



## Perbandingan Kinerja Algoritma Ensemble dan Boosting untuk Prediksi Diabetes pada Data Tidak Seimbang Menggunakan SMOTE

<sup>1,\*</sup>Cicin Hardiyanti P, <sup>2</sup>Dita Danianti

<sup>1,2</sup>Jurusan Informatika, Universitas Alma Ata, Jl. Brawijaya 99, Yogyakarta 55183, Indonesia

**Abstrak** — Sebagai salah satu penyakit kronis yang terus meningkat secara global, diagnosis dini diabetes sangat penting untuk mencegah komplikasi yang lebih serius. Untuk memprediksi risiko diabetes, metode pembelajaran mesin telah banyak digunakan. Namun, masalah utama yang sering muncul adalah ketidakseimbangan distribusi kelas pada dataset medis, yang dapat menghambat model klasifikasi. Penelitian ini bertujuan untuk melakukan analisis komparatif terhadap beberapa algoritma machine learning dalam memprediksi risiko diabetes pada dataset yang tidak seimbang serta mengevaluasi dampak penerapan teknik oversampling *Synthetic Minority Oversampling Technique* (SMOTE). Dataset yang digunakan berasal dari Diabetes Prediction Dataset yang tersedia di Kaggle dengan jumlah sekitar 100.000 data dan distribusi kelas positif sekitar 8,5%. Penelitian ini membandingkan lima algoritma klasifikasi, yaitu *Logistic Regression*, *Random Forest*, *XGBoost*, *LightGBM*, dan *CatBoost*. Evaluasi model dilakukan menggunakan metode *Stratified 5-fold Cross Validation* dengan metrik evaluasi *Precision*, *Recall*, *F1-score*, *ROC-AUC*, dan *PR-AUC*. Hasil penelitian menunjukkan bahwa model berbasis boosting memiliki performa yang lebih unggul dibandingkan model lainnya. *LightGBM* memperoleh nilai *PR-AUC* tertinggi sebesar sekitar 0,89 dengan *ROC-AUC* sebesar 0,9785, diikuti oleh *CatBoost* dan *XGBoost* dengan performa yang sangat mendekati. Menurut analisis confusion matrix, model terbaik mampu mendeteksi kasus diabetes dengan akurasi sekitar 97%, ketepatan sekitar 93,6%, dan recall sekitar 70%. Hasil penelitian ini menunjukkan bahwa algoritma boosting, khususnya *LightGBM*, dapat memprediksi diabetes dengan baik pada dataset medis yang tidak seimbang.

**Kata Kunci:** *LightGBM*; pembelajaran mesin; prediksi diabetes; SMOTE

**Abstract** — As a chronic disease with a continuously increasing global prevalence, early diagnosis of diabetes is essential to prevent severe complications. Machine learning methods have been widely applied to predict diabetes risk; however, a major challenge lies in the class imbalance commonly found in medical datasets, which can degrade classification performance. This study aims to conduct a comparative analysis of several machine learning algorithms for diabetes prediction on an imbalanced dataset and to evaluate the impact of the *Synthetic Minority Oversampling Technique* (SMOTE). The dataset used is the Diabetes Prediction Dataset from Kaggle, consisting of approximately 100,000 records with around 8.5% positive cases. Five classification algorithms were evaluated: *Logistic Regression*, *Random Forest*, *XGBoost*, *LightGBM*, and *CatBoost*. Model performance was assessed using *Stratified 5-fold Cross-Validation* with evaluation metrics including *Precision*, *Recall*, *F1-score*, *ROC-AUC*, and *PR-AUC*. The results indicate that boosting-based models outperform other approaches. *LightGBM* achieved the highest *PR-AUC* ( $\approx 0.89$ ) and *ROC-AUC* (0.9785), followed closely by *CatBoost* and *XGBoost*. Based on confusion matrix analysis, the best model achieved approximately 97% accuracy, 93.6% precision, and 70% recall. These findings demonstrate that boosting algorithms, particularly *LightGBM*, are highly effective for diabetes prediction in imbalanced medical datasets.

**Keywords:** *LightGBM*; machine learning; diabetes prediction; SMOTE

\* Corresponding author :

Cicin Hardiyanti P

Program Studi Informatika, Universitas Alma Ata, Yogyakarta, Indonesia

cicinhardiyanti@almaata.ac.id

### 1. PENDAHULUAN

Diabetes mellitus adalah penyakit kronis yang menjadi salah satu tantangan serius dalam sistem kesehatan global. Penyakit ini ditandai dengan tingginya kadar glukosa dalam darah, yang

disebabkan oleh masalah dalam produksi insulin. Berdasarkan data tahun 2024, jumlah populasi dewasa di Indonesia mencapai sekitar 185.217.400 orang. Dari jumlah tersebut, prevalensi diabetes pada populasi dewasa diperkirakan sebesar 11,3%, yang setara dengan sekitar 20.426.400 kasus diabetes pada orang dewasa di Indonesia [1]. Di Indonesia, jumlah kasus diabetes terus meningkat, menjadikan penyakit ini sebagai salah satu masalah kesehatan yang serius. Hal tersebut berkontribusi pada tingginya angka morbiditas dan mortalitas akibat komplikasi jangka panjang seperti penyakit kardiovaskular, nefropati, neuropati, dan retinopati. Keadaan ini tidak hanya mempengaruhi kesehatan seseorang, tetapi juga dapat menyebabkan beban ekonomi yang signifikan bagi sistem pelayanan kesehatan. Oleh karena itu, melakukan deteksi awal terhadap risiko diabetes sangatlah vital agar penanganan medis bisa dilakukan dengan lebih cepat dan efisien. Dalam beberapa tahun terakhir, kemajuan dalam teknologi analisis data serta kecerdasan buatan telah memainkan peran penting, terutama dalam bidang pembelajaran mesin, telah tercipta peluang baru untuk menciptakan sistem prediksi penyakit yang lebih tepat dan efisien. Algoritma *machine learning* dapat menemukan pola kompleks yang terkait dengan risiko diabetes dengan menggunakan data klinis [2] seperti usia, indeks massa tubuh (BMI), kadar glukosa darah, dan riwayat gaya hidup.

Walaupun demikian, penerapan *machine learning* di sektor medis mengalami sejumlah kendala. Salah satu masalah yang dihadapi adalah ketidakseimbangan kelas dalam dataset kesehatan [3]. Dalam berbagai dataset medis, jumlah pasien yang sebenarnya mengidap penyakit tertentu sering kali jauh lebih rendah dibandingkan dengan mereka yang tidak terkena penyakit tersebut. Ketidakseimbangan ini dapat menyebabkan model klasifikasi lebih condong kepada kelas mayoritas, sehingga mengurangi kemampuannya dalam mendeteksi kasus positif yang sebenarnya. Dalam dunia medis, kesalahan seperti ini sangat berbahaya karena dapat mengakibatkan pasien yang membutuhkan perawatan justru tidak teridentifikasi oleh sistem prediksi. Oleh sebab itu, pengembangan teknik *machine learning* yang dapat mengatasi masalah ketidakseimbangan kelas menjadi isu penelitian yang krusial dalam sektor analitik kesehatan.

Salah satu metode yang umum diterapkan untuk mengatasi ketidakseimbangan data adalah teknik oversampling, terutama menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE) [4]. Metode ini berfungsi dengan menciptakan contoh sintesis tambahan untuk kelas yang kurang terwakili, sehingga distribusi data menjadi lebih merata [5]. Dengan menambahkan representasi dari kelompok minoritas, diharapkan model klasifikasi dapat lebih baik dalam memahami karakteristik pasien yang berisiko. Berbagai studi sebelumnya menunjukkan bahwa penerapan SMOTE dapat meningkatkan kinerja model klasifikasi dalam mendeteksi penyakit pada dataset yang tidak seimbang. Namun, efektivitas teknik ini sering kali tergantung pada jenis algoritma pembelajaran mesin yang diterapkan serta ciri-ciri dataset yang sedang dianalisis.

Beberapa penelitian sebelumnya telah berupaya menerapkan berbagai algoritma *machine learning* untuk memprediksi risiko diabetes. Penelitian yang dilakukan oleh Latuconsina dan Rahardi (2026) [6] membandingkan algoritma *LightGBM* dan *CatBoost* dalam memprediksi diabetes menggunakan dataset klinis berukuran besar. Penelitian tersebut mengindikasikan bahwa kedua algoritma *boosting* modern menunjukkan kinerja yang sangat baik dalam memodelkan hubungan kompleks antar variabel kesehatan, dengan *CatBoost* menunjukkan stabilitas prediksi yang sedikit lebih unggul dibandingkan *LightGBM*. Namun, penelitian tersebut hanya terfokus pada dua algoritma *boosting* tanpa melakukan perbandingan yang lebih komprehensif dengan pendekatan *machine learning* lainnya. Selain itu, meskipun SMOTE

diterapkan untuk mengatasi ketidakseimbangan data, penelitian tersebut belum secara menyeluruh mengeksplorasi dampak teknik ini terhadap berbagai jenis model yang berbeda.

Studi lain oleh Ali et al. (2025) [7] secara khusus mengeksplorasi penerapan algoritma *Random Forest* dalam prediksi diabetes dengan mengintegrasikan teknik penyeimbangan kelas seperti SMOTE dan SMOTE-ENN. Hasil penelitian menunjukkan bahwa penggunaan teknik penyeimbangan data dapat meningkatkan kemampuan model dalam mendeteksi pasien diabetes secara signifikan. Namun, penelitian tersebut hanya terfokus pada satu jenis algoritma, sehingga tidak memberikan gambaran komprehensif mengenai kinerja metode lain seperti boosting atau model linier. Penelitian oleh Pratama dan Utomo (2026) [2] menunjukkan bahwa algoritma *boosting*, seperti *XGBoost* dan *LightGBM*, memiliki kemampuan superior dalam memodelkan hubungan non-linear pada data klinis. Namun, penelitian tersebut tidak secara eksplisit mengkaji teknik penyeimbangan data sintesis, sehingga potensi bias akibat ketidakseimbangan kelas belum dianalisis secara mendalam.

Pendekatan lain juga telah dikembangkan untuk meningkatkan performa prediksi diabetes. Sidiq et al. (2025) [8] mengusulkan penggunaan algoritma *AdaBoost* yang dikombinasikan dengan teknik oversampling SMOTE untuk meningkatkan kemampuan model dalam mendeteksi kelas minoritas. Penelitian tersebut menunjukkan bahwa penerapan SMOTE secara signifikan meningkatkan nilai *recall*, yang berarti lebih banyak kasus diabetes dapat teridentifikasi dengan akurat. Namun, algoritma *AdaBoost* memiliki keterbatasan dalam efisiensi komputasi dibandingkan dengan varian boosting yang lebih kontemporer. Sebaliknya, penelitian yang dilakukan oleh Salam et al. (2025) [9] menawarkan pendekatan *hybrid* yang menggabungkan *clustering* K-Means dan SMOTE sebelum menggunakan *Logistic Regression* untuk klasifikasi. Metode ini memiliki kemampuan untuk meningkatkan kinerja model linier dalam beberapa metrik evaluasi. Namun, metode ini masih memiliki beberapa keterbatasan dalam menangkap pola data yang sangat kompleks.

Studi lain telah membandingkan berbagai algoritma *machine learning* untuk prediksi diabetes. Setyawan dan Wakhidah (2025) [10] membandingkan metode *ensemble* dan *Random Forest*, *Gradient Boosting*, dan *Logistic Regression* untuk memprediksi diabetes. Mereka menemukan bahwa metode *ensemble* cenderung lebih baik daripada model statistik tradisional. Namun demikian, penelitian tersebut belum memasukkan algoritma *boosting* modern seperti *LightGBM* dan *CatBoost* yang saat ini banyak digunakan dalam berbagai penelitian *machine learning*. Selain itu, penelitian yang dilakukan oleh Maulana dan Ernawati (2025) [11] mencadangkan pendekatan *ensemble* yang lebih kompleks yang menggunakan metode *soft voting* yang menggabungkan beberapa algoritma sekaligus. Meskipun metode ini menghasilkan akurasi yang tinggi, kompleksitas model yang tinggi membuatnya tidak sesuai untuk digunakan pada sistem dengan keterbatasan sumber daya komputasi.

Secara keseluruhan, penelitian sebelumnya menunjukkan bahwa algoritma *ensemble* dan *boosting* sangat berpotensi untuk meningkatkan akurasi prediksi pada data medis yang tidak seimbang. Studi oleh Asif et al. (2025) [12] menunjukkan bahwa metode *boosting* secara konsisten menghasilkan hasil yang lebih baik dibandingkan metode tradisional ketika dikombinasikan dengan teknik *oversampling* seperti SMOTE. Studi oleh Ayodele (2023) [13] dan Kavitha dan Kasthuri (2024) [14] juga menunjukkan bahwa penggabungan teknik penyeimbangan data dan algoritma kelompok dapat meningkatkan kemampuan model untuk menangani dataset dengan rasio ketidakseimbangan yang tinggi. Meskipun demikian, sebagian besar penelitian masih menghadapi keterbatasan dalam hal cakupan algoritma yang

dibandingkan, kompleksitas model yang digunakan, atau kurangnya analisis komparatif yang sistematis pada dataset klinis berukuran besar.

Ada beberapa celah penelitian yang dapat diidentifikasi berdasarkan tinjauan pustaka tersebut. Pertama, banyak penelitian hanya membandingkan satu atau dua algoritma, sehingga kurang memberikan gambaran komprehensif kinerja berbagai model *machine learning*. Kedua, dampak teknik penyeimbangan data seperti SMOTE pada berbagai algoritma masih kurang dianalisis. Ketiga, kompleksitas model pada beberapa pendekatan menyulitkan evaluasi yang sistematis. Oleh karena itu, diperlukan penelitian yang dapat mempertimbangkan performa prediksi, stabilitas, dan efisiensi komputasi dalam analisis komparatif.

Penelitian ini bertujuan untuk melakukan analisis komparatif terhadap beberapa algoritma *Machine Learning* memprediksi diabetes, seperti *Logistic Regression*, *Random Forest*, *XGBoost*, *LightGBM*, dan *CatBoost*, berdasarkan masalah tersebut, serta mengevaluasi pengaruh SMOTE dengan membandingkan kinerja sebelum dan sesudah penyeimbangan data. Evaluasi dilakukan menggunakan *precision*, *recall*, *F1-score*, *ROC-AUC*, dan *PR-AUC*. Hasilnya diharapkan memberikan pemahaman komprehensif serta rekomendasi model terbaik untuk deteksi dini diabetes pada data tidak seimbang.

## 2. METODOLOGI PENELITIAN

### 2.1. Dataset dan Preprocessing

Dataset prediksi diabetes yang diperoleh dari platform Kaggle melalui tautan <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset> berisi sekitar 100.000 data pasien yang berhubungan dengan risiko diabetes. Dataset ini berfokus pada diabetes, yang menunjukkan apakah seseorang memiliki diabetes atau tidak. Dataset terdiri dari beberapa fitur penting yang umum digunakan dalam analisis kesehatan, seperti usia (usia), indeks massa tubuh (BMI), jenis kelamin (jenis kelamin), kadar HbA1c (kadar HbA1c), riwayat penyakit jantung (riwayat penyakit jantung), kadar glukosa darah (kadar glukosa darah), dan riwayat merokok. Distribusi kelas pada data ini bersifat tidak seimbang, yaitu sekitar 8,5% data termasuk dalam kelas positif mengidap diabetes dan 91,5% merupakan kelas negatif. Ketidakseimbangan ini menjadi tantangan utama dalam proses klasifikasi karena model yang cenderung bias terhadap kelas mayoritas.

Pada tahap eksplorasi awal data, dilakukan pemeriksaan terhadap kualitas data, termasuk pengecekan *missing values*, tipe data, dan distribusi fitur. Hasil analisis menunjukkan bahwa dataset tidak memiliki *missing values*, sehingga tidak diperlukan proses imputasi. Selanjutnya, data dibagi menjadi label target (y) dan fitur (x). Setiap aspek lainnya digunakan sebagai variabel prediktor, sedangkan variabel target diabetes digunakan sebagai label klasifikasi. Selanjutnya, identifikasi tipe fitur, yang terdiri dari fitur numerik dan kategorikal, dilakukan. Dalam penelitian ini, fitur numerik termasuk usia, hipertensi, penyakit jantung, bmi, tingkat HbA1c, dan tingkat glukosa darah; fitur kategorikal termasuk gender dan riwayat merokok.

Tahap *preprocessing* dilakukan untuk memastikan algoritma *machine learning* dapat menggunakan data secara optimal. Metode *StandardScaler* digunakan untuk normalisasi fitur numerik untuk mengubah distribusi data sehingga rata-rata nol dan standar deviasi satu. Proses ini sangat penting untuk menjaga stabilitas model, terutama pada algoritma yang sensitif terhadap skala data. Sementara itu, fitur kategorikal dikonversi menjadi representasi numerik menggunakan *One-Hot Encoding*, yaitu teknik mengubah setiap kategori menjadi variabel biner untuk menghindari asumsi hubungan ordinal antar kategori.

Dalam penelitian ini, *ColumnTransformer* digunakan untuk secara efektif mengintegrasikan proses transformasi fitur numerik dan kategorikal dalam satu pipeline pemrosesan data. Dengan cara ini, seluruh tahap preprocessing dapat diterapkan secara konsisten pada setiap iterasi pelatihan model.

## 2.2. Penanganan *Class Imbalance*

Ketidakeimbangan kelas menjadi masalah utama pada dataset medis, di mana hanya sekitar 8,5% data merupakan pasien diabetes. Hal ini dapat menyebabkan model bias terhadap kelas mayoritas dan menurunkan kemampuan mendeteksi kasus diabetes. Untuk mengatasinya, penelitian ini menggunakan SMOTE, yaitu teknik *oversampling* yang menghasilkan sampel sintetis dari kelas minoritas berdasarkan tetangga terdekat. Metode ini membuat distribusi data lebih seimbang sehingga model dapat mempelajari karakteristik pasien diabetes dengan lebih baik. Untuk mencegah data *leakage*, SMOTE hanya diterapkan pada data pelatihan dan diintegrasikan dalam pipeline selama proses *cross-validation*. Penelitian ini membandingkan dua skenario, yaitu model tanpa SMOTE dan dengan SMOTE, guna menganalisis pengaruh penyeimbangan data terhadap kinerja model..

## 2.3. Model *Machine Learning*

Penelitian ini menggunakan lima algoritma *machine learning* yaitu *Logistic Regression*, *Random Forest*, *XGBoost*, *LightGBM*, dan *CatBoost*. *Logistic Regression* dijadikan sebagai *baseline* model karena sederhana dan mudah diinterpretasikan dalam memodelkan probabilitas penyakit. *Random Forest* merupakan metode *ensemble* berbasis *bagging* yang menggabungkan banyak pohon keputusan untuk menangani hubungan non-linear dan mengurangi *overfitting*. *XGBoost* adalah algoritma *boosting* yang populer karena mampu meningkatkan akurasi dengan meminimalkan kesalahan secara iteratif. *LightGBM* merupakan varian *gradient boosting* yang lebih efisien dalam penggunaan memori dan komputasi, sehingga cocok untuk dataset besar. Sedangkan *CatBoost* digunakan untuk menangani fitur kategorikal secara efektif serta meningkatkan stabilitas prediksi dan mengurangi *overfitting*. Dalam penelitian ini, seluruh model dikonfigurasi menggunakan parameter dasar (*default parameters*) dengan beberapa penyesuaian ringan untuk memastikan stabilitas pelatihan. *Logistic Regression* menggunakan *max\_iter = 1000* untuk memastikan konvergensi, sementara *Random Forest*, *XGBoost*, dan *LightGBM* menggunakan *n\_estimators = 100*. *CatBoost* dikonfigurasi dengan *iterations = 100*. Penelitian ini tidak melakukan proses *tuning* lanjutan agar fokus penelitian tetap pada analisis perbandingan performa antar algoritma serta dampak teknik penyeimbangan data SMOTE terhadap masing-masing model.

Dengan beragam kombinasi model ini, penelitian dapat mengevaluasi secara komperhensif perbedaan performa antara *Logistic Regression*, *Random Forest*, serta algoritma *boosting* modern (*XGBoost*, *LightGBM*, dan *CatBoost*) dalam konteks prediksi diabetes pada dataset yang tidak seimbang.

## 2.4. Desain Eksperimen dan Evaluasi

Penelitian ini menggunakan metode *Stratified K-Fold Cross Validation* (5-fold) untuk memastikan distribusi kelas tetap seimbang pada setiap lipatan, sehingga evaluasi lebih stabil dan representatif. Eksperimen dilakukan dalam dua skenario, yaitu dengan SMOTE dan tanpa penyeimbangan data SMOTE. Pada setiap fold, model dilatih pada data pelatihan dan diuji pada data pengujian, serta waktu pelatihan dicatat untuk menilai efisiensi komputasi. Evaluasi kinerja menggunakan metrik untuk data tidak seimbang, yaitu *precision*, *recall*, *F1-score*, *ROC-AUC*, dan *PR-AUC*. *Precision* mengukur ketepatan prediksi positif, *recall* menilai kemampuan mendeteksi seluruh kasus positif, dan *F1-score* menyeimbangkan keduanya. *ROC-*

*AUC* digunakan untuk mengukur kemampuan diskriminasi model secara keseluruhan, sedangkan *PR-AUC* lebih sensitif terhadap kelas minoritas. Selain itu, *confusion matrix* digunakan untuk memberikan gambaran distribusi prediksi benar dan salah pada setiap kelas. Penelitian diharapkan dapat memberikan evaluasi yang komprehensif terhadap performa berbagai algoritma *machine learning* dalam memprediksi risiko diabetes pada dataset yang tidak seimbang serta mengidentifikasi model yang paling efektif untuk mendukung sistem deteksi dini berbasis data klinis.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Perbandingan Kinerja Model Machine Learning

Penelitian ini mengevaluasi lima algoritma *machine learning*, yaitu *Logistic Regression*, *Random Forest*, *XGBoost*, *LightGBM*, dan *CatBoost* dalam memprediksi risiko diabetes pada dataset yang tidak seimbang. Evaluasi dilakukan menggunakan skema *5-fold stratified cross-validation* dengan dua skenario eksperimen, yaitu model tanpa penyeimbangan data dan model dengan penerapan teknik *oversampling* SMOTE. Hasil evaluasi menunjukkan bahwa performa model berbeda secara signifikan tergantung pada jenis algoritma yang digunakan.

Tabel 1 menunjukkan hasil evaluasi performa model *machine learning*. Evaluasi dilakukan menggunakan metrik *Precision*, *Recall*, *F1-score*, *ROC-AUC*, dan *PR-AUC*, dan metode *cross-validation* lima kali lipatan. Secara umum, model berbasis *ensemble* dan *boosting* menunjukkan performa yang lebih tinggi dibandingkan model linear.

Tabel 1. Hasil Evaluasi Model Machine Learning (Mean  $\pm$  Std)

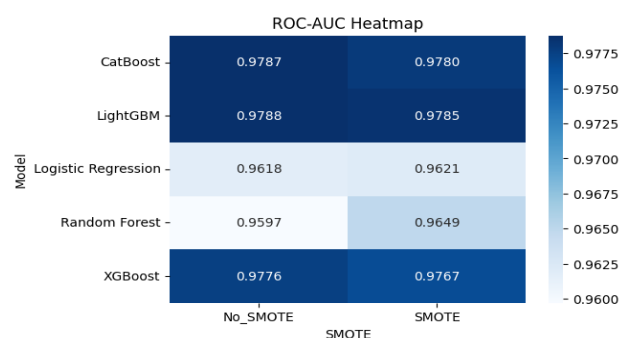
Model	SMOTE	Precision	Recall	F1-Score	ROC-AUC	PR-AUC	Waktu Training (detik)
<i>CatBoost</i>	No	0.9639 $\pm$	0.6902 $\pm$	0.8044 $\pm$	0.9787 $\pm$	0.8826 $\pm$	1.5505
	SMOTE	0.0092	0.0083	0.0080	0.0012	0.0041	
<i>CatBoost</i>	SMOTE	0.9326 $\pm$	0.7045 $\pm$	0.8026 $\pm$	0.9780 $\pm$	0.8806 $\pm$	3.7408
<i>LightGBM</i>	No	0.9762 $\pm$	0.6853 $\pm$	0.8052 $\pm$	0.9788 $\pm$	0.8839 $\pm$	1.3119
	SMOTE	0.0082	0.0106	0.0053	0.0012	0.0039	
<i>LightGBM</i>	SMOTE	0.9361 $\pm$	0.7016 $\pm$	0.8020 $\pm$	0.9785 $\pm$	0.8827 $\pm$	2.3900
<i>Logistic Regression</i>	No	0.8677 $\pm$	0.6292 $\pm$	0.7293 $\pm$	0.9618 $\pm$	0.8173 $\pm$	1.5360
	SMOTE	0.0149	0.0128	0.0094	0.0015	0.0068	
<i>Logistic Regression</i>	SMOTE	0.4245 $\pm$	0.8808 $\pm$	0.5729 $\pm$	0.9621 $\pm$	0.8166 $\pm$	1.9021
<i>Random Forest</i>	No	0.9438 $\pm$	0.6858 $\pm$	0.7943 $\pm$	0.9597 $\pm$	0.8535 $\pm$	6.6970
	SMOTE	0.0049	0.0080	0.0055	0.0027	0.0036	
<i>Random Forest</i>	SMOTE	0.7563 $\pm$	0.7502 $\pm$	0.7532 $\pm$	0.9649 $\pm$	0.8550 $\pm$	17.3684
<i>XGBoost</i>	No	0.9555 $\pm$	0.6931 $\pm$	0.8033 $\pm$	0.9776 $\pm$	0.8794 $\pm$	0.7708
	SMOTE	0.0053	0.0085	0.0047	0.0015	0.0047	
<i>XGBoost</i>	SMOTE	0.9133 $\pm$	0.7059 $\pm$	0.7963 $\pm$	0.9767 $\pm$	0.8756 $\pm$	2.3427

Secara umum, algoritma berbasis *ensemble* dan *boosting* bekerja lebih baik daripada model linier konvensional. Jika dibandingkan dengan algoritma lainnya, model regresi logistik memiliki performa terendah, terutama pada skor F1 dan PR-AUC. Ini menunjukkan bahwa model linier tidak dapat menangkap hubungan non-linear antara variabel klinis seperti kadar glukosa darah, indeks massa tubuh, dan riwayat penyakit yang memiliki risiko diabetes. Sebaliknya, algoritma *ensemble* seperti *Random Forest* serta metode *boosting* seperti *XGBoost*, *LightGBM*, dan *CatBoost* mampu memodelkan interaksi kompleks antar fitur dengan lebih efektif.

Secara keseluruhan, algoritma *boosting* kontemporer, terutama *LightGBM* dan *CatBoost*, memiliki nilai ROC-AUC dan PR-AUC yang lebih besar daripada model lainnya. Hal ini menunjukkan bahwa metode *gradient boosting* sangat efektif untuk mengidentifikasi pola kompleks dalam data klinis yang sangat besar. Studi sebelumnya menunjukkan bahwa algoritma *boosting* seringkali memiliki hasil terbaik dalam berbagai tugas klasifikasi medis karena kemampuan mereka untuk mengoptimalkan kesalahan prediksi secara iteratif.

### 3.2. Analisis ROC-AUC

Untuk menganalisis kemampuan model dalam membedakan kelas positif dan negatif, Gambar 1 menyajikan visualisasi heatmap ROC-AUC untuk setiap algoritma, baik sebelum maupun setelah penerapan SMOTE. Hasil eksperimen menunjukkan bahwa algoritma *LightGBM* dan *CatBoost* memiliki nilai ROC-AUC tertinggi dari semua model, berdasarkan ROC-AUC, yang mengukur kemampuan model untuk membedakan antara pasien diabetes kelas positif dan non-pasien diabetes kelas negatif. Ini menunjukkan bahwa kedua model memiliki kemampuan yang sangat baik untuk mengidentifikasi pasien yang berisiko terkena diabetes.

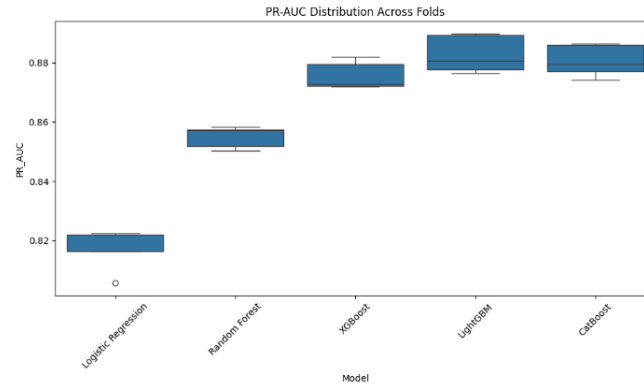


Gambar 1. Heatmap ROC-AUC Model dengan dan tanpa SMOTE

Baik sebelum maupun setelah SMOTE, visualisasi *heatmap* ROC-AUC menunjukkan bahwa algoritma *boosting* relatif stabil, yang menunjukkan bahwa model *boosting* sudah cukup kuat terhadap ketidakseimbangan kelas pada dataset ini. Sebaliknya, *Random Forest* menunjukkan peningkatan nilai ROC-AUC setelah SMOTE, yang menunjukkan bahwa metode *bagging* lebih peka terhadap distribusi data. Mekanisme pembelajaran bertahap yang dimiliki algoritma tersebut menjelaskan nilai ROC-AUC yang tinggi pada model *boosting*. Pada setiap iterasi, model *boosting* berusaha mengurangi kesalahan yang dibuat oleh model sebelumnya, yang menghasilkan model yang lebih akurat dalam membedakan dua kelas. Metode ini sangat efektif pada data klinis di mana hubungan non-linear antara variabel prediktor dan target ditemukan.

### 3.3. Analisis PR-AUC pada Data Tidak Seimbang

Walaupun ROC-AUC adalah metrik evaluasi yang umum, pada dataset dengan ketidakseimbangan kelas yang signifikan, metrik ini sering kali kurang sensitif dalam menilai kemampuan model untuk mendeteksi kelas minoritas. Oleh karena itu, penelitian ini juga mengadopsi PR-AUC sebagai metrik utama untuk mengevaluasi kinerja model.

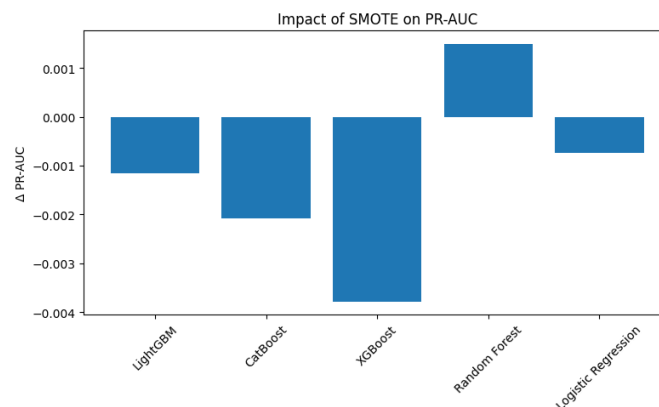


Gambar 2. Distribusi PR-AUC pada Setiap Model

Secara keseluruhan *LightGBM* memiliki kinerja terbaik berdasarkan distribusi nilai PR-AUC pada setiap fold, kemudian diikuti oleh algoritma *CatBoost* dan *XGBoost*. Algoritma *Random Forest* memiliki kinerja yang lebih rendah dibandingkan algoritma *boosting*, tetapi masih lebih unggul daripada Algoritma *Logistic Regression*. Di sisi lain, algoritma *Logistic Regression* menunjukkan nilai PR-AUC yang paling rendah, yang menunjukkan keterbatasannya dalam mendeteksi pasien diabetes secara akurat. Visualisasi boxplot PR-AUC pada Gambar 2, menunjukkan bahwa model *boosting* memiliki distribusi performa yang relatif stabil antar fold, dengan variasi nilai yang kecil. Hal ini menunjukkan bahwa algoritma tersebut tidak hanya memiliki kinerja yang tinggi tetapi juga konsisten dalam berbagai subset data. Untuk aplikasi medis, stabilitas model sangat penting karena model yang stabil lebih dapat diandalkan dalam praktik klinis

### 3.4. Dampak Penerapan SMOTE

Untuk memahami pengaruh teknik penyeimbangan data terhadap kinerja model, Gambar 3 menyajikan perbandingan nilai PR-AUC sebelum dan sesudah penerapan SMOTE pada setiap algoritma. Secara umum, penerapan SMOTE meningkatkan nilai *recall* pada sebagian besar model, yang berarti model menjadi lebih mampu mendeteksi pasien diabetes. Ini sangat penting untuk diagnosis medis karena kesalahan *false negative* dapat menyebabkan pasien yang sebenarnya sakit tidak terlihat. Namun demikian, peningkatan *recall* seringkali diikuti oleh penurunan ketepatan. Ini karena dengan bertambahnya jumlah sampel sintetis pada kelas minoritas, model menjadi lebih agresif dalam memprediksi kelas positif, meningkatkan kemungkinan kesalahan prediksi pada kelas negatif.



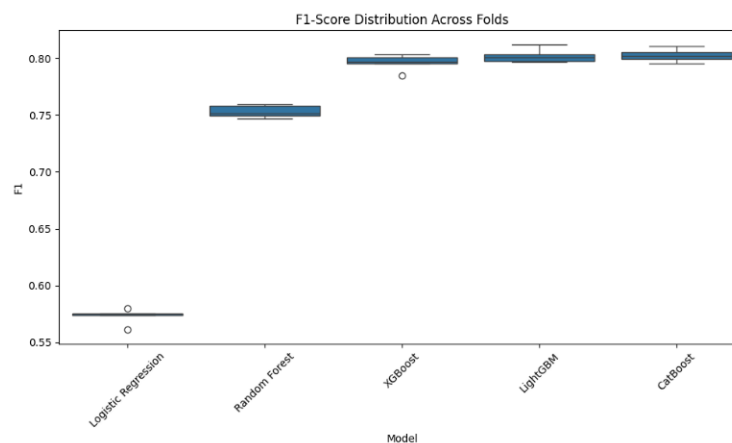
Gambar 3. Dampak Penerapan SMOTE terhadap Nilai PR-AUC

Hasil eksperimen menunjukkan bahwa SMOTE tidak selalu memiliki dampak positif terhadap PR-AUC. Dengan beberapa algoritma *boosting* seperti *XGBoost*, *LightGBM*, dan *CatBoost*,

penggunaan SMOTE justru menghasilkan penurunan nilai PR-AUC yang sedikit. Ini menunjukkan bahwa algoritma *boosting* modern sebenarnya sudah cukup efektif dalam menangani ketidakseimbangan kelas tanpa menggunakan teknik *oversampling* tambahan. Sebaliknya, *Random Forest* lebih baik setelah SMOTE digunakan, menunjukkan bahwa metode *bagging* memiliki distribusi data yang lebih seimbang. Temuan ini menunjukkan bahwa jenis algoritma yang digunakan sangat memengaruhi seberapa efektif metode penyeimbangan data. Oleh karena itu, penggunaan SMOTE tidak selalu menghasilkan peningkatan kinerja pada semua model, dan setiap penggunaan harus dievaluasi secara empiris.

### 3.5. Stabilitas Model Berdasarkan F1-Score

Stabilitas performa model pada berbagai *fold cross-validation* dianalisis menggunakan distribusi *F1-score* yang ditampilkan pada Gambar 4.



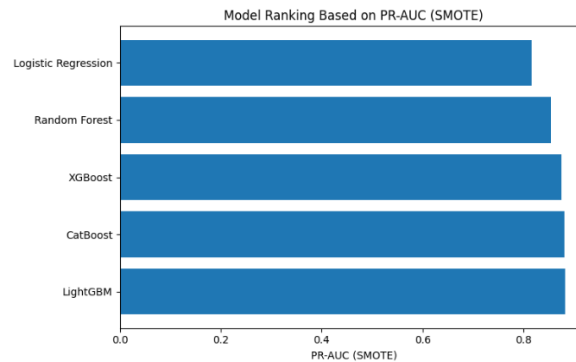
Gambar 4. Distribusi F1-Score pada Setiap model

Hasil visualisasi menunjukkan bahwa algoritma *boosting* memiliki distribusi *F1-score* yang lebih baik dan lebih stabil dibandingkan model lainnya. Nilai *F1-score* *LightGBM* dan *CatBoost* relatif konsisten dengan variasi yang kecil antar fold, menunjukkan bahwa model tersebut mampu menghasilkan prediksi yang stabil meskipun dilatih pada berbagai subset data. Sebaliknya, *Logistic Regression* menunjukkan nilai *F1-score* yang jauh lebih rendah serta variasi yang lebih besar, yang menandakan bahwa model linier kurang mampu menangkap kompleksitas data klinis.

Dalam penerapan *machine learning* di bidang kesehatan, stabilitas performa sangat penting karena model yang tidak stabil dapat membuat prediksi yang berbeda-beda tergantung pada subset data yang digunakan. Akibatnya, model *boosting* yang memiliki stabilitas tinggi disarankan untuk sistem prediksi medis.

### 3.6. Ranking Model Berdasarkan PR-AUC

Perbandingan performa beberapa algoritma *machine learning* berdasarkan nilai PR-AUC setelah penerapan SMOTE ditampilkan pada Gambar 5. PR-AUC digunakan sebagai metrik utama karena lebih representatif dalam mengevaluasi model pada dataset yang tidak seimbang. Hasil penelitian menunjukkan bahwa *LightGBM* memiliki nilai PR-AUC tertinggi sekitar 0.89, diikuti oleh *CatBoost* dan *XGBoost* dengan nilai yang hampir sama. Sementara itu, *Random Forest* menunjukkan kinerja yang cukup baik dengan nilai PR-AUC sekitar 0.86. *Logistic Regression* memiliki nilai PR-AUC paling rendah, sekitar 0.82, menunjukkan bahwa model linear memiliki keterbatasan dalam menangkap hubungan kompleks antar fitur.

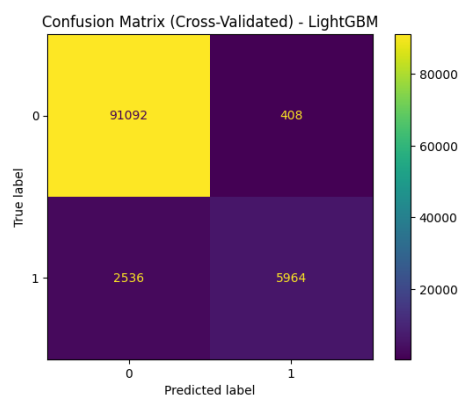


Gambar 5. Ranking Model Berdasarkan PR-AUC

*LightGBM* dipilih sebagai model terbaik untuk tahap evaluasi lanjutan karena dominasi model *boosting* pada peringkat teratas menunjukkan bahwa pendekatan *ensemble* berbasis *boosting* lebih efektif dalam memodelkan hubungan non-linear dalam dataset medis.

### 3.7. Analisis Confusion Matrix Model Terbaik

Evaluasi lebih lanjut terhadap model terbaik dilakukan menggunakan *confusion matrix* yang ditampilkan pada Gambar. 6.



Gambar 6. Confusion Matrix Model Terbaik

Hasil ini menunjukkan bahwa model sangat baik dalam menemukan kelas negatif, dengan nilai *specificity* sekitar 99.55% dan nilai *precision* sebesar 93.6%, yang menunjukkan bahwa sebagian besar prediksi positif yang dihasilkan oleh model benar. Namun demikian, dari 2.536 kasus false negatif, nilai *recall* yang diperoleh sekitar 70.16% menunjukkan bahwa model masih mendeteksi beberapa kasus positif yang tidak berhasil. Kondisi ini menunjukkan bahwa model mungkin lebih konservatif dalam memprediksi kelas positif.

Namun, dengan nilai *F1-score* sebesar 0,80, model *LightGBM* menunjukkan keseimbangan yang cukup baik antara *precision* dan *recall*. Dengan tingkat akurasi keseluruhan mencapai sekitar 97%, model *LightGBM* dapat dianggap memiliki performa yang sangat baik dalam memprediksi risiko diabetes pada dataset yang tidak seimbang.

### 3.8. Analisis Efisiensi Komputasi

Penelitian ini menganalisis performa prediksi dan efisiensi komputasi algoritma dengan mengukur waktu pelatihan model. Hasil eksperimen menunjukkan bahwa model *XGBoost* memiliki waktu pelatihan tercepat, diikuti oleh *LightGBM* dan *Logistic Regression*. Sebaliknya, *Random Forest* membutuhkan waktu yang paling lama untuk dilatih karena harus membangun

banyak pohon keputusan secara mandiri. *LightGBM* mencapai keseimbangan yang baik antara akurasi dan efisiensi komputasi. Algoritma ini dapat membuat prediksi yang sangat baik dengan waktu pelatihan yang cepat. Keunggulan ini dihasilkan dari pendekatan pembelajaran berbasis histogram *LightGBM* yang mempercepat proses pembentukan pohon keputusan. Model seperti *LightGBM* dan *XGBoost*, yang menggabungkan efisiensi komputasi yang tinggi dengan waktu pelatihan yang efisien, menjadi pilihan yang sangat menarik untuk aplikasi prediksi medis dalam konteks implementasi praktis, terutama ketika model harus digunakan pada sistem dengan sumber daya terbatas atau pada aplikasi yang membutuhkan prediksi secara *real-time*.

### 3.9. Diskusi Temuan Penelitian

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma *boosting* seperti *LightGBM*, *CatBoost*, dan *XGBoost* sangat baik untuk memprediksi risiko diabetes pada dataset klinis yang tidak seimbang karena mereka memiliki performa prediksi yang tinggi dan stabilitas model yang baik. Studi ini mendukung temuan penelitian sebelumnya yang menunjukkan bahwa metode *ensemble* berbasis *boosting* lebih baik dalam memodelkan hubungan yang kompleks antara variabel kesehatan. Selain itu, penelitian ini menemukan bahwa teknik *oversampling* seperti SMOTE tidak selalu meningkatkan kinerja model, terutama pada algoritma *boosting* yang sudah memiliki mekanisme internal untuk menangani ketidakseimbangan data. Oleh karena itu, memilih algoritma *machine learning* yang tepat sangat penting untuk membangun sistem prediksi diabetes yang akurat dan dapat diandalkan. Metode berbasis *boosting* yang dirancang dengan baik dapat mendukung deteksi diabetes dini berdasarkan data klinis.

## 4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja lima algoritma *machine learning*, yaitu *Logistic Regression*, *Random Forest*, *XGBoost*, *LightGBM*, dan *CatBoost*, dalam memprediksi risiko diabetes pada dataset klinis yang tidak seimbang, serta mengevaluasi pengaruh teknik *oversampling* SMOTE terhadap performa model. Eksperimen dilakukan menggunakan *stratified 5-fold cross-validation* dengan evaluasi berbasis metrik *ROC-AUC*, *PR-AUC*, *F1-score*, serta waktu pelatihan. Hasil penelitian menunjukkan bahwa algoritma berbasis *gradient boosting*, khususnya *LightGBM* dan *CatBoost*, secara konsisten memberikan performa terbaik dengan nilai *ROC-AUC* dan *PR-AUC* yang tinggi serta stabil di setiap *fold*, yang mengindikasikan kemampuan unggul dalam menangkap pola *non-linear* pada data klinis. Penerapan SMOTE terbukti meningkatkan *recall* pada kelas minoritas, namun tidak selalu meningkatkan performa secara keseluruhan, bahkan cenderung menurunkan nilai *PR-AUC* pada model *boosting*, sehingga menunjukkan bahwa algoritma tersebut telah cukup robust terhadap ketidakseimbangan data. Dari sisi efisiensi komputasi, *LightGBM* menunjukkan keseimbangan terbaik antara akurasi dan waktu pelatihan, menjadikannya model yang paling potensial untuk implementasi skala besar. Dengan demikian, penelitian ini menegaskan bahwa pemilihan algoritma memiliki pengaruh yang lebih signifikan dibandingkan penggunaan teknik penyeimbangan data secara umum, serta memberikan kontribusi dalam bentuk analisis komprehensif terhadap efektivitas model *ensemble* dan *boosting* pada data medis tidak seimbang. Untuk pengembangan selanjutnya, penelitian dapat diarahkan pada optimasi hyperparameter, eksplorasi metode penyeimbangan data yang lebih kompleks, penerapan teknik *explainable AI (XAI)*, serta pengujian pada dataset yang lebih beragam guna meningkatkan generalisasi model.

## DAFTAR PUSTAKA

- [1] “Indonesia - International Diabetes Federation,” *International Diabetes Federation*, 2024. <https://idf.org/our-network/regions-and-members/western-pacific/members/indonesia/>.
- [2] N. A. Pratama and D. W. Utomo, “Deteksi Diabetes Mellitus dengan Menggunakan Teknik Ensemble XGBoost dan LightGBM,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 11, no. 1, pp. 1–12, Jan. 2026, doi: 10.14421/jiska.4908.
- [3] C. Hardiyanti P, “Optimizing breast cancer classification using SMOTE, Boruta, and XGBoost,” *Sci. Inf. Technol. Lett.*, vol. 6, no. 1, pp. 16–33, May 2025, doi: 10.31763/sitech.v6i1.2109.
- [4] F. S. Pratiwi, M. A. Barata, and A. D. Ardianti, “Implementasi Metode Smote Dan Random Over-Sampling Pada Algoritma Machine Learning Untuk Prediksi Customer Churn Di Sektor Perbankan,” *J. Sist. Inf. dan Inform.*, vol. 8, no. 1, pp. 87–98, Jan. 2025, doi: 10.47080/simika.v8i1.3678.
- [5] Amanda Prawita Ningrum, Sri Winarno, and Vincentius Praskatama, “Klasifikasi Kualitas Biji Kedelai Menggunakan Transfer Learning Convolutional Neural Network Dan SMOTE,” *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 2, pp. 155–164, Dec. 2024, doi: 10.52158/jacost.v5i2.1002.
- [6] M. S. Latuconsina and M. Rahardi, “Comparison of LightGBM and CatBoost Algorithms for Diabetes Prediction Based on Clinical Data,” *J. Appl. Informatics Comput.*, vol. 10, no. 1, pp. 1058–1065, 2026.
- [7] H. Ali, A. Rahim, A. Ahmed, U. Tanveer, W. S. Khan, and S. Bibi, “Machine Learning for Diabetes Prediction Using Random Forest : A Comprehensive Analysis with Class Balancing Techniques,” *Spectr. Eng. Sci.*, vol. 3138, pp. 1129–1142, 2025, [Online]. Available: <https://zenodo.org/records/18030488>.
- [8] S. Sidiq, P. Korespondensi, and N. Shobi Mabru, “Pengembangan model prediksi risiko diabetes menggunakan pendekatan AdaBoost dan Teknik Oversampling SMOTE,” *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 4, pp. 13–23, 2025, [Online]. Available: <https://doi.org/10.58602/jima-ilkom.v4i1.41>.
- [9] A. Salam, L. Azhari, R. S. Septarini, and N. Heriyani, “Pendekatan Hybrid K-Means SMOTE dan Logistic Regression Untuk Deteksi Dini Diabetes Mellitus Pada Imbalanced Data,” *Bull. Comput. Sci. Res.*, vol. 5, no. 3, pp. 219–227, Apr. 2025, doi: 10.47065/bulletincsr.v5i3.502.
- [10] N. H. Setyawan and N. Wakhidah, “Analisis Perbandingan Metode Logistic Regression, Random Forest, Gradient Boosting Untuk Prediksi Diabetes,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 10, no. 1, pp. 150–162, Jan. 2025, doi: 10.29100/jupi.v10i1.5743.
- [11] S. Ernawati and I. Maulana, “Meningkatkan Klasifikasi Penyakit Diabetes Menggunakan Metode Ensemble Softvoting Dengan SMOTE-ENN dan Optimasi Bayesian,” *Evolusi J. Sains dan Manaj.*, vol. 13, no. 1, pp. 71–86, 2025, doi: 10.31294/evolusi.v13i1.8267.
- [12] R. Asif, D. Upadhyay, M. Zaman, and S. Sampalli, “Enhancing diabetes risk prediction: A comparative evaluation of bagging, boosting, and ensemble classifiers with SMOTE oversampling,” *Informatics Med. Unlocked*, vol. 57, no. February, p. 101661, 2025, doi: 10.1016/j.imu.2025.101661.
- [13] Adefemi Ayodele, “A comparative study of ensemble learning techniques for imbalanced classification problems,” *World J. Adv. Res. Rev.*, vol. 19, no. 2, pp. 1633–1643, Aug. 2023, doi: 10.30574/wjarr.2023.19.1.1202.
- [14] M. Kavitha, “Comparative Analysis of SMOTE Techniques and Machine Learning Models for Imbalanced Medical Datasets,” *IEEE Conf. Proc.*, no. June, pp. 1–9, 2024, [Online]. Available: <https://www.researchgate.net/publication/381805587>.