**Journal of Applied Informatics and Computing (JAIC)**
Vol.10, No.1, February 2026, pp. 955~963
e-ISSN: 2548-6861


# Suicidal Ideation Detection in Social Media using Optimized CNN-BiLSTM Architecture

**Hestiana Putri Novitasari [1]*, M. Arief Soeleman [2]*, Sifa Ayu Rosita Sari [3]*, Mamay Maida [4]***
* Teknik Informatika, Universitas Dian Nuswantoro
111202214243@mhs.dinus.ac.id [1], arief2802@dsn.dinus.ac.id [2], 111202214204@mhs.dinus.ac.id [3], 111202214216@mhs.dinus.ac.id [4]

**Article Info**

***Article history:***

Received 2025-12-03
Revised 2026-01-09
Accepted 2026-01-13

***Keyword:***

*CNN-BiLSTM,*
*Deep Learning,*
*Suicidal Ideation,*
*Text Classification,*
*Social Media.*

**ABSTRACT**

This research aims to develop an optimized hybrid deep learning model for detecting suicidal ideation from social media text. The growing volume of online discussions, particularly on platforms such as Reddit, provides valuable signals for early identification of individuals at risk; however, the linguistic characteristics of user-generated content are highly diverse and often noisy. To address this challenge, this study proposes an Optimized CNN-BiLSTM architecture enhanced with a dropout rate of 0.6 and a strategic training approach utilizing Early Stopping (patience=3) and a Learning Rate Scheduler (ReduceLROnPlateau) to prevent local minima and ensure convergence stability. The dataset used consists of 232,074 text entries with a balanced class distribution (50% suicide, 50% non-suicide) to ensure the validity of evaluation metrics and eliminate majority class bias. Experimental results demonstrate that the optimized model achieves an accuracy of 94.96%, precision of 95.70%, recall of 94.15%, and an F1-score of 94.92%, indicating a significant improvement over the baseline CNN-BiLSTM and single BiLSTM models. Furthermore, interpretability analysis via keyword visualization (Word Cloud) validates that the model effectively captures semantically relevant emotional expressions of despair. These findings suggest that the optimized hybrid architecture provides a robust and operationally viable approach for supporting real-time early-warning systems on social media platforms to facilitate timely mental health interventions.



## I. PENDAHULUAN

Media sosial telah menjadi ruang ekspresi bagi individu untuk membagikan pengalaman, perasaan, maupun tekanan psikologis yang mereka alami. *Platform* seperti Reddit, Twitter, dan forum daring lainnya menyediakan kanal bagi pengguna untuk menyampaikan emosi secara spontan dan anonim. Fenomena tersebut membuka peluang besar bagi para peneliti untuk menganalisis pola bahasa yang mencerminkan kondisi mental, termasuk indikasi ide bunuh diri. Mengingat tingginya angka kasus bunuh diri serta dampak serius yang ditimbulkannya, deteksi dini melalui pemanfaatan data digital menjadi langkah penting dalam upaya pencegahan dan intervensi secara cepat.

Proses identifikasi ide bunuh diri pada teks media sosial secara manual menghadapi berbagai tantangan, seperti volume data yang sangat besar, variasi bahasa informal, penggunaan metafora dan *slang*, serta ekspresi implisit yang sulit ditafsirkan. Keterbatasan tersebut menyebabkan metode tradisional kurang efektif dalam memproses data dalam skala besar. Oleh karena itu, pemanfaatan *machine learning* dan *deep learning* dalam deteksi kesehatan mental berbasis teks menjadi semakin relevan dan banyak dikembangkan dalam beberapa tahun terakhir.

Penelitian-penelitian awal mengenai deteksi ide bunuh diri umumnya menggunakan metode *machine learning* tradisional seperti Naïve Bayes, *Logistic Regression*, dan *Support Vector Machine* (SVM). Salah satu penelitian yang menjadi acuan adalah studi oleh Goni *et al.* [1], yang menerapkan SVM dengan fitur TF-IDF dan berhasil mencapai akurasi sebesar 93.85% dalam mendeteksi ide bunuh diri pada data media sosial. Meskipun metode tersebut cukup efektif, pendekatan

*machine learning* tradisional memiliki keterbatasan dalam memahami konteks kalimat yang bersifat sekuensial maupun hubungan semantik mendalam antar kata. Hal ini membuat model sulit menangkap makna implisit yang sering muncul dalam ekspresi pengguna media sosial.

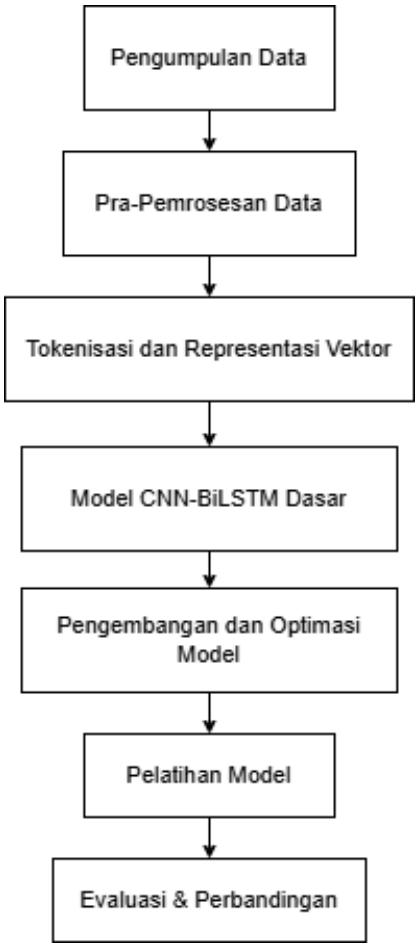
Perkembangan *deep learning* memberikan solusi terhadap keterbatasan tersebut melalui kemampuan dalam mengekstraksi representasi fitur yang lebih kompleks. Berbagai penelitian telah membuktikan bahwa model-model seperti *Convolutional Neural Network* (CNN), *Long Short-Term Memory* (LSTM), *Bidirectional LSTM* (BiLSTM), dan model *hybrid* mampu menghasilkan performa yang lebih baik dalam klasifikasi teks *suicidal*. Aldhyani *et al.* [2] menunjukkan bahwa pendekatan gabungan antara model *deep learning* dan *machine learning* mampu meningkatkan akurasi deteksi *suicidal ideation* secara signifikan. Penelitian oleh Renjith *et al.* [5] juga mengusulkan metode *ensemble deep learning* yang menangkap pola linguistik kompleks secara lebih efektif. Selain itu, penelitian terbaru oleh Bhuiyan *et al.* [8] mengonfirmasi bahwa integrasi CNN dan BiLSTM dapat memproses informasi spasial dan sekuensial secara lebih optimal. Meskipun penelitian-penelitian tersebut telah menetapkan fondasi kuat, tantangan mengenai stabilitas pelatihan pada dataset berskala besar dan risiko *overfitting* akibat karakteristik teks media sosial yang penuh *noise* masih menjadi kendala utama.

Penelitian ini hadir untuk mengisi celah tersebut (*research gap*) dengan mengusulkan arsitektur Optimized CNN-BiLSTM. Berbeda dengan pendekatan hibrida standar yang dilakukan oleh Bhuiyan et al. [8] atau Renjith et al. [5], studi ini melakukan optimasi eksplisit pada mekanisme regularisasi dan strategi pelatihan. Perbaikan tersebut mencakup penyederhanaan jumlah lapisan BiLSTM untuk efisiensi, peningkatan nilai *dropout* menjadi 0.6 untuk memperkuat ketahanan terhadap *noise*, serta penerapan *Learning Rate Scheduler* dan *Early Stopping* guna mencegah model terjebak dalam *local minima* dan mengalami *overfitting*. Strategi ini dirancang untuk menciptakan model hibrida yang lebih stabil dan memiliki kemampuan generalisasi lebih baik dibandingkan pendekatan berbasis SVM oleh Goni et al. [1] maupun model *baseline* internal.

Penelitian ini bertujuan meningkatkan kemampuan deteksi ide bunuh diri pada data media sosial melalui pemanfaatan arsitektur *deep learning* yang lebih adaptif. Selain itu, penelitian ini juga berupaya mengevaluasi sejauh mana strategi optimasi yang diterapkan dapat memperbaiki performa model dibandingkan pendekatan sebelumnya. Hasil penelitian diharapkan memberikan kontribusi akademis dengan memperkuat literatur mengenai deteksi *suicidal ideation* berbasis *deep learning* dan memberikan manfaat praktis sebagai dasar pengembangan sistem pemantauan kesehatan mental secara otomatis. Dengan pendekatan yang lebih *robust* dan adaptif terhadap variasi bahasa pengguna media sosial, penelitian ini berupaya menghadirkan solusi yang lebih efektif dan reliabel untuk identifikasi risiko ide bunuh diri di lingkungan digital.

## II. METODE

Metodologi penelitian ini dirancang untuk menghasilkan model deteksi ide bunuh diri berbasis Optimized CNN-BiLSTM yang mampu melakukan klasifikasi teks secara akurat pada data media sosial. Proses penelitian dilakukan melalui lima tahapan utama, yaitu pengumpulan *dataset*, pra-pemrosesan teks, tokenisasi dan representasi vektor, perancangan arsitektur model dasar, serta pengembangan model Optimized CNN-BiLSTM dengan parameter yang ditingkatkan. Seluruh proses mengikuti pendekatan umum dalam riset deteksi ide bunuh diri dengan NLP sebagaimana digunakan oleh Goni *et al.* [1], Aldhyani *et al.* [2], serta Bhuiyan *et al.* [8].




Gambar 1. Diagram Alur Penelitian

### A. Dataset

*Dataset* yang digunakan dalam penelitian ini diperoleh dari *platform* Kaggle, yaitu *dataset* Suicide Watch yang dikurasi oleh Nikhileswar Komati. *Dataset* ini merupakan kumpulan teks dari *subreddit* r/SuicideWatch, r/depression, dan r/teenagers, yang dikumpulkan melalui Pushshift API. Pemilihan subreddit r/SuicideWatch sebagai sumber utama kelas "suicide" didasarkan pada karakteristik platform tersebut sebagai komunitas daring di mana pengguna secara

terbuka mengekspresikan pikiran tentang mengakhiri hidup atau mencari dukungan krisis. Sementara itu, label "non-suicide" diambil dari r/teenagers dan r/depression untuk mencakup variasi bahasa emosional yang bersifat umum tanpa adanya indikasi niat bunuh diri yang eksplisit. Kriteria seleksi data dilakukan dengan menyaring postingan yang memiliki label kelas yang jelas dan menghapus entri yang mengandung konten multimedia atau hanya terdiri dari judul tanpa isi teks guna menjamin kualitas data yang diolah. Goni et al. [1], Aldhyani et al. [2], Chatterjee et al. [3], dan Renjith et al. [5] menekankan bahwa dataset ini menjadi pilihan utama karena ukurannya yang besar dan anotasi yang terstruktur dengan baik. Dataset ini berisi total 232.074 entri teks dengan distribusi kelas yang seimbang (50% per kelas).

TABEL I
DISTRIBUSI KELAS DATASET

| Label | Jumlah Data | Persentase (%) |
|---|---|---|
| Suicide (1) | 116.037 | 50% |
| Non-Suicide (0) | 116.037 | 50% |
| Total | 232.074 | 100% |

Setiap entri memiliki empat kolom utama, yaitu *text* sebagai isi posting asli dari Reddit, *class* sebagai label kategori awal (*suicide*/ *non-suicide*), *cleaned_text* sebagai teks hasil pra-pemrosesan, dan *label* sebagai representasi numerik dari kelas. Struktur lengkap *dataset*, termasuk tipe data dan jumlah nilai *non-null* pada masing-masing kolom, ditampilkan pada Tabel II untuk memberikan gambaran tentang integritas data dan memastikan *dataset* siap untuk tahap pemrosesan lanjutan.

TABEL II
STRUKTUR DATASET SUICIDE DETECTION

| No | Nama | Tipe Data | Jumlah Non-Nul | Deskripsi Singkat |
|---|---|---|---|---|
| 1 | text | Object | 232.074 | Teks asli dari posting Reddit |
| 2 | class | Object | 232.074 | Label awal (suicide / non-suicide) |
| 3 | cleaned_text | Object | 232.074 | Hasil preprocessing (lowercase, cleaning, stopwords) |
| 4 | label | Int64 | 232.074 | Label numerik (1=suicide, 0=non-suicide) |

Renjith *et al*. [5], Bhuiyan *et al*. [6], dan Vipin Kumar Sahu & Gupta [12] menunjukkan bahwa pemisahan dataset menggunakan stratified split, dengan 80% data untuk pelatihan dan 20% untuk pengujian, membantu menjaga proporsi kelas tetap seimbang selama pelatihan dan evaluasi model deteksi *suicidal ideation.*

Meskipun isu kesehatan mental di media sosial sering kali menghasilkan data yang tidak seimbang (*imbalanced*), dataset ini telah dikurasi secara sistematis untuk memiliki distribusi kelas yang seimbang (*perfectly balanced*) dengan proporsi 50:50. Hal ini krusial agar metrik akurasi yang dihasilkan tidak bias dan benar-benar merepresentasikan kemampuan diskriminatif model.

### *B. Pra-pemrosesan Data*

Tahap pra-pemrosesan teks dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam pelatihan model. Data yang berasal dari media sosial seperti Reddit umumnya mengandung banyak *noise* berupa URL, simbol, huruf acak, serta penggunaan bahasa informal yang dapat menurunkan efektivitas model apabila tidak dibersihkan. Goni et al. [1], Aldhyani et al. [2], Bhuiyan et al. [6], dan Vipin Kumar Sahu & Gupta [12] menunjukkan bahwa proses pra-pemrosesan yang sistematis merupakan langkah penting dalam deteksi *suicidal ideation* melalui analisis teks sosial.

Setiap teks pada kolom *text* dibersihkan dari URL menggunakan ekspresi reguler. Selanjutnya, dilakukan penghapusan karakter non-alfabet dan tanda baca untuk menangani *noise* berupa simbol-simbol yang tidak relevan. Langkah ini secara teknis menangani penggunaan kontraksi atau *slang* yang umum di media sosial (seperti mengubah "don't" menjadi "dont" atau menghapus tanda kutip) guna menyeragamkan variasi linguistik agar model dapat memproses kata dasar dengan lebih konsisten. Seluruh teks kemudian dikonversi ke huruf kecil (*lowercase*) untuk menyeragamkan format dan mengurangi variansi kosakata.

Tahap tokenisasi dilakukan secara spesifik menggunakan *Keras Tokenizer* yang memecah teks menjadi unit-unit kata (token) berdasarkan pemisahan spasi. Bersamaan dengan itu, dilakukan penghapusan *stopwords* bahasa Inggris menggunakan korpus NLTK untuk menghilangkan kata-kata fungsional yang memiliki frekuensi tinggi namun minim makna semantik. Kata-kata dengan panjang hanya satu karakter juga dihapus karena tidak memberikan kontribusi signifikan terhadap konteks klasifikasi. Hasil akhir dari tahap ini disimpan dalam kolom *cleaned_text*, yang menjadi representasi teks utama untuk proses *embedding* dan pembelajaran mesin. Tahap pra-pemrosesan ini memastikan bahwa model menerima data yang lebih bersih, ringkas, dan bebas dari elemen yang tidak relevan, sehingga meningkatkan kemampuan model dalam menangkap pola linguistik terkait ide bunuh diri.

### *C. Tokenisasi dan Padding*

Tahap tokenisasi dilakukan menggunakan *Keras Tokenizer* dengan membatasi jumlah kosakata maksimal pada 5.000 kata untuk menjaga efisiensi komputasi dan mencegah *overfitting* pada kata-kata yang jarang muncul. Setiap teks dikonversi menjadi urutan indeks numerik dan distandarkan panjangnya melalui teknik *padding* dan *truncation*. Panjang sekuens maksimum ditetapkan sebesar 100 token dengan konfigurasi *post-padding*, yang dipilih berdasarkan distribusi panjang teks pada dataset untuk memastikan informasi krusial

di akhir postingan tidak terpotong. Detail konfigurasi ini dirangkum dalam Tabel III.

### *D. Model CNN-BiLSTM Dasar*

Tahap awal pengembangan model dimulai dengan menetapkan batasan pada proses tokenisasi untuk mengelola kompleksitas fitur. Bhuiyan et al. [8] dan Vipin Kumar Sahu & Gupta [12] menekankan bahwa pembatasan jumlah kosakata, seperti penggunaan maksimal 5.000 kata unik pada *Keras Tokenizer*, sangat krusial untuk menjaga efisiensi komputasi model serta memitigasi risiko *overfitting* yang disebabkan oleh kemunculan kata-kata langka yang tidak representatif terhadap pola ide bunuh diri secara umum. Teks yang telah melalui tahap pembersihan kemudian ditransformasikan menjadi urutan indeks numerik dan distandarkan dimensinya melalui teknik *padding* serta *truncation* hingga mencapai panjang tetap 100 token. Sebagaimana dijelaskan oleh Renjith et al. [5] serta Vipin Kumar Sahu & Gupta [12], standarisasi panjang input ini merupakan prosedur standar dalam pemodelan saraf berbasis CNN dan LSTM guna memastikan konsistensi bentuk data saat diproses oleh lapisan *embedding*. Detail parameter ini dirangkum dalam Tabel III.

TABEL III
PARAMETER TOKENISASI DAN PADDING

| Parameter | Nilai |
|---|---|
| Vocabulary Size | 5.000 |
| Maksimal Panjang | 100 token |
| Padding | Post |
| Truncating | Post |
| OOV Token | <unk> |

Model dasar CNN-BiLSTM ini dikembangkan untuk berfungsi sebagai *baseline* atau standar perbandingan guna mengukur secara objektif sejauh mana peningkatan performa yang dapat dihasilkan melalui proses optimasi yang diusulkan. Dalam arsitektur hibrida ini, lapisan *Convolutional Neural Network* (CNN) berperan vital dalam mengekstraksi fitur spasial dan menangkap pola lokal, seperti kombinasi kata kunci atau *n-gram* yang sensitif terhadap konten emosional. Hasil ekstraksi tersebut kemudian diteruskan ke lapisan *Bidirectional Long Short-Term Memory* (BiLSTM) yang memiliki kemampuan untuk mempelajari konteks ketergantungan jangka panjang dari dua arah secara simultan (maju dan mundur). Integrasi ini memungkinkan model untuk tidak hanya mengenali kata-kata negatif secara terisolasi, tetapi juga memahami alur narasi dan konteks sekuensial dari pernyataan pengguna. Renjith et al. [5], Bhuiyan et al. [8], serta Vipin Kumar Sahu & Gupta [12] menunjukkan bahwa sinergi antara kemampuan ekstraksi fitur lokal CNN dan pemahaman konteks global BiLSTM sangat efektif dalam mendeteksi kompleksitas linguistik pada teks yang mengandung *suicidal ideation*. Detail konfigurasi arsitektur dasar tersebut disajikan pada Tabel IV.

TABEL IV
ARSITEKTUR MODEL CNN-BILSTM DASAR

| Layer | Konfigurasi | Deskripsi Fungsi |
|---|---|---|
| Embedding | 100 dim | Representasi vektor kata dari 5.000 kosakata. |
| Conv1D | 128 filter, kernel 5, ReLU | Ekstraksi fitur spasial dan pola lokal *n-gram*. |
| MaxPooling1D | Pool size 4 | Reduksi dimensi fitur dan pengambilan informasi dominan. |
| BiLSTM | 64 unit | Menangkap dependensi konteks sekuensial dua arah. |
| Dropout | 0.5 | Mekanisme regularisasi awal untuk mencegah *overfitting*. |
| Dense | 32 unit, ReLU | Lapisan terhubung penuh untuk integrasi fitur. |
| Output | 1 unit, Sigmoid | Klasifikasi biner untuk probabilitas ide bunuh diri. |

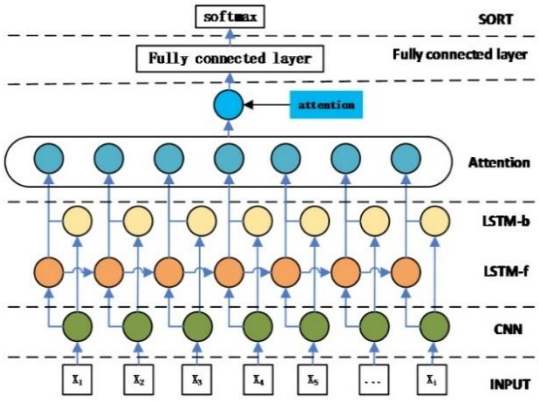
### *E. Optimasi Arsitektur*

Optimasi dilakukan untuk meningkatkan stabilitas dan performa model dengan mengatasi kendala konvergensi yang sering muncul pada arsitektur hibrida yang kompleks, terutama saat menangani karakteristik dataset media sosial yang memiliki tingkat ambiguitas linguistik dan *noise* yang tinggi. Proses ini mencakup peningkatan nilai *dropout* menjadi 0.6 untuk memperkuat mekanisme regulasi, yang secara teknis memaksa model agar tidak bergantung pada neuron tertentu secara berlebihan sehingga tercipta representasi fitur yang lebih kuat dan mampu tergeneralisasi dengan baik pada data uji. Selain itu, dilakukan penyederhanaan pada lapisan BiLSTM untuk mengurangi beban komputasi tanpa mengurangi kemampuan penangkapan konteks sekuensial, yang dipadukan dengan strategi *Early Stopping* (*patience* = 3) untuk menghentikan pelatihan secara otomatis saat performa pada data validasi mencapai titik jenuh, guna menghindari risiko *overfitting*. Aspek krusial lainnya dalam optimasi ini adalah pengintegrasian *Learning Rate Scheduler* melalui mekanisme *ReduceLROnPlateau* (faktor 0.1) yang berfungsi mencegah model terjebak dalam *local minima* selama proses pencarian bobot optimal; teknik ini memungkinkan transisi informasi dari ekstraksi pola lokal oleh lapisan CNN menuju pemahaman urutan temporal oleh BiLSTM berjalan lebih harmonis tanpa risiko fluktuasi gradien yang ekstrem. Pendekatan sistematis ini, sebagaimana didukung oleh penelitian Renjith et al. [5], Bhuiyan et al. [8], serta Vipin Kumar Sahu & Gupta [12], memastikan bahwa model yang dihasilkan tidak hanya unggul dalam akurasi statistik tetapi juga memiliki stabilitas operasional yang lebih reliabel dalam mendeteksi *suicidal ideation*. Melalui integrasi berbagai teknik penyempurnaan ini, model akhir mampu mencapai titik konvergensi yang lebih cepat dengan nilai *error* yang minimal, sehingga menjamin konsistensi klasifikasi pada

berbagai variasi teks emosional yang heterogen. Perbandingan detail antara arsitektur *baseline* dan *optimized* dirangkum dalam Tabel V.

TABEL V
PERBANDINGAN ARSITEKTUR BASELINE VS OPTIMIZED CNN-BILSTM

| Komponen | Baseline | Optimasi |
|---|---|---|
| Dropout | 0.5 | 0.6 |
| Jumlah BiLSTM | 1 layer | 1 layer (lebih ringan) |
| Dense Layer | 1 hidden | 1 hidden (32 unit) |
| Early Stopping | Tidak ada | Ada (patience = 3) |
| Stabilitas Training | Sedang | Lebih Stabil |
| LR Scheduler | Tidak ada | ReduceLROnPlateau (Factor 0.1) |

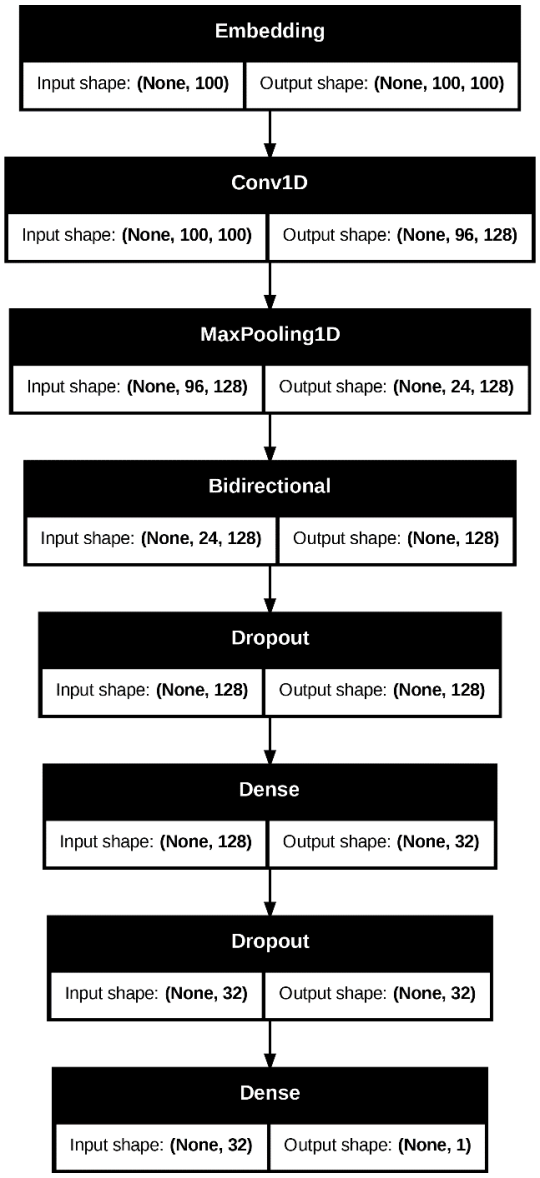



Gambar 2. Diagram Arsitektur Model Hybrid Optimized CNN-BiLSTM oleh Jin, Liu & Dai [29]

## F. Pelatihan Model

Goni et al. [1] dan Aldhyani et al. [2] menegaskan bahwa penggunaan *optimizer* Adam dan fungsi *binary crossentropy* merupakan praktik standar dalam pelatihan model klasifikasi biner berbasis NLP. Pelatihan dilakukan menggunakan *batch size* 64 untuk menyeimbangkan stabilitas *gradient* dan efisiensi memori, serta *validation split* sebesar 10% untuk pemantauan performa secara *real-time*.

Aspek utama dalam strategi pelatihan ini adalah pengintegrasian *Early Stopping* yang dikonfigurasi untuk memulihkan bobot terbaik (*restore best weights*). Mekanisme ini memastikan bahwa model akhir yang disimpan adalah model dengan nilai *validation loss* terendah, sehingga generalisasi terhadap data baru tetap terjaga. Konfigurasi lengkap mengenai *hyperparameter* dan strategi optimasi yang diterapkan dalam penelitian ini dirangkum secara mendalam pada Tabel VI.




Gambar 3. Detail Layer dan Transformasi Shape Output pada Arsitektur CNN-BiLSTM yang Dioptimalkan.

TABEL VI
KONFIGURASI HYPERPARAMETER DAN STRATEGI OPTIMASI MODEL

| Parameter | Konfigurasi / Nilai | Fungsi Optimasi |
|---|---|---|
| Optimizer | Adam | Optimasi *stochastic gradient descent* adaptif. |
| Loss Function | Binary Crossentropy | Standar klasifikasi biner untuk teks *suicidal*. |
| Batch Size | 64 | Penyeimbang efisiensi memori dan stabilitas *gradient*. |
| Dropout Rate | 0.6 | Regulasi ketat untuk mencegah *overfitting*. |
| Learning Rate Scheduler | ReduceLROn Plateau | Menurunkan *learning rate* (faktor 0.1) saat *loss* stagnan. |

| | | |
|---|---|---|
| Early Stopping | Patience = 3 | Menghentikan *training* dan memulihkan bobot terbaik. |
| Max Sequence Length | 100 token | Standarisasi input melalui *padding* dan *truncation*. |
| Validation Split | 0.1 (10%) | Pemantauan performa pada data yang tidak terlihat. |

### G. Evaluasi Model

Evaluasi model dilakukan menggunakan akurasi, *presisi*, *recall*, dan F1-score. Aldhyani *et al*. [2], Renjith *et al*. [5], serta Levkovich & Omar [10] dan Vipin Kumar Sahu & Gupta [12] menekankan bahwa recall merupakan metrik utama dalam deteksi suicidal ideation, karena model harus mampu meminimalkan kesalahan dalam mengidentifikasi kelas positif (suicide).

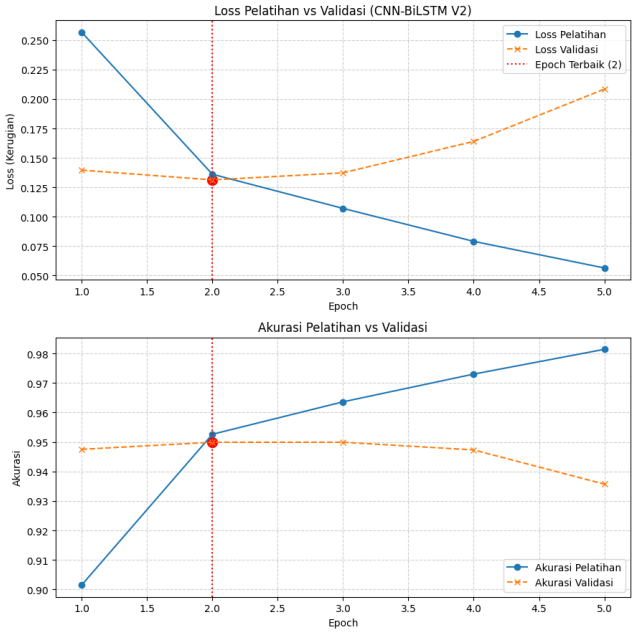
## III. HASIL DAN PEMBAHASAN

Hasil eksperimen yang diperoleh dari proses pelatihan model pada *dataset* Suicide Watch, diikuti analisis mendalam terhadap performa model, perilaku selama proses pelatihan, dan keunggulan arsitektur Optimized CNN-BiLSTM dibandingkan model *baseline*. Pembahasan ini juga dihubungkan dengan temuan penelitian terdahulu untuk menunjukkan kontribusi dan konsistensi model terhadap pendekatan yang digunakan di bidang deteksi ide bunuh diri berbasis teks.

### A. Hasil Pelatihan Model

Proses pelatihan dilakukan menggunakan konfigurasi *hyperparameter* yang telah dijelaskan sebelumnya, dengan pembagian data 80% untuk pelatihan dan 20% untuk pengujian. Selama proses pelatihan, model menunjukkan penurunan *training loss* yang konsisten pada setiap *epoch*, namun *validation loss* mulai meningkat setelah *epoch*-2. Pola ini menandakan bahwa model mulai memasuki fase *overfitting* setelah titik tersebut. Untuk mengatasi hal ini, mekanisme *Early Stopping* diterapkan dan secara otomatis menghentikan proses pelatihan pada *epoch* ke-5, serta mengembalikan bobot terbaik dari *epoch* ke-2, yaitu titik ketika *validation loss* berada pada nilai terendah

Pola konvergensi ini sejalan dengan temuan Renjith *et al.* [5], yang menjelaskan bahwa kombinasi CNN dan BiLSTM mampu mencapai stabilitas pembelajaran lebih cepat dibandingkan model RNN yang lebih kompleks. Selain itu, fenomena pelatihan yang berhenti lebih awal juga konsisten dengan penelitian Bhuiyan *et al.* [8], yang melaporkan bahwa arsitektur *hybrid* CNN-BiLSTM mampu menangkap pola spasial dan temporal secara efisien sehingga bobot optimal sering ditemukan pada *epoch* awal.




Gambar 4. Learning Curve (Training vs Validation Loss vs Accuracy)

### B. Hasil Evaluasi Model

Evaluasi terhadap model dilakukan menggunakan metrik akurasi, *presisi*, *recall*, dan F1-score. Nilai evaluasi menunjukkan bahwa model Optimized CNN-BiLSTM mencapai akurasi sebesar 94.96%, disertai nilai *presisi* sebesar 95.70%, *recall* 94.15%, serta F1-score 94.92%. Nilai *recall* yang tinggi mengindikasikan bahwa model memiliki kemampuan untuk mendeteksi sebagian besar teks yang benar-benar masuk dalam kategori *suicidal ideation*, sesuatu yang sangat penting karena kesalahan dalam bentuk *false negative* dapat berimplikasi serius dalam konteks intervensi kesehatan mental. Keseimbangan antara *presisi* dan *recall* tercermin dari nilai F1-score yang tinggi.

Kinerja model ini menunjukkan peningkatan yang cukup signifikan dibanding *baseline* CNN-BiLSTM yang sebelumnya digunakan. Hal ini sejalan dengan temuan Aldhyani *et al.* [2] dan Sharma & Karwasra [15], yang menekankan bahwa pendekatan *hybrid* biasanya lebih stabil dan efektif pada data teks emosional yang tidak beraturan, terutama ketika data bersumber dari *platform* seperti Reddit atau Twitter.

TABEL VII
EVALUASI MODEL

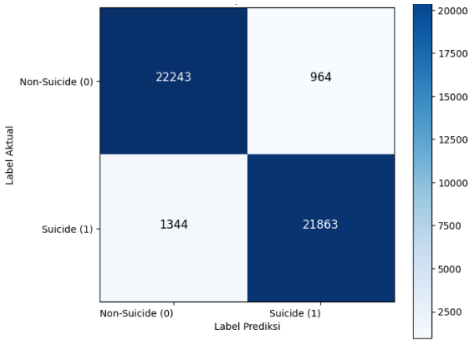
| **Metrik** | **Baseline SVM** | **BiLSTM Tunggal** | **Optimized CNN BiLSTM** |
|---|---|---|---|
| Akurasi | 93.85% | 94.81% | 94.96% |
| Presisi | 93.86% | 95.72% | 95.70% |
| Recall | 93.85% | 93.81% | 94.15% |
| F1-Score | 93.85% | 94.76% | 94.92% |

### C. Analisis Confusion Matrix

Analisis *confusion matrix* memberikan gambaran lebih rinci mengenai kemampuan model dalam membedakan teks *suicidal* dan *non-suicidal*. Model menunjukkan jumlah *True*

*Positive* yang tinggi, menandakan sensitivitas yang baik terhadap teks berbahaya. Selain itu, jumlah *True Negative* juga cukup besar sehingga model tidak cenderung memberikan prediksi positif secara berlebihan. Jumlah *False Negative* yang rendah menjadi indikator penting bahwa model berhasil mengenali sebagian besar teks dengan risiko tinggi yang seharusnya tidak terlewatkan.

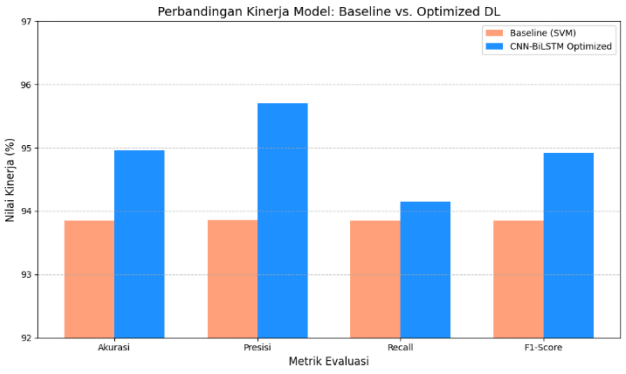
Perilaku *confusion matrix* seperti ini sesuai dengan studi Chatterjee *et al.* [3] dan Renjith *et al.* [5], yang menyatakan bahwa struktur bimodal CNN-BiLSTM sangat efektif dalam mengidentifikasi pola linguistik yang mengekspresikan kondisi psikologis tertentu. Dalam konteks deteksi ide bunuh diri, tingkat kesalahan *false negative* yang rendah sangat signifikan karena memastikan bahwa sistem dapat diandalkan untuk memberikan peringatan dini. Dominasi nilai pada diagonal utama *matrix* ini membuktikan bahwa model memiliki tingkat kekeliruan yang sangat rendah dalam mengkategorikan sentimen krisis, sehingga meminimalkan risiko pengabaian terhadap individu yang benar-benar membutuhkan bantuan darurat.




Gambar 5. Confusion Matrix

## D. Perbandingan Baseline vs Optimasi CNN-BiLSTM

Perbandingan kinerja antara model *baseline* CNN-BiLSTM dan model Optimized CNN-BiLSTM dilakukan untuk melihat dampak dari penyesuaian arsitektur dan *hyperparameter* yang diterapkan dalam penelitian ini. Renjith *et al.* [5] dan Bhuiyan *et al.* [8] menunjukkan bahwa konfigurasi umum CNN dan BiLSTM banyak diterapkan sebagai model baseline dalam penelitian deteksi ide bunuh diri di media sosial. Sementara itu, model yang dioptimasi memanfaatkan *dropout* yang lebih tinggi, penyederhanaan jumlah unit BiLSTM, serta *Early Stopping* untuk mencegah *overfitting*. Hasil evaluasi menunjukkan bahwa model yang dioptimasi memberikan peningkatan akurasi dan stabilitas prediksi dibanding *baseline*. Seperti ditampilkan pada Tabel VII, akurasi meningkat dari 93.85% menjadi 94.96%, sedangkan *presisi* dan F1-score juga mengalami peningkatan yang cukup signifikan. Temuan ini menunjukkan bahwa optimasi arsitektur efektif dalam meningkatkan kemampuan model dalam mendeteksi teks yang mengandung indikasi bunuh diri.

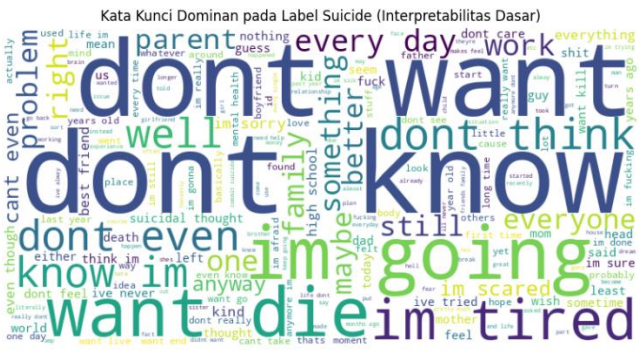



Gambar 6. Diagram Perbandingan Baseline vs Optimasi

## E. Pembahasa Hasil Penelitian

Hasil pengujian menunjukkan bahwa model Optimized CNN-BiLSTM memberikan peningkatan performa signifikan dibandingkan model baseline, terutama pada akurasi, presisi, dan F1-score. Peningkatan ini terlihat konsisten pada dataset berukuran besar dan heterogen, sehingga menunjukkan bahwa optimasi arsitektur berhasil mengurangi pengaruh noise dan variasi bahasa dalam teks. Renjith et al. [5] dan Vipin Kumar Sahu & Gupta [12] menegaskan bahwa model hybrid berbasis CNN-BiLSTM efektif dalam menganalisis risiko bunuh diri, sementara Goni et al. [1] dan Chatterjee et al. [3] menunjukkan bahwa ketidakseragaman karakteristik pada data media sosial turut memengaruhi performa model.

Selain evaluasi kuantitatif, pengamatan kualitatif menunjukkan bahwa model mampu menangkap ekspresi emosional yang tersirat, termasuk bentuk bahasa implisit yang sering digunakan oleh pengguna *platform* seperti Reddit. Model juga menunjukkan kemampuan membedakan antara kalimat bernada negatif yang bersifat umum dengan pesan yang mengandung indikasi ide bunuh diri, sehingga *presisi* dapat dipertahankan dengan baik. Hal ini mendukung temuan Grimland *et al.* [23] mengenai pentingnya sensitivitas terhadap konteks linguistik dalam deteksi risiko bunuh diri.




Gambar 7. Visualisasi Kata Kunci Dominan pada Label Suicide (Interpretabilitas Dasar)

Analisis interpretabilitas melalui visualisasi kata kunci pada Gambar 7 memperkuat temuan kualitatif tersebut. Kata-kata yang mendominasi seperti *"don't want"*, *"im tired"*, dan

*"die"* menunjukkan bahwa model secara cerdas mengenali pola linguistik yang merepresentasikan keputusasaan. Kemampuan model untuk memprioritaskan kata-kata bermakna emosional ini, dan bukan sekadar kata negatif umum, menjelaskan mengapa metrik *Recall* dan *F1-Score* pada model yang dioptimasi lebih unggul dibandingkan model *baseline*.

Secara keseluruhan, model Optimized CNN-BiLSTM terbukti efektif dan stabil dalam mendeteksi ide bunuh diri berbasis teks. Levkovich & Omar [10] menunjukkan bahwa model ini memiliki kinerja yang kompetitif dibandingkan beberapa pendekatan mutakhir seperti BERT, namun tetap mempertahankan efisiensi komputasi yang lebih baik. Hasil ini menunjukkan bahwa model *hybrid* yang dioptimasi dapat menjadi alternatif yang layak untuk implementasi sistem deteksi dini pada *platform* media sosial berskala besar.

### *F. Implikasi Praktis, Isu Etika, dan Keterbatasan*

*1) Interpretasi Fitur dan Validasi Semantik:* Penalaran ilmiah di balik pemilihan arsitektur hybrid ini didasarkan pada kemampuan model dalam mengenali pola bahasa yang sensitif. Berdasarkan visualisasi pada Gambar 7 (Word Cloud), model terbukti tidak hanya mengenali kata kunci negatif secara acak, tetapi fokus pada ekspresi keputusasaan yang mendalam seperti *"tired"*, *"don't want"*, dan *"die"*. Hal ini memberikan validasi semantik bahwa model memiliki interpretasi yang tepat terhadap konteks kesehatan mental, bukan sekadar menghafal dataset.

*2) Implikasi Praktis dan Operasional:* Secara operasional, model *Optimized CNN-BiLSTM* ini dapat diintegrasikan sebagai sistem peringatan dini (*early warning system*) pada moderator platform media sosial. Dengan nilai *Recall* sebesar 94.15%, sistem ini mampu memprioritaskan postingan dengan risiko tinggi untuk segera mendapatkan penanganan. Secara teknis, integrasi dapat dilakukan melalui API platform yang mengirimkan teks postingan ke model untuk diklasifikasikan secara *real-time*, di mana skor probabilitas di atas ambang batas tertentu akan memicu notifikasi otomatis kepada tim penanganan krisis atau layanan kesehatan mental terkait.

*3) Aspek Etika dan Privasi:* Penelitian ini dilakukan dengan menjunjung tinggi prinsip privasi digital. Dataset yang digunakan bersifat anonim dan tidak mencantumkan identitas pengguna (*username*) atau metadata pribadi lainnya. Selain itu, sistem ini dirancang sebagai alat bantu pendukung (*decision support tool*) bagi pakar kesehatan mental, bukan sebagai pengganti diagnosa medis profesional. Penggunaan sistem otomatis ini harus tetap berada di bawah pengawasan manusia guna menghindari dampak psikologis dari kesalahan prediksi (*false positive/negative*), di mana intervensi manusia (*human-in-the-loop*) menjadi prasyarat sebelum tindakan medis dilakukan.

*4) Keterbatasan Penelitian:* Meskipun mencapai performa yang tinggi, penelitian ini memiliki keterbatasan pada ketergantungan dataset berbahasa Inggris dan platform Reddit yang memiliki karakteristik anonimitas tinggi. Generalisasi model pada platform lain seperti Twitter atau Instagram yang menggunakan bahasa lebih informal dan multimedia, serta adaptasi pada bahasa lokal (seperti Bahasa Indonesia), merupakan tantangan yang memerlukan penelitian lebih lanjut.

## IV. KESIMPULAN

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi kinerja arsitektur *Deep Learning* hibrida yang dioptimalkan, yaitu Optimized CNN-BiLSTM, untuk tugas krusial deteksi ide bunuh diri dari data tekstual media sosial. Kontribusi utama dari penelitian ini terletak pada implementasi strategi regularisasi yang ketat dan optimasi pelatihan, khususnya mekanisme Early Stopping dan peningkatan *Dropout* tinggi, yang berhasil mengatasi masalah *overfitting* yang umum terjadi pada model *Deep Learning* awal (CNN-BiLSTM V1).

Hasil eksperimen menunjukkan bahwa arsitektur yang diusulkan mencapai kinerja *State-of-the-Art* (SOTA) baru pada *benchmark* yang digunakan. Model *Optimized* CNN-BiLSTM V2 mencatat metrik evaluasi tertinggi: Akurasi 94.96%, Presisi 95.70%, *Recall* 94.15%, dan F1-*Score* 94.92%. Kinerja ini menunjukkan peningkatan signifikan, yaitu 1.11 percentage points (pp), melampaui kinerja tertinggi yang ditetapkan oleh model *Machine Learning* tradisional (SVM) pada 93.85%.

Secara metodologis, efektivitas optimasi terbukti dari *Learning Curve*, di mana mekanisme *Early Stopping* berhasil mengidentifikasi dan memulihkan bobot terbaik model pada *epoch* ke-2, sebelum *validation loss* mulai meningkat tajam. Hal ini memvalidasi bahwa model mampu belajar fitur kontekstual dan sekuensial yang kuat dengan sangat cepat sambil mempertahankan kemampuan generalisasi yang tinggi. Untuk penelitian di masa mendatang, disarankan untuk melanjutkan eksplorasi ke arah model *hybrid* lanjutan, seperti penggabungan fitur psikolinguistik eksternal (misalnya, LIWC) dengan representasi fitur dari CNN-BiLSTM, serta fokus pada teknik *Explainable AI* (XAI) untuk memberikan interpretasi linguistik yang lebih mendalam terhadap prediksi model.

## DAFTAR PUSTAKA

[1] Goni, A., Jahangir, Md. U. F., Chowdhury, R. R., Akter, F., & Hussain, K. (2024). Detection of Suicidal Ideation on Social Media using Machine Learning Approaches. *International Journal of Computer Applications*, *186*(51), 8–14. https://doi.org/10.5120/ijca2024924161.

[2] Aldhyani, T. H. H., Alsubari, S. N., Alshebami, A. S., Alkahtani, H., & Ahmed, Z. A. T. (2022). Detecting and Analyzing Suicidal Ideation on Social Media Using Deep Learning and Machine Learning Models. *International Journal of Environmental Research and Public Health*, *19*(19), 12635. https://doi.org/10.3390/ijerph191912635.

[3] M. Chatterjee, P. Samanta, P. Kumar and D. Sarkar, "Suicide Ideation Detection using Multiple Feature Analysis from Twitter Data," 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-6, doi: 10.1109/DELCON54057.2022.9753295.

[4] Babulal, K. S., & Nayak, B. K. (2022). *Suicidal Analysis on Social Networks Using Machine Learning* (pp. 230–247). https://doi.org/10.4018/978-1-6684-3533-5.ch012

[5] Renjith, S., Abraham, A., Jyothi, S. B., Chandran, L., & Thomson, J. (2022). An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms. *Journal of King Saud University - Computer and Information Sciences*, *34*(10), 9564–9575. https://doi.org/10.1016/j.jksuci.2021.11.010.

[6] Islam Bhuiyan, M., Pahang Al-Sultan Abdullah Pahang, M., Nur Shazwani Kamarudin, M., & Hafieza Ismail, N. (n.d.). *Detecting Suicidal Ideation in Text with Interpretable Deep Learning: A CNN-BiGRU with Attention Mechanism*.

[7] G. Ananthakrishnan, A. K. Jayaraman, T. E. Trueman, S. Mitra, A. A. K and A. Murugappan, "Suicidal Intention Detection in Tweets Using BERT-Based Transformers," 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2022, pp. 322-327, doi: 10.1109/ICCCIS56430.2022.10037677.

[8] Islam Bhuiyan, M., Shazwani Kamarudin, N., Hafieza Ismail, N., & Tasnim Raya, J. (2025). Enhanced Suicidal Ideation Detection from Social Media using a CNN-BiLSTM Hybrid Model. *Semarak International Journal of Applied Psychology Journal Homepage*, *8*, 31–45. https://doi.org/10.37934/sijap.8.1.3145.

[9] Dzakwan Diash, H., Nathania, V., & Idhom, M. (2025). Application of CNN-BiLSTM Algorithm for Ethereum Price Prediction. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 9, Issue 4). http://jurnal.polibatam.ac.id/index.php/JAIC.

[10] Levkovich, I., & Omar, M. (2024). Evaluating of BERT-based and Large Language Mod for Suicide Detection, Prevention, and Risk Assessment: A Systematic Review. *Journal of Medical Systems*, *48*(1), 113. https://doi.org/10.1007/s10916-024-02134-3.

[11] Yang, Z., Leonard, R., Tran, H., Driscoll, R., & Davis, C. (2025). *Detection of Suicidal Risk on Social Media: A Hybrid Model*. http://arxiv.org/abs/2505.23797.

[12] Vipin Kumar Sahu, & Dr. Brijendra Gupta. (2025). Prediction of Suicidal Ideation Using Deep Learning Models. *International Journal of Advanced Research in Science, Communication and Technology*, 228–235. https://doi.org/10.48175/IJARSCT-29632.

[13] Sanders, A., Strzalkowski, T., Si, M., Chang, A., Dey, D., Braasch, J., & Wang, D. (2022). *Towards a Progression-Aware Autonomous Dialogue Agent*. http://arxiv.org/abs/2205.03692.

[14] Bunnell, B. E., Tsalatsanis, A., Chaphalkar, C., Robinson, S., Klein, S., Cool, S., Szwast, E., Heider, P. M., Wolf, B. J., & Obeid, J. S. (2025). Automated detection and prediction of suicidal behavior from clinical notes using deep learning. *PLOS One*, *20*(9), e0331459. https://doi.org/10.1371/journal.pone.0331459.

[15] N. Sharma and P. Karwasra, "Suicidal Text Detection on Social Media for Suicide Prevention Using Deep Learning Models," TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON), Chiang Mai, Yin, J., Hou, B., Dai, J., & Zu, Y. (2024). A CNN-BiLSTM Method Based on Attention Mechanism for Class-imbalanced Abnormal Traffic Detection. *Proceedings of the International Conference on Computer Vision and Deep Learning*, 1–6. https://doi.org/10.1145/3653781.3653807

[16] Thailand, 2023, pp. 25-30, doi: 10.1109/TENCON58879.2023.10322499.

[17] Hu, Y.-H., Wu, R.-Y., Su, M.-Y., Lin, I.-L., & Shen, C.-C. (2025). Multimodal Multitask Learning for Predicting Depression Severity and Suicide Risk Using Pretrained Audio and Text Embeddings: Methodology Development and Application. *JMIR Medical Informatics*, *13*, e66907–e66907. https://doi.org/10.2196/66907.

[18] Zevallos, R., Schoene, A., & Ortega, J. E. (2024). *The First Multilingual Model For The Detection of Suicide Texts*. http://arxiv.org/abs/2412.15498.

[19] Hansani, P., & Arachchi, K. (n.d.). *Mental Health Prediction Using Social Media Posts: A Nature-Inspired Approach*. https://doi.org/10.13140/RG.2.2.20212.21123.

[20] Hansani, P., & Arachchi, K. (n.d.). Mental Health Prediction Using Social Media Posts: A Nature-Inspired Approach. https://doi.org/10.13140/RG.2.2.20212.21123.

[21] Holmes, G., Tang, B., Gupta, S., Venkatesh, S., Christensen, H., & Whitton, A. (2025). Applications of Large Language Models in the Field of Suicide Prevention: Scoping Review. *Journal of Medical Internet Research*, *27*, e63126. https://doi.org/10.2196/63126

[22] Amiriparian, S., Gerczuk, M., Lutz, J., Strube, W., Papazova, I., Hasan, A., Kathan, A., & Schuller, B. W. (2024). *Non-Invasive Suicide Risk Prediction Through Speech Analysis*. http://arxiv.org/abs/2404.12132.

[23] Cui, Z., Lei, C., Wu, W., Duan, Y., Qu, D., Wu, J., Chen, R., & Zhang, C. (2024). *Spontaneous Speech-Based Suicide Risk Detection Using Whisper and Large Language Models*. https://doi.org/10.21437/Interspeech.2024-1895.

[24] Grimland, M., Benatov, J., Yeshayahu, H., Izmaylov, D., Segal, A., Gal, K., & Levi-Belz, Y. (2024). Predicting suicide risk in real-time crisis hotline chats integrating machine learning with psychological factors: Exploring the black box. *Suicide and Life-Threatening Behavior*, *54*(3), 416–424. https://doi.org/10.1111/sltb.13056.

[25] Chen, Y., Li, J., Song, C., Zhao, Q., Tong, Y., & Fu, G. (2024). *Deep Learning and Large Language Models for Audio and Text Analysis in Predicting Suicidal Acts in Chinese Psychological Support Hotlines*. http://arxiv.org/abs/2409.06164.

[26] Iyer, R., Nedeljkovic, M., & Meyer, D. (2022). Using Voice Biomarkers to Classify Suicide Risk in Adult Telehealth Callers: Retrospective Observational Study. *JMIR Mental Health*, *9*(8), e39807. https://doi.org/10.2196/39807.

[27] Thakur, R. S., & Singh, J. (2025). AI and NLP for Mental Health Prediction from Social Media: A Decade of Progress, Challenges, and Explainability (2015–2025). *International Journal for Research in Applied Science and Engineering Technology*, *13*(8), 1707–1715. https://doi.org/10.22214/ijraset.2025.73857.

[28] Chatterjee, M., Kumar, P., Samanta, P., & Sarkar, D. (2022). Suicide ideation detection from online social media: A multi-modal feature based technique. *International Journal of Information Management Data Insights*, *2*(2), 100103. https://doi.org/10.1016/j.jjimei.2022.100103.

[29] Jin, J., Liu, Y., & Dai, Y. (2025). *Multimodal Speech and Text Models to Detect Suicidal Risks in Adolescents*. https://doi.org/10.1101/2025.07.31.25332367.