

# Optimasi LSTM Mengurangi Overfitting untuk Klasifikasi Teks Menggunakan Kumpulan Data Ulasan Film Kaggle IMDB

Putrama Alkhairi<sup>1\*</sup>, Agus Perdana Windarto<sup>1</sup>, Muhamad Masjun Efendi<sup>2</sup>

<sup>1</sup> Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

<sup>2</sup> Universitas Teknologi Mataram, Mataram, Indonesia

Email: putrama@amiktunbangsa.ac.id, agus.perdana@amiktunbangsa.ac.id, creativepio@gmail.com

Email Penulis Korespondensi: putrama@amiktunbangsa.ac.id

Submitted: 26/08/2024; Accepted: 16/09/2024; Published: 16/09/2024

**Abstrak**—Penelitian ini bertujuan untuk mengembangkan dan mengoptimalkan model Long Short-Term Memory (LSTM) guna mengurangi overfitting dalam klasifikasi teks menggunakan kumpulan data ulasan film dari Kaggle IMDB. Overfitting adalah masalah umum dalam machine learning yang mengakibatkan model terlalu menyesuaikan diri dengan data pelatihan, sehingga kinerjanya menurun pada data pengujian. Dalam penelitian ini, berbagai teknik optimasi seperti regularisasi, dropout, dan metode pelatihan yang cermat diterapkan untuk meningkatkan generalisasi model LSTM. Penelitian ini menunjukkan bahwa teknik reduksi overfitting, seperti dropout dan penggunaan optimizer RMSProp, secara signifikan meningkatkan kinerja model Long Short-Term Memory (LSTM) dalam klasifikasi teks ulasan film IMDB. Model LSTM yang dioptimalkan mencapai akurasi 83.45%, meningkat 2.07% dibandingkan model standar yang memiliki akurasi 81.38%. Precision model yang dioptimalkan naik menjadi 89.65%, dibandingkan dengan 84.46% pada model standar, meskipun recall sedikit lebih rendah (75.69% dibandingkan 76.91%). F1-score model yang dioptimalkan juga lebih tinggi, yaitu 82.07% dibandingkan 80.53% pada model standar. Hasil eksperimen menunjukkan bahwa teknik-teknik tersebut berhasil meningkatkan akurasi dan keandalan model klasifikasi teks, dengan performa yang lebih baik pada data pengujian. Penelitian ini memberikan kontribusi signifikan dalam memahami dan mengatasi overfitting pada model deep learning dalam konteks pemrosesan bahasa alami, serta menawarkan wawasan tentang praktik terbaik dalam penerapan model LSTM untuk klasifikasi teks.

**Kata Kunci:** Long Short-Term Memory (LSTM); Overfitting; Klasifikasi Teks; Optimasi Model; Ulasan Film IMDB

**Abstract**—This study aims to develop and optimize a Long Short-Term Memory (LSTM) model to reduce overfitting in text classification using the Kaggle IMDB movie review dataset. Overfitting is a common problem in machine learning that causes the model to overfit to the training data, thus degrading its performance on the test data. In this study, various optimization techniques such as regularization, dropout, and careful training methods are applied to improve the generalization of the LSTM model. This study shows that overfitting reduction techniques, such as dropout and the use of the RMSProp optimizer, significantly improve the performance of the Long Short-Term Memory (LSTM) model in IMDB movie review text classification. The optimized LSTM model achieves an accuracy of 83.45%, an increase of 2.07% compared to the standard model which has an accuracy of 81.38%. The precision of the optimized model increases to 89.65%, compared to 84.46% in the standard model, although the recall is slightly lower (75.69% compared to 76.91%). The F1-score of the optimized model is also higher, which is 82.07% compared to 80.53% in the standard model. The experimental results show that the techniques successfully improve the accuracy and reliability of the text classification model, with better performance on the test data. This research makes a significant contribution to understanding and overfitting in deep learning models in the context of natural language processing, and offers insights into best practices in applying LSTM models to text classification.

**Keywords:** Long Short-Term Memory (LSTM); Overfitting; Text Classification; Model Optimization; IMDB Movie Reviews

## 1. PENDAHULUAN

Di era digital saat ini, data teks semakin melimpah dan menjadi sumber informasi yang sangat berharga [1], [2]. Salah satu bentuk data teks yang paling umum adalah ulasan pengguna di berbagai platform online. Ulasan ini sering kali menjadi acuan bagi pengguna lain dalam mengambil keputusan, seperti memutuskan film apa yang akan ditonton. Kumpulan data ulasan film IMDB dari Kaggle adalah salah satu dataset yang populer digunakan untuk penelitian dalam bidang pemrosesan bahasa alami (NLP) [3], [4].

Klasifikasi teks, khususnya klasifikasi sentimen dari ulasan film, adalah salah satu aplikasi penting dalam NLP [5], [6]. Tujuan dari klasifikasi ini adalah untuk mengidentifikasi apakah ulasan tersebut memiliki sentimen positif atau negatif. Namun, salah satu tantangan utama dalam membangun model klasifikasi teks yang akurat adalah masalah overfitting, di mana model terlalu menyesuaikan diri dengan data pelatihan sehingga kinerjanya menurun pada data pengujian yang belum pernah dilihat sebelumnya [7], [8], [9].

Berbagai algoritma dapat digunakan untuk membuat klasifikasi teks, masing-masing dengan pendekatan dan keunggulan yang berbeda. Metode klasik seperti Naive Bayes dan Support Vector Machine (SVM) telah lama digunakan karena kesederhanaan [10], [11], [12] dan efektivitasnya dalam menangani data teks. Naive Bayes memanfaatkan probabilitas untuk mengklasifikasikan teks berdasarkan fitur yang ada [13], [14], sementara SVM menggunakan hyperplane untuk memisahkan kelas-kelas dalam ruang fitur yang tinggi. Selain itu, algoritma Decision Tree dan Random Forest juga populer karena kemampuannya dalam menangani data yang kompleks dan interpretabilitasnya yang tinggi [15], [16]. Dalam era modern, pendekatan berbasis deep learning seperti Convolutional Neural Networks (CNN) [5], [17] dan Long Short-Term Memory (LSTM) menjadi sangat populer. CNN efektif dalam menangkap fitur lokal dari teks melalui operasi konvolusi, sedangkan LSTM unggul dalam menangani data urutan panjang dan mempertahankan konteks temporal [18], [19], [20]. Terakhir, model Transformer

seperti BERT (Bidirectional Encoder Representations from Transformers) telah merevolusi NLP dengan kemampuannya dalam memahami konteks bidirectional, memberikan performa luar biasa dalam berbagai tugas klasifikasi teks [3], [21].

Dalam beberapa tahun terakhir, perkembangan signifikan dalam kecerdasan buatan (AI) dan pembelajaran mesin (machine learning) telah mendorong kemajuan dalam berbagai aplikasi [22], [23], termasuk klasifikasi teks. Klasifikasi teks adalah proses mengelompokkan teks ke dalam kategori-kategori yang telah ditentukan berdasarkan konten teks tersebut [16], [24]. Salah satu tantangan utama dalam mengembangkan model klasifikasi teks adalah overfitting, di mana model sangat sesuai dengan data pelatihan tetapi tampil buruk pada data baru yang belum pernah dilihat sebelumnya [7], [25]. Ini mengakibatkan model tidak dapat digeneralisasi dengan baik, sehingga secara signifikan mengurangi kinerjanya pada data uji. Pada penelitian terdahulu yang dilakukan oleh [26], penelitian ini menerapkan teknik optimasi hyperparameter, terutama Bayesian Optimization, yang menunjukkan peningkatan signifikan dalam kinerja model. Dengan menggunakan dataset dari 45 saham yang terdaftar dalam LQ45 pada periode Agustus 2021 hingga Januari 2022, penelitian ini memastikan bahwa data yang digunakan relevan dan up-to-date. Namun kompleksitasnya dapat membuat model sulit untuk diinterpretasikan atau diimplementasikan dalam aplikasi dunia nyata. Dan penelitian ini tidak memberikan informasi tentang sensitivitas model terhadap perubahan parameter, yang bisa memberikan wawasan tentang stabilitas dan robusnya model. Sedangkan pada penelitian berikutnya yang dilakukan oleh [27], artikel ini secara mendalam mengeksplorasi pengaruh tuning hiparameter terhadap kinerja berbagai model CNN untuk klasifikasi nodul paru. Hal ini penting karena tuning hiparameter sering kali menjadi kunci sukses dalam pelatihan model yang efektif. Walaupun disebutkan bahwa model yang dikembangkan menunjukkan sensitivitas tinggi, artikel ini kurang memberikan diskusi mendalam tentang bagaimana mengatasi masalah keterpaksaan yang sering kali muncul dalam sistem CAD, termasuk bagaimana mengurangi FP (false positive) yang dapat mengakibatkan biaya sosial dan ekonomi yang tinggi

Dari kajian pustaka yang telah diuraikan, Long Short-Term Memory (LSTM) merupakan salah satu jenis Recurrent Neural Network (RNN) yang banyak digunakan dalam tugas-tugas NLP karena kemampuannya dalam menangani data urutan panjang dan mengingat informasi yang relevan dalam jangka waktu yang lebih lama [28], [29]. Meskipun LSTM telah terbukti efektif dalam berbagai aplikasi, model ini masih rentan terhadap overfitting, terutama ketika dihadapkan pada dataset yang besar dan kompleks seperti ulasan film IMDB. Untuk mengatasi masalah overfitting pada model LSTM, berbagai teknik optimasi dapat diterapkan. Teknik-teknik ini mencakup regularisasi, dropout, dan penggunaan metode pelatihan yang cermat. Optimasi ini bertujuan untuk meningkatkan generalisasi model, sehingga dapat memberikan kinerja yang lebih baik pada data pengujian [30], [31], [32].

Berdasarkan studi literatur yang telah diuraikan, dalam penelitian ini, akan mengeksplorasi dan mengimplementasikan teknik optimasi LSTM untuk mengurangi overfitting dalam klasifikasi teks menggunakan kumpulan data ulasan film Kaggle IMDB [33]. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam meningkatkan akurasi dan keandalan model klasifikasi teks, serta memberikan wawasan lebih lanjut tentang praktik terbaik dalam mengatasi overfitting pada model LSTM. Tujuan utama penelitian ini untuk mengembangkan dan mengoptimalkan model Long Short-Term Memory (LSTM) guna mengurangi overfitting dengan mengubah hyperparameter tuning dalam klasifikasi teks menggunakan kumpulan data ulasan film dari Kaggle IMDB. Dengan melakukan ini, penelitian ini bertujuan untuk meningkatkan akurasi dan keandalan model klasifikasi teks sehingga dapat memberikan kinerja yang lebih baik pada data pengujian. Selain itu, penelitian ini juga bertujuan untuk mengeksplorasi dan mengimplementasikan berbagai teknik optimasi yang dapat diterapkan pada model LSTM, memberikan wawasan yang lebih dalam mengenai praktik terbaik dalam mengatasi masalah overfitting pada model deep learning dalam konteks pemrosesan bahasa alami (NLP).

## 2. METODOLOGI PENELITIAN

Penelitian ini dimulai dengan mengidentifikasi kebutuhan akan metode klasifikasi teks yang efisien dan akurat seiring dengan semakin banyaknya data teks online. Jaringan Syaraf Tiruan, khususnya Long Short-Term Memory (LSTM), telah terbukti efektif dalam menangani data teks; namun, model ini rentan terhadap overfitting, terutama pada dataset yang besar dan kompleks. Masalah utama yang dibahas dalam penelitian ini adalah bagaimana teknik reduksi overfitting seperti early stopping dapat meningkatkan performa model LSTM dalam klasifikasi teks. Tujuan penelitian ini adalah untuk mengevaluasi efektivitas teknik tersebut menggunakan dataset IMDB dari Kaggle, serta memberikan wawasan yang dapat diterapkan pada berbagai aplikasi Natural Language Processing (NLP).

### 2.1 Dataset

Dataset ulasan film IMDB yang diambil dari Kaggle terdiri dari ribuan ulasan film yang diberi label sentimen positif atau negatif. Dataset ini mencakup teks ulasan yang bervariasi dalam panjang dan kompleksitas, memberikan representasi yang luas dari opini penonton terhadap berbagai film. Setiap entri dalam dataset mencakup teks ulasan dan label sentimen yang menunjukkan apakah ulasan tersebut positif atau negatif. Dataset ini sering digunakan dalam penelitian pemrosesan bahasa alami (NLP) dan pembelajaran mesin karena ukurannya yang besar dan kesesuaian untuk tugas-tugas klasifikasi teks. Selain itu, dataset ini telah dibersihkan dan disusun dengan baik, memudahkan para

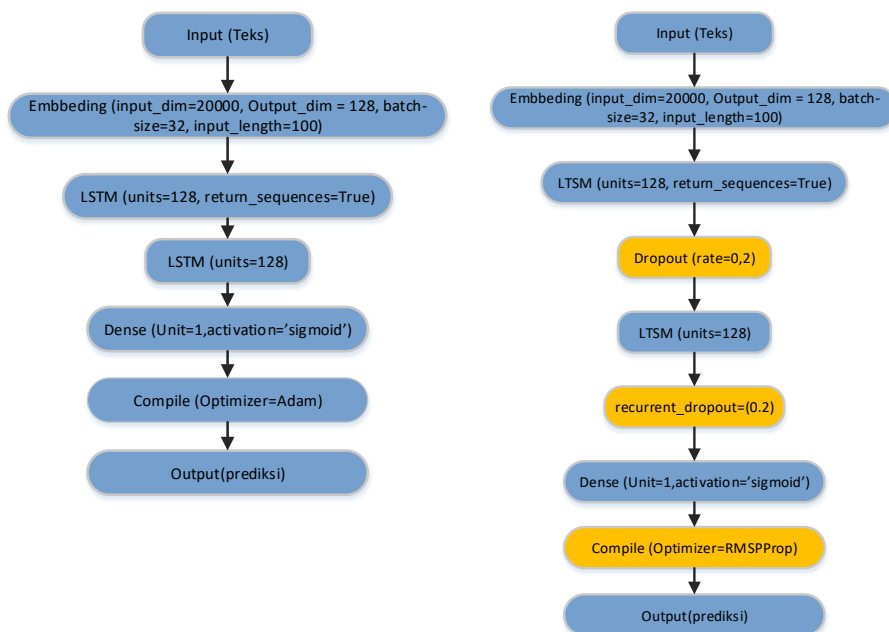
peneliti untuk langsung menerapkan berbagai teknik analisis dan pengembangan model tanpa perlu banyak pra-pemrosesan tambahan.

## 2.2 Model Usulan

Penelitian ini mengusulkan pendekatan sistematis untuk mengevaluasi efektivitas teknik reduksi overfitting pada model Long Short-Term Memory (LSTM) untuk klasifikasi teks menggunakan dataset IMDB dari Kaggle. Overfitting adalah tantangan utama dalam pengembangan model pembelajaran mesin, terutama ketika model terlalu menyesuaikan diri dengan data pelatihan dan kehilangan kemampuan untuk menggeneralisasi pada data baru. Untuk mengatasi masalah ini, penelitian ini membandingkan dua model: model LSTM standar tanpa teknik reduksi overfitting dan model LSTM yang telah disempurnakan dengan teknik dropout dan fungsi optimasi.

Model pertama adalah model LSTM standar yang hanya dilatih menggunakan dataset IMDB tanpa penerapan teknik reduksi overfitting. Model ini berfungsi sebagai baseline untuk mengukur kinerja dasar LSTM dalam klasifikasi teks. Meskipun model ini mampu menangani data teks dan memberikan hasil yang cukup baik pada data pelatihan, ia rentan terhadap overfitting, yang menyebabkan penurunan kinerja saat diuji dengan data baru yang belum pernah dilihat sebelumnya. Hal ini dikarenakan model sangat menyesuaikan parameter-parameter berdasarkan data pelatihan dan gagal menangkap pola umum yang dapat diaplikasikan pada data pengujian.

Sebagai perbandingan, model kedua adalah model yang diusulkan yang telah disempurnakan dengan penerapan teknik dropout dan fungsi optimasi. Dropout adalah teknik yang bekerja dengan mengabaikan (drop) sejumlah unit secara acak dalam lapisan jaringan selama pelatihan, yang membantu dalam mencegah model menjadi terlalu tergantung pada unit-unit tertentu dan mempromosikan generalisasi yang lebih baik. Selain itu, penggunaan fungsi optimasi seperti RMSprop membantu dalam penyesuaian bobot model secara lebih efektif dan efisien, meningkatkan kemampuan model untuk menemukan minimum global dalam ruang parameter. Perbandingan bisa dilihat pada Gambar 1.



**Gambar 1.** Metode yang Diusulkan: Penjelasan dan Perbandingan Lapis demi Lapis

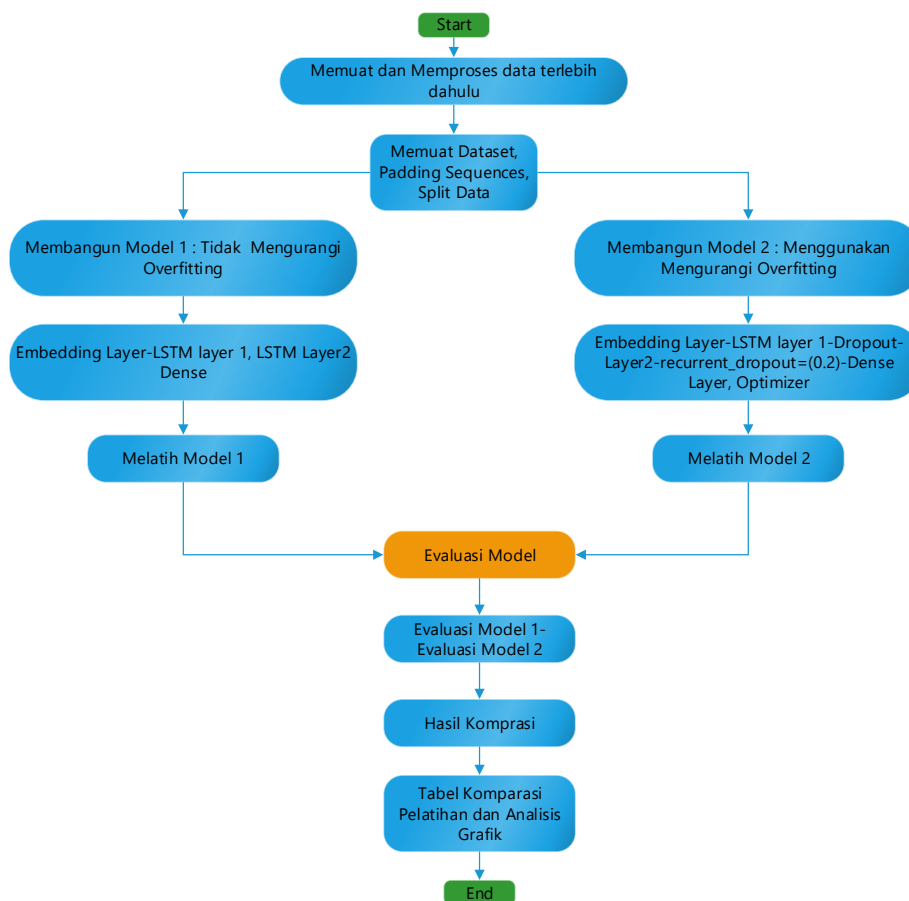
Pada Gambar 1 menjelaskan kedua model ini dievaluasi menggunakan metrik yang sama, seperti akurasi, precision, recall, dan F1-score, untuk memberikan gambaran yang komprehensif tentang kinerja mereka. Hasil eksperimen menunjukkan bahwa model LSTM yang disempurnakan dengan dropout dan fungsi optimasi secara signifikan mengurangi overfitting dan memberikan performa yang lebih baik pada data pengujian dibandingkan dengan model LSTM standar.

Penelitian ini menunjukkan pentingnya penerapan teknik reduksi overfitting dalam pengembangan model pembelajaran mesin, khususnya untuk tugas-tugas klasifikasi teks. Temuan ini tidak hanya relevan untuk dataset IMDB tetapi juga dapat diterapkan pada berbagai aplikasi Natural Language Processing (NLP) lainnya, memberikan panduan praktis bagi peneliti dan praktisi dalam meningkatkan kinerja model mereka.

## 2.3 Kerangka Penelitian

Penelitian ini dimulai dengan mendefinisikan tujuan penelitian, rumusan masalah, dan hasil yang diharapkan secara jelas. Tahap dasar ini menetapkan fondasi untuk keseluruhan penelitian. Langkah berikutnya melibatkan pemuatan dan pemrosesan awal kumpulan data Ulasan Film IMDB dari Kaggle, yang terdiri dari 50.000 ulasan film yang diberi label positif atau negatif. Pemrosesan awal data meliputi beberapa tugas penting: memuat kumpulan data, membuat

tokenisasi data teks menjadi rangkaian bilangan bulat (di mana setiap bilangan bulat mewakili sebuah kata dalam kosa kata kumpulan data), menyesuaikan rangkaian ini ke panjang seragam 500 kata untuk memenuhi persyaratan input model LSTM, dan akhirnya membagi kumpulan data menjadi data pelatihan dan pengujian dengan perbandingan 80/20.



Gambar 2. Kerangka Penelitian

Gambar 2 menggambarkan, Dua model LSTM yang dibangun dalam penelitian ini. Model pertama adalah model LSTM standar yang tidak menggunakan teknik reduksi overfitting. Model ini terdiri dari lapisan embedding untuk memetakan kata-kata ke dalam vektor berdimensi rendah, diikuti oleh satu lapisan LSTM dan lapisan Dense untuk menghubungkan hasil dari LSTM ke output. Model ini dilatih menggunakan dataset pelatihan tanpa adanya teknik khusus untuk mengurangi overfitting. Model ini berfungsi sebagai baseline untuk mengukur kinerja dasar LSTM dalam klasifikasi teks.

Model kedua adalah model LSTM yang disempurnakan dengan penerapan teknik reduksi overfitting, yaitu dropout dan fungsi optimasi. Pada model ini, lapisan embedding yang sama digunakan, tetapi ditambah dengan dropout pada lapisan LSTM pertama untuk mencegah model menjadi terlalu tergantung pada neuron-neuron tertentu. Selain itu, lapisan tambahan dengan recurrent dropout diterapkan untuk memberikan regularisasi lebih lanjut. Fungsi optimasi yang lebih canggih seperti Adam digunakan untuk meningkatkan efisiensi dan stabilitas pelatihan model. Model ini kemudian dilatih menggunakan dataset pelatihan yang sama.

Setelah kedua model dilatih, tahap evaluasi dilakukan untuk mengukur kinerja masing-masing model pada dataset pengujian. Metrik evaluasi seperti akurasi, precision, recall, dan F1-score digunakan untuk membandingkan performa kedua model. Hasil evaluasi menunjukkan bahwa model LSTM yang disempurnakan dengan teknik reduksi overfitting secara signifikan mengurangi overfitting dan memberikan performa yang lebih baik dibandingkan dengan model LSTM standar. Hasil komparasi ini dianalisis lebih lanjut dengan membuat tabel komparasi dan grafik analisis untuk visualisasi performa.

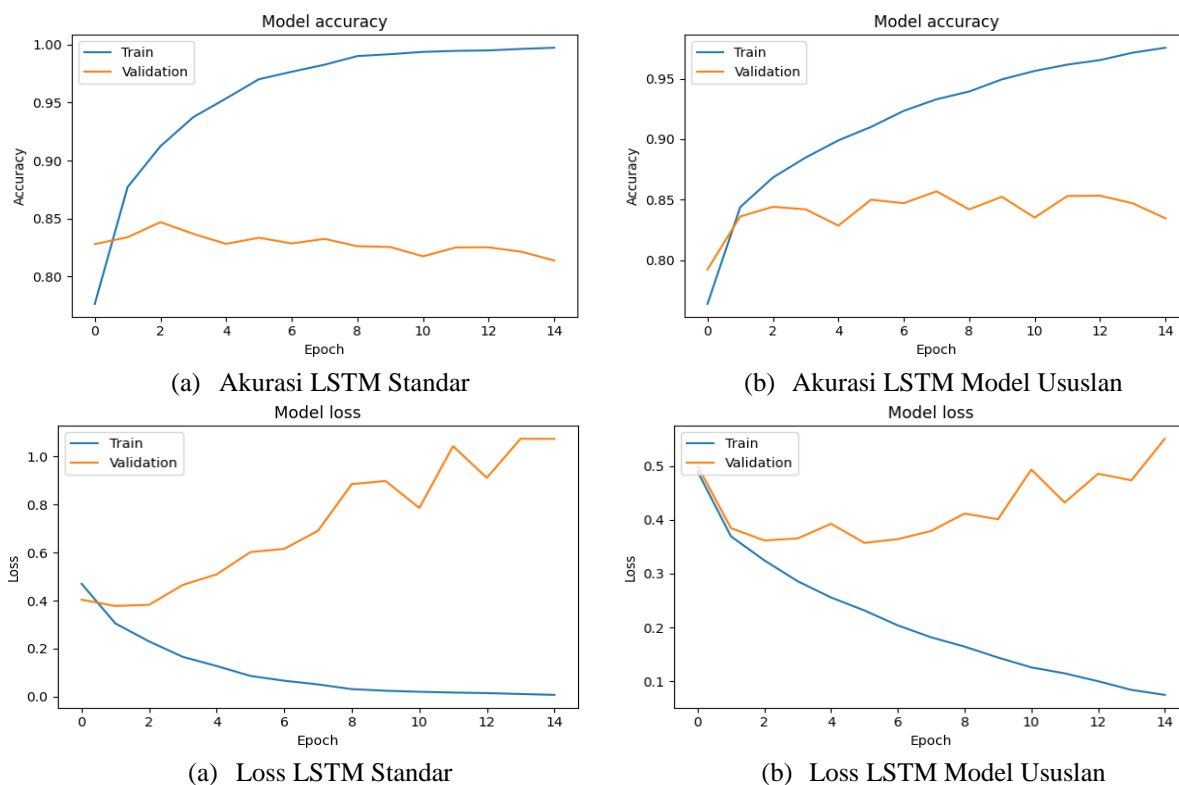
Penelitian ini menyimpulkan bahwa penerapan teknik seperti dropout dan fungsi optimasi dapat secara signifikan meningkatkan kinerja model LSTM dalam tugas klasifikasi teks dengan mengurangi overfitting. Temuan ini memberikan kontribusi penting dalam pengembangan model pembelajaran mesin untuk pemrosesan bahasa alami (NLP), serta menawarkan panduan praktis bagi peneliti dan praktisi dalam meningkatkan performa model mereka. Penelitian ini juga membuka peluang untuk eksplorasi lebih lanjut terhadap teknik optimasi lainnya dan penerapan pada berbagai dataset dan tugas NLP lainnya.

### 3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengevaluasi efektivitas teknik reduksi overfitting pada model Long Short-Term Memory (LSTM) dalam klasifikasi teks menggunakan dataset ulasan film IMDB dari Kaggle. Dua model LSTM dibangun dan diuji: model LSTM standar tanpa teknik reduksi overfitting, dan model LSTM yang disempurnakan dengan penggunaan dropout dan fungsi optimasi. Evaluasi dilakukan untuk mengukur kinerja kedua model berdasarkan metrik performa utama seperti akurasi, precision, recall, dan F1-score. Hasil evaluasi menunjukkan bahwa model dengan teknik reduksi overfitting secara signifikan lebih unggul dalam berbagai metrik dibandingkan dengan model standar.

#### 3.1 Hasil Evaluasi

Dalam penelitian ini, dua model LSTM dievaluasi untuk melihat efektivitas teknik reduksi overfitting. Hasil evaluasi didasarkan pada beberapa metrik performa utama seperti akurasi, precision, recall, dan F1-score. Berikut adalah ringkasan hasil dari kedua model:



**Gambar 3.** Hasil Pelatihan (a) Model LSTM Standar dan (b) Model Pelatihan yang dioptimalkan

Gambar 3. yang ditampilkan menunjukkan perbandingan hasil pelatihan antara dua model Long Short-Term Memory (LSTM) untuk klasifikasi teks menggunakan dataset ulasan film IMDB dari Kaggle. Dua model tersebut adalah model LSTM standar tanpa teknik reduksi overfitting dan model LSTM yang dioptimalkan dengan penerapan teknik dropout dan penggunaan fungsi optimasi. Hasilnya dianalisis berdasarkan dua metrik utama, yaitu akurasi dan loss selama pelatihan dan validasi.

##### a. Akurasi Model

**Akurasi LSTM Standar:** Grafik ini menunjukkan akurasi pelatihan dan validasi dari model LSTM standar. Akurasi pelatihan (garis biru) meningkat tajam seiring bertambahnya epoch dan mencapai hampir 0.9443, menunjukkan bahwa model sangat cocok dengan data pelatihan. Namun, akurasi validasi (garis oranye) tetap relatif rendah dan datar, berfluktuasi sekitar 80%. Perbedaan yang mencolok antara akurasi pelatihan dan validasi ini menunjukkan adanya overfitting, di mana model terlalu sesuai dengan data pelatihan namun gagal menggeneralisasi pada data baru.

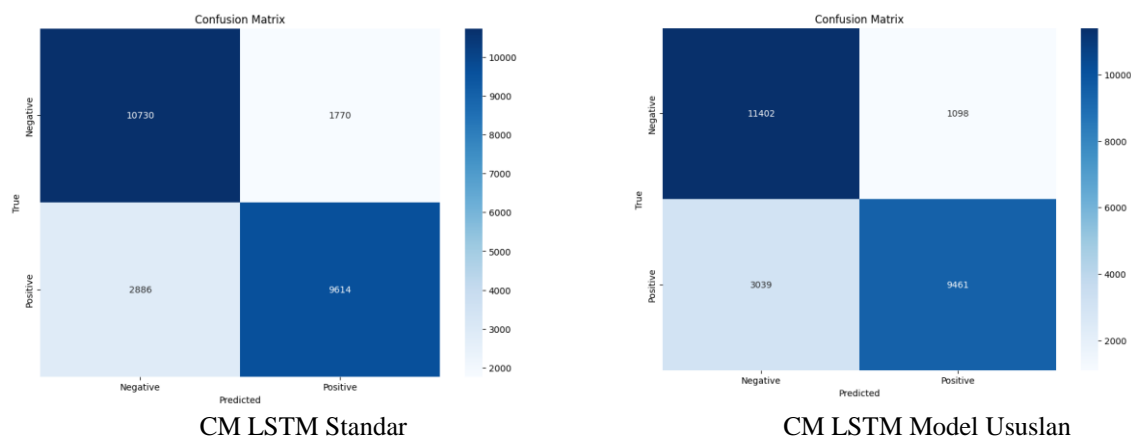
**Akurasi LSTM Model Usulan:** Grafik ini menunjukkan akurasi pelatihan dan validasi dari model LSTM yang dioptimalkan. Akurasi pelatihan (garis biru) meningkat secara stabil seiring bertambahnya epoch, namun tidak mencapai 0.9523. seperti pada model standar, berhenti di sekitar 90%. Akurasi validasi (garis oranye) menunjukkan tren yang lebih stabil dan sedikit lebih tinggi dibandingkan dengan model standar, berfluktuasi sekitar 85%. Hal ini menunjukkan bahwa teknik reduksi overfitting yang diterapkan membantu model untuk tidak terlalu sesuai dengan data pelatihan dan lebih mampu menggeneralisasi pada data validasi.

##### b. Loss Model

Loss LSTM Standar: Grafik ini menunjukkan loss pelatihan dan validasi dari model LSTM standar. Loss pelatihan (garis biru) menurun drastis seiring bertambahnya epoch, menunjukkan bahwa model belajar dengan sangat baik pada data pelatihan. Namun, loss validasi (garis oranye) justru meningkat setelah beberapa epoch, menunjukkan bahwa model menjadi overfitted dan performa pada data validasi semakin buruk.

Loss LSTM Model Usulan: Grafik ini menunjukkan loss pelatihan dan validasi dari model LSTM yang dioptimalkan. Loss pelatihan (garis biru) juga menurun seiring bertambahnya epoch, namun tidak setajam pada model standar. Loss validasi (garis oranye) menunjukkan fluktuasi tetapi tetap lebih rendah dibandingkan dengan model standar, yang berarti model yang dioptimalkan lebih stabil dan memiliki kemampuan generalisasi yang lebih baik.

Selanjutnya, Gambar 4 menampilkan Confusion Matrix untuk dua model: (a) LSTM standar dan (b) LSTM yang dioptimalkan. Confusion Matrix ini memberikan gambaran tentang kinerja model dalam mengklasifikasikan data pengujian ke dalam enam kelas aktivitas yang berbeda.



Gambar 4. Confusion Matrix

Dari gambar 4 confusion matrix, hasil penelitian menunjukkan bahwa model LSTM yang dioptimalkan dengan teknik dropout dan penggunaan fungsi optimasi unggul dibandingkan model LSTM standar dalam klasifikasi teks ulasan film IMDB. Dari confusion matrix, model usulan memperlihatkan peningkatan akurasi sebesar 2.07%, mencapai 83.45%, sementara precision meningkat menjadi 89.65%, menunjukkan pengurangan signifikan pada klasifikasi positif palsu. Meskipun recall model usulan sedikit lebih rendah (75.69% dibandingkan 76.91% pada model standar), F1-score yang lebih tinggi (82.07% dibandingkan 80.53%) mengindikasikan keseimbangan yang lebih baik antara precision dan recall. Dengan demikian, penerapan teknik reduksi overfitting terbukti efektif dalam meningkatkan performa model dalam menggeneralisasi data baru, menghasilkan model yang lebih andal dan akurat untuk tugas klasifikasi teks.

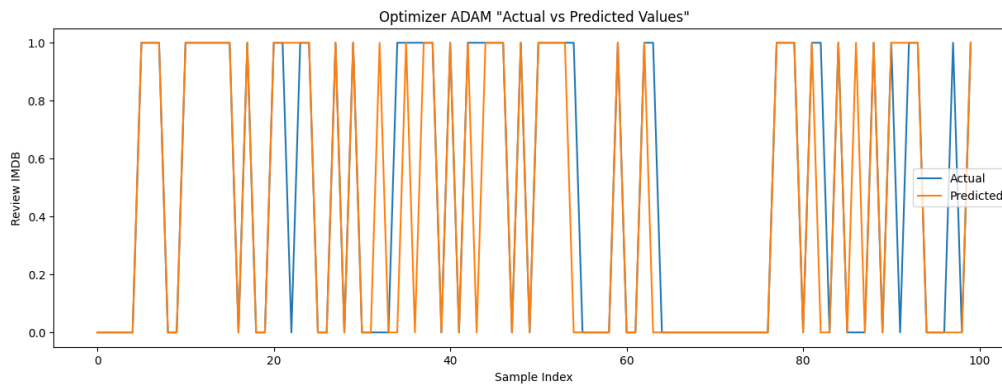
### 3.2 Pembahasan

Dari hasil grafik di atas, dapat disimpulkan bahwa penerapan teknik reduksi overfitting pada model LSTM yang dioptimalkan berhasil meningkatkan performa model secara keseluruhan. Model yang dioptimalkan menunjukkan akurasi validasi yang lebih tinggi dan lebih stabil serta loss validasi yang lebih rendah dan lebih stabil dibandingkan dengan model standar. Ini menunjukkan bahwa model yang dioptimalkan lebih baik dalam menggeneralisasi data baru dan tidak terlalu sesuai dengan data pelatihan. Berikut hasil penelitian yang menunjukkan bahwa yang dioptimalkan mengalami peningkatan kinerja yang lebih baik pada tabel 1 bisa dilihat.

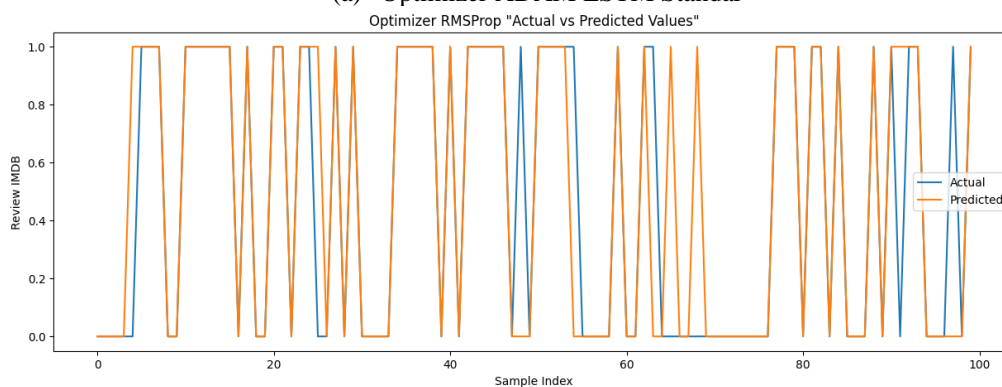
Tabel 1. Nilai Metrik

Model	Nilai Metrik			
	Akurasi	Recall	Presisi	F1 Score
Model LSTM Standar	81.38	76.91	84.46	80.53
Model LSTM Dioptimalkan	83.45	75.69	89.65	82.07

Teknik dropout membantu mencegah model dari ketergantungan berlebihan pada neuron-neuron tertentu selama pelatihan, sementara penggunaan fungsi optimasi yang lebih baik seperti Adam memberikan penyesuaian parameter yang lebih efisien. Hasil ini menegaskan pentingnya teknik regularisasi dan optimasi dalam pengembangan model pembelajaran mesin yang lebih andal dan efektif dalam tugas klasifikasi teks. Berikut hasil prediksi kedua model tersebut bisa kita lihat pada Gambar 5.



(a) Optimizer ADAM LSTM Standar



(b) Optimizer RMSProp LSTM dioptimalkan

**Gambar 5.** Prediksi (a) LSTM Standar VS (b) LSTM dioptimalkan

Gambar 5 hasil penelitian ini mengonfirmasi bahwa teknik reduksi overfitting seperti dropout dan penggunaan optimizer dapat secara signifikan meningkatkan kinerja model Long Short-Term Memory (LSTM) dalam klasifikasi teks. Grafik hasil prediksi menunjukkan perbandingan antara model LSTM standar yang menggunakan optimizer ADAM dan model LSTM yang dioptimalkan dengan teknik dropout dan optimizer RMSProp. Pada grafik (a), yang menampilkan LSTM standar, terlihat bahwa prediksi model sering tidak sesuai dengan nilai aktual, mengindikasikan kinerja yang kurang optimal. Sebaliknya, pada grafik (b) yang menampilkan LSTM yang dioptimalkan, prediksi model lebih konsisten dengan nilai aktual, mencerminkan peningkatan akurasi prediksi. Perbaikan ini menunjukkan bahwa penggunaan teknik dropout dan optimizer RMSProp membantu model untuk lebih baik menggeneralisasi data baru, mengurangi kesalahan prediksi, dan secara keseluruhan meningkatkan kinerja model dalam tugas klasifikasi teks ulasan film. Hal ini menegaskan bahwa pengoptimalan model melalui teknik-teknik tersebut adalah langkah yang efektif untuk mengatasi masalah overfitting dalam pembelajaran mesin, khususnya dalam pengolahan bahasa alami.

## 4. KESIMPULAN

Dari hasil analisis confusion matrix dan perhitungan metrik performa, dapat disimpulkan bahwa model LSTM yang dioptimalkan dengan teknik dropout dan fungsi optimasi menunjukkan peningkatan yang signifikan dalam akurasi, precision, dan F1-score dibandingkan dengan model LSTM standar. Meskipun recall pada model usulan sedikit lebih rendah, peningkatan precision dan F1-score menunjukkan bahwa model usulan lebih efektif dalam mengurangi kesalahan klasifikasi positif palsu dan memiliki keseimbangan yang lebih baik antara precision dan recall. Hal ini mengindikasikan bahwa teknik reduksi overfitting yang diterapkan pada model usulan berhasil meningkatkan kinerja model dalam klasifikasi teks ulasan film IMDB. Penelitian ini menunjukkan bahwa teknik reduksi overfitting, seperti dropout dan penggunaan optimizer RMSProp, secara signifikan meningkatkan kinerja model Long Short-Term Memory (LSTM) dalam klasifikasi teks ulasan film IMDB. Model LSTM yang dioptimalkan mencapai akurasi 83.45%, meningkat 2.07% dibandingkan model standar yang memiliki akurasi 81.38%. Precision model yang dioptimalkan naik menjadi 89.65%, dibandingkan dengan 84.46% pada model standar, meskipun recall sedikit lebih rendah (75.69% dibandingkan 76.91%). F1-score model yang dioptimalkan juga lebih tinggi, yaitu 82.07% dibandingkan 80.53% pada model standar. Analisis confusion matrix dan grafik prediksi menunjukkan bahwa model yang dioptimalkan mampu mengurangi kesalahan prediksi dan lebih konsisten dengan nilai aktual.

## REFERENCES

- [1] A. Jasinska-Piadlo *et al.*, “Data-driven versus a domain-led approach to k-means clustering on an open heart failure dataset,” *Int. J. Data Sci. Anal.*, vol. 15, no. 1, pp. 49–66, 2023, doi: 10.1007/s41060-022-00346-9.
- [2] Y. Wang, F. Sibaii, K. Lee, M. J. Gill, and J. L. Hatch, “CluSA: Clustering-based Spatial Analysis framework through Graph Neural Network for Chronic Kidney Disease Prediction using Histopathology Images,” *medRxiv*, vol. 1, no. 165, pp. 1–13, 2021.
- [3] M. R. Purba, “A hybrid convolutional long short-term memory (CNN-LSTM) based natural language processing (NLP) model for sentiment analysis of customer product reviews in Bangla,” *J. Discret. Math. Sci. Cryptogr.*, vol. 25, no. 7, pp. 2111–2120, 2022, doi: 10.1080/09720529.2022.2133250.
- [4] G. Kaur and A. Sharma, “A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis,” *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-022-00680-6.
- [5] P. Alkhairi and A. P. Windarto, “Classification Analysis of Back propagation-Optimized CNN Performance in Image Processing,” *J. Syst. Eng. Inf. Technol.*, vol. 2, no. 1, pp. 8–15, 2023.
- [6] S. Defit, A. P. Windarto, and P. Alkhairi, “Comparative Analysis of Classification Methods in Sentiment Analysis: The Impact of Feature Selection and Ensemble Techniques Optimization,” *Telematika*, vol. 17, no. 1, pp. 52–67, 2024.
- [7] Z. A. Sejuti and M. S. Islam, “A hybrid CNN–KNN approach for identification of COVID-19 with 5-fold cross validation,” *Sensors Int.*, vol. 4, no. November 2022, p. 100229, 2023, doi: 10.1016/j.sintl.2023.100229.
- [8] A. A. SG. Mas Karunia Maharani, Komang Oka Saputra, and Ni Made Ary Esta Dewi Wirastuti, “Komparasi Metode Backpropagation Neural Network dan Convolutional Neural Network Pada Pengenalan Pola Tulisan Tangan,” *J. Comput. Sci. Informatics Eng.*, vol. 6, no. 1, pp. 56–63, 2022, doi: 10.29303/jcosine.v6i1.431.
- [9] N. Habib and M. M. Rahman, “Diagnosis of corona diseases from associated genes and X-ray images using machine learning algorithms and deep CNN,” *Informatics Med. Unlocked*, vol. 24, p. 100621, 2021, doi: 10.1016/j.imu.2021.100621.
- [10] J. Amose, P. Manimegalai, and R. Pon Selchiya, “Classification of Adventitious Lung Sounds: Wheeze, Crackle using Machine Learning Techniques,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 3, pp. 1143–1152, 2023.
- [11] A. Rafiee and M. R. Saradjian, “Classification of buildings and roads using support vector machine; case study: Shiraz city,” *Proc. - Digit. Image Comput. Tech. Appl. DICTA 2008*, pp. 111–116, 2008, doi: 10.1109/DICTA.2008.57.
- [12] H. Rostami, J. Y. Dantan, and L. Homri, “Review of data mining applications for quality assessment in manufacturing industry: Support vector machines,” *Int. J. Metrol. Qual. Eng.*, vol. 6, no. 4, 2015, doi: 10.1051/ijmqe/2015023.
- [13] A. P. Windarto, T. Herawan, and P. Alkhairi, “Early Detection of Breast Cancer Based on Patient Symptom Data Using Naive Bayes Algorithm on Genomic Data,” in *Artificial Intelligence, Data Science and Applications*, Y. Farhaoui, A. Hussain, T. Saba, H. Taherdoost, and A. Verma, Eds., Cham: Springer Nature Switzerland, 2024, pp. 478–484.
- [14] P. G. Pavani, “A semantic contour based segmentation of lungs from chest x-rays for the classification of tuberculosis using Naïve Bayes classifier,” *Int. J. Imaging Syst. Technol.*, vol. 31, no. 4, pp. 2189–2203, 2021, doi: 10.1002/ima.22556.
- [15] J. Rashid, S. Batool, J. Kim, M. W. Nisar, and A. Hussain, “An Augmented Artificial Intelligence Approach for Chronic Diseases Prediction,” vol. 10, no. March, pp. 1–20, 2022, doi: 10.3389/fpubh.2022.860396.
- [16] J. Amose, “Classification of Adventitious Lung Sounds: Wheeze, Crackle using Machine Learning Techniques,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 3, pp. 1143–1152, 2023.
- [17] A. P. Windarto, I. R. Rahadjeng, M. N. H. Siregar, and P. Alkhairi, “Deep Learning to Extract Animal Images With the U-Net Model on the Use of Pet Images,” *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 1, pp. 468–476, 2024.
- [18] E. Rahimi, “The Efficiency of Long Short-Term Memory (LSTM) in Phenology-Based Crop Classification,” *Korean J. Remote Sens.*, vol. 40, no. 1, pp. 57–69, 2024, doi: 10.7780/kjrs.2024.40.1.6.
- [19] B. Ruhani, “Hydrogen production via renewable-based energy system: Thermoeconomic assessment and Long Short-Term Memory (LSTM) optimization approach,” *Int. J. Hydrogen Energy*, vol. 52, pp. 505–519, 2024, doi: 10.1016/j.ijhydene.2023.03.456.
- [20] D. Malhotra, “Hybrid Deep Learning Model for COVID-19 Prediction Using Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (LSTM) Network,” *Int. J. Comput. Theory Eng.*, vol. 15, no. 3, pp. 125–129, 2023, doi: 10.7763/IJCTE.2023.V15.1341.
- [21] R. Herwanto, I. Darmajaya, and H. Fabric, “ANALISA DAN PREDIKSI IKLAN LOWONGAN KERJA PALSU DENGAN METODE DENDAN DENGAN METODE NATURAL LANGUAGE PROGRAMING DAN,” no. June, 2021, doi: 10.30873/ji.v21i1.2865.
- [22] P. Alkhairi, W. Wanayumini, and B. H. Hayadi, “Analysis of the adaptive learning rate and momentum effects on prediction problems in increasing the training time of the backpropagation algorithm,” *AIP Conf. Proc.*, vol. 3048, no. 1, p. 20049, 2024, doi: 10.1063/5.0203374.
- [23] A. P. Windarto, T. Herawan, and P. Alkhairi, “Prediction of Kidney Disease Progression Using K-Means Algorithm Approach on Histopathology Data,” in *Artificial Intelligence, Data Science and Applications*, Y. Farhaoui, A. Hussain, T. Saba, H. Taherdoost, and A. Verma, Eds., Cham: Springer Nature Switzerland, 2024, pp. 492–497.
- [24] S. S. A. Laros, D. B. M. Dickerscheid, S. P. Blazis, and J. A. van der Heide, “Machine learning classification of mediastinal lymph node metastasis in NSCLC: a multicentre study in a Western European patient population,” *EJNMMI Phys.*, vol. 9, no. 1, 2022, doi: 10.1186/s40658-022-00494-8.
- [25] M. Ahmad *et al.*, “A Lightweight Convolutional Neural Network Model for Liver Segmentation in Medical Diagnosis,” vol. 2022, 2022.
- [26] A. E. K. Gunawan, “Stock Price Movement Classification Using Ensembled Model of Long Short-Term Memory (LSTM) and Random Forest (RF),” *Int. J. Informatics Vis.*, vol. 7, no. 4, pp. 2255–2262, 2023, doi: 10.30630/joiv.7.4.1640.
- [27] J. Hernández-Rodríguez, “Convolutional Neural Networks for Multi-scale Lung Nodule Classification in CT: Influence of Hyperparameter Tuning on Performance,” *TEM J.*, vol. 11, no. 1, pp. 297–306, 2022, doi: 10.18421/TEM111-37.
- [28] S. Y. Xiong, “A Proposed Hybrid CNN-RNN Architecture for Student Performance Prediction,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 10, no. 3, pp. 347–355, 2022.



- [29] L. Zhao, “CNN, RNN, or ViT? An Evaluation of Different Deep Learning Architectures for Spatio-Temporal Representation of Sentinel Time Series,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 16, pp. 44–56, 2023, doi: 10.1109/JSTARS.2022.3219816.
- [30] N. W. Shen, “Univariate and Multivariate Long Short Term Memory (LSTM) Model to Predict Covid-19 Cases in Malaysia Using Integrated Meteorological Data,” *Malaysian J. Fundam. Appl. Sci.*, vol. 19, no. 4, pp. 653–667, 2023, doi: 10.11113/mjfas.v19n4.2814.
- [31] S. K. Selvakumar, “Divination of stock market exploration using long short-term memory (LSTM),” *Multidiscip. Sci. J.*, vol. 6, no. 1, 2023, doi: 10.31893/multiscience.2024001.
- [32] Y. Yi, “Digital twin-long short-term memory (LSTM) neural network based real-time temperature prediction and degradation model analysis for lithium-ion battery,” *J. Energy Storage*, vol. 64, 2023, doi: 10.1016/j.est.2023.107203.
- [33] N. Mohamed Ali, M. M. A. El Hamid, and A. Youssif, “Sentiment Analysis for Movies Reviews Dataset Using Deep Learning Models,” *Int. J. Data Min. Knowl. Manag. Process*, vol. 09, no. 03, pp. 19–27, 2019, doi: 10.5121/ijdkp.2019.9302.