perovskite karena sifat kuantum intrinsik dari sistem material dapat direpresentasikan lebih natural dalam *quantum feature space*, serta kemampuannya untuk mengeksplorasi ruang fitur yang lebih luas dengan jumlah parameter yang lebih efisien, terutama relevan untuk menangkap korelasi elektron *many-body* dan interaksi non-linear kompleks antar deskriptor material yang menentukan stabilitas termodinamika [24].

3) *Hybrid SVM (HSVM)*: Hybrid SVM menggabungkan kekuatan komputasi klasik dan kuantum dengan menggunakan *variational quantum circuits* sebagai *kernel function* yang parameternya dioptimasi secara klasik. Pendekatan hybrid ini memungkinkan fleksibilitas dalam menyeimbangkan ekspresivitas *quantum feature map* dengan skalabilitas komputasi klasik, di mana *quantum circuit* digunakan untuk *feature extraction* dan transformation, sementara optimasi parameter dan *training* dilakukan menggunakan algoritma klasik seperti gradient descent.

Hybrid SVM sangat cocok untuk era NISQ (*Noisy Intermediate-Scale Quantum*) saat ini karena dapat beradaptasi dengan keterbatasan jumlah qubit dan *noise quantum hardware*, memungkinkan eksplorasi berbagai ansatz *quantum circuit* untuk menemukan representasi optimal dari hubungan fitur-target dalam dataset perovskite, serta memberikan *trade-off* yang dapat dikustomisasi antara kompleksitas kuantum dan performa prediksi sesuai dengan sumber daya komputasi yang tersedia [25].

G. Evaluasi

Pada tahap evaluasi model regresi, tujuan utamanya adalah mengukur kemampuan model dalam memprediksi nilai kontinu dengan menggunakan beberapa metrik evaluasi. R^2 (koefisien determinasi) digunakan untuk menilai sejauh mana variabilitas data yang dapat dijelaskan oleh model. Nilai R^2 yang lebih tinggi menunjukkan model yang lebih baik dalam menjelaskan data. [26]

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

Root Mean Squared Error (RMSE) mengukur akar kuadrat dari rata-rata kesalahan kuadrat antara nilai yang diprediksi dan nilai sebenarnya, dengan nilai yang lebih kecil menunjukkan kesalahan prediksi yang lebih rendah. [26]

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

Mean Absolute Error (MAE) mengukur rata-rata selisih absolut antara prediksi dan nilai aktual, memberikan gambaran yang jelas mengenai seberapa besar kesalahan rata-rata model dalam unit asli data [26].

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

dimana:

y_i : Nilai sebenarnya (*true value*)

\hat{y}_i : Nilai yang diprediksi (*predicted value*)

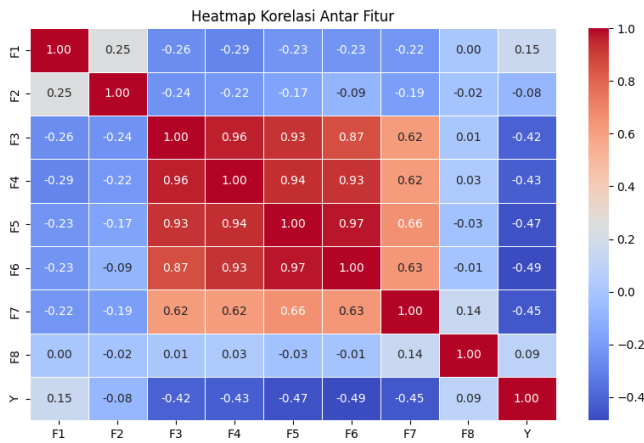
\bar{y} : Rata-rata dari nilai sebenarnya

n : Jumlah data

III. HASIL DAN PEMBAHASAN

A. Exploratory Data Analysis (EDA)

Dalam subbagian ini, analisis statistika deskriptif akan dilakukan untuk memberikan gambaran umum tentang karakteristik sifat fisikokimia material perovskite dan distribusi FE dalam *dataset*.

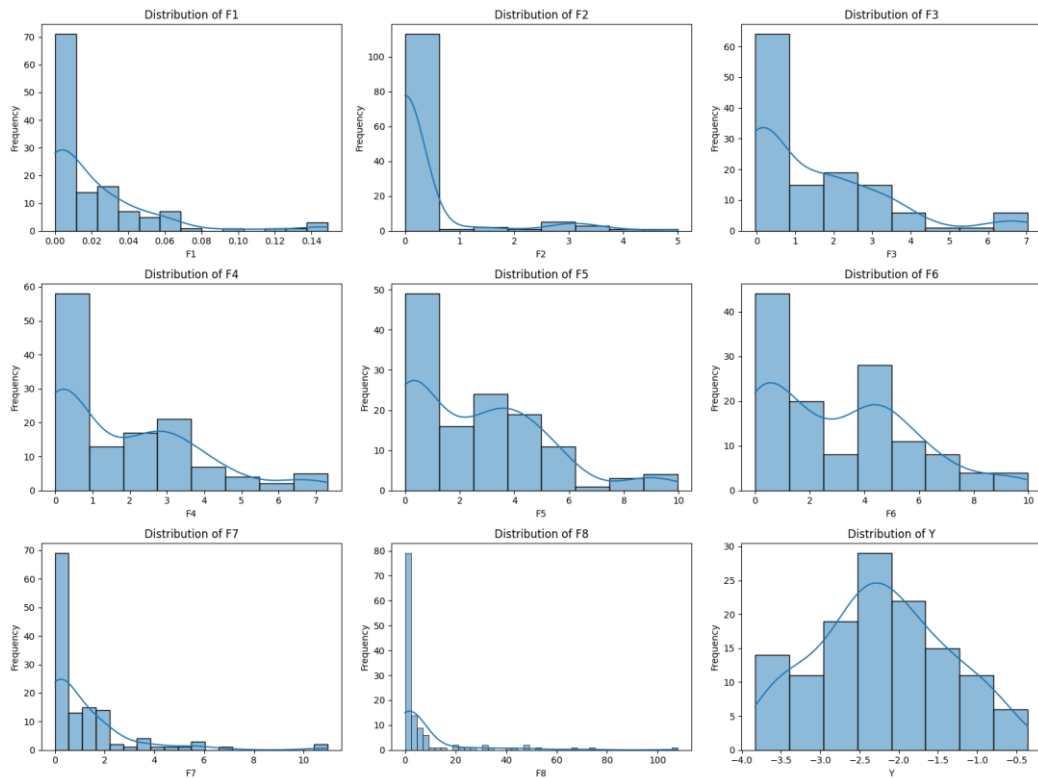


Gambar 3. Heatmap Korelasi Tiap Variabel

Berdasarkan heatmap korelasi pada Gambar 3, terlihat bahwa mayoritas deskriptor fisikokimia material perovskite (F2-F7) menunjukkan korelasi negatif terhadap variabel target FE (Y). Dari delapan fitur prediktor, hanya dua fitur

F6 ($r = -0,49$), F5 ($r = -0,47$), F7 ($r = -0,45$), F4 ($r = -0,43$), dan F3 ($r = -0,42$), yang mengindikasikan bahwa peningkatan nilai deskriptor-deskriptor ini akan menurunkan nilai FE (menjadi lebih negatif atau lebih stabil secara termodinamika). Fenomena ini menunjukkan bahwa material perovskite dengan nilai yang lebih tinggi pada fitur-fitur seperti bandgap (PBE/HSE gaps), massa efektif hole (m^*h), atau parameter struktural lainnya cenderung memiliki stabilitas termodinamika yang lebih baik, yang merupakan karakteristik *desirable* untuk aplikasi sel surya perovskite yang tahan lama.

Berdasarkan histogram distribusi pada Gambar 4, terlihat bahwa sebagian besar fitur fisikokimia material perovskite menunjukkan pola distribusi yang tidak seimbang (*skewed*). Fitur E-hull (F1), μ/μ_B (F2), dan m^*h (F7) menunjukkan *right-skewed distribution (positively skewed)* yang sangat kuat, di mana mayoritas data terkonsentrasi pada nilai rendah dengan ekor distribusi yang memanjang ke arah nilai tinggi. Pola *skewness* positif ini mengindikasikan bahwa sebagian besar material perovskite dalam dataset memiliki nilai E-hull



Gambar 4. Distribusi Data Tiap Variabel

yang menunjukkan korelasi positif lemah yaitu F8 ($r = 0,09$) dan F1 ($r = 0,15$), mengindikasikan bahwa peningkatan nilai kedua deskriptor ini berkontribusi pada peningkatan nilai FE (menjadi kurang negatif atau kurang stabil secara termodinamika).

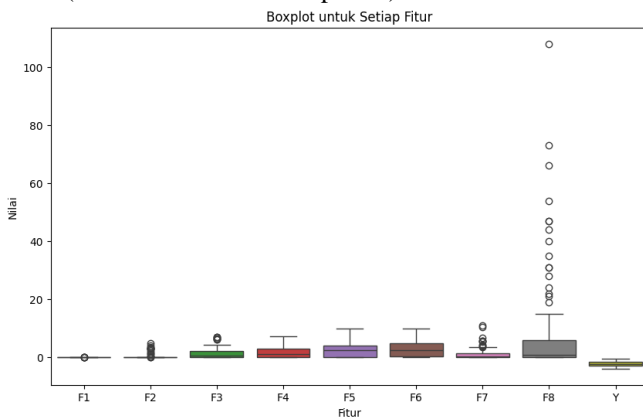
Sebaliknya, enam fitur lainnya (F2, F3, F4, F5, F6, F7) menunjukkan korelasi negatif dengan kekuatan yang bervariasi. Korelasi negatif yang paling kuat ditemukan pada

mendekati nol (stabil pada convex hull), octahedral factor yang kecil, dan massa efektif hole yang rendah, sementara hanya sebagian kecil material yang memiliki nilai ekstrem tinggi pada parameter-parameter ini.

Distribusi PBE gap (indirect) (F3) dan PBE gap (direct) (F4) juga menunjukkan *right-skewness* moderat, dengan konsentrasi data pada rentang bandgap rendah hingga menengah (0-3 eV) dan distribusi yang lebih tersebar pada

nilai yang lebih tinggi. Sebaliknya, HSE gap (direct) (F5) dan HSE gap (indirect) (F6) menunjukkan distribusi yang lebih mendekati normal meskipun masih memiliki *slight right-skewness*, mengindikasikan bahwa koreksi HSE menghasilkan distribusi bandgap yang lebih seimbang dibandingkan PBE. Fitur P(LDA) (F8) menampilkan distribusi yang sangat tidak simetris dengan mayoritas material memiliki polarisasi rendah dan beberapa *outlier* dengan nilai polarisasi sangat tinggi.

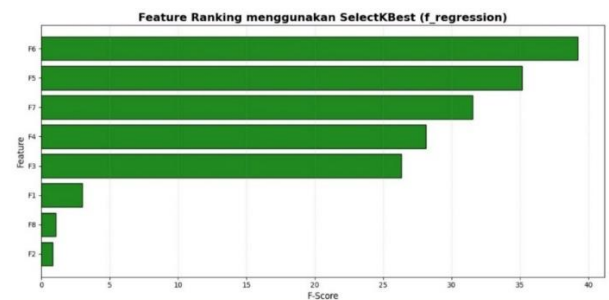
Yang paling menarik adalah distribusi FE sebagai variabel target yang menunjukkan distribusi mendekati normal (*approximately normal distribution*) dengan *slight left-skewness*, di mana data terkonsentrasi pada rentang -2,0 hingga -3,0 eV/atom dengan ekor yang sedikit memanjang ke arah nilai yang lebih negatif (lebih stabil). Pola distribusi ini mengindikasikan bahwa mayoritas material perovskite dalam dataset memiliki stabilitas termodinamika yang moderat hingga tinggi, dengan sebagian kecil material yang memiliki stabilitas sangat tinggi (FE sangat negatif) atau relatif tidak stabil (FE mendekati nol atau positif).



Gambar 5. Visualisasi *Boxplot* Tiap Variabel

Boxplot pada Gambar 5 mengungkapkan keberadaan *outlier* yang signifikan pada beberapa fitur fisikokimia material perovskite. Fitur F1 (E-hull), F2 (miu/miuB), F3 (PBE gap indirect), F4 (PBE gap direct), F5 (HSE gap direct), dan F6 (HSE gap indirect) menunjukkan distribusi yang relatif kompak dengan median mendekati nilai rendah dan *interquartile range* (IQR) yang sempit, namun memiliki beberapa *outlier* pada nilai tinggi yang konsisten dengan pola *right-skewness* yang telah diidentifikasi sebelumnya. Fitur F7 (m*h) menampilkan keberadaan *outlier* ekstrem dengan nilai mencapai sekitar 10-12, jauh melebihi *upper fence* dari boxplot, mengindikasikan adanya material dengan massa efektif hole yang sangat tinggi yang mungkin memiliki karakteristik transport elektron yang unik. Yang paling mencolok adalah F8 (P(LDA)) yang memiliki jumlah *outlier* terbanyak dengan nilai yang sangat tersebar hingga mencapai lebih dari 100 $\mu\text{C}/\text{cm}^2$, menunjukkan bahwa sebagian kecil material perovskite dalam dataset memiliki sifat ferroelektrik atau piezoelektrik yang sangat kuat. Variabel target Y (FE) menunjukkan distribusi yang lebih seimbang dengan median sekitar -2,5 eV/atom dan IQR yang moderat, meskipun

terdapat beberapa *outlier* pada nilai yang lebih negatif (material sangat stabil) dan nilai mendekati nol (material kurang stabil), yang perlu diperhatikan dalam proses modeling karena dapat mempengaruhi pembelajaran pola oleh algoritma ML. Dilakukan penghapusan seluruh *outlier* yang terdeteksi pada semua variabel menggunakan metode IQR untuk memastikan data yang digunakan berada dalam rentang distribusi normal dan mengurangi *noise* yang dapat mengganggu pembelajaran pola oleh model, terutama penting untuk Quantum SVM yang sensitif terhadap variasi data ekstrem dalam *quantum feature space*, serta melakukan standarisasi data menggunakan metode Min-Max Scaler untuk memastikan setiap fitur berada pada skala yang sebanding (rentang 0 sampai 1).



Gambar 6. *Ranking* Variabel Berdasarkan SelectKBest

Untuk mengidentifikasi fitur yang paling relevan dalam memprediksi FE, dilakukan seleksi fitur menggunakan metode SelectKBest dengan *f_regression* sebagai *scoring function*. Berdasarkan Gambar 6, terlihat bahwa fitur F6 (HSE gap indirect) memiliki F-Score tertinggi (~39), diikuti oleh F5 (HSE gap direct) (~35), F7 (m*h) (~31), F4 (PBE gap direct) (~28), dan F3 (PBE gap indirect) (~26), mengindikasikan bahwa kelima fitur ini memiliki hubungan linear yang kuat dan signifikan secara statistik dengan variabel target. Sebaliknya, F1 (E-hull), F8 (P(LDA)), dan F2 (miu/miuB) menunjukkan F-Score yang sangat rendah (< 5), menandakan kontribusi yang minimal terhadap prediksi FE. Berdasarkan hasil analisis ini, F1 dan F8 dihapus dari dataset karena F-Score mereka yang sangat rendah menunjukkan bahwa kedua fitur ini tidak memberikan informasi prediktif yang cukup signifikan, sementara F2 dipertahankan sebagai representasi parameter struktural meskipun memiliki F-Score rendah. Reduksi fitur dari 8 menjadi 6 variabel prediktor ini tidak hanya meningkatkan efisiensi komputasi tetapi juga mengurangi *noise* dalam model, yang sangat penting untuk implementasi Quantum SVM di mana setiap fitur tambahan memerlukan qubit dan quantum gate tambahan yang dapat meningkatkan kompleksitas *circuit* dan sensitivitas terhadap *noise* pada *quantum hardware*.

B. Model Regresi Energi Formasi

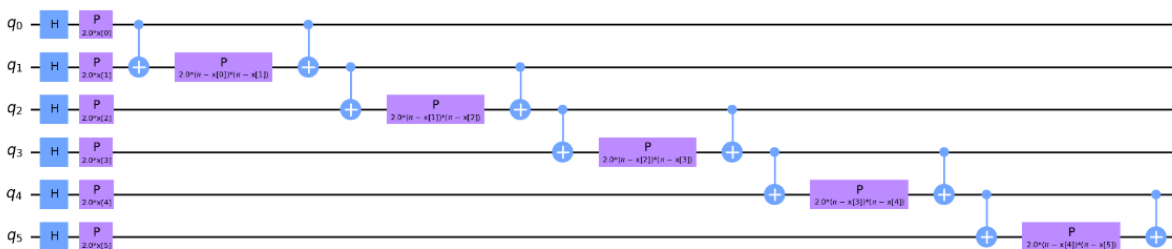
Ketiga model (QSVR, Hybrid SVM, dan SVM klasik) diimplementasikan dengan parameter yang konsisten untuk memastikan *fairness* komparasi. *ZZFeatureMap* dikonfigurasi dengan *reps*=1 dan *entanglement*="linear"

untuk menghasilkan quantum circuit yang efisien dengan entangling gates antar qubit bertetangga, di mana jumlah qubit disesuaikan dengan dimensi fitur input (6 variabel prediktor). Parameter SVR diatur dengan $C=5$ untuk regularisasi yang mengontrol kompleksitas model, $\epsilon=0,2$ untuk lebar epsilon-insensitive tube yang menentukan margin toleransi *error*, dan $\gamma='scale'$ yang menghitung γ berdasarkan *variance data* untuk menyeimbangkan pengaruh fitur. Perbedaan utama terletak pada kernel function: QSVM menggunakan quantum kernel melalui FidelityQuantumKernel dengan StatevectorSampler, Hybrid SVM menggunakan `quantum_kernel.evaluate` yang mengintegrasikan quantum kernel dengan classical optimizer, dan SVM klasik menggunakan Radial Basis Function (RBF) kernel sebagai baseline.

Encoding fitur klasik ke dalam quantum state (Gambar 7) dilakukan menggunakan ZZFeatureMap dengan 5 layer repetition yang mengimplementasikan entangling gates $U_{ZZ}(\theta) = e^{-i\theta(Z \otimes Z)}$ antar qubit untuk menangkap interaksi non-linear antar fitur dalam *quantum feature space*, di mana setiap layer terdiri dari Hadamard gates untuk menciptakan superposisi, rotation gates untuk encoding nilai fitur, dan CNOT gates untuk membentuk *entanglement* yang memungkinkan representasi kompleks dari korelasi antar deskriptor fisikokimia material. *Feature map* ini akan diimplementasikan ke dalam quantum kernel yang digunakan oleh QSVM dan hybrid SVM.

TABEL II EVALUASI MODEL PADA DATA UJI

Metrik	QSVM	HSVM	SVM
R²	0,6966	0,7221	0,6674
RMSE	0,4392	0,4203	0,4598
MAE	0,3306	0,3228	0,3572
Waktu Eksekusi (detik)	9,91	10,35	0,0005



Gambar 7 Feature Encoding menggunakan ZZFeatureMap.

Berdasarkan Tabel II, pada data uji, Hybrid SVM menunjukkan performa terbaik dengan $R^2 = 0,7221$, $RMSE = 0,4203$, dan $MAE = 0,3228$, mengindikasikan bahwa model ini mampu menjelaskan sekitar 72% variabilitas FE dengan *error* prediksi yang paling rendah. QSVM menempati posisi kedua dengan $R^2 = 0,6966$, $RMSE = 0,4392$, dan $MAE = 0,3306$, menunjukkan performa yang kompetitif meskipun sedikit di bawah Hybrid SVM. SVM klasik memiliki performa terendah dengan $R^2 = 0,6674$, $RMSE = 0,4598$, dan $MAE = 0,3572$, mengindikasikan bahwa classical kernel

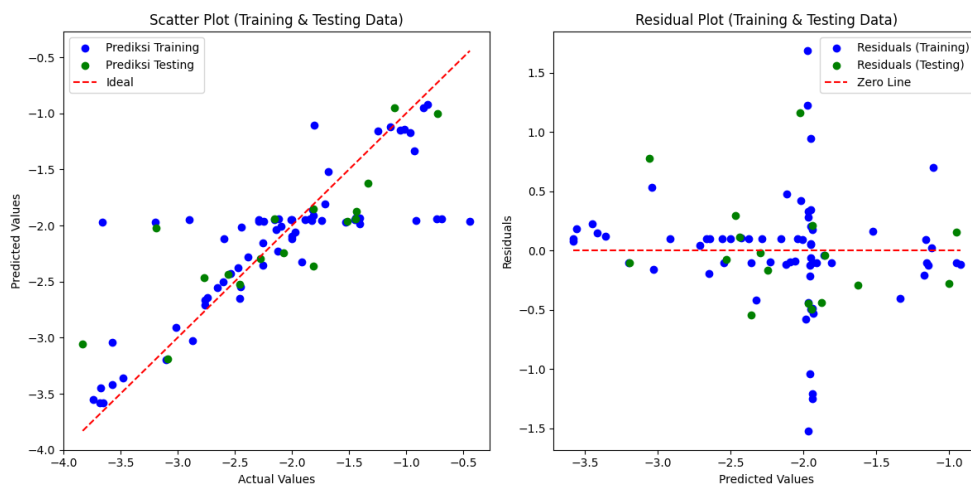
(RBF) kurang efektif dalam menangkap kompleksitas hubungan non-linear antara deskriptor fisikokimia dengan FE dibandingkan dengan *quantum-enhanced approaches*.

TABEL III EVALUASI MODEL PADA DATA LATIH

Metrik	QSVM	HSVM	SVM
R²	0,6700	0,6905	0,6505
RMSE	0,4804	0,4653	0,4945
MAE	0,3064	0,3071	0,3466
Waktu Eksekusi (detik)	20,59	20,35	0,00135

Pada data latih (Tabel III), pola yang konsisten terlihat di mana Hybrid SVM tetap unggul ($R^2 = 0,6905$, $RMSE = 0,4653$, $MAE = 0,3071$), diikuti oleh QSVM ($R^2 = 0,6700$, $RMSE = 0,4804$, $MAE = 0,3064$), dan SVM klasik ($R^2 = 0,6505$, $RMSE = 0,4945$, $MAE = 0,3466$). Yang menarik adalah selisih performa antara data latih dan data uji yang relatif kecil pada ketiga model (R^2 gap < 0,04), mengindikasikan bahwa tidak terjadi *overfitting* yang signifikan dan model memiliki kemampuan generalisasi yang baik. QSVM menunjukkan MAE terendah pada data latih (0,3064), bahkan lebih rendah dari Hybrid SVM, yang menunjukkan bahwa *quantum feature space* mampu merepresentasikan pola *training data* dengan sangat baik, meskipun pada data uji Hybrid SVM lebih unggul karena optimasi parameter variational yang lebih adaptif.

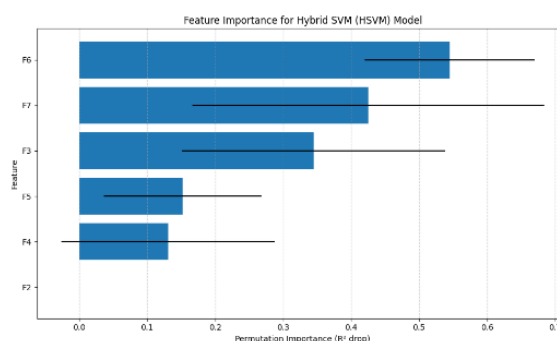
Perbandingan dengan metode DFT konvensional menunjukkan keunggulan signifikan dari pendekatan ML dalam hal efisiensi komputasi. Kalkulasi DFT untuk FE material perovskite menuntut biaya komputasi yang sangat tinggi dan waktu eksekusi yang masif pada *cluster* HPC, seringkali memakan waktu hingga puluhan jam, sangat bergantung pada ukuran *supercell* dan kompleksitas fungsional yang digunakan [14], sedangkan Hybrid SVM

Gambar 8. Visualisasi *Scatter Plot* dan *Residual Plot*

dalam penelitian ini mampu memprediksi FE untuk seluruh dataset testing hanya dalam 10,35 detik, menghasilkan percepatan komputasi lebih dari 3000 kali lipat. Bahkan model tercepat (SVM klasik dengan waktu inferensi 0,0005 detik), memvalidasi potensi ML sebagai solusi cerdas untuk *high-throughput screening* ribuan kandidat material perovskite yang sebelumnya tidak praktis dilakukan dengan DFT.

Visualisasi *scatter plot* dan *residual plot* pada Gambar 8 memberikan *insight* tambahan tentang kualitas prediksi model. *Scatter plot* (kiri) menunjukkan bahwa mayoritas prediksi baik pada data *training* (biru) maupun *testing* (hijau) berada dekat dengan garis ideal (merah), mengindikasikan korespondensi yang baik antara nilai aktual dan nilai prediksi FE, dengan sebaran titik yang relatif seragam di sepanjang rentang $-4,0$ hingga $-0,5$ tanpa pola sistematis yang menunjukkan bias prediksi pada region tertentu. *Residual plot* (kanan) menampilkan distribusi *error* yang terpusat di sekitar garis nol dengan mayoritas residual berada dalam rentang $\pm 0,5$, mengkonfirmasi bahwa model tidak memiliki bias sistematis dan *error* bersifat *random*, meskipun terdapat beberapa *outlier* pada data *testing* (hijau) dengan residual hingga $\pm 1,0$ yang mengindikasikan adanya material perovskite dengan karakteristik unik yang sulit diprediksi oleh model. Pola residual yang homoskedastik (*variance konstan*) dan tidak menunjukkan trend tertentu memvalidasi asumsi regresi linear dan mengkonfirmasi bahwa model Hybrid SVM telah menangkap hubungan fundamental antara deskriptor fisikokimia dengan FE tanpa *underfitting* atau *overfitting* yang signifikan.

Untuk meningkatkan transparansi model, dilakukan analisis interpretabilitas menggunakan teknik *Explainable AI* (XAI) melalui *permutation feature importance* untuk mengukur kontribusi deskriptor fisikokimia terhadap performa model Hybrid SVM.

Gambar 9. *Feature Importance* HSVM

Berdasarkan Gambar 9, fitur F6 (HSE gap indirect) menunjukkan *importance* tertinggi dengan penurunan R^2 sekitar 0,53-0,60 saat dipermutasi, menandakan bahwa bandgap indirect dari HSE adalah prediktor utama stabilitas termodinamika material perovskite. F7 ($m \cdot h$) dan F3 (PBE gap indirect) masing-masing memiliki *importance* sekitar 0,42-0,68 dan 0,33-0,54, menunjukkan bahwa massa efektif hole dan bandgap indirect PBE juga berperan signifikan. Fitur F5 (HSE gap direct) dan F4 (PBE gap direct) memberikan kontribusi moderat ($\sim 0,12-0,28$), sedangkan F2 (μ/μ_B) hampir tidak berkontribusi, menunjukkan bahwa fitur ini lebih representatif struktural daripada prediktif. Analisis ini mengungkapkan bahwa bandgap dan mobilitas pembawa muatan mendominasi stabilitas termodinamika perovskite, sementara parameter struktural memiliki pengaruh minimal. Hal ini mengarah pada rekomendasi desain material untuk mengoptimalkan bandgap dan mobilitas untuk mencapai perovskite dengan FE rendah (stabil), serta mengonfirmasi bahwa model Hybrid SVM mencerminkan pola yang konsisten dengan pemahaman fisika material yang ada.

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa Hybrid SVM (HSVM) memiliki performa yang paling superior dibandingkan dengan metode QSVM maupun SVM klasik. Keunggulan ini dibuktikan secara kuantitatif melalui pencapaian nilai R^2 tertinggi pada data uji, yakni sebesar 0,7221. Tingginya akurasi ini mengindikasikan bahwa arsitektur hybrid yang menggabungkan keunggulan komputasi klasik dan kuantum lebih efektif dalam mengekstraksi dan memodelkan pola non-linear yang kompleks pada dataset material perovskite. Selain itu, model ini terbukti memiliki stabilitas generalisasi yang lebih baik dibandingkan jika hanya menggunakan pendekatan murni klasik atau murni kuantum.

Penelitian ini memvalidasi peran strategis pendekatan *Quantum Machine Learning* (QML) sebagai alternatif solusi yang efisien dalam simulasi material. Metode ini terbukti mampu menanggulangi kendala utama metode konvensional *Density Functional Theory* (DFT), yaitu tingginya beban dan biaya komputasi. Dengan menawarkan prediksi yang akurat namun lebih ringan secara komputasional, pendekatan ini berpotensi besar untuk mengakselerasi siklus penemuan (*discovery*) dan pengembangan material sel surya Perovskite ABX_3 , memungkinkan proses skrining material kandidat dilakukan dengan lebih cepat dan ekonomis.

Meskipun penelitian ini menunjukkan hasil yang menjanjikan, terdapat beberapa keterbatasan yang perlu dipertimbangkan untuk pengembangan lebih lanjut. Pertama, model dikembangkan menggunakan *StatevectorSampler simulator* yang menyediakan simulasi ideal tanpa *noise*, sehingga performa pada *quantum hardware* nyata di era NISQ (*Noisy Intermediate-Scale Quantum*) saat ini mungkin berbeda akibat *error gate*, *decoherence*, dan keterbatasan jumlah qubit yang tersedia pada perangkat komersial seperti IBM Quantum atau Rigetti. Kedua, skalabilitas model perlu dievaluasi pada dataset yang lebih besar (ribuan hingga jutaan material) untuk memastikan efisiensi komputasi dan akurasi tetap konsisten, mengingat kompleksitas *quantum kernel computation* yang meningkat secara signifikan dengan ukuran data. Penelitian lanjutan perlu mengintegrasikan model prediksi ini dengan *workflow eksperimental* untuk validasi empiris dan optimasi komposisi material secara iteratif, serta mengembangkan *decision support system* yang dapat memberikan rekomendasi konkret kepada *materials scientists* dalam memilih kandidat material paling menjanjikan untuk fabrikasi sel surya perovskite generasi berikut.

DAFTAR PUSTAKA

- [1] Nunuk Febrianingsih, "Tata Kelola Energi Terbarukan di Sektor Ketenagalistrikan dalam Kerangka Pembangunan Hukum Nasional," *Majalah Hukum Nasional*, vol. 49, no. 2, pp. 29–56, Nov. 2019, doi: 10.33331/mhn.v49i2.31.
- [2] T. Tian *et al.*, "Bandgap prediction of ABX_3 -type perovskites using Broad Learning System," *Mater. Today Commun.*, vol. 37, p. 107026, Dec. 2023, doi: 10.1016/j.mtcomm.2023.107026.
- [3] M. Shafiq, B. Amin, M. A. Jehangir, A. R. Chaudhry, and G. Murataza, "First-principle calculations to investigate mechanical and acoustical properties of predicted stable halide Perovskite ABX_3 ," *J. Mol. Graph. Model.*, vol. 133, p. 108861, Dec. 2024, doi: 10.1016/j.jmgm.2024.108861.
- [4] H. Wang, Q. Zhang, M. Qiu, and B. Hu, "Synthesis and application of perovskite-based photocatalysts in environmental remediation: A review," *J. Mol. Liq.*, vol. 334, p. 116029, Jul. 2021, doi: 10.1016/j.molliq.2021.116029.
- [5] M. Akrom, S. Rustad, and H. K. Dipojono, "A machine learning approach to predict the efficiency of corrosion inhibition by natural product-based organic inhibitors," *Phys. Scr.*, vol. 99, no. 3, p. 36006, Aug. 2024, doi: 10.1088/1402-4896/ad28a9.
- [6] W. Hu and L. Zhang, "High-throughput calculation and machine learning of two-dimensional halide perovskite materials: Formation energy and band gap," *Mater. Today Commun.*, vol. 35, p. 105841, Jun. 2023, doi: 10.1016/j.mtcomm.2023.105841.
- [7] S. Badini, S. Regondi, and R. Pugliese, "Unleashing the Power of Artificial Intelligence in Materials Design," *Materials*, vol. 16, no. 17, p. 5927, Aug. 2023, doi: 10.3390/ma16175927.
- [8] F. Emmert-Streib and M. Dehmer, "Taxonomy of machine learning paradigms: A data-centric perspective," *WIREs Data Mining and Knowledge Discovery*, vol. 12, no. 5, Sep. 2022, doi: 10.1002/widm.1470.
- [9] S. B. Akinpelu *et al.*, "Interpretable machine learning methods to predict the mechanical properties of ABX_3 perovskites," *Results Phys.*, vol. 65, p. 107978, Oct. 2024, doi: 10.1016/j.rinp.2024.107978.
- [10] D. O. Obada, E. Okafor, S. A. Abolade, A. M. Ukpong, D. Doodoo-Arhin, and A. Akande, "Explainable machine learning for predicting the band gaps of ABX_3 perovskites," *Mater. Sci. Semicond. Process.*, vol. 161, Jul. 2023, doi: 10.1016/j.mssp.2023.107427.
- [11] A. Alhashmi, M. B. Kanoun, and S. Goumri-Said, "Machine Learning for Halide Perovskite Materials ABX_3 ($B = Pb, X = I, Br, Cl$) Assessment of Structural Properties and Band Gap Engineering for Solar Energy," *Materials*, vol. 16, no. 7, p. 2657, Mar. 2023, doi: 10.3390/ma16072657.
- [12] Q. Tao, P. Xu, M. Li, and W. Lu, "Machine learning for perovskite materials design and discovery," *NPJ Comput. Mater.*, vol. 7, no. 1, p. 23, Jan. 2021, doi: 10.1038/s41524-021-00495-8.
- [13] T. Zhang, "Bandgap prediction of MXene materials via multiple machine learning approaches," *J. Phys. Conf. Ser.*, vol. 3112, no. 1, p. 012093, Sep. 2025, doi: 10.1088/1742-6596/3112/1/012093.
- [14] Z. Deng *et al.*, "Prediction of ABX_3 Perovskite Formation Energy Using Machine Learning," *Materials*, vol. 18, no. 13, p. 2927, Jun. 2025, doi: 10.3390/ma18132927.
- [15] R. M. Devadas and S. T., "Quantum machine learning: A comprehensive review of integrating AI with quantum computing for computational advancements," *MethodsX*, vol. 14, p. 103318, Jun. 2025, doi: 10.1016/j.mex.2025.103318.
- [16] M. Schuld and N. Killoran, "Quantum Machine Learning in Feature Hilbert Spaces," *Phys. Rev. Lett.*, vol. 122, no. 4, p. 040504, Feb. 2019, doi: 10.1103/PhysRevLett.122.040504.
- [17] J. Wei *et al.*, "Machine learning in materials science," *InfoMat*, vol. 1, no. 3, pp. 338–358, Sep. 2019, doi: 10.1002/inf2.12028.
- [18] S. Körbel, M. A. L. Marques, and S. Botti, "Stability and electronic properties of new inorganic perovskites from high-throughput ab initio calculations," *J. Mater. Chem. C Mater.*, vol. 4, no. 15, pp. 3157–3167, 2016, doi: 10.1039/C5TC04172D.
- [19] C. Zuccarini, K. Ramachandran, and D. D. Jayaseelan, "Material discovery and modeling acceleration via machine learning," *APL Mater.*, vol. 12, no. 9, Sep. 2024, doi: 10.1063/5.0230677.
- [20] A. Kesorn *et al.*, "Formation energy prediction of neutral single-atom impurities in 2D materials using tree-based machine learning," *Mach. Learn. Sci. Technol.*, vol. 5, no. 3, p. 035039, Sep. 2024, doi: 10.1088/2632-2153/ad66ae.
- [21] Z. Krunic, F. Flother, G. Seegan, N. Earnest-Noble, and S. Omar, "Quantum Kernels for Real-World Predictions Based on Electronic Health Records," *IEEE Transactions on Quantum Engineering*, vol. 3, pp. 1–11, 2022, doi: 10.1109/TQE.2022.3176806.
- [22] Z. Özpölat, Ö. Yıldırım, and M. Karabatak, "The Effect of Linear Discriminant Analysis and Quantum Feature Maps on QSVM

- Performance for Obesity Diagnosis,” *Balkan Journal of Electrical and Computer Engineering*, vol. 12, no. 3, pp. 206–213, Sep. 2024, doi: 10.17694/bajece.1475896.
- [23] K.-L. Du, B. Jiang, J. Lu, J. Hua, and M. N. S. Swamy, “Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions,” *Mathematics*, vol. 12, no. 24, p. 3935, Dec. 2024, doi: 10.3390/math12243935.
- [24] V. S. Naresh and S. Reddi, “Quantum SVM-based predictive analytics: transforming classification methods in healthcare and beyond,” *Quantum Inf. Process.*, vol. 24, no. 9, p. 272, Aug. 2025, doi: 10.1007/s11128-025-04892-6.
- [25] A. Callison and N. Chancellor, “Hybrid quantum-classical algorithms in the noisy intermediate-scale quantum era and beyond,” *Phys. Rev. A (Coll. Park)*, vol. 106, no. 1, p. 010101, Jul. 2022, doi: 10.1103/PhysRevA.106.010101.
- [26] B. Sekeroglu, Y. K. Ever, K. Dimililer, and F. Al-Turjman, “Comparative Evaluation and Comprehensive Analysis of Machine Learning Models for Regression Problems,” *Data Intell.*, vol. 4, no. 3, pp. 620–652, Jul. 2022, doi: 10.1162/dint_a_00155.