JURNAL ILMIAH SISTEM INFORMASI (JUSI)
VOL 4, No. 3, November 2025, pp. 417 – 430
p-ISSN : 2803-1507 (media cetak) | e-ISSN : 2803-1531 (media online)
Link Publikasi: https://ejurnal.provisi.ac.id/index.php/JUISI/issue/view/151

Systematic Literature Review: Analysis of AI Implementation for Document Verification

Fedhira Syaila Putri Azzahra*[1], Cahyo Printo[2]
[1,2] Teknik Informatika, Universitas Logistik dan Bisnis Internaisonal
Jl. Sariasih No. 54, Sarijadi, 40151, Kota Bandung, Jawa Barat, Indonesia
e-mail*[1]: pedirasyaila@gmail.com

ARTICLE INFO

Article History:

Received: 01-06-2025
Revised: 12-07-2025
Accepted: 08-08-2025
Published: 21-11-2025

ABSTRACT

The growing complexity of digital document processing has led to the increased adoption of artificial intelligence (AI) to enhance the accuracy and efficiency of document verification systems. This study presents a systematic literature review that explores various AI-based methods used for automated document inspection and validation. A total of 42 scientific articles published between 2020 and 2025 were analyzed to identify the most commonly used AI techniques, major implementation challenges, and their respective advantages and limitations. The results indicate that natural language processing (NLP) is the most widely applied method, followed by machine learning (ML) and deep learning (DL). These approaches have proven effective in reducing manual workloads, accelerating document processing, and adapting to various document types, such as legal contracts, financial records, academic certificates, and medical documents. However, several challenges persist, including inconsistent data formats, limited model generalization, high computational requirements, and difficulties in interpreting model outputs. This review offers a comprehensive understanding of current trends in AI-based document verification and provides practical insights for designing more robust and scalable systems. The findings are expected to support future research and the implementation of intelligent document analysis technologies across different sectors, including finance, education, healthcare, and legal administration. By addressing the current limitations, organizations can better leverage AI to streamline document validation processes, improve operational efficiency, and ensure higher levels of accuracy and reliability in document handling.

Keywords: *Artificial intelligence, Deep Learning, Document Verification, Natural Language Processing.*

Abstrak

Pemrosesan dokumen digital yang semakin kompleks mendorong pemanfaatan kecerdasan buatan *Artificial Intelligence* (AI) untuk meningkatkan akurasi dan efisiensi dalam sistem verifikasi dokumen. Penelitian ini menyajikan sebuah tinjauan literatur sistematis untuk mengkaji berbagai metode berbasis *AI* yang digunakan dalam inspeksi dan validasi dokumen secara otomatis. Sebanyak 42 artikel ilmiah yang diterbitkan antara

DOI: https://doi.org/10.51903/kjjwk708

p-ISSN : 2803-1507 e-ISSN : 2803-1531

tahun 2020 hingga 2025 dianalisis untuk mengidentifikasi pendekatan AI yang paling dominan, tantangan utama implementasi, serta kelebihan dan keterbatasan dari masing-masing metode. Hasil kajian menunjukkan bahwa *Natural Language Processing* (NLP) merupakan teknik yang paling banyak digunakan, diikuti oleh *Machine Learning* (ML) dan *Deep Learning* (DL). Ketiga pendekatan ini menunjukkan efektivitas dalam mengurangi beban kerja manual, mempercepat waktu pemrosesan, serta memiliki fleksibilitas untuk diterapkan pada berbagai jenis dokumen, seperti dokumen keuangan, kontrak hukum, dokumen pendidikan, dan catatan medis. Meskipun demikian, penerapan metode AI dalam verifikasi dokumen masih menghadapi berbagai tantangan, termasuk format data yang tidak seragam, kemampuan generalisasi model yang terbatas, kebutuhan komputasi yang tinggi, dan kesulitan dalam menafsirkan hasil model. Tinjauan ini memberikan pemahaman yang lebih mendalam mengenai tren terkini dalam verifikasi dokumen berbasis AI dan menawarkan wawasan praktis untuk merancang sistem yang andal dan dapat diskalakan. Diharapkan, hasil dari studi ini dapat menjadi landasan bagi penelitian lanjutan dan penerapan teknologi analisis dokumen cerdas di berbagai sektor seperti keuangan, pendidikan, kesehatan, dan administrasi hukum.
Kata Kunci: *Kecerdasan Buatan, Machine Learning, Verifikasi Dokumen, Natural language processing*.

1. PENDAHULUAN

Transformasi digital yang terjadi secara masif di berbagai sektor industri telah mendorong adopsi teknologi cerdas, termasuk *Artificial Intelligence* (*AI*), untuk mengotomatisasi proses bisnis berbasis dokumen. Salah satu sektor yang paling terdampak oleh evolusi ini adalah sektor perbankan, di mana pengolahan dokumen secara manual dianggap tidak lagi efisien. Studi sebelumnya menunjukkan bahwa *Intelligent Document Processing* (*IDP*) memainkan peran sentral dalam meningkatkan efisiensi operasional perbankan, khususnya dalam manajemen dokumen pinjaman [1]. Namun, pengetahuan sektoral yang spesifik mengenai *IDP* masih terbatas, sehingga penelitian lebih lanjut diperlukan untuk memahami tantangan dan peluang implementasi *AI* dalam konteks dokumen perbankan secara menyeluruh.

Sejalan dengan itu, sebagian besar data dalam dunia industri, baik dari sektor publik maupun swasta, bersifat tidak terstruktur, seperti teks cetak dan tulisan tangan [2]. Dalam studi terkait, metode konvensional berbasis aturan tidak lagi memadai untuk mengekstraksi informasi dari dokumen yang kompleks, khususnya dokumen dengan format variatif atau dokumen hasil pindai [3]. Oleh karena itu, pendekatan berbasis *AI*, khususnya yang menggabungkan teknik *Natural Language Processing* (*NLP*), *Computer Vision*, dan *Large Language Models* (*LLM*), menjadi solusi utama yang dikembangkan untuk meningkatkan akurasi dan efisiensi ekstraksi informasi dari dokumen tidak terstruktur [4]. Pendekatan ini juga memungkinkan integrasi otomatis ke dalam sistem manajemen dokumen digital yang lebih luas di berbagai sektor industri.

Selain sektor perbankan dan dokumen tidak terstruktur secara umum, penerapan *AI* dalam pemeriksaan dokumen juga mulai merambah sektor konstruksi, publik, dan biomedis. Misalnya, studi lainnya [5] menyoroti bagaimana *Document Layout Analysis* (*DLA*) digunakan untuk mengidentifikasi komponen utama dokumen dalam domain pemerintahan, termasuk pengolahan dokumen resmi dari lebih dari 24 sumber data publik di Spanyol. Studi lain menunjukkan bahwa pada sektor biomedis, tantangan yang dihadapi dalam klasifikasi dokumen meliputi kualitas data, kompleksitas terminologi, dan kebutuhan akan representasi semantik yang presisi tinggi dalam dokumen klinis [6]. Hal ini menekankan bahwa pengembangan model *AI* tidak dapat bersifat generik, melainkan harus disesuaikan dengan konteks domain dokumen yang diproses.

Mengacu pada berbagai temuan tersebut, maka menjadi sangat relevan untuk melakukan tinjauan literatur secara sistematis (*Systematic Literature Review*/*SLR*) terhadap berbagai pendekatan *AI* yang digunakan dalam pemeriksaan dan pemrosesan dokumen secara otomatis. Penelitian ini didasari oleh belum adanya gambaran yang jelas dan menyeluruh mengenai metode *AI* yang telah digunakan, tantangan yang sering muncul dalam penerapannya, serta kelebihan dan keterbatasan dari penggunaan *AI* dalam sistem pemeriksaan dokumen otomatis berbasis *web.* Oleh karena itu, penelitian ini bertujuan untuk menjawab tiga fokus utama, yaitu: Pertama, apa saja metode yang digunakan oleh AI dalam pemeriksaan atau validasi dokumen secara otomatis. Kedua, apa tantangan utama dalam mengimplementasikan AI untuk pemeriksaan dokumen secara otomatis, Ketiga, apa kelebihan dan keterbatasan penerapan AI dalam sistem pemeriksaan dokumen. Penelitian ini diharapkan dapat menjadi acuan bagi peneliti, pengembang, dan praktisi dalam merancang sistem pemeriksaan dokumen yang lebih cerdas, efisien, dan dapat diandalkan di berbagai sektor, seperti keuangan, pemerintahan, kesehatan, dan pendidikan.

2. TINJAUAN PUSTAKA

p-ISSN : 2803-1507 e-ISSN : 2803-1531

2.1. *Artificial Intelligence*

Artificial Intelligence (AI) adalah teknologi berbasis komputasi cerdas yang digunakan untuk menganalisis data, memprediksi pola, dan mendukung pengambilan keputusan otomatis. Tidak hanya digunakan untuk analisis data, tetapi juga untuk mendukung pengambilan keputusan dalam e-procurement. Misalnya, sistem berbasis AI yang dikembangkan dnegan menggabungkan NLP, pencarian semantik, dan ekstraksi informasi terbuka untuk mengelola dokumen pengadaan secara efisien [7]. Pendekatan ini dinilai mampu mempercepat proses pengadaan dan mengurangi kesalahan manusia.

2.2. *Machine Learning*

Machine Learning dipahami sebagai pendekatan pembelajaran berbasis data yang digunakan untuk menganalisis dan mengenali dokumen, termasuk untuk klasifikasi, ekstraksi informasi, dan pengenalan struktur. Dalam studi analisis dokumen historis, ML digunakan untuk menggabungkan fitur visual dan tekstual guna meningkatkan akurasi pemrosesan dokumen kompleks [3]. Metode ini memungkinkan sistem untuk belajar dari data sebelumnya tanpa pemrograman eksplisit. Hasilnya, sistem dapat secara otomatis mengenali pola dalam dokumen yang beragam.

2.3. *Deep Learning*

Deep Learning adalah cabang dari *Machine Learning* yang menggunakan jaringan saraf berlapis untuk secara otomatis mempelajari representasi fitur dari data. Dalam konteks analisis dokumen, pendekatan ini telah digunakan untuk tugas seperti segmentasi layout, deteksi objek dalam dokumen, serta klasifikasi dokumen historis berbasis multimodal data [3]. Jaringan saraf dalam *deep learning* mampu menangkap hubungan kompleks antar fitur dalam data. Hal ini memungkinkan proses pemrosesan dokumen menjadi lebih presisi dan efisien.

2.4. *Natural Language Processing (NLP)*

Natural Language Processing (NLP) adalah cabang dari kecerdasan buatan (AI) yang berfokus pada interaksi antara komputer dan manusia melalui bahasa alami. Dalam studi analisis dokumen, NLP digunakan untuk mengotomatiskan ekstraksi informasi penting seperti subjek, tindakan, dan objek dalam dokumen administratif, seperti surat dinas dan laporan resmi [8]. NLP juga digunakan dalam sistem pencarian cerdas dan analisis sentimen. Dengan NLP, pemrosesan dokumen dalam jumlah besar menjadi lebih efisien dan akurat.

2.5. *Systematic Literature Review (SLR)*

Systematic Literature Review (SLR) adalah metode kajian pustaka yang dilakukan secara sistematis dan terstruktur dengan tujuan untuk memberikan pemahaman menyeluruh terhadap perkembangan penelitian pada suatu topik tertentu. SLR juga memungkinkan transparansi, objektivitas, dan replikasi dalam proses telaah literatur, menjadikannya pendekatan yang banyak digunakan untuk memperoleh landasan teoretis yang kuat dalam penelitian [9]. Teknik ini banyak diterapkan di bidang ilmu komputer dan ilmu sosial. Melalui SLR, peneliti dapat mengidentifikasi tren, kesenjangan, dan arah masa depan penelitian.

2.6. *Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)*

PRISMA merupakan sebuah pedoman pelaporan yang dirancang untuk meningkatkan transparansi dan kualitas sistematik review. Pedoman ini banyak digunakan dalam penelitian berbasis bukti untuk memastikan bahwa proses kajian pustaka dilakukan secara komprehensif dan dapat direplikasi [10]. PRISMA menyediakan diagram alur seleksi studi yang mempermudah visualisasi proses review. Dengan adanya pedoman ini, integritas ilmiah dari hasil tinjauan pustaka dapat lebih terjamin.

2.7. *Watase Uake*

Watase merupakan sistem penelitian kolaboratif antar peneliti secara online yang dikembangkan untuk memudahkan pencarian literatur secara sistematis dengan metode PRISMA, melakukan meta-analisis sederhana, klasifikasi artikel, serta visualisasi data. *Watase* mendukung proses penelitian bersama dengan fitur berbagi informasi antar peneliti secara daring [11]. Sistem ini mempermudah koordinasi penelitian lintas institusi dan wilayah geografis. Dengan integrasi fitur yang lengkap, Watase mempercepat proses kajian literatur berbasis bukti.

3. METODOLOGI PENELITIAN

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Untuk mengarahkan fokus penelitian, dirumuskan tiga research question utama yang menjadi landasan dalam proses telaah literatur. Ketiga pertanyaan ini mencerminkan aspek penting yang dikaji, yaitu metode yang digunakan, tantangan implementasi, serta kelebihan dan keterbatasan penerapan AI dalam pemeriksaan dokumen. Rumusan pertanyaan ini disusun agar mampu menggambarkan secara komprehensif ruang lingkup dan tujuan penelitian. Rincian research question tersebut disajikan pada Tabel 1.

Tabel 1. Research Question

Research Questions
RQ1: Apa saja metode AI yang digunakan dalam pemeriksaan atau validasi dokumen secara otomatis?
RQ2: Apa tantangan utama dalam mengimplementasikan AI untuk pemeriksaan dokumen secara otomatis?
RQ3: Apa kelebihan dan keterbatasan penerapan AI dalam sistem pemeriksaan dokumen?

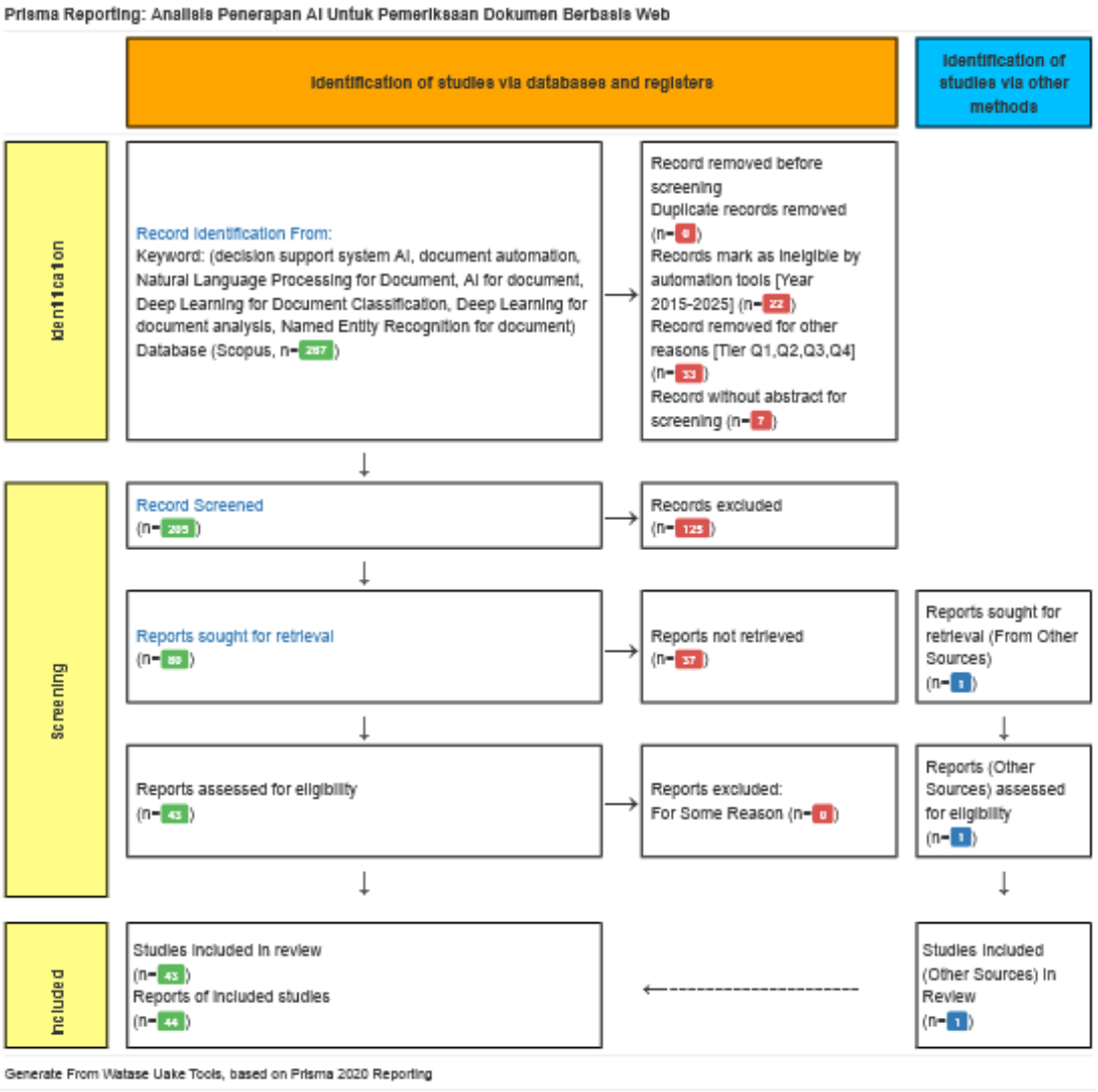
Untuk menjawab fokus kajian dalam penelitian ini, digunakan metodologi *Systematic Literature Review (SLR)* dengan metode *Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)*. Metode ini dipilih karena memungkinkan peneliti untuk mengumpulkan, mengevaluasi, dan mensintesis literatur yang relevan secara sistematis dan transparan, sehingga hasil yang diperoleh dapat dipertanggungjawabkan serta direplikasi oleh studi selanjutnya [9]. Proses SLR dalam penelitian ini mengacu pada tiga tahapan utama, yaitu: *planning, conducting, dan reporting.* Untuk mendukung proses SLR, peneliti menggunakan platform berbasis *web* bernama *Watase. Watase* merupakan sistem penelitian kolaboratif antar peneliti secara online [11], alat bantu yang dirancang untuk mempermudah proses tinjauan sistematis, mulai dari manajemen referensi, penyaringan artikel, hingga ekstraksi data.

Pada tahap *planning*, peneliti menyusun pertanyaan penelitian, mengidentifikasi kata kunci, memilih database pencarian, serta menentukan kriteria inklusi dan eksklusi. Selanjutnya, tahap conducting dilakukan dengan mengikuti panduan PRISMA yang terdiri dari empat langkah utama: *identification, screening, eligibility,* dan *included* [10]. Tahapan *identification* dimulai dengan pencarian artikel dari *database* yang relevan menggunakan kata kunci yang telah ditentukan. Tujuan dari tahap ini adalah mengumpulkan seluruh studi yang berpotensi relevan dengan topik kajian. PRISMA menekankan bahwa proses ini harus terdokumentasi dengan baik agar dapat direplikasi secara ilmiah [10].

Pada tahap *screening*, artikel diseleksi berdasarkan judul dan abstrak. Artikel yang tidak sesuai dengan kriteria inklusi awal akan dikeluarkan dari analisis untuk menjaga kualitas dan fokus tinjauan. Tahap berikutnya adalah *eligibility*, di mana artikel yang lolos proses screening dibaca secara menyeluruh *(full-text review)* untuk memastikan kesesuaiannya dengan kriteria penelitian. Pada tahap ini juga dijelaskan secara eksplisit alasan eksklusi setiap artikel yang tidak memenuhi syarat. Terakhir, artikel yang memenuhi seluruh kriteria dimasukkan dalam tahap included untuk dianalisis lebih lanjut.

PRISMA menyarankan agar seluruh proses seleksi ditampilkan secara visual dalam diagram alir guna menunjukkan transparansi metodologi yang digunakan [10]. PRISMA terdiri atas empat tahapan utama yaitu *Identification, Screening, Eligibility,* dan *Included* [10]. Visualisasi ini membantu menjelaskan proses penyaringan studi dari tahap identifikasi hingga artikel yang akhirnya dianalisis. Visualisasi alur seleksi literatur dalam penelitian ini menggambarkan secara sistematis proses identifikasi, penyaringan, penilaian kelayakan, hingga pemilihan akhir studi berdasarkan pedoman PRISMA. Diagram ini juga berfungsi untuk menunjukkan transparansi metodologi dan justifikasi pemilihan artikel yang dianalisis dalam penelitian ini, sebagaimana ditampilkan pada Gambar 1.

p-ISSN : 2803-1507 e-ISSN : 2803-1531



Gambar 1. *Systematic Literature Review (SLR) Report*

3.1 *Identification*

Pada tahap ini dilakukan pencarian awal artikel dari database yang relevan dengan menggunakan kata kunci yang telah ditentukan sebelumnya. Tujuannya adalah mengumpulkan semua studi yang berpotensi relevan. Proses ini penting karena menjadi dasar dari kelengkapan kajian sistematis. PRISMA menekankan bahwa pencarian literatur harus menggunakan istilah yang telah didefinisikan sebelumnya serta kriteria inklusi yang eksplisit agar dapat direplikasi oleh peneliti lain [10].

3.2 *Screening*

Setelah identifikasi, dilakukan proses penyaringan terhadap hasil pencarian berdasarkan judul dan abstrak. Artikel yang tidak relevan atau tidak memenuhi kriteria awal akan dikeluarkan. Tahapan ini penting untuk menghindari bias dalam seleksi studi dan menjaga kualitas hasil tinjauan sistematis [10]. Proses ini juga memungkinkan peneliti memfokuskan kajian hanya pada literatur yang sesuai dengan tujuan penelitian.

3.3 *Eligibility*

Studi yang lolos tahap screening kemudian dinilai lebih lanjut melalui pembacaan penuh (full-text) untuk menentukan apakah mereka benar-benar memenuhi syarat untuk dimasukkan dalam ulasan. PRISMA menekankan bahwa pada tahap ini harus jelas alasan eksklusi agar proses seleksi transparan dan terdokumentasi dengan baik [10]. Proses ini membantu meningkatkan akurasi seleksi studi yang relevan.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Selain itu, dokumentasi alasan eksklusi dapat memperkuat kredibilitas dan replikasi penelitian oleh pihak lain.

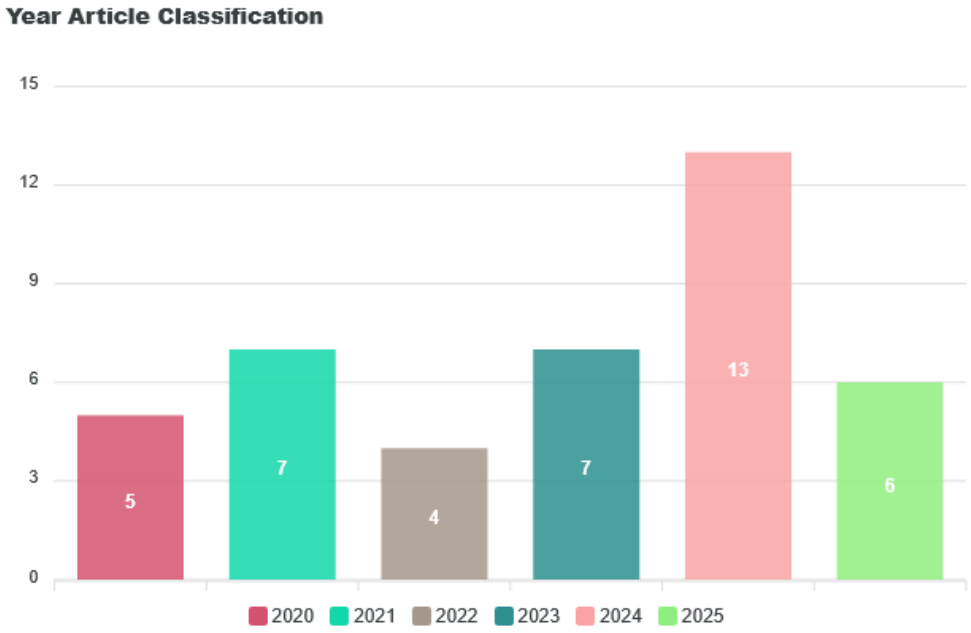
3.4 *Included*

Studi yang memenuhi seluruh kriteria akan dimasukkan ke dalam analisis akhir. PRISMA menyarankan agar jumlah studi yang disertakan serta alur seleksi ditampilkan dalam bentuk diagram alir seperti yang telah ditampilkan dalam penelitian ini, untuk menunjukkan transparansi proses seleksi [10]. Diagram ini menggambarkan jumlah artikel pada tiap tahap seleksi. Penyajian visual ini membantu pembaca memahami proses dan keputusan seleksi studi secara lebih jelas.

4. HASIL DAN PEMBAHASAN

Gambar 1 menunjukkan alur *Systematic Literature Review (SLR)* berdasarkan pedoman PRISMA yang digunakan dalam penelitian ini. Pada tahap awal *(identification),* diperoleh sebanyak 297 artikel dari database Scopus dengan menggunakan kombinasi kata kunci seperti *"decision support system AI", "document automation", "Natural Language Processing for Document"*, dan istilah lain yang relevan dengan topik pemeriksaan dokumen berbasis AI. Selanjutnya, dilakukan proses penyaringan awal terhadap artikel-artikel tersebut. Sebanyak 23 artikel dikeluarkan karena tidak termasuk dalam rentang tahun yang ditentukan (2015–2025), 13 artikel dikecualikan karena alasan konten yang tidak sesuai tema kajian, dan 7 artikel lainnya dikecualikan karena tidak memiliki abstrak yang dapat ditinjau. Dengan demikian, sebanyak 254 artikel lolos ke tahap *screening*.

Pada tahap *screening*, dilakukan seleksi berdasarkan judul dan abstrak. Dari 254 artikel yang ditinjau, sebanyak 124 artikel dikeluarkan karena tidak memenuhi kriteria inklusi. Sisanya, 130 artikel dilanjutkan ke proses pengambilan dokumen penuh *(full-text retrieval)*, namun 37 di antaranya tidak berhasil diakses dan dikeluarkan dari proses seleksi. Tahap berikutnya adalah *eligibility*, di mana 93 artikel yang berhasil diperoleh secara penuh dievaluasi lebih dalam untuk menilai kesesuaian metodologi, fokus kajian, dan relevansi terhadap topik penelitian. Hasil dari tahap ini menunjukkan bahwa 52 artikel tidak memenuhi kriteria kelayakan, dan hanya 42 artikel yang dinyatakan layak untuk dianalisis lebih lanjut. Ke-42 artikel inilah yang kemudian dijadikan sebagai bahan dalam proses sintesis literatur sistematis untuk menjawab pertanyaan penelitian terkait penerapan teknologi AI dalam pemeriksaan dokumen. Proses seleksi ini menggambarkan prinsip transparansi dan keterlacakan yang menjadi inti dari metode SLR dengan pedoman PRISMA. Distribusi 42 artikel terpilih berdasarkan tahun publikasi ditampilkan pada Gambar 2, yang menunjukkan tren peningkatan jumlah publikasi secara signifikan pada tahun 2025.

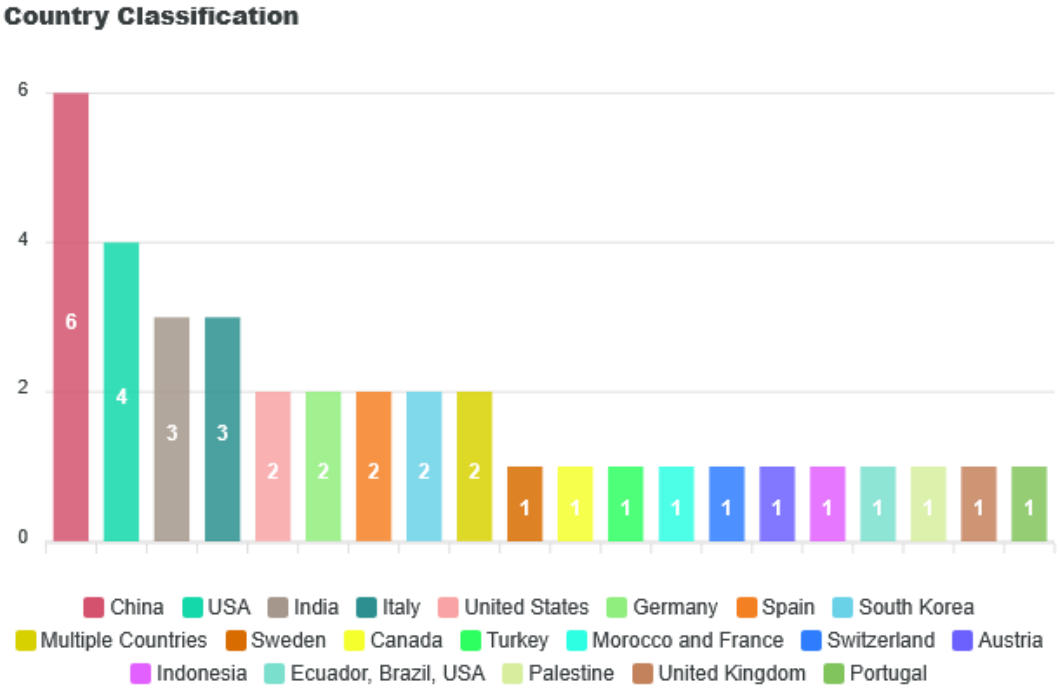


Gambar 2. Jumlah artikel berdasarkan tahun publikasi

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Berdasarkan Gambar 2, terlihat distribusi artikel yang disertakan dalam tinjauan sistematis ini berdasarkan tahun publikasi. Artikel dengan tahun publikasi tertinggi berasal dari tahun 2025, yaitu sebanyak 13 artikel. Kemudian, terdapat masing-masing 7 artikel dari tahun 2021 dan 2023, serta 6 artikel dari tahun 2024. Adapun artikel dari tahun 2020 tercatat sebanyak 5 artikel, dan artikel tertua yang dianalisis berasal dari tahun 2022 dengan jumlah 4 artikel. Hal ini menunjukkan bahwa tren penelitian mengenai penerapan AI dalam pemeriksaan dokumen cenderung meningkat dalam beberapa tahun terakhir, terutama pada tahun 2025 yang menunjukkan puncak kontribusi.

Seluruh artikel yang telah lolos seleksi dan termasuk dalam tahapan included kemudian dianalisis dan diekstraksi informasinya. Informasi yang diperoleh dari artikel tersebut meliputi tahun publikasi, negara asal, pendekatan AI yang digunakan, metode utama, serta konteks penerapannya. Informasi ini menjadi dasar penyusunan sintesis tematik yang disajikan dalam pembahasan selanjutnya. Visualisasi sebaran negara asal penelitian yang terlibat dalam studi ini dapat dilihat pada Gambar 3.



Gambar 3. Klasifikasi Negara

Berdasarkan Gambar 3, dapat diketahui bahwa artikel-artikel yang direview dalam studi ini berasal dari berbagai negara dengan distribusi yang cukup beragam. China menjadi negara yang paling dominan dalam kontribusi publikasi, dengan total 6 artikel yang berasal dari sana. Disusul oleh USA dan India, masing-masing menyumbangkan 4 dan 3 artikel. Negara-negara seperti Italy, United States (terpisah label dari USA), Germany, Spain, dan South Korea masing-masing menyumbangkan 2 artikel, menunjukkan kontribusi signifikan dari negara-negara dengan kekuatan riset teknologi menengah hingga tinggi. Selanjutnya, terdapat sejumlah negara yang masing-masing menyumbang 1 artikel, antara lain: Multiple Countries (studi lintas negara), Sweden, Canada, Turkey, Morocco and France (kolaborasi), Switzerland, Austria, Indonesia, Ecuador-Brazil-USA (kolaborasi), Palestine, United Kingdom, dan Portugal. Distribusi negara ini mencerminkan bahwa penelitian tentang AI dalam domain dokumentasi digital telah menjadi perhatian global, dengan kontribusi signifikan dari negara-negara dengan infrastruktur teknologi yang kuat seperti Amerika Serikat, Tiongkok, dan Italia.

Berikut adalah studi yang membahas penerapan kecerdasan buatan (AI) dalam proses pemeriksaan dan validasi dokumen. Tabel 2 merangkum beberapa referensi terkait model, metode yang digunakan, serta tujuan dari implementasi AI pada konteks dokumen digital. Informasi ini menjadi dasar awal untuk mengidentifikasi pendekatan umum yang digunakan. Analisis lebih lanjut terhadap studi-studi ini akan memperlihatkan tren dan efektivitas dari pendekatan yang diterapkan dalam konteks autentikasi dan pengolahan dokumen.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Tabel 2. Studi Penerapan AI pada Pemeriksaan Dokumen

Ref	Tahun	Metode/Pendekatan AI	Model	*Task*	Domain Aplikasi
[12]	2024	*Natural Language Processing (NLP)*	*Text mining with OUV annotation*	Evaluasi teks konservasi lokal	Dokumen konservasi lokal
[4]	2024	*Document Layout Analysis (DLA)*	*QCD, LLM-based NLP*	Analisis layout dokumen publik	Dokumen publik
[1]	2025	*Machine Learning & Natural Language Processing (NLP)*	*SVM, RF, Ensemble*	Otomatisasi proses dokumen perbankan	Dokumen perbankan
[13]	2025	*Deep Learning*	*Multimodal Fusion & CNN*	Analitik dokumen multimodal	Dokumen multimodal
[14]	2025	*Machine Learning & Natural Language Processing (NLP)*	*Naïve Bayes, Decision Tree*	Penyaringan relevansi dokumen	Dokumen umum
[15]	2024	*Machine Learning & Natural Language Processing (NLP)*	*Naïve Bayes, Random Forest, SVM*	Ekstraksi kebutuhan non-fungsional	Dokumen perangkat lunak
[16]	2024	*Deep Learning*	*CNN, Bi-LSTM, Hybrid CNN-BiLSTM*	Klasifikasi dokumen teknis	Dokumen Teknis
[17]	2024	*Supervised Machine Learning*	*RF, SVM, Bio Clinical BERT*	Identifikasi gangguan ADL/IADL dari catatan klinis	Catatan klinis
[18]	2024	*Natural Language Processing (NLP) & Semantic Analysis*	*Rule-based NLP & Knowledge graphs*	Memahami semantik dokumen	Dokumen umum
[19]	2024	*Named Entity Recognition (NER) & Graph Construction*	*NN with Triplet Loss*	Pembangunan *knowledge graph*	Knowledge Graph
[20]	2024	*Content-based indexing & LLM*	*GPT-4o*	Peningkatan pengindeksan	Dokumen umum
[21]	2024	*Natural Language Processing (NLP) & AI Pipeline*	*LLM, NLP APIs*	Penanganan dokumen otomatis	Dokumen umum
[22]	2023	*Natural Language Processing (NLP)*	*BERT, CNN, LSTM*	Prediksi kelangsungan hidup *(survival prediction)*	Dokumen medis
[23]	2023	*Deep Learning*	*ETEDL-THDR (MF + MobileNet+ BiGRU + WSO)*	*Recognition & Classification*	Dokumen teknis
[24]	2023	*Ontology-based Deep Learning & Machine Learning*	*WordNet + BERT*	*Multi-class Classification*	*Multi-label documents*
[25]	2023	*Natural Language Processing (NLP)*	*BERT variants*	*Named Entity Recognition* dan ekstraksi informasi	Dokumen dengan Informasi Entitas
[7]	2023	*AI-based Decision Support*	*Rule-based System & NLP pipeline*	Klasifikasi pengadaan publik	Pengadaan publik
[26]	2023	*Deep Learning*	*CNN & BiLSTM*	Analisis sentimen berbasis dokumen	Dokumen opini
[27]	2021	*Deep Learning*	*CNN, BOVW, Zernike*	Tata letak dokumen	Tata letak dokumen
[28]	2021	*NLP, Clustering*	*Doc2vec, Clustering*	Pengelompokan/klastering dokumen	Dokumen umum
[29]	2022	*Deep Learning Regression*	*Custom regression-based DL Model*	*ICO success prediction*	*ICO documents*
[30]	2022	*Multi-label DL & Attention*	*Label-attention DL & domain-specific pre-training*	*Multi-label classification*	Dokumen Multi-Kategori
[31]	2021	*Deep Learning*	*YOLO, CNN-based methods*	Pemrosesan tata letak faktur	
[32]	2020	*Supervised Learning & NLP Pipeline*	*Word Embeddings & NER*	Kategorisasi dokumen dan ekstraksi informasi	Dokumen umum
[3]	2020	*Deep Learning*	*CNN, Siamese Network, LSTM*	Pengenalan dan klasifikasi dokumen	Dokumen umum
[33]	2021	*Real-time AI-based decision- making system*	*Knowledge-based AI & NLP*	Pengambilan keputusan berbasis informasi	Dokumen umum
[34]	2021	*Visual sensing deep learning*	*CNN, ResNet, Gabor filter fusion*	Klasifikasi gambar dokumen	Dokumen gambar

p-ISSN : 2803-1507 e-ISSN : 2803-1531

[35]	2020	*Benchmarking ML/DL*	*SVM, CNN, BiLSTM, Random Forest*	Klasifikasi teks	Teks
[36]	2020	*AI & NLP Pipeline*	*Rule-based & ML*	Percepatan penyusunan dokumen	Dokumen Teknis
[8]	2020	*NLP Model*	*TF-IDF & Classification*	Klasifikasi dokumen keamanan siber	Dokumen keamanan siber

Dari Tabel 2 dapat dilihat bahwa berdasarkan 42 artikel yang dianalisis, pendekatan berbasis *Natural Language Processing (NLP)* dan *Machine Learning* masih mendominasi, terutama untuk ekstraksi informasi, klasifikasi teks, dan pengolahan dokumen publik. Sementara itu, *Deep Learning* lebih banyak diterapkan pada analisis dokumen multimodal yang melibatkan teks dan citra, dengan model seperti CNN dan Bi-LSTM. Tren terbaru menunjukkan peningkatan penggunaan *Large Language Model (LLM),* integrasi multimodal *fusion*, serta kombinasi ML dan DL untuk meningkatkan akurasi. Temuan ini dirangkum secara lebih ringkas dalam Tabel 3 untuk memberikan ikhtisar yang lebih terstruktur.

Tabel 3. Ringkasan Studi Terkait

Kategori	Sintesis Temuan
Pendekatan AI Dominan	NLP, *Machine Learning &Deep Learning*
Model Paling Umum	*Naïve Bayes, Random Forest, SVM, CNN, Bi-LSTM, Ensemble, LLM-based NLP, Rule-based*
Domain Aplikasi Terbanyak	Dokumen publik, dokumen umum, perbankan, klinis
Tren Tahun 2023–2025	Peningkatan penggunaan LLM-based NLP, multimodal fusion, dan kombinasi ML+DL

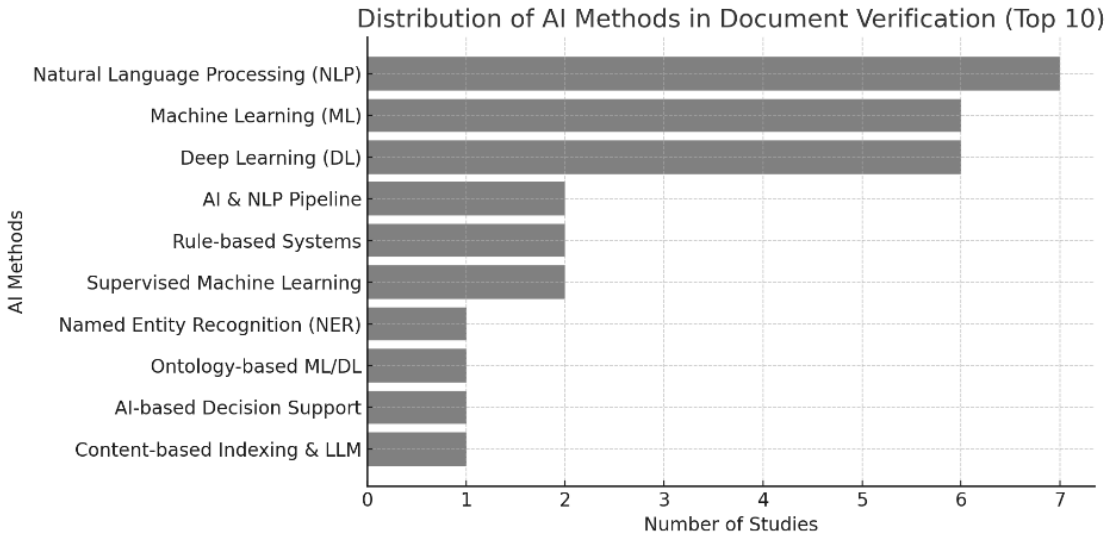
Hasil penelitian ini merupakan sintesis dari proses telaah literatur dan ekstraksi informasi terhadap 42 artikel yang dianalisis sebelumnya. Analisis ini bertujuan untuk memberikan gambaran menyeluruh mengenai pendekatan-pendekatan kecerdasan buatan (AI) yang digunakan dalam pengolahan dokumen tidak terstruktur. Untuk memandu analisis yang lebih terfokus, penelitian ini merujuk pada tiga pertanyaan penelitian utama yang telah dirumuskan sebelumnya. Bagian berikut akan menguraikan hasil temuan yang menjawab masing-masing pertanyaan tersebut secara sistematis dan berdasarkan bukti dari berbagai studi.

4.1 *(RQ1) Apa saja metode yang digunakan oleh AI dalam pemeriksaan atau validasi dokumen secara otomatis?*

Berdasarkan tinjauan pustaka yang telah dilakukan, studi ini berhasil mengidentifikasi berbagai pendekatan *Artificial Intelligence* (AI) yang digunakan dalam sistem verifikasi dokumen otomatis. Temuan menunjukkan bahwa pendekatan yang paling sering digunakan adalah *Natural Language Processing* (NLP), yang ditemukan dalam 7 studi. Sebagai contoh, dalam studi oleh Vaira [37], penggunaan *Large Language Models* (LLMs) seperti ChatGPT-4 menunjukkan kinerja yang tinggi dalam menghasilkan dokumen informed consent yang akurat dan komprehensif. Studi lain [38] yang mengembangkan model *Named Entity Recognition* (NER) berbasis *multi-task learning* dan *cross-document attention* yang mampu secara konsisten mengidentifikasi entitas kunci di berbagai dokumen. Studi lainnya, menerapkan pencarian berbasis NLP semantik untuk meningkatkan efisiensi pengambilan informasi medis dalam *electronic health records*, yang secara signifikan mempercepat alur kerja klinis [39]. Studi lain juga menyoroti peran NER dalam memproses teks-teks historis untuk keperluan digitalisasi [40].

Selain itu, metode *Machine Learning* (ML) dan *Deep Learning* (DL) masing-masing digunakan dalam 6 studi. *Supervised Machine Learning* ditemukan dalam 2 studi, sementara metode seperti *Document Layout Analysis* (DLA), *Named Entity Recognition* (NER), *Ontology-based ML/DL*, dan *AI-based Decision Support* masing-masing muncul dalam 1 studi. Pendekatan lainnya, meskipun digunakan lebih jarang, meliputi *Content-based Indexing & LLM*, *Multi-label DL & Attention*, *Visual Sensing Deep Learning*, *Benchmarking ML/DL*, *Clustering with NLP*, *Regression-based DL*, *NLP & Semantic Analysis*, *NER & Graph Construction*, *NLP Model (TF-IDF)*, *AI & NLP Pipeline*, dan *Rule-based Systems*. Distribusi sepuluh pendekatan AI teratas dalam sistem verifikasi dokumen sebagaimana diidentifikasi dalam tinjauan ini, diilustrasikan pada Gambar 4.

p-ISSN : 2803-1507 e-ISSN : 2803-1531



Gambar 4. Distribusi Metode AI dalam Pemeriksaan Dokumen

4.2 *(RQ2) Apa tantangan utama dalam mengimplementasikan AI untuk pemeriksaan dokumen secara otomatis?*

Berdasarkan tinjauan terhadap berbagai studi, penerapan AI untuk verifikasi dokumen menghadapi sejumlah tantangan teknis dan praktis yang signifikan. Salah satu tantangan utama adalah keterbatasan dalam kualitas dan konsistensi data masukan. Dokumen yang tidak terstruktur atau memiliki format yang bervariasi dapat menghambat proses ekstraksi informasi, seperti yang disoroti oleh studi Liu, yang menekankan bahwa perbedaan dalam struktur dan semantik antar dokumen dapat mengurangi kinerja algoritma NLP dan membutuhkan pendekatan semantik yang lebih canggih, seperti *knowledge graph* dan ekstraksi berbasis aturan (*rule-based extraction*) [18]. Masalah ini menjadi lebih kompleks ketika sistem harus menangani dokumen multibahasa atau hasil pemindaian berkualitas rendah yang memperburuk akurasi ekstraksi informasi.

Selain itu, beberapa studi menunjukkan bahwa pendekatan multimodal atau berbasis tata letak (misalnya pada dokumen legal atau dokumen domain publik) sering kali memerlukan anotasi manual atau intervensi manusia untuk menjaga akurasi, sehingga mengurangi potensi otomatisasi penuh. Misalnya, menekankan pentingnya peran "*human-in-the-loop*" dalam analisis tata letak dokumen karena keterbatasan AI dalam menafsirkan konteks visual yang kompleks [4]. Studi lain menunjukkan bagaimana variasi dalam data klinis, struktur rekam medis elektronik, dan kompleksitas konteks penggunaan menyebabkan hasil yang tidak konsisten dari sistem AI [41]. Demikian pula, studi lainnya mencatat bahwa variasi terminologi medis dan kesalahan ketik menyulitkan proses pengambilan informasi secara otomatis [19].

Tantangan lainnya terletak pada keterbatasan kemampuan generalisasi model AI ke konteks atau domain yang berbeda. Model yang dilatih pada satu jenis dokumen sering kali gagal menunjukkan performa yang efektif pada dokumen dengan format atau konten yang berbeda. Seperti yang dikemukakan, model NER cenderung hanya berkinerja baik dalam domain atau dataset tertentu [38]. Secara konseptual, menyoroti pentingnya standar hukum dan etika ketika mengintegrasikan AI ke dalam sistem hukum dan pemerintahan, mengingat adanya risiko penyalahgunaan atau bias dalam pengambilan keputusan otomatis [42], [43]. Studi lainnya, juga menekankan bahwa keterjelasan (*interpretability*) keluaran model AI dalam konteks klinis tetap menjadi tantangan penting, khususnya dalam menjaga kepercayaan pengguna akhir serta memastikan bahwa keputusan yang dihasilkan AI bersifat transparan dan dapat dijelaskan [17].

Dari perspektif teknis, penggunaan model besar seperti BERT dan GPT-4 juga menghadirkan tantangan terkait tingginya kebutuhan sumber daya komputasi, masalah latensi, dan kesulitan dalam mengintegrasikan model-model tersebut ke dalam sistem verifikasi dokumen berbasis *web* secara *real-time* [20], [21]. Beberapa organisasi yang mencoba mengadopsi model besar sering kali menghadapi keterbatasan infrastruktur, terutama dalam hal kapasitas GPU dan manajemen memori. Selain itu, proses pelatihan ulang (fine-tuning) untuk penyesuaian domain tertentu membutuhkan waktu yang lama dan biaya yang tinggi, sehingga tidak

p-ISSN : 2803-1507 e-ISSN : 2803-1531

semua institusi dapat melakukannya secara berkelanjutan. Masalah lainnya adalah kurangnya interoperabilitas antara model besar dan sistem lama (legacy systems) yang masih banyak digunakan dalam sektor pemerintahan atau layanan publik.

4.3 *(RQ3) Apa kelebihan dan keterbatasan penerapan AI dalam sistem pemeriksaan dokumen?*

Berdasarkan literatur yang telah dianalisis, penerapan *Artificial Intelligence* (AI) dalam sistem verifikasi dokumen menawarkan sejumlah keuntungan yang signifikan. Sejumlah studi menunjukkan bahwa AI, khususnya melalui model berbasis *Machine Le*arning dan *Natural Language Processing* (NLP), dapat meningkatkan efisiensi, mengurangi biaya operasional, serta mempercepat proses validasi dokumen. Sebagai contoh, sistem otomatis berbasis NLP mampu mengklasifikasikan dan mengekstraksi informasi dari ribuan dokumen dengan kecepatan dan akurasi yang lebih tinggi dibandingkan dengan tenaga manusia [15], [16]. Model *deep learning* seperti CNN dan Bi-LSTM juga terbukti efektif dalam memahami struktur dan konteks dokumen yang kompleks [23], [26]. Selain itu, sistem yang mengintegrasikan data multimodal, seperti gambar dan teks dalam dokumen, menunjukkan hasil yang menjanjikan dalam tugas-tugas seperti analisis tata letak dokumen dan ekstraksi informasi visual [4], [13].

Namun, implementasi AI dalam bidang ini juga menghadapi sejumlah keterbatasan. Salah satu tantangan utama adalah ketergantungan pada data pelatihan yang berkualitas tinggi dan dalam jumlah besar, yang sering kali tidak memadai atau memiliki anotasi yang buruk [20]. Selain itu, beberapa model masih bersifat tidak transparan (*black-box*), sehingga sulit untuk diinterpretasikan dan menimbulkan masalah kepercayaan, terutama dalam sektor yang sangat sensitif seperti keuangan dan pemerintahan [44]. Pada tingkat kebijakan, studi menunjukkan bahwa banyak pemerintah daerah masih berada pada tahap awal dalam menerapkan prinsip AI yang bertanggung jawab, sehingga kerangka etika dan tata kelola yang ada masih terbatas [42]. Tantangan lainnya mencakup keterbatasan kemampuan generalisasi model AI terhadap dokumen dengan format dan bahasa yang beragam, serta kebutuhan sumber daya komputasi yang tinggi [17], [18].

5. KESIMPULAN DAN SARAN

5.1 *Kesimpulan*

Studi ini menyajikan sintesis komprehensif dari penelitian-penelitian terbaru mengenai penerapan kecerdasan buatan (*artificial intelligence*/AI) dalam sistem verifikasi dokumen otomatis. Melalui tinjauan sistematis terhadap 42 studi akademik yang dipublikasikan antara tahun 2020 hingga 2025, ditemukan bahwa *Natural Language Processing* (NLP), *Machine Learning* (ML), dan *Deep Learning* (DL) merupakan metode yang paling umum digunakan. Teknologi-teknologi ini menawarkan manfaat nyata dalam meningkatkan efisiensi, akurasi, dan skalabilitas sistem analisis dokumen. Namun demikian, sejumlah tantangan penting masih belum teratasi, seperti inkonsistensi data, keterbatasan generalisasi model terhadap berbagai jenis dokumen, serta sifat tidak transparan (*opaque*) dari beberapa model AI. Dampak praktis dan operasional dari temuan ini sangat relevan, khususnya di sektor-sektor yang sangat bergantung pada pengelolaan dokumen skala besar, seperti layanan kesehatan, pendidikan, administrasi hukum, dan pelayanan publik. Dengan meningkatkan otomatisasi di bidang-bidang tersebut, organisasi dapat mengurangi beban kerja manual dan keterlambatan operasional.

5.2 *Saran*

Berdasarkan temuan dalam studi ini, disarankan agar pengembangan sistem verifikasi dokumen otomatis ke depan difokuskan pada pembuatan model AI yang lebih adaptif dan dapat dijelaskan (*interpretable*), sehingga mampu menangani data heterogen serta bekerja secara konsisten pada dokumen dengan berbagai format dan bahasa. Peneliti selanjutnya diharapkan dapat mengeksplorasi pendekatan yang menggabungkan NLP dengan model multimodal secara lebih optimal untuk mengatasi tantangan struktur dan konteks dokumen yang kompleks. Selain itu, penting bagi pengembang sistem dan pembuat kebijakan untuk mempertimbangkan aspek transparansi dan etika dalam implementasi AI, terutama pada sektor-sektor sensitif seperti kesehatan, hukum, dan pemerintahan. Mengatasi keterbatasan seperti ketergantungan pada data pelatihan yang berkualitas tinggi dan isu *black-box* pada model AI dapat menjadi landasan penting bagi penelitian lanjutan dalam membangun sistem yang lebih andal dan dapat dipercaya.

DAFTAR PUSTAKA

[1] J. L. Alonso-Rocha, A. Martínez-Rojas, J. G. Enríquez, and J. M. Sánchez-Oliva, “From manual to automated: a state-of-the-art review to examine the impact of intelligent document processing in

p-ISSN : 2803-1507 e-ISSN : 2803-1531

banking automation," *Expert Syst. Appl.*, vol. 285, no. April, 2025, doi: 10.1016/j.eswa.2025.127958.

[2] S. V. Mahadevkar, S. Patil, K. Kotecha, L. W. Soong, and T. Choudhury, "Exploring AI-driven approaches for unstructured document analysis and future horizons," *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00948-z.

[3] F. Lombardi and S. Marinai, "Deep Learning for Historical Document Analysis and Recognition—A Survey," *J. Imaging*, vol. 6, no. 10, 2020, doi: 10.3390/jimaging6100110.

[4] A. Peña *et al.*, "Continuous document layout analysis: Human-in-the-loop AI-based data curation, database, and evaluation in the domain of public affairs," *Inf. Fusion*, vol. 108, no. April, p. 102398, 2024, doi: 10.1016/j.inffus.2024.102398.

[5] L. Jacques de Sousa, J. Poças Martins, L. Sanhudo, and J. Santos Baptista, "Automation of text document classification in the budgeting phase of the Construction process: a Systematic Literature Review," *Constr. Innov.*, vol. 24, no. 7, pp. 292–318, 2024, doi: 10.1108/CI-12-2022-0315.

[6] C. Y. Y. Kesiku, A. Chaves-Villota, and B. Garcia-Zapirain, "Natural Language Processing Techniques for Text Classification of Biomedical Documents: A Systematic Review," *Inf.*, vol. 13, no. 10, 2022, doi: 10.3390/info13100499.

[7] L. Siciliani, V. Taccardi, P. Basile, M. Di Ciano, and P. Lops, "AI-based decision support system for public procurement," *Inf. Syst.*, vol. 119, p. 102284, 2023, doi: 10.1016/j.is.2023.102284.

[8] T. M. Georgescu, "Natural language processing model for automatic analysis of cybersecurity-related documents," *Symmetry (Basel).*, vol. 12, no. 3, 2020, doi: 10.3390/sym12030354.

[9] P. Zare, C. Pettit, S. Leao, and O. Gudes, "Digital Bicycling Planning: A Systematic Literature Review of Data-Driven Approaches," *Sustain.*, vol. 14, no. 23, 2022, doi: 10.3390/su142316319.

[10] A. Schniedermann, "Shaping the Qualities, Values and Standards of Science. How Reporting Guidelines Improve the Transparency of Biomedical Research," *Front. Res. Metrics Anal.*, vol. 7, no. June, pp. 1–12, 2022, doi: 10.3389/frma.2022.846822.

[11] S. Simamora and V. Gaffar, "Systematic Literatur Review Dengan Metode Prisma: Dampak Teknologi Blockchain Terhadap Periklanan Digital," *J. Ilm. M-Progress*, vol. 14, no. 1, 2024, doi: 10.35968/m-pu.v14i1.1182.

[12] Y. Chen, L. Zhang, and Q. Dong, "Using natural language processing to evaluate local conservation text: A study of 624 documents from 303 sites of the World Heritage Cities Programme," *J. Cult. Herit.*, vol. 70, no. October, pp. 259–270, 2024, doi: 10.1016/j.culher.2024.09.011.

[13] C. Gerling and S. Lessmann, "Multimodal Document Analytics for Banking Process Automation," *Inf. Fusion*, vol. 118, no. January, p. 102973, 2023, doi: 10.1016/j.inffus.2025.102973.

[14] R. Bridgelall, "Document Relevance Filtering by Natural Language Processing and Machine Learning: A Multidisciplinary Case Study of Patents," *Appl. Sci.*, vol. 15, no. 5, 2025, doi: 10.3390/app15052357.

[15] Q. A. Shreda and A. A. Hanani, "Identifying Non-functional Requirements from Unconstrained Documents using Natural Language Processing and Machine Learning Approaches," *IEEE Access*, vol. 4, pp. 1–22, 2021, doi: 10.1109/ACCESS.2021.3052921.

[16] S. Jiang, J. Hu, C. L. Magee, and J. Luo, "Deep Learning for Technical Document Classification," *IEEE Trans. Eng. Manag.*, vol. 71, no. March, pp. 1163–1179, 2024, doi: 10.1109/TEM.2022.3152216.

[17] J. Laurentiev *et al.*, "Identifying Functional Status Impairment in People Living With Dementia Through Natural Language Processing of Clinical Documents: Cross-Sectional Study," *J. Med. Internet Res.*, vol. 26, no. 1, pp. 1–12, 2024, doi: 10.2196/47739.

[18] Y. Liu, "Role of Natural Language Processing in Document Understanding and Semantic Analysis: A Chinese Perspective," *Prof. la Inf.*, vol. 33, no. 3, pp. 1–11, 2024, doi: 10.3145/epi.2025.ene.0324.

[19] S. H. Jeon, H. J. Lee, J. Park, and S. Cho, "Building knowledge graphs from technical documents using named entity recognition and edge weight updating neural network with triplet loss for entity normalization," *Intell. Data Anal.*, vol. 28, no. 1, pp. 331–355, 2024, doi: 10.3233/IDA-227129.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

[20] S. Bouzid and L. Piron, "Leveraging Generative AI in Short Document Indexing," *Electron.*, vol. 13, no. 17, pp. 1–24, 2024, doi: 10.3390/electronics13173563.

[21] G. Di Marzo Serugendo *et al.*, "Streamlining Tax and Administrative Document Management with AI-Powered Intelligent Document Management System," *Inf.*, vol. 15, no. 8, 2024, doi: 10.3390/info15080461.

[22] J. J. Nunez, B. Leung, C. Ho, A. T. Bates, and R. T. Ng, "Predicting the Survival of Patients With Cancer From Their Initial Oncology Consultation Document Using Natural Language Processing," *JAMA Netw. open*, vol. 6, no. 2, p. e230813, 2023, doi: 10.1001/jamanetworkopen.2023.0813.

[23] C. Vinotheni and S. L. Pandian, "End-To-End Deep-Learning-Based Tamil Handwritten Document Recognition and Classification Model," *IEEE Access*, vol. 11, no. May, pp. 43195–43204, 2023, doi: 10.1109/ACCESS.2023.3270895.

[24] I. Yelmen, A. Gunes, and M. Zontul, "Multi-Class Document Classification Using Lexical Ontology-Based Deep Learning †," *Appl. Sci.*, vol. 13, no. 10, 2023, doi: 10.3390/app13106139.

[25] A. Turchin, S. Masharsky, and M. Zitnik, "Comparison of BERT implementations for natural language processing of narrative medical documents," *Informatics Med. Unlocked*, vol. 36, no. November 2022, p. 101139, 2023, doi: 10.1016/j.imu.2022.101139.

[26] P. Atandoh, F. Zhang, D. Adu-Gyamfi, P. H. Atandoh, and R. E. Nuhoho, "Integrated deep learning paradigm for document-based sentiment analysis," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 7, p. 101578, 2023, doi: 10.1016/j.jksuci.2023.101578.

[27] F. Grijalva, E. Santos, B. Acuna, J. C. Rodriguez, and J. C. Larco, "Deep Learning in Time-Frequency Domain for Document Layout Analysis," *IEEE Access*, vol. 9, pp. 151254–151265, 2021, doi: 10.1109/ACCESS.2021.3125913.

[28] I. Ö. Arnarsson, O. Frost, E. Gustavsson, M. Jirstrand, and J. Malmqvist, "Natural language processing methods for knowledge management—Applying document clustering for fast search and grouping of engineering documents," *Concurr. Eng. Res. Appl.*, vol. 29, no. 2, pp. 142–152, 2021, doi: 10.1177/1063293X20982973.

[29] J. Wang, R. Chen, W. Xu, Y. Tang, and Y. Qin, "A document analysis deep learning regression model for initial coin offerings success prediction," *Expert Syst. Appl.*, vol. 210, no. August, p. 118367, 2022, doi: 10.1016/j.eswa.2022.118367.

[30] D. Song, A. Vold, K. Madan, and F. Schilder, "Multi-label legal document classification: A deep learning-based approach with label-attention and domain-specific pre-training," *Inf. Syst.*, vol. 106, no. xxxx, p. 101718, 2022, doi: 10.1016/j.is.2021.101718.

[31] Di. Baviskar, S. Ahirrao, and K. Kotecha, "Multi-Layout Unstructured Invoice Documents Dataset: A Dataset for Template-Free Invoice Processing and Its Evaluation Using AI Approaches," *IEEE Access*, vol. 9, pp. 101494–101512, 2021, doi: 10.1109/ACCESS.2021.3096739.

[32] I. Bordino, A. Ferretti, F. Gullo, and S. Pascolutti, "GarNLP: A Natural Language Processing Pipeline for Garnishment Documents," *Inf. Syst. Front.*, vol. 23, no. 1, pp. 101–114, 2021, doi: 10.1007/s10796-020-09997-0.

[33] A. Islam and K. Chang, "Real-time AI-based informational decision-making support system utilizing dynamic text sources," *Appl. Sci.*, vol. 11, no. 13, 2021, doi: 10.3390/app11136237.

[34] K. Mohsenzadegan, V. Tavakkoli, and K. Kyamakya, "A deep-learning based visual sensing concept for a robust classification of document images under real-world hard conditions," *Sensors*, vol. 21, no. 20, 2021, doi: 10.3390/s21206763.

[35] M. N. Asim, M. U. Ghani, M. A. Ibrahim, W. Mahmood, A. Dengel, and S. Ahmed, *Benchmarking performance of machine and deep learning-based methodologies for Urdu text document classification*, vol. 33, no. 11. Springer London, 2021. doi: 10.1007/s00521-020-05321-8.

[36] S. Viswanath, J. W. Fennell, K. Balar, and P. Krishna, "An Industrial Approach to Using Artificial

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Intelligence and Natural Language Processing for Accelerated Document Preparation in Drug Development," *J. Pharm. Innov.*, vol. 16, no. 2, pp. 302–316, 2021, doi: 10.1007/s12247-020-09449-x.

[37] L. A. Vaira *et al.*, "Evaluating AI-Generated informed consent documents in oral surgery: A comparative study of ChatGPT-4, Bard gemini advanced, and human-written consents," *J. Cranio-Maxillofacial Surg.*, vol. 53, no. June 2024, pp. 18–23, 2024, doi: 10.1016/j.jcms.2024.10.002.

[38] D. Wang, H. Fan, and J. Liu, "Learning with joint cross-document information via multi-task learning for named entity recognition," *Inf. Sci. (Ny).*, vol. 579, pp. 454–467, 2021, doi: 10.1016/j.ins.2021.08.015.

[39] E. H. Park *et al.*, "Evaluating the Impact on Clinical Task Efficiency of a Natural Language Processing Algorithm for Searching Medical Documents: Prospective Crossover Study," *JMIR Med. Informatics*, vol. 10, no. 10, pp. 1–11, 2022, doi: 10.2196/39616.

[40] M. Humbel, J. Nyhan, A. Vlachidis, K. Sloan, and A. Ortolja-Baird, "Named-entity recognition for early modern textual documents: a review of capabilities and challenges with strategies for the future," *J. Doc.*, vol. 77, no. 6, pp. 1223–1247, 2021, doi: 10.1108/JD-02-2021-0032.

[41] G. D. Giebel *et al.*, "Problems and Barriers Related to the Use of AI-Based Clinical Decision Support Systems: Interview Study," *J. Med. Internet Res.*, vol. 27, 2025, doi: 10.2196/63377.

[42] A. David *et al.*, "Understanding local government responsible AI strategy: An international municipal policy document analysis," *Cities*, vol. 155, no. April, p. 105502, 2024, doi: 10.1016/j.cities.2024.105502.

[43] O. R. Goodenough and P. J. Carlson, "Words or code first? Is the legacy document or a code statement the better starting point for complexity-reducing legal automation?," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 382, no. 2270, 2024, doi: 10.1098/rsta.2023.0160.

[44] S. Ackerhans, K. Wehkamp, R. Petzina, D. Dumitrescu, and C. Schultz, "Perceived Trust and Professional Identity Threat in AI-Based Clinical Decision Support Systems: Scenario-Based Experimental Study on AI Process Design Features," *JMIR Form. Res.*, vol. 9, pp. 1–17, 2025, doi: 10.2196/64266.