CyberSecurity dan Forensik Digital e-ISSN: 2615-8442
Vol. 9, No. 1, Mei 2026, hlm. 35-42

Perbandingan Kinerja dan Efisiensi Full Fine-Tuning dan LoRA untuk Deteksi Email Phishing pada Model Transformer

David Suharjanto[1], Bambang Sugiantoro[2]

[1,2]Informatika, Universitas Islam Negeri Sunan Kalijaga, Yogyakarta, Indonesia
Email: [1]24206052003@student.uin-suka.ac.id, [2]bambang.sugiantoro@uin-suka.ac.id

Abstrak

Serangan phishing merupakan ancaman keamanan siber yang terus berkembang dan semakin adaptif, sehingga pendekatan deteksi berbasis machine learning klasik menjadi kurang efektif. Model deep learning berbasis Transformer telah terbukti unggul dalam memahami semantik teks, namun penerapannya melalui skema full fine-tuning memerlukan sumber daya komputasi yang tinggi. Keterbatasan ini mendorong kebutuhan akan metode yang lebih efisien tanpa mengorbankan kinerja deteksi. Penelitian ini mengevaluasi efektivitas Parameter-Efficient Fine-Tuning (PEFT) menggunakan metode Low-Rank Adaptation (LoRA) untuk deteksi email phishing. Eksperimen dilakukan pada dataset publik yang terdiri dari 18.644 email dengan membandingkan lima arsitektur Transformer encoder, yaitu RoBERTa, BERT, ELECTRA, DeBERTa, dan DistilBERT. Evaluasi berfokus pada analisis trade-off antara kinerja klasifikasi, yang diukur menggunakan akurasi, presisi, recall, dan F1-score, serta efisiensi komputasi berdasarkan penggunaan VRAM dan waktu pelatihan. Hasil eksperimen menunjukkan bahwa LoRA mampu mempertahankan performa deteksi yang kompetitif dengan penurunan performa rata-rata kurang dari 1% dibandingkan full fine-tuning. BERT dengan full fine-tuning mencapai F1-score tertinggi sebesar 98,23%. Menariknya, pada DeBERTa, penerapan LoRA justru menghasilkan sedikit peningkatan performa hingga 98,13% dibandingkan versi full fine-tuning sebesar 98,02%, yang mengindikasikan efek regularisasi yang cukup efektif. Dari sisi efisiensi, LoRA mampu menurunkan konsumsi memori pada seluruh model, dengan penghematan tertinggi pada DistilBERT hingga 40%. Berdasarkan temuan ini, penggunaan full fine-tuning direkomendasikan jika prioritas utama adalah akurasi maksimal, sedangkan LoRA lebih sesuai untuk efisiensi memori.

Kata kunci: deteksi phishing, Transformer, Low-Rank Adaptation (LoRA), efisiensi komputasi, keamanan siber

Comparison of Performance and Efficiency between Full Fine-Tuning and LoRA for Phishing Email Detection using Transformer Models

Abstract

Phishing attacks represent an evolving and increasingly adaptive cybersecurity threat, rendering classical machine learning-based detection approaches less effective. Transformer-based deep learning models have demonstrated superiority in comprehending textual semantics, but their training via full fine-tuning schemes demands substantial computational resources. These limitations necessitate more efficient methods that do not compromise detection performance. This study evaluates the effectiveness of Parameter-Efficient Fine-Tuning (PEFT) using the Low-Rank Adaptation (LoRA) method for phishing email detection. Experiments were conducted on a public dataset comprising 18,644 emails, comparing five Transformer encoder architectures: RoBERTa, BERT, ELECTRA, DeBERTa and DistilBERT. The evaluation focuses on analyzing the trade-off between classification performance, measured using accuracy, precision, recall, & F1-score, and computational efficiency based on VRAM usage and training time. Experimental results demonstrate that LoRA is capable of maintaining competitive detection performance with an average performance degradation of less than 1% compared to full fine-tuning. The BERT model with full fine-tuning achieved the highest F1-score of 98.23%. Notably, in DeBERTa, the application of LoRA yielded a slight performance improvement to 98.13% compared to the full fine-tuning version (98.02%), indicating an effective regularization effect. In terms of efficiency, LoRA reduced memory consumption across all models, with the highest saving observed in DistilBERT, reaching up to 40%. Based on these findings, the use of full fine-tuning is recommended if the primary priority is maximum accuracy, whereas LoRA is more suitable for memory efficiency.

***Keywords**: phishing detection, Transformer, Low-Rank Adaptation (LoRA), computational efficiency, cybersecurity*

1. PENDAHULUAN

Perkembangan teknologi informasi yang semakin pesat telah mendorong penggunaan email sebagai komunikasi digital, baik untuk keperluan personal maupun organisasi (Safi and Singh, 2023). Seiring dengan hal tersebut, risiko terhadap keamanan siber juga semakin meningkat, salah satunya adalah ancaman serangan phishing melalui layanan email (Febriyani et al., 2023). Dengan menyamar sebagai pihak yang tepercaya, serangan rekayasa sosial ini bertujuan untuk menjebak pengguna agar memberikan informasi sensitif seperti data pribadi (Nakamura and Dobashi, 2019). Dalam perkembangannya, serangan phishing tidak hanya meningkat dari segi jumlah, tetapi juga menjadi semakin kompleks dan adaptif, sehingga menjadikannya salah satu ancaman utama dalam bidang keamanan siber saat ini (Osamor et al., 2025).

Berbagai pendekatan berbasis machine learning telah banyak dikembangkan untuk mendeteksi phishing email (Wosah and Win, 2021). Patil, Rane and Bhalekar (2017) menerapkan algoritma Support Vector Machine (SVM) yang diintegrasikan dengan pendekatan MapReduce untuk meningkatkan efisiensi dalam deteksi email phishing. Yahya et al. (2021) melakukan evaluasi terhadap beberapa algoritma klasifikasi, yaitu Decision Tree, K-Nearest Neighbor (KNN), dan Random Forest, untuk mendeteksi phishing email. Hasil evaluasi menunjukkan bahwa model KNN memberikan performa terbaik dengan akurasi mencapai 97,6%. Moorthy and Pabitha (2020) mengusulkan pendekatan Sine Cosine Algorithm dengan K-Nearest Neighbor (SCAK-NN) yang mampu mencapai akurasi sebesar 97,18%, menggungguli metode Decision Tree (95,88%) dan Naive Bayes (92,98%).

Meskipun berbagai metode tersebut mampu mencapai tingkat akurasi yang relatif tinggi, pendekatan machine learning klasik masih memiliki keterbatasan dari sisi representasi fitur. Model-model tersebut sangat bergantung pada proses rekayasa fitur manual, yang sering kali tidak mampu menangkap representasi semantik isi email secara mendalam (Meléndez, Ptaszynski and Masui, 2024). Akibatnya, metode deteksi phishing tradisional rentan mengalami penurunan kinerja saat menghadapi email phishing dengan pola baru yang tidak tercakup dalam data pelatihan.

Seiring dengan kemajuan di bidang Natural Language Processing (NLP), model deep learning berbasis Transformer telah menunjukkan performa yang unggul dalam berbagai tugas pemrosesan teks, salah satunya pada klasifikasi email phishing (Otieno, Siami Namin and Jones, 2023). Arsitektur Transformer menggunakan mekanisme self-attention untuk menangkap hubungan kontekstual antar kata secara efektif, sehingga menghasilkan representasi semantik yang kuat (Vaswani et al., 2017). Model Transformer pre-trained seperti BERT telah banyak digunakan dan menunjukkan kinerja yang kompetitif dalam deteksi phishing berbasis teks (Kyaw, Gutierrez and Ghobakhlou, 2024).

Meskipun menunjukkan kinerja yang menjanjikan, penerapan model Transformer melalui metode full fine-tuning menghadapi tantangan berupa keterbatasan sumber daya komputasi, karena membutuhkan kapasitas GPU besar dan waktu pelatihan yang lama, sehingga kurang optimal untuk diterapkan pada lingkungan dengan sumber daya terbatas (Blake, 2025). Untuk mengatasi permasalahan tersebut, pendekatan Parameter-Efficient Fine-Tuning (PEFT) telah diperkenalkan sebagai solusi alternatif yang bertujuan untuk mengurangi biaya komputasi tanpa mengorbankan performa model secara signifikan (Parthasarathy et al., 2024). Di antara berbagai metode PEFT yang dikembangkan, Low-Rank Adaptation (LoRA) merupakan salah satu pendekatan yang banyak diadopsi. Metode ini bekerja dengan menambahkan matriks berperingkat rendah pada lapisan tertentu dalam arsitektur Transformer, sehingga hanya sebagian kecil parameter tambahan yang dilatih, sementara parameter utama model tetap dibekukan (Hu et al., 2021). Dengan demikian, LoRA memungkinkan proses fine-tuning Transformer dilakukan secara lebih efisien pada lingkungan dengan sumber daya komputasi terbatas.

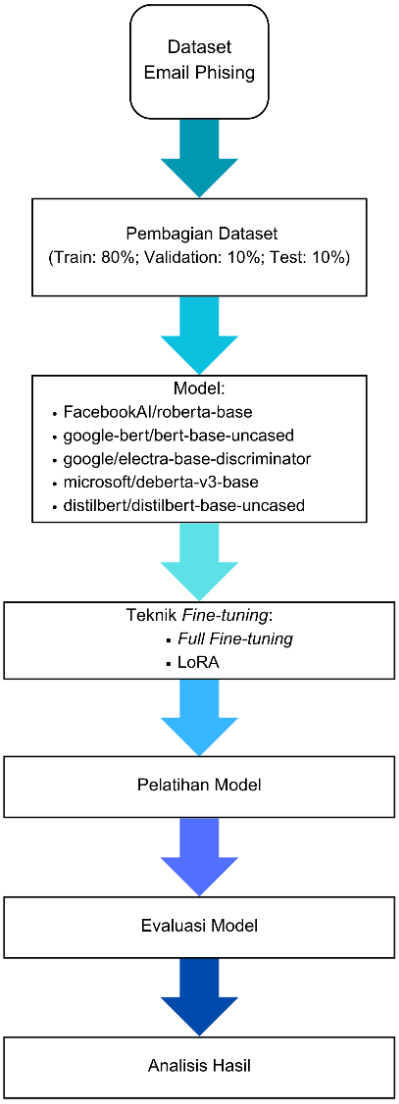
Berdasarkan uraian latar belakang, penelitian ini bertujuan untuk melakukan analisis komparatif terhadap performa dan efisiensi metode full fine-tuning dan LoRA pada tugas deteksi email phishing. Secara spesifik, kontribusi utama dari penelitian ini adalah sebagai berikut:

1. Menyajikan evaluasi komparatif yang komprehensif pada lima arsitektur Transformer encoder yang berbeda untuk menentukan trade-off terbaik antara akurasi deteksi dan biaya komputasi.
2. Memberikan bukti empiris terkait efektivitas LoRA sebagai regularisasi pada model DeBERTa, yang terbukti mampu meningkatkan generalisasi model dibandingkan metode full fine-tuning.
3. Memberikan rekomendasi praktis pemilihan metode fine-tuning berdasarkan ketersediaan sumber daya komputasi.

2. METODE PENELITIAN

Bagian ini menyajikan penjelasan mengenai desain eksperimen, dataset, arsitektur model, konfigurasi pelatihan, metrik evaluasi, dan infrastruktur pelatihan yang digunakan. Alur kerja penelitian secara keseluruhan ditunjukkan pada Gambar 1. Penelitian ini diawali dengan persiapan dataset yang berisi data Phishing Email dan Safe

Email. Dataset kemudian dibagi menjadi tiga subset, yaitu 80% train, 10% validation, dan 10% test. Selanjutnya, lima model Transformer pre-trained digunakan untuk klasifikasi email phishing. Proses pelatihan dilakukan selama 10 epoch dengan menerapkan dua pendekatan fine-tuning, yaitu full fine-tuning dan Low-Rank Adaptation (LoRA). Setelah pelatihan, kinerja model dievaluasi menggunakan metrik standar klasifikasi, meliputi akurasi, presisi, recall, F1-score, dan confusion matrix. Hasil evaluasi tersebut kemudian dianalisis secara komparatif sebagai dasar penarikan kesimpulan penelitian.



Gambar 1. Diagram Alir Penelitian

2.1 Dataset

Dataset yang digunakan dalam penelitian ini merupakan dataset deteksi email phishing berbahasa Inggris yang dikembangkan oleh Subhadeep Chakraborty[1]. Dataset ini berisi 18,644 email lalu dibagi menjadi 80% data pelatihan, 10% data validasi, dan 10% data pengujian. Detail dataset disajikan pada Tabel 1.

Tabel 1. Dataset Deteksi Email Phishing

Kategori	Train	Validation	Test
Safe Email	9.048	1.125	1.149
Phishing Email	5.872	740	716
Total	14.920	1.865	1.865

2.2 Arsitektur Model

Penelitian ini memanfaatkan lima model Transformer pre-trained yang berbeda untuk tugas deteksi email phishing. Model-model tersebut meliputi RoBERTa-base dengan 125 juta parameter (Liu et al., 2019), BERT-base dengan 110 juta parameter (Devlin et al., 2018), ELECTRA-base dengan 110 juta parameter (Clark et al., 2020), DeBERTa-V3 dengan 86 juta parameter (He et al., 2020), serta DistilBERT dengan 66 juta parameter (Sanh et al., 2019). Kelima model encoder ini dipilih karena kemampuannya dalam memahami nuansa semantik teks berbahasa Inggris yang kompleks serta performa yang telah terbukti dalam berbagai tugas NLP.

2.3 Konfigurasi Pelatihan

Proses pelatihan model Transformer dilakukan melalui dua pendekatan, yaitu full fine-tuning (FFT) dan Low-Rank Adaptation (LoRA). Konfigurasi umum hyperparameter untuk kedua pendekatan tersebut dirangkum pada Tabel 2. Pada skema LoRA, digunakan parameter peringkat rendah dengan nilai r atau rank sebesar 8, LoRA alpha sebesar 16, serta LoRA dropout sebesar 0,2. Sementara itu, learning rate pada LoRA ditetapkan sebesar 5e-5, sedangkan pada FFT digunakan learning rate sebesar 2e-5. Rincian konfigurasi spesifik untuk masing-masing eksperimen disajikan pada Tabel 3.

Tabel 2. Konfigurasi Umum Hyperparameter

Hyperparameter	Nilai
Optimizer	AdamW
LR scheduler	Cosine
Training batch size	16
Validation batch size	16
Weight decay	0,01
Max length	256
Epoch	10

Tabel 3. Konfigurasi Spesifik Hyperparameter

Teknik	Learning rate	r	Lora alpha	Lora dropout
LoRA	5e-5	8	16	0.2
FFT	2e-5	–	–	–

[1]https://www.kaggle.com/datasets/subhajournal/phishingemails/data

2.4 Metrik Evaluasi

Kinerja seluruh model dinilai menggunakan sejumlah metrik evaluasi, yaitu akurasi, presisi, recall, dan F1-score. Seluruh metrik tersebut dihitung dengan pendekatan *macro-average*, di mana nilai akhir diperoleh dari rata-rata performa masing-masing kelas tanpa memperhatikan distribusi jumlah data pada setiap kelas. Selain itu, confusion matrix digunakan untuk memberikan analisis kesalahan klasifikasi secara lebih mendalam.

Secara umum, confusion matrix terdiri atas empat komponen utama, yaitu True Positive (TP) yang merepresentasikan data positif yang diprediksi dengan benar, True Negative (TN) yang menunjukkan data negatif yang berhasil diklasifikasikan secara tepat, False Positive (FP) yang menggambarkan data negatif yang keliru diprediksi sebagai positif, serta False Negative (FN) yang merupakan data positif yang salah diklasifikasikan sebagai negatif.

Berdasarkan keempat komponen tersebut, metrik evaluasi dihitung menggunakan rumus-rumus sebagai berikut:

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

$$\text{Presisi} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F1} - \text{score} = \frac{2 \times \text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (4)$$

2.5 Infrastruktur Pelatihan

Seluruh eksperimen dan proses pelatihan model dilaksanakan pada infrastruktur pelatihan berbasis Google Colaboratory yang menyediakan akselerasi GPU NVIDIA Tesla T4 dengan kapasitas VRAM sebesar 15 GB. Proses pengembangan dan pelatihan model diimplementasikan menggunakan framework deep learning PyTorch, dengan dukungan pustaka Hugging Face Transformers untuk pemodelan arsitektur Transformer dan proses tokenisasi teks. Selain itu, pustaka Parameter-Efficient Fine-Tuning (PEFT) digunakan untuk mengimplementasikan metode Low-Rank Adaptation (LoRA) pada skema pelatihan yang efisien secara komputasi.

3. PEMBAHASAN

Bagian ini membahas hasil eksperimen dari penerapan teknik Low-Rank Adaptation (LoRA) dan full fine-tuning (FFT) pada lima arsitektur Transformer dalam tugas deteksi email phishing. Analisis difokuskan pada perbandingan kinerja klasifikasi berdasarkan metrik evaluasi utama, efisiensi penggunaan sumber daya komputasi yang meliputi konsumsi VRAM dan durasi selama pelatihan, serta analisis kesalahan prediksi melalui confusion matrix. Selain itu, hasil yang diperoleh juga dibandingkan dengan temuan penelitian sebelumnya guna memberikan analisis terhadap kelebihan dan keterbatasan metode yang diterapkan.

3.1 Perbandingan Kinerja Klasifikasi

Hasil eksperimen menunjukkan bahwa integrasi teknik LoRA pada model Transformer mampu mempertahankan kinerja deteksi phishing yang kompetitif, bahkan mendekati performa FFT. Berdasarkan Tabel 4, model BERT dengan metode FFT mencatat kinerja tertinggi secara keseluruhan dengan akurasi dan F1-score mencapai 98,23%. Hal ini mengonfirmasi bahwa arsitektur BERT masih relevan dan tangguh dalam menangkap isi semantik teks pada dataset email phishing.

Tabel 4. Komparasi Kinerja Klasifikasi

Model	Metode	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
RoBERTa	FFT	98,07	98,10	98,07	98,07
	LoRA	97,43	97,43	97,43	97,43
BERT	FFT	**98,23**	**98,26**	**98,23**	**98,23**
	LoRA	98,02	98,07	98,02	98,02
ELECTRA	FFT	98,02	98,05	98,02	98,02
	LoRA	97,69	97,73	97,69	97,70
DeBERTa	FFT	98,02	98,06	98,02	98,02
	LoRA	**98,12**	**98,14**	**98,12**	**98,13**
DistilBERT	FFT	98,02	98,05	98,02	98,02
	LoRA	97,64	97,69	97,64	97,65

Temuan paling menarik terlihat pada model DeBERTa. Berbeda dengan model lain yang mengalami sedikit penurunan performa saat menggunakan LoRA, DeBERTa justru menunjukkan sedikit peningkatan performa, di mana akurasi dan F1-score, masing-masing sebesar 98,12% dan 98,13% lebih tinggi dibandingkan versi FFT-nya (98,02%). Pola yang berbeda ini mengindikasikan bahwa pada model dengan arsitektur yang lebih kompleks, seperti DeBERTa, pembatasan parameter latih melalui LoRA justru bertindak sebagai regularisasi yang efektif, sehingga mampu mencegah overfitting pada data latih, dan meningkatkan generalisasi pada data uji email phishing.

Sementara itu, penurunan performa akibat penggunaan LoRA pada model lain tergolong relatif kecil. Model RoBERTa menunjukkan penurunan terbesar, yaitu sebesar 0,64%, sedangkan model lainnya hanya mengalami penurunan performa di bawah 0,4%. Hasil ini mengindikasikan bahwa pendekatan LoRA dapat menjadi salah satu alternatif adaptasi model Tranformer untuk tugas spesifik, di mana pembaruan parameter secara terbatas pada matriks attention berperingkat rendah masih mampu menangkap karakteristik email phishing secara memadai.

3.2 Analisis Penggunaan VRAM dan Waktu Pelatihan

Efisiensi komputasi menjadi salah satu keunggulan utama dari penerapan LoRA. Berdasarkan hasil yang disajikan pada Tabel 5, teknik ini secara konsisten mampu menurunkan penggunaan VRAM pada proses pelatihan di seluruh arsitektur model yang diuji. Penurunan konsumsi VRAM paling besar terjadi pada model DistilBERT, dengan pengurangan mencapai 40%, diikuti oleh DeBERTa sebesar 29%, RoBERTa sebesar 21%, BERT sebesar 17%, dan ELECTRA sebesar 14%.

Tabel 5. Komparasi Efisiensi Komputasi

Model	Metode	Parameter latih (%)	VRAM (GB)	Durasi pelatihan (per epoch)
RoBERTa	FFT	100	3,7	3:42
	LoRA	1,51	2,9	3:58
BERT	FFT	100	3,4	3:41
	LoRA	1,21	2,8	4:05
ELECTRA	FFT	100	4,3	4:34
	LoRA	1,72	3,7	4:52
DeBERTa	FFT	100	7,1	6:45
	LoRA	**0,48**	5	5:40
DistilBERT	FFT	100	2,0	1:51
	LoRA	1,31	**1,2**	**1:32**

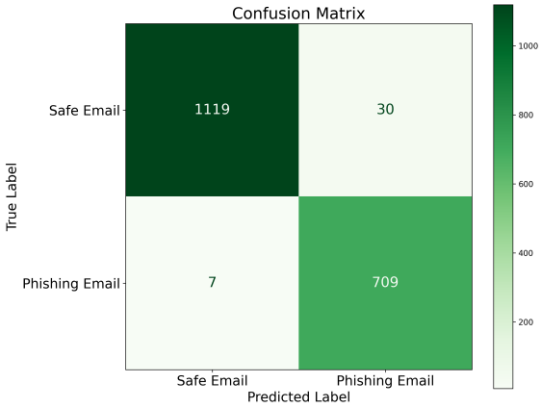
Dari sisi efisiensi parameter, DeBERTa mencatatkan rasio parameter latih terkecil, yaitu hanya 0,48% dari total parameter model. Meskipun hanya melatih kurang dari setengah persen parameter, model ini mampu menghasilkan akurasi yang kompetitif. Ini menunjukkan efektivitas arsitektur DeBERTa dalam memanfaatkan metode LoRA.

Dari sisi durasi pelatihan, penerapan LoRA menunjukkan adanya trade-off yang bervariasi antar arsitektur model. Meskipun jumlah parameter yang dilatih jauh lebih sedikit, waktu pelatihan per epoch tidak selalu lebih singkat dibandingkan dengan FFT. Hal ini terlihat pada model berukuran lebih dari 100 juta parameter seperti RoBERTa, ELECTRA, dan BERT yang justru mengalami sedikit peningkatan waktu pelatihan per epoch. Fenomena ini menunjukkan bahwa meskipun jumlah parameter latih berkurang secara signifikan, keberadaan operasi matriks tambahan pada lapisan adapter LoRA selama proses *forward* dan *backward pass* masih berkontribusi terhadap overhead komputasi pada model berukuran besar.

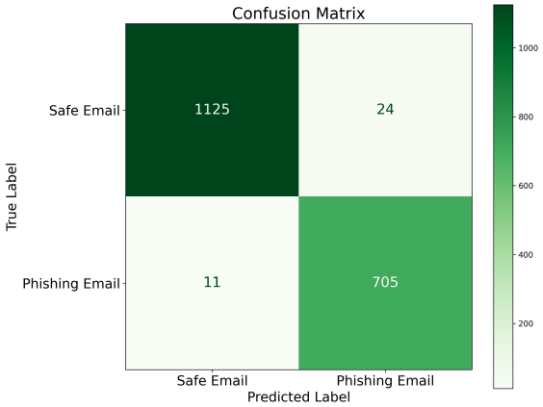
Sebaliknya, pada model dengan jumlah parameter yang lebih kecil, seperti DeBERTa dengan 86 juta parameter dan DistilBERT dengan 66 juta parameter, penerapan LoRA mampu mengurangi durasi pelatihan secara signifikan. Pada DeBERTa, waktu pelatihan per epoch berkurang dari 6:45 menit menjadi 5:40 menit, sementara pada DistilBERT terjadi penurunan dari 1:51 menit menjadi 1:32 menit per epoch. Hasil ini mengindikasikan bahwa pada model dengan ukuran parameter yang lebih ringkas dan arsitektur yang lebih efisien, pengurangan beban komputasi gradien akibat pembatasan parameter latih melalui LoRA lebih dominan dibandingkan dengan overhead operasi matriks tambahan, sehingga menghasilkan semacam percepatan proses pelatihan.

3.3 Analisis Confusion Matrix

Untuk melengkapi evaluasi kinerja berbasis metrik statistik, dilakukan analisis lebih lanjut terhadap confusion matrix guna mengidentifikasi karakteristik kesalahan klasifikasi secara lebih rinci. Analisis ini difokuskan pada model DeBERTa, karena model tersebut menunjukkan temuan menarik di mana penerapan LoRA menghasilkan kinerja yang sedikit lebih baik dibandingkan metode FFT. Visualisasi confusion matrix untuk metode FFT dan LoRA masing-masing disajikan pada Gambar 2 dan Gambar 3.



Gambar 2. Confusion Matrix DeBERTa dengan FFT



Gambar 3. Confusion Matrix DeBERTa dengan LoRA

Model dengan pendekatan FFT menunjukkan kinerja yang lebih baik dalam meminimalkan False Negative (FN), yaitu kondisi ketika email phishing gagal terdeteksi. Metode FFT mencatatkan 7 sampel FN, sedangkan penerapan LoRA menghasilkan 11 sampel FN. Temuan ini menunjukkan bahwa FFT memiliki tingkat sensitivitas (recall) yang sedikit lebih tinggi dalam mendeteksi seluruh potensi

ancaman, sehingga cenderung lebih ketat dari sisi pencegahan risiko.

Sebaliknya, penerapan LoRA menunjukkan keunggulan pada aspek presisi, yang tercermin dari jumlah False Positive (FP) yang lebih rendah. Metode LoRA mencatatkan 24 sampel FP, lebih sedikit dibandingkan FFT yang menghasilkan 30 sampel FP. Penurunan jumlah FP ini mengindikasikan bahwa LoRA lebih jarang mengklasifikasikan email aman sebagai phishing, sehingga berpotensi mengurangi gangguan terhadap komunikasi email yang valid. Dari sudut pandang pengalaman pengguna, karakteristik ini dapat memberikan keuntungan dengan menekan tingkat alarm palsu pada sistem deteksi email phishing.

Secara keseluruhan, meskipun penerapan LoRA menghasilkan jumlah FN yang sedikit lebih tinggi dibandingkan FFT, pendekatan ini menawarkan keseimbangan yang lebih baik antara sensitivitas dan presisi. Dengan mempertimbangkan efisiensi komputasi yang diperoleh melalui LoRA, perbedaan sensitivitas yang relatif kecil tersebut dapat dipandang sebagai trade-off yang wajar untuk memperoleh model yang lebih ringan dan efisien, tanpa penurunan signifikan pada nilai F1-score secara keseluruhan.

3.4 Perbandingan dengan Penelitian Sebelumnya

Untuk memvalidasi efektivitas model yang diusulkan, hasil penelitian ini dikomparasikan dengan studi terdahulu yang menggunakan dataset serupa. Sebagai acuan utama, digunakan penelitian Sajad (2025) yang menerapkan model DistilBERT dengan teknik Fast Gradient Method (FGM). Rangkuman perbandingan kinerja dengan studi sebelumnya disajikan pada Tabel 5.

Tabel 5. Perbandingan Kinerja dengan Studi Sebelumnya

Model	Metode	Akurasi (%)	F1-Score (%)
BERT (penelitian ini)	FFT	**98,23**	**98,23**
	LoRA	98,02	98,02
DeBERTa (penelitian ini)	FFT	98,02	98,02
	LoRA	**98,12**	**98,13**
DistilBERT (penelitian ini)	FFT	98,02	98,02
	LoRA	97,64	97,65
DistilBERT (Sajad, 2025)	FFT	93	91,50
DistilBERT + FGM (Sajad, 2025)	FFT	96,50	96

Analisis komparatif menunjukkan beberapa temuan penting yang menegaskan keunggulan metode yang digunakan dalam penelitian ini. Pada arsitektur yang sama, yaitu DistilBERT, hasil penelitian ini menunjukkan performa lebih tinggi dibandingkan studi sebelumnya. Model DistilBERT (FFT) mencapai akurasi 98,02%, melampaui model DistilBERT + FGM pada penelitian terdahulu yang menghasilkan akurasi 96,50%. Bahkan, model DistilBERT (LoRA), yang hanya melatih 1,3% parameter untuk efisiensi, masih mampu memperoleh akurasi 97,64% dan F1-score 97,65%, lebih tinggi dibandingkan metode FGM yang mencapai F1-Score 96%. Hasil ini menunjukkan bahwa strategi fine-tuning yang diterapkan efektif dalam mengekstraksi fitur penting dalam email phishing, bahkan dalam skenario sumber daya komputasi terbatas.

Selain model DistilBERT, penelitian ini juga mengeksplorasi arsitektur Transformer yang lebih canggih. Misalnya, model BERT menunjukkan performa terbaik untuk metode FFT dengan F1-Score 98,23%, sedangkan DeBERTa mencatat kinerja tertinggi untuk metode LoRA dengan F1-Score 98,13%, keduanya melampaui hasil DistilBERT pada studi sebelumnya. Kenaikan performa hampir 2% dibandingkan riset sebelumnya menunjukkan bahwa pemilihan arsitektur model yang tepat memiliki pengaruh signifikan terhadap akurasi deteksi. Secara keseluruhan, komparasi ini menegaskan bahwa pendekatan yang diusulkan, baik melalui FFT maupun LoRA, menawarkan solusi kompetitif dalam deteksi email phishing yang lebih akurat dan andal dibandingkan pendekatan sebelumnya.

4. KESIMPULAN DAN SARAN

Penelitian ini telah berhasil mencapai tujuan utamanya dalam mengevaluasi kinerja dan efisiensi antara Full Fine-Tuning (FFT) dan Low-Rank Adaptation (LoRA) pada lima model Transformer encoder untuk deteksi email phishing. Hasil eksperimen menunjukkan bahwa LoRA mampu mempertahankan kinerja klasifikasi yang kompetitif, dengan penurunan performa rata-rata kurang dari 1% dibandingkan FFT. Model BERT dengan metode FFT mencapai akurasi tertinggi 98,23%, sedangkan DeBERTa dengan metode LoRA dapat melampaui FFT (98,12% vs 98,02%) berkat efek regularisasi dari efisiensi parameter yang mencegah overfitting.

Dari sisi efisiensi komputasi, LoRA terbukti efektif dalam mengurangi penggunaan sumber daya komputasi. Model DistilBERT dengan pendekatan LoRA menjadi solusi optimal untuk lingkungan dengan sumber daya komputasi terbatas, karena mampu menurunkan penggunaan VRAM hingga 40% sambil tetap mempertahankan akurasi yang kompetitif. Berdasarkan temuan ini, penggunaan FFT direkomendasikan jika prioritas utama adalah akurasi maksimal, sedangkan LoRA lebih sesuai untuk efisiensi memori.

Meskipun tujuan penelitian telah tercapai, penelitian ini masih memiliki beberapa keterbatasan yang perlu diperhatikan. Pertama, evaluasi performa teknik LoRA belum sepenuhnya konsisten mengungguli teknik FFT pada seluruh arsitektur model karena sangat bergantung pada kesesuaian konfigurasi hyperparameter dengan karakteristik model. Kedua, penelitian ini masih berfokus pada dataset berbahasa Inggris, sehingga generalisasi terhadap serangan phishing lintas bahasa belum teruji.

Untuk pengembangan di masa mendatang, direkomendasikan untuk melakukan kajian mendalam mengenai optimasi hyperparameter krusial dalam LoRA, seperti rank (r), Lora alpha, dan Lora dropout. Selain itu, penelitian selanjutnya juga dapat memperluas eksplorasi pada teknik Parameter-Efficient Fine-Tuning (PEFT) lainnya, seperti QloRA (Dettmers et al., 2023), DoRA (Liu et al., 2024), ReFT (Wu et al., 2024), atau kombinasi dari teknik-teknik tersebut guna mencapai keseimbangan yang lebih baik antara performa dan biaya komputasi.

DAFTAR PUSTAKA

Blake, S.E., 2025. *Phishsense-1B: A Technical Perspective on an AI-Powered Phishing Detection Model*. Available at: <https://arxiv.org/abs/2503.10944>.

Clark, K., Luong, M.-T., Le, Q. V and Manning, C.D., 2020. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. *CoRR*, [online] abs/2003.10555. Available at: <https://arxiv.org/abs/2003.10555>.

Dettmers, T., Pagnoni, A., Holtzman, A. and Zettlemoyer, L., 2023. QLORA: efficient finetuning of quantized LLMs. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23. Red Hook, NY, USA: Curran Associates Inc.

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K., 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR*, [online] abs/1810.04805. Available at: <http://arxiv.org/abs/1810.04805>.

Febriyani, W., Fathia, D., Widjajarto, A. and Lubis, M., 2023. Security Awareness Strategy for Phishing Email Scams: A Case Study One of a Company in Singapore. *International Journal on Informatics Visualization (JOIV)*, 7(3), pp.808–814.

He, P., Liu, X., Gao, J. and Chen, W., 2020. DeBERTa: Decoding-enhanced BERT with Disentangled Attention. *CoRR*, [online] abs/2006.03654. Available at: <https://arxiv.org/abs/2006.03654>.

Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S. and Chen, W., 2021. LoRA: Low-Rank Adaptation of Large Language Models. *CoRR*, [online] abs/2106.09685. Available at: <https://arxiv.org/abs/2106.09685>.

Kyaw, P.H., Gutierrez, J. and Ghobakhlou, A., 2024. A Systematic Review of Deep Learning Techniques for Phishing Email Detection. *Electronics*, [online] 13(19). https://doi.org/10.3390/electronics13193823.

Liu, S.-Y., Wang, C.-Y., Yin, H., Molchanov, P., Wang, Y.-C.F., Cheng, K.-T. and Chen, M.-H., 2024. DoRA: weight-decomposed low-rank adaptation. In: *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V., 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR*, [online] abs/1907.11692. Available at: <http://arxiv.org/abs/1907.11692>.

Meléndez, R., Ptaszynski, M. and Masui, F., 2024. Comparative Investigation of Traditional Machine-Learning Models and Transformer Models for Phishing Email Detection. *Electronics*, [online] 13(24). https://doi.org/10.3390/electronics13244877.

Moorthy, R.S. and Pabitha, P., 2020. Optimal Detection of Phising Attack using SCA based K-NN. *Procedia Computer Science*, [online] 171, pp.1716–1725. https://doi.org/https://doi.org/10.1016/j.procs.2020.04.184.

Nakamura, A. and Dobashi, F., 2019. Proactive Phishing Sites Detection. In: *IEEE/WIC/ACM International Conference on Web Intelligence*, WI '19. [online] New York, NY, USA: Association for Computing Machinery. pp.443–448. https://doi.org/10.1145/3350546.3352565.

Osamor, J., Ashawa, M., Shahrabi, A., Philip, A. and Iwendi, C., 2025. The Evolution of Phishing and Future Directions: A Review. In: *Proceedings of the 20th International Conference on Cyber Warfare and Security (ICCWS 2025)*. [online] Academic Conferences & Publishing International Limited. pp.361–368. https://doi.org/10.34190/iccws.20.1.3366.

Otieno, D.O., Siami Namin, A. and Jones, K.S., 2023. The Application of the BERT Transformer Model for Phishing Email Classification. In: *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*. pp.1303–1310. https://doi.org/10.1109/COMPSAC57700.2023.00198.

Parthasarathy, V.B., Zafar, A., Khan, A. and Shahid, A., 2024. *The Ultimate Guide to Fine-Tuning LLMs from Basics to Breakthroughs: An Exhaustive Review of Technologies, Research, Best Practices, Applied Research Challenges and*

Opportunities. Available at: <https://arxiv.org/abs/2408.13296>.

Patil, P., Rane, R. and Bhalekar, M., 2017. Detecting spam and phishing mails using SVM and obfuscation URL detection algorithm. In: *2017 International Conference on Inventive Systems and Control (ICISC)*. pp.1–4. https://doi.org/10.1109/ICISC.2017.8068633.

Safi, A. and Singh, S., 2023. A systematic literature review on phishing website detection techniques. *Journal of King Saud University - Computer and Information Sciences*, [online] 35(2), pp.590–611. https://doi.org/https://doi.org/10.1016/j.jksuci.2023.01.004.

Sajad, 2025. *Explainable Transformer-Based Email Phishing Classification with Adversarial Robustness*. Available at: <https://arxiv.org/abs/2511.12085>.

Sanh, V., Debut, L., Chaumond, J. and Wolf, T., 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *CoRR*, [online] abs/1910.01108. Available at: <http://arxiv.org/abs/1910.01108>.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, \L{}ukasz and Polosukhin, I., 2017. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17. Red Hook, NY, USA: Curran Associates Inc. pp.6000–6010.

Wosah, N.P. and Win, T., 2021. Phishing Mitigation Techniques: A Literature Survey. *CoRR*, [online] abs/2104.06989. Available at: <https://arxiv.org/abs/2104.06989>.

Wu, Z., Arora, A., Wang, Z., Geiger, A., Jurafsky, D., Manning, C.D. and Potts, C., 2024. *ReFT: Representation Finetuning for Language Models*. Available at: <https://arxiv.org/abs/2404.03592>.

Yahya, F., W Mahibol, R.I., Ying, C.K., Anai, M. Bin, Frankie, S.A., Nin Wei, E.L. and Utomo, R.G., 2021. Detection of Phising Websites using Machine Learning Approaches. In: *2021 International Conference on Data Science and Its Applications (ICoDSA)*. pp.40–47. https://doi.org/10.1109/ICoDSA53588.2021.9617482.