OPTIMIZATION OF BIOBERT MODEL FOR MEDICAL ENTITY RECOGNITION THROUGH BILSTM AND CNN-CHAR INTEGRATION

OPTIMALISASI MODEL BIOBERT UNTUK PENGENALAN ENTITAS MEDIS MELALUI INTEGRASI BILSTM DAN CNN-CHAR

Gilang Djati Prinantyo[1], Abu Salam[2]
[1,2]Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, Semarang, Indonesia
111202113919@mhs.dinus.ac.id[1], abu.salam@dsn.dinus.ac.id[2]

Abstract – *Biomedical Named Entity Recognition (NER) is essential for extracting structured information from medical texts. However, existing models like BioBERT face challenges when dealing with complex biomedical entities, particularly those with intricate morphological structures. This research enhances the BioBERT model by integrating BiLSTM and character-level CNN (CNN-Char), aiming to improve the recognition of Chemical and Disease entities. The proposed models were trained and evaluated on the BC5CDR dataset sourced from the official BioCreative V CDR Corpus. The modified model achieved an F1-score of 0.8678, indicating a significant improvement compared to the standard BioBERT model, which scored 0.8597. This increase is primarily observed in the recognition of complex entity structures, particularly those requiring character-level representation. Despite this improvement, the model is limited to Chemical and Disease entities and may not generalise to other biomedical categories. Future work should focus on expanding the entity types and exploring other model architectures, such as SciBERT or BioALBERT, to further enhance performance.*

Keywords – *Named Entity Recognition (NER), BioBERT, BiLSTM, CNN-Char, BC5CDR Dataset.*

Abstrak – Pengenalan Entitas Bernama (NER) biomedis sangat penting untuk ekstraksi informasi terstruktur dari teks medis. Namun, model seperti BioBERT masih menghadapi kesulitan dalam menangani entitas biomedis kompleks, terutama yang memiliki struktur morfologi rumit. Penelitian ini mengoptimalkan model BioBERT dengan mengintegrasikan BiLSTM dan CNN tingkat karakter (CNN-Char) untuk meningkatkan kemampuan pengenalan entitas Kimia dan Penyakit. Model ini dilatih dan dievaluasi menggunakan dataset BC5CDR yang berasal dari publikasi resmi BioCreative V CDR Corpus. Model modifikasi mencapai F1-Score sebesar 0.8678, menunjukkan peningkatan signifikan dibandingkan model BioBERT standar dengan skor 0.8597. Peningkatan ini terutama terlihat pada pengenalan entitas dengan struktur kompleks yang membutuhkan representasi tingkat karakter. Meskipun demikian, model ini terbatas pada entitas Kimia dan Penyakit dan mungkin tidak berlaku pada kategori biomedis lainnya. Penelitian selanjutnya disarankan untuk memperluas jenis entitas dan mengeksplorasi arsitektur model lainnya, seperti SciBERT atau BioALBERT, guna meningkatkan performa.

Kata Kunci: Named Entity Recognition (NER), BioBERT, BiLSTM, CNN-Char, Dataset BC5CDR.

I. PENDAHULUAN

Kemajuan teknologi biomedis menghasilkan banyak literatur ilmiah, termasuk jurnal, rekam medis elektronik, dan laporan klinis. Tantangan utama adalah pengenalan istilah medis melalui Named Entity Recognition (NER), yang berperan penting dalam ekstraksi informasi medis untuk membangun knowledge graph[1],[2],[3]. BioBERT unggul dalam NER biomedis namun mengalami kendala dalam mengenali entitas kompleks akibat keterbatasan morfologis dan kontekstual [4]. Integrasi metode BiLSTM dan CNN menawarkan solusi untuk masalah ini. Metode deep learning, seperti BiLSTM-CRF, mendominasi NER berkat kemampuannya mengekstraksi fitur tanpa rekayasa manual, sementara embedding karakter meningkatkan generalisasi model[5], [6]. Pendekatan klasik berbasis RNN seperti BiLSTM, di sisi lain, memiliki beberapa keterbatasan. Ini termasuk lambatnya waktu pelatihan karena ketergantungan sekuensial dan ketidakmampuan untuk memahami konteks global kalimat secara menyeluruh. Selain itu, performa pendekatan tersebut sangat dipengaruhi oleh parameter tuning dan desain embedding[7].

Model transformer seperti BERT (Bidirectional Encoder Representations from Transformers) dan penerusannya, seperti BioBERT dan SciBERT, dibuat untuk mengatasi keterbatasan ini dalam domain biomedis. BioBERT, model yang dilatih khusus dengan literatur PubMed dan PMC, mengungguli model sebelumnya pada benchmark NER medis seperti BC5CDR, NCBI Disease, dan JNLPBA [8]. Meskipun BioBERT sangat baik, model ini belum ideal untuk menangani struktur morfologis unik dari entitas medis dengan banyak bentuk gabungan atau token [9]. Dalam hal ini, metode integrasi BioBERT dengan BiLSTM dan karakter-level CNN (CNN-Char) menjadi relevan.

Beberapa penelitian mengusulkan pendekatan integratif yang memanfaatkan sensitivitas urutan BiLSTM dan kemampuan CNN-Char untuk mengidentifikasi struktur morfologi kata, khususnya untuk entitas medis yang memiliki bentuk kompleks dan variatif [10], [11]. Kombinasi ini diharapkan dapat memperbaiki kelemahan model berbasis BioBERT tunggal, yang biasanya hanya bergantung pada penerapan tingkat huruf tanpa karakter tambahan.

Pendekatan hybrid, seperti integrasi BiLSTM dan CNN-Char, telah terbukti meningkatkan kinerja NER biomedis [12]. Oleh karena itu, penelitian ini difokuskan untuk meningkatkan kemampuan BioBERT dalam mengenali entitas medis kompleks pada dataset BC5CDR dengan mengadopsi metode tersebut. Penelitian ini bertujuan untuk meningkatkan kemampuan BioBERT dalam mengenali entitas medis kompleks melalui integrasi BiLSTM dan CNN-Char, dengan fokus pada pengenalan entitas Chemical dan Disease.

Penelitian ini berfokus pada dua masalah utama, yaitu:

1. Bagaimana model BioBERT mampu mendeteksi entitas medis dengan menggunakan dataset BC5CDR?
2. Bagaimana peningkatan akurasi NER pada model BioBERT dipengaruhi oleh integrasi BiLSTM dan CNN-Char?

Untuk mengoptimalkan pengenalan entitas medis, penelitian ini bertujuan untuk merancang serta menguji performa arsitektur BioBERT yang dikombinasikan dengan BiLSTM dan CNN-Char, dengan tujuan akhir memberikan kontribusi terhadap pengembangan model NER yang lebih presisi dan efisien.

II. SIGNIFIKANSI STUDI

Penelitian ini meningkatkan kinerja BioBERT melalui integrasi BiLSTM dan CNN-Char, mengatasi tantangan OOV dan memperbaiki generalisasi entitas medis[13],[14],[15]. Model ini berpotensi diintegrasikan ke EHR untuk ekstraksi entitas medis real-time dan mendukung knowledge graph biomedis[16],[17],[18].

A. Studi Literatur

CNN mengenali morfologi kata, sementara BiLSTM menangkap konteks sekuensial, menjadikan kombinasi ini ideal untuk NER biomedis[19], [20]. Ini terutama berlaku saat menggabungkan model bahasa yang diatur seperti BERT dengan BiLSTM dan lapisan tambahan seperti CRF atau CNN-Char. Namun, sebagian besar model sebelumnya tidak secara eksplisit menangani masalah entitas OOV dalam teks medis, menyebabkan kesalahan pengenalan istilah yang tidak biasa atau tidak ada dalam korpus pelatihan. BioBERT efektif untuk NER biomedis, namun terbatas dalam menangani variasi morfologi, sehingga karakter-level direkomendasikan[21].

B. Bahan dan Lokasi Penelitian

Referensi [22] menjelaskan penggunaan dataset BC5CDR, yang berfokus pada entitas Chemical dan Disease yang ditemukan dalam literatur PubMed dan dibagi menjadi tiga subset: train, validasi, dan tes. Dataset ini digunakan dalam benchmark BioNER dan dilabeli dengan metode BIO tagging. Penelitian dilakukan di Google Colab (GPU Tesla T4) menggunakan dataset BC5CDR dari BioCreative V CDR Corpus melalui PubTator. Tabel I dan Tabel II menunjukkan konfigurasi pelatihan dan desain model untuk mendukung proses fine-tuning. Penyesuaian ini mencakup parameter pelatihan seperti learning rate dan batch size. Integrasi BiLSTM bertujuan meningkatkan generalisasi model pada datset BC5CDR.

TABEL I
KONFIGURASI BASED MODEL

Parameter	BioBERT
Scheduler	Linear
Learning Rate	5e-5
Batch Size	16
Optimizer	AdamW
Dropout Rate	0.15
Weight Decay	0.05
Epoch	25
Early Stopping	3

TABEL II
KONFIGURASI MODIFIED MODEL

Parameter	BioBERT + BiLSTM - CNN-Char
Scheduler	Cosine with Restart
Learning Rate	2e-5
Batch Size	16
Optimizer	AdamW
Dropout Rate	0.5
Weight Decay	0.01
Epoch	30
Early Stopping	5
CNN-Char Filter Size	3
CNN-Char Embedding Dimensi	50
BiLSTM Hidden Size	256

Untuk menjaga kestabilan proses fine-tuning pada dataset berukuran kecil, parameter pelatihan disesuaikan secara konservatif. Namun, penyesuaian ini masih memungkinkan model menunjukkan peningkatan signifikan dalam kinerja.

C. Metode Penelitian dan Evaluasi

Proses penelitian dilakukan melalui langkah-langkah seperti preprocessing teks (tokenisasi, normalisasi), pelatihan model dengan kombinasi BioBERT + BiLSTM + CNN-Char, dan evaluasi menggunakan Metrik evaluasi berupa Precision, Recall, dan F1-Score. Penyempurnaan model dilakukan dengan mengatur ulang parameter seperti tingkat pembelajaran (learning rate), ukuran batch, serta dropout. Metrik umum seperti Precision, Recall, dan F1-Score diterapkan sebagai indikator untuk mengevaluasi hasil kinerja model. Setelah pelatihan selesai, setiap model diuji pada dataset uji BC5CDR. Untuk menilai efektivitas integrasi BiLSTM dan CNN-Char ke dalam arsitektur BioBERT, perbandingan performa antar model dilakukan secara deskriptif berdasarkan nilai metrik tersebut. Strategi pelatihan dan modifikasi arsitektur sangat memengaruhi fine-tuning model BioBERT. Penambahan CNN-Char dan BiLSTM meningkatkan fitur representasi sensitivitas entitas kompleks.

D. Evaluasi dan Metrik Pengukuran

Evaluasi kinerja model dalam tugas Named Entity Recognition (NER) dilakukan dengan menggunakan beberapa metrik pengukuran, antara lain Accuracy, Precision, Recall, dan F1-Score. Metrik-metrik ini membantu mengevaluasi seberapa baik model dalam mengidentifikasi entitas yang relevan dalam teks medis. Setiap metrik memiliki fungsi spesifik, memberikan pemahaman kuantitatif mengenai keakuratan dan keefektifan model dalam melakukan klasifikasi entitas medis. Berikut ini adalah rumus perhitungan dari setiap metrik tersebut:

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

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

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

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

Di mana:

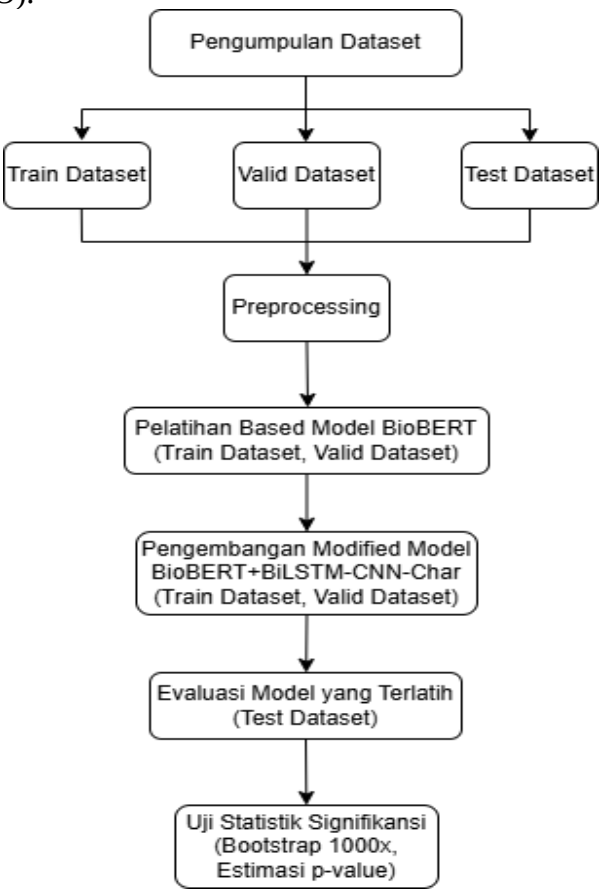
1. TP (True Positive) merupakan banyaknya prediksi yang benar untuk label yang tepat.
2. FP (False Positive) mencerminkan jumlah prediksi yang keliru terhadap label yang seharusnya tidak terdeteksi.
3. FN (False Negative) mengacu pada kegagalan model dalam mengidentifikasi label yang sebenarnya terdapat dalam data.
4. TN (True Negative) menunjukkan jumlah prediksi tepat terhadap kasus yang memang bukan merupakan label target.
5. Accuracy mengukur keseluruhan ketepatan klasifikasi model.
6. Precision menunjukkan ketepatan model dalam mengidentifikasi entitas.
7. Recall menunjukkan seberapa lengkap model dalam mengidentifikasi semua entitas yang relevan.
8. F1-Score adalah harmonisasi antara Precision dan Recall.

E. Diagram Proses dan Struktur Dataset

TABEL III
DISTRIBUSI LABEL DATASET

Label	Train	Validation	Test
B-Chemical	5.101	5.270	5.282
B-Disease	4.033	4.094	4.321
I-Disease	2.208	2.119	2.204
I-Chemical	519	484	354
O	99.324	98.681	106.194

Tabel berikut menunjukkan jumlah token berlabel untuk masing-masing subset (train, validasi, dan tes). Label ini menggunakan format BIO dan terdiri dari entitas Chemical dan Disease serta token non-entitas (O). Tabel ini menunjukkan karakteristik data yang digunakan untuk pelatihan dan evaluasi model. Ketidakseimbangan data ditunjukkan oleh distribusi label, dimana mayoritas token adalah non-entitas (label O).



Gambar 1. Diagram Tahapan Penelitian

Penelitian ini diawali dengan pengumpulan dataset BC5CDR, yang terdiri dari tiga bagian: data Latihan (CDR_Train), validasi (CDR_Dev), dan pengujian (CDR_Test). Diagram ini menggambarkan alur metodologi yang digunakan dalam penelitian. Setiap dataset melalui proses pelabelan ketat menggunakan skema BIO dan disesuaikan dengan tokenizer BioBERT pada tahap preprocessing. Model BioBERT diterapkan sebagai model awal dan dievaluasi setelah data diubah ke format HuggingFace. Selanjutnya, model ditingkatkan dengan integrasi arsitektur BiLSTM dan CNN-Char. Kedua model diuji dengan data pengujian untuk menilai efektivitasnya. Pada evaluasi akhir, metrik pengukuran seperti Precision, Recall, dan F1-Score dihitung, dan kinerja dari dua model yang dibandingkan dianalisis. Untuk memastikan bahwa perbedaan kinerja ini signifikan, dilakukan uji bootstrap. Semua hasil dicatat sebagai bagian dari laporan akhir penelitian, termasuk model yang dihasilkan dan grafik evaluasi.

III. HASIL DAN PEMBAHASAN

A. Hasil Preprocessing dan Distribusi Label

Preprocessing data dilakukan pada tiga file utama (CDR_TrainigSet, CDR_DevelopmentSet, CDR_TestSet) dengan metode BIO Tagging ketat, tokenization whitespace, allignment subword berbasis BioBERT, dan parsing PubTator, menghasilkan dataset terstruktur siap pakai. Tabel III di BAB II menunjukkan distribusi label setelah preprocessing, di mana sebagian besar token diberi label O. Presentase label B-Chemical, I-Chemical, B-Disease, dan I-Disease lebih sedikit dibandingkan label O, namun tetap mencukupi untuk keperluan pelatihan model.

B. Hasil Pelatihan dan Evaluasi Based Model

BioBERT dilatih selama 25 epoch dengan early stopping, dievaluasi menggunakan dataset CDR_TestSet. Hasil evaluasi model disajikan dalam bentuk laporan klasifikasi, dengan penekanan pada empat label utama: B-Chemical, I-Chemical, B-Disease, dan I-Disease, label-label ini mengindikasikan seberapa baik model dalam mendeteksi entitas medis.

$$F1_{entitas} = \frac{F1_{B-Chemical}+F1_{B-Disease}+F1_{I-Disease}+F1_{I-Chemical}}{4} \quad (5)$$

Berdasarkan hasil evaluasi pada data testing, nilai F1-Score untuk masing-masing label adalah sebagai berikut:

TABEL IV
NILAI F1-SCORE ENTITAS UTAMA

Label	F1-Score
B-Chemical	0.9456
B-Disease	0.8654
I-Disease	0.8107
I-Chemical	0.7069

Sehingga, rata-rata nilai F1-Score untuk entitas utama adalah:

$$F1_{entitas} = \frac{0.9456+0.8654+0.8107+0.7069}{4} = 0.8322 \quad (6)$$

Hasil ini menunjukkan bahwa model BioBERT dapat mendeteksi entitas Chemical dan Disease dengan baik. Namun, model ini masih menghadapi masalah dengan pola morfologi yang tidak konsisten dan struktur entitas medis multi-token yang terdiri dari banyak token. Faktor utama yang mendorong pengembangan model modifikasi yang menggabungkan konteks karakter dan urutan kata adalah ini.

C. Pengembangan dan Pelatihan Modified Model

Model modifikasi memanfaatkan arsitektur BiLSTM dan CNN-Char yang terintegrasi dengan BioBERT. CNN-Char menggunakan filter ukuran 3 dan output 50 untuk menangkap pola karakter, sementara BiLSTM memiliki dua layer dengan hidden size 256 untuk memahami konteks sekuensial. Dropout 0.5 diterapkan untuk mencegah overfitting, dan Linear Layer menghasilkan prediksi akhir. Modifikasi ini bertujuan untuk meningkatkan pengenalan entitas medis kompleks.

D. Evaluasi Test Modified Model

Model modifikasi dievaluasi menggunakan dataset uji untuk mengukur peningkatan kinerja dibandingkan model dasar BioBERT, terutama pada pengenalan entitas Chemical dan Disease berdasarkan nilai F1-Score untuk setiap label.

TABEL V
PERBANDINGAN F1-SCORE BERDASARKAN ENTITAS

Entitas	BioBERT (Based Model)	BioBERT + BiLSTM + CNN (Modified Model)
Chemical	0.8262	0.8414
Disease	0.8380	0.8421

Tabel berikut menunjukkan F1-Score model Based Model dan Modified Model untuk entitas Chemical dan Disease pada dataset test.

F1-Score rata-rata untuk model modifikasi dihitung dengan menggunakan rumus berikut berdasarkan nilai F1-Score untuk setiap label entitas utama:

$$F1_{entitas} = \frac{F1_{B-Chemical}+F1_{B-Disease}+F1_{I-Disease}+F1_{I-Chemical}}{4} \quad (7)$$

Dengan nilai:

$$F1_{entitas} = \frac{0.9510+0.8685+0.8157+0.7319}{4} = 0.8418 \quad (8)$$

Sesuai dengan hasil evaluasi sebelumnya, hasil ini menunjukkan bahwa model modifikasi memiliki F1-Score rata-rata yang lebih baik daripada model dasar.

Arsitektur CNN-Char, yang mampu menangkap pola morfologis dari token panjang dan tidak umum, seperti bahan kimia atau istilah medis gabungan, menyebabkan peningkatan F1-Score pada label I-Chemical dan I-Disease. Selain itu, BiLSTM mampu mengenali pola linguistik yang menunjukkan keberadaan entitas kompleks melalui pemahaman kontekstual dua arah terhadap urutan token. Sensitivitas model terhadap struktur multi-token meningkat dengan kombinasi dua lapisan ini, ini merupakan kelemahan pada model berbasis BioBERT murni.

TABEL VI
NILAI F1-SCORE ENTITAS UTAMA (MODIFIED MODEL)

Label	F1-Score
B-Chemical	0.9510
B-Disease	0.8685
I-Disease	0.8157
I-Chemical	0.7319

Meskipun F1-Score entitas I-Chemical meningkat dari 0.7069 pada Based Model menjadi 0.7319 pada Modified Model, nilainya tetap rendah dibandingkan label lainnya. Ini menunjukkan bahwa I-Chemical masih menjadi masalah dalam pekerjaan NER biomedis. Salah satu penyebabnya adalah struktur berbagai token, terutama pada frasa seperti “lithium carbonate toxicity”, yang sering menyebabkan penandaan token lanjutan yang salah. Pemodelan konteks pada posisi I- juga sulit karena karakteristik OOV (Out-of-Vocabulary), format yang kompleks, dan keragaman morfologis.

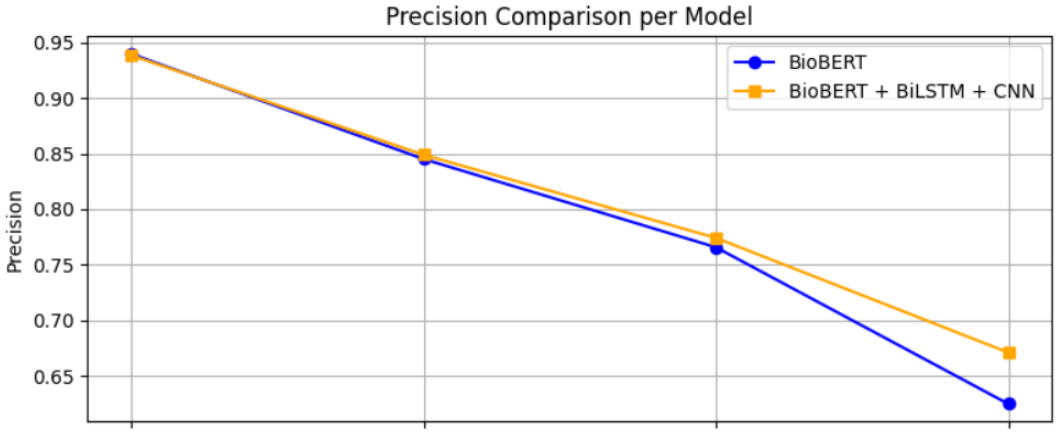
E. Evaluasi Akhir dan Perbandingan Model

Untuk evaluasi akhir, dataset CDR_Test digunakan. Model yang dilatih dievaluasi dengan menghasilkan laporan klasifikasi untuk masing-masing model. Hasil menunjukkan bahwa model modifikasi tetap mempertahankan kinerja yang unggul, meskipun Recall dan Precision entitas bertingkat kompleks sedikit meningkat.

TABEL VII
HASIL EVALUASI MODEL PADA DATA UJI

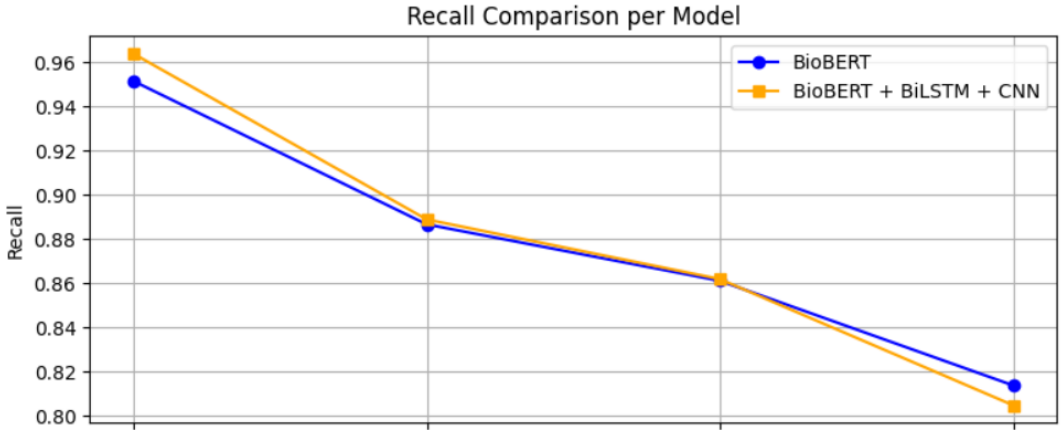
Entitas	Model	Precision	Recall	F1-Score
Chemical	BioBERT	0.782	0.883	0.826
Chemical	BioBERT + BiLSTM + CNN-Char	0.805	0.884	0.841
Disease	BioBERT	0.806	0.874	0.838
Disease	BioBERT + BiLSTM + CNN-Char	0.812	0.875	0.842

F. Visualisasi Perbandingan Metrik per Entitas



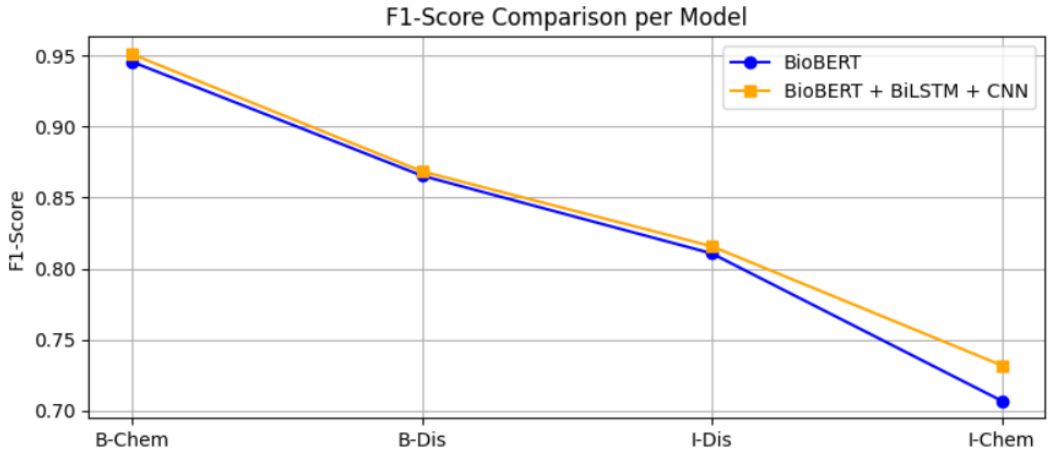
Gambar 2. Grafik Perbandingan Precision per Entitas antara Model BioBERT dan BioBERT + BiLSTM + CNN-Char

Gambar 2 menunjukkan perbandingan Precision antara model BioBERT dan BioBERT + BiLSTM + CNN-Char. Terlihat bahwa model modifikasi memiliki tingkat Precision yang lebih tinggi pada hampir semua label, terutama pada entitas I-Chemical, yang biasanya terdiri dari banyak token kompleks.



Gambar 3. Grafik Perbandingan Recall per Entitas antara Model BioBERT dan BioBERT + BiLSTM + CNN-Char

Gambar 3 menunjukkan perbandingan Recall. Model modifikasi menunjukkan peningkatan pada entitas bertingkat seperti I-Disease, menunjukkan kemampuan model untuk mendeteksi entitas yang lebih kompleks.



Gambar 4. Grafik Perbandingan F1-Score per Entitas antara Model BioBERT dan BioBERT + BiLSTM + CNN-Char

Gambar 4 mengilustrasikan adanya peningkatan nilai F1-Score pada model yang dimodifikasi untuk setiap label dibandingkan model standar. Hasil ini selaras dengan uji statistik yang mengonfirmasi adanya perbedaan signifikan dalam performa model.

Setiap label entitas utama divisualisasikan menggunakan tiga indikator evaluasi, yaitu Precision, Recall, dan F1-Score. Tujuan dari visualisasi ini adalah untuk menyediakan pemahaman yang lebih komprehensif terkait performa kedua model. Berdasarkan visualisasi yang ditampilkan, penerapan BiLSTM dan CNN-Char terbukti meningkatkan kemampuan model dalam menggeneralisasi dan mengenali entitas medis, terutama pada struktur token yang lebih kompleks dan beragam.

Gambar 2, 3, dan 4 menunjukkan peningkatan Precision, Recall, dan F1-Score pada model modifikasi (BioBERT + BiLSTM + CNN-Char) dibandingkan model dasar. Precision meningkat, terutama pada label I-Chemical, menandakan kemampuan model mengenali token kompleks dengan lebih akurat. Recall juga meningkat signifikan, terutama pada label I-Disease, yang menunjukkan sensitivitas model terhadap entitas tersembunyi dalam teks panjang. F1-Score yang lebih tinggi pada semua label mengonfirmasi bahwa kombinasi BiLSTM dan CNN-Char secara efektif meningkatkan keseimbangan antara Precision dan Recall, bukan hanya fluktuasi sementara.

G. Ringkasan Evaluasi Kinerja Model

Kinerja model dasar dan modifikasi dibandingkan menggunakan Accuracy, Precision, Recall, dan F1-Score. Tabel VIII berikut menguraikan hasil evaluasi.

TABEL VIII
HASIL EVALUASI AKHIR BERDASARKAN MODEL

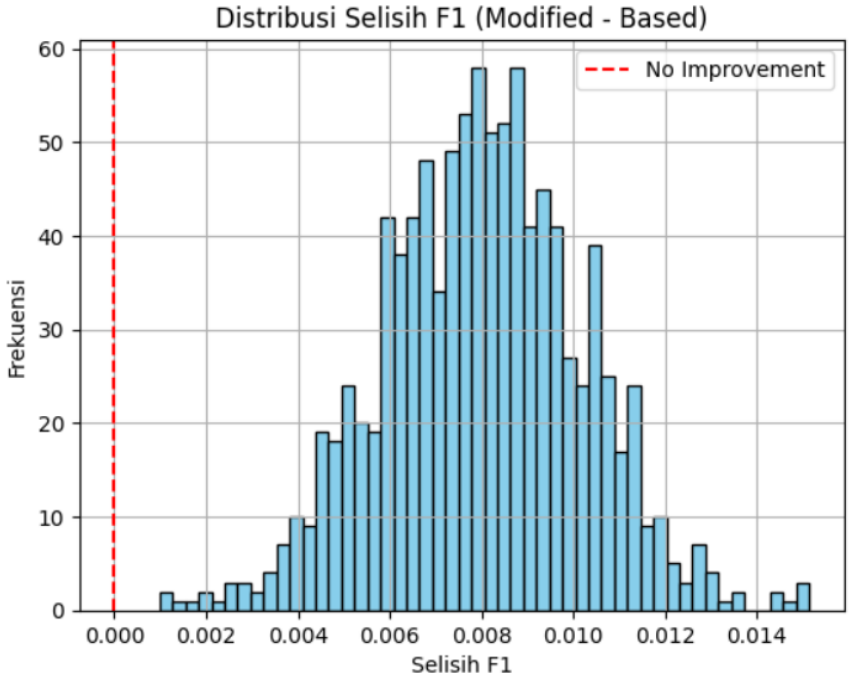
Model	Accuracy	Precision	Recall	F1-Score
BioBERT (Based Model)	0.9475	0.8310	0.8948	0.8597
BioBERT + BiLSTM-CNN-Char (Modified Model)	0.9503	0.8429	0.8964	0.8678

Model modifikasi menunjukkan peningkatan signifikan pada Accuracy, Precision, Recall, dan F1-Score dibanding BioBERT.

H. Uji Statistik Signifikansi

Pengujian bootstrap 1.000 iterasi digunakan untuk mengukur signifikansi peningkatan kinerja. Gambar 5 menunjukkan hasil distribusi selisih antara Modified Model dan Based Model. Pada tiap iterasi, dilakukan pemilihan acak dengan pengembalian terhadap label data uji, dan dihitung selisih F1-Score antara keduanya.

Menurut distribusi histogram, semua nilai selisih F1 berada di atas nol, dengan selisih rata-rata 0.00801. Ini menunjukkan bahwa Modified Model (BioBERT + BiLSTM-CNN-Char) secara konsisten lebih baik daripada model dasar. Selain itu, peningkatan kinerja adalah signifikan secara statistik pada tingkat signifikansi $\alpha = 0.05$, karena tidak ada iterasi bootstrap yang menghasilkan kinerja Modified Model lebih rendah dari Based Model. Hasil pengujian menunjukkan p-value = 0.00000.



Gambar 5. Distribusi Selisih F1-Score antara Modified Model dan Based Model. Berdasarkan uji Bootstrap sebanyak 1.000 iterasi. Garis merah vertikal menunjukkan batas "No Improvement"

I. Keterbatasan dan Diskusi

Meskipun temuan penelitian ini menunjukkan integrasi BiLSTM dan CNN-Char ke dalam BioBERT berhasil, ada beberapa keterbatasan:

1. Dataset BC5CDR terbatas pada dua jenis entitas, sehingga hasil mungkin tidak berlaku untuk skenario entitas lain seperti Gen atau Protein.
2. Evaluasi hanya dilakukan berdasarkan metrik F1-Score tanpa metrik tambahan seperti waktu inferensi atau konsumsi memori.
3. Uji signifikansi hanya dilakukan antar dua model, belum termasuk eksperimen tambahan ablation (misalnya hanya BiLSTM tanpa CNN-Char).

Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan integrasi dapat memperkuat arsitektur bahasa yang diajarkan model, terutama di bidang biomedis.

Dalam penelitian ini, fokus perbandingan dilakukan antara model BioBERT dan versi modifikasinya yang menggunakan arsitektur BiLSTM dan CNN-Char. Meskipun demikian, metode alternatif seperti SciBERT, BioALBERT, dan BiLSTM-CRF juga telah terbukti efektif dalam penerapan Named Entity Recognition (NER) untuk domain biomedis. Pemilihan BioBERT sebagai model dasar didasarkan pada stabilitas kinerjanya serta kesesuaiannya dengan dataset BC5CDR yang digunakan. Untuk memperoleh pemahaman yang lebih komprehensif terkait keunggulan dan keterbatasan masing-masing pendekatan dalam konteks medis yang kompleks, disarankan agar penelitian selanjutnya mengeksplorasi model lain dengan arsitektur yang lebih mutakhir.

Studi ini menciptakan model yang dapat diintegrasikan ke dalam sistem Electronic Health Record (EHR) untuk membantu proses ekstraksi entitas medis secara otomatis dan real-time. Hal ini akan

sangat membantu proses pengambilan keputusan medis, pembuatan indeks informasi pasien, dan pengolahan catatan klinis. Selain itu, peningkatan akurasi entitas kompleks seperti I-Chemical dan I-Disease memungkinkan hasil ekstraksi entitas digunakan dalam pembuatan knowledge graph di bidang biomedis. Ini dapat menjadi dasar untuk pengembangan sistem cerdas dalam riset medis lanjutan. Selain itu, struktur ini fleksibel dan dapat disesuaikan untuk mengidentifikasi berbagai jenis entitas dalam berbagai domain biomedis, seperti protein, gen, dan perawatan.

IV. KESIMPULAN

Penelitian ini bertujuan untuk meningkatkan kemampuan model BioBERT dalam mengenali entitas medis, khususnya entitas Chemical dan Disease, melalui integrasi BiLSTM dan CNN-Char. Hasil penelitian menunjukkan bahwa model modifikasi (BioBERT + BiLSTM + CNN-Char) secara signifikan meningkatkan performa model dalam pengenalan entitas medis dibandingkan model dasar (BioBERT). Peningkatan performa ini terlihat dari nilai Precision (dari 0.8310 menjadi 0.8429), Recall (dari 0.8948 menjadi 0.8964), dan F1-Score (dari 0.8597 menjadi 0.8678). Integrasi BiLSTM dan CNN-Char memberikan nilai tambah pada model dengan menangkap pola karakteristik dan konteks sekuensial antar token. Efektivitasnya terutama terlihat pada pengenalan entitas dengan struktur morfologis kompleks, seperti Chemical dan Disease. Penelitian ini berkontribusi pada komunitas NER biomedis dengan menyediakan pendekatan yang dapat diimplementasikan dalam sistem Electronic Health Record (EHR) untuk ekstraksi informasi medis secara otomatis, mendukung pembuatan knowledge graph biomedis, dan mengoptimalkan analisis teks medis. Model ini juga memiliki potensi untuk diadaptasi ke domain medis lainnya, seperti protein, gen, atau perawatan klinis. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi model lain, seperti SciBERT atau BioALBERT, dan memperluas cakupan entitas ke domain lain, serta menguji metode lain seperti CRF untuk meningkatkan performa model.

REFERENSI

[1] S. Rizal Sidiq and A. Salam, "SciBERT Optimisation for Named Entity Recognition on NCBI Disease Corpus with Hyperparameter Tuning," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 2, pp. 432–441, 2025.

[2] C. H. Wei, A. Allot, R. Leaman, and Z. Lu, "PubTator central: automated concept annotation for biomedical full text articles," *Nucleic Acids Res*, vol. 47, no. W1, pp. W587–W593, Jul. 2019.

[3] N. Perera, M. Dehmer, and F. Emmert-Streib, "Named Entity Recognition and Relation Detection for Biomedical Information Extraction," *Front Cell Dev Biol*, vol. 8, 2020.

[4] V. Yadav and S. Bethard, "A Survey on Recent Advances in Named Entity Recognition from Deep Learning models," *Association for Computational Linguistics*, pp. 2145–2158, Aug. 2018.

[5] S. Gajendran, M. D, and V. Sugumaran, "Character level and word level embedding with bidirectional LSTM – Dynamic recurrent neural network for biomedical named entity recognition from literature," *J Biomed Inform*, vol. 112, no. C, Dec. 2020.

[6] B. Song, F. Li, Y. Liu, and X. Zeng, "Deep learning methods for biomedical named entity recognition: a survey and qualitative comparison," *Brief Bioinform*, vol. 22, no. 6, Nov. 2021.

[7] L. Weber, M. Sänger, J. Münchmeyer, M. Habibi, U. Leser, and A. Akbik, "HunFlair: An Easy-to-Use Tool for State-of-the-Art Biomedical Named Entity Recognition," *Bioinformatics*, vol. 37, no. 17, pp. 2792–2794, Sep. 2021.

[8] J. Lee *et al.*, "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, Feb. 2020.

[9] I. Nejadgholi, K. C. Fraser, and B. De Bruijn, "Extensive Error Analysis and a Learning-Based Evaluation of Medical Entity Recognition Systems to Approximate User Experience," *Association for Computational Linguistics*, pp. 177–186, Jul. 2020.

[10] U. Naseem, M. Khushi, V. Reddy, S. Rajendran, I. Razzak, and J. Kim, "*BioALBERT: A Simple and Effective Pre-trained Language Model for Biomedical Named Entity Recognition*," in 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 2021, pp. 1–7.

[11] Y. Xu and Y. Chen, "Attention-based interactive multi-level feature fusion for named entity recognition," *Sci Rep*, vol. 15, no. 3069, Jan. 2025.

[12] V. Kocaman and D. Talby, "Accurate Clinical and Biomedical Named Entity Recognition at Scale," *Software Impacts*, vol. 13, 2022.

[13] Z. Han, S. Lin, Z. Huang, and C. Guo, "*Named Entity Recognition for Long COVID Biomedical Literature by Using Bert-BiLSTM-IDCNN-ATT-CRF Approach*," in Proceedings of the 2023 4th International Symposium on Artificial Intelligence for Medicine Science, Chengdu, China, 2024, pp. 1200–1205.

[14] M. Cho, J. Ha, C. Park, and S. Park, "Combinatorial feature embedding based on CNN and LSTM for biomedical named entity recognition," *J Biomed Inform*, vol. 103, 2020.

[15] X. Wang *et al.*, "Cross-type biomedical named entity recognition with deep multi-task learning," *Bioinformatics*, vol. 35, no. 10, pp. 1745–1752, May 2019.

[16] M. C. Durango, E. A. Torres-Silva, and A. Orozco-Duque, "Named Entity Recognition in Electronic Health Records: A Methodological Review," *Healthc Inform Res*, vol. 29, no. 4, pp. 286–300, Oct. 2023.

[17] L. Wang, H. Hao, X. Yan, T. H. Zhou, and K. H. Ryu, "From biomedical knowledge graph construction to semantic querying: a comprehensive approach," *Sci Rep*, vol. 15, no. 8523, Mar. 2025.

[18] D. N. Nicholson and C. S. Greene, "Constructing knowledge graphs and their biomedical applications," *Comput Struct Biotechnol J*, vol. 18, pp. 1414–1428, 2020.

[19] Á. Alonso Casero, Ó. Corcho, and C. Badenes-Olmedo, "Named Entity Recognition and Normalization in Biomedical Literature: A Practical Case in SARS-CoV-2 Literature," *E.T.S. de Ingenieros Informáticos (UPM)*, Jul. 2021.

[20] J. Li, A. Sun, J. Han, and C. Li, "*A Survey on Deep Learning for Named Entity Recognition*," in 2023 IEEE 39th International Conference on Data Engineering (ICDE), Anaheim, CA, USA, 2023, pp. 3817–3818.

[21] R. Prasanna Kumar, G. Bharathi Mohan, P. Srinivasan, and R. Venkatakrishnan, "*Transformer-Based Models for Named Entity Recognition: A Comparative Study*," in 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1–5.

[22] C. D. Nafanda and A. Salam, "Optimalisasi Model BioBERT untuk Pengenalan Entitas pada Teks Medis dengan Conditional Random Fields (CRF)," *Technology and Science (BITS)*, vol. 6, no. 4, pp. 2525–2534, Mar. 2025.