A Claim-Centric Multi-Source Verification Architecture for Reducing Hallucinations in Retrieval-Augmented Generation

Prabhjot Singh
Software Engineering
Delhi Technological University
Delhi, India
prabhjotsingh_mc22a14_53@dtu.ac.in

***Abstract*—Retrieval-Augmented Generation (RAG) has become an effective approach for grounding large language models (LLMs) in external knowledge bases. However, conventional RAG pipelines are still susceptible to hallucinations due to their dependence on single-pass retrieval and the implicit confidence placed in the retrieved information, which restricts their dependability in scientific and critical domains.**

This paper proposes ARA, a claim-centric verification architecture that decomposes documents into atomic, verifiable claims and evaluates each claim using a structured multi-stage validation pipeline. The framework integrates claim extraction, query reformulation, multi-source web retrieval, and an evidence-grounded verification mechanism that bases decisions solely on externally retrieved sources, thereby reducing reliance on internal model priors and improving interpretability.

We evaluated ARA on a curated dataset of 496 claims extracted from scientific documents and compared its performance with that of a standard RAG baseline under identical retrieval settings. Our experimental results demonstrate that ARA achieves an accuracy of 61.09%, outperforming the baseline accuracy of 40.52%, and improves the macro F1-score from 0.210 to 0.377. Notably, ARA reduces hallucination rates from 58.27% to 11.29%, representing a reduction of more than five times in false-positive assertions.

These results highlight the effectiveness of structured claim-level verification and multi-source grounding in effectively reducing hallucinations in LLM-based systems. Our proposed architecture provides a practical and extensible framework for improving factual reliability in generative AI applications in the future.

Keywords—Retrieval-Augmented Generation, Hallucination Reduction, Claim Verification, Evidence Retrieval, Large Language Models

I. INTRODUCTION

Large language models (LLMs) have shown strong performance across a wide range of natural language processing tasks, including question answering, summarization, and scientific reasoning. However, despite these capabilities, LLMs still tend to generate factually incorrect or unsupported statements, commonly referred to as hallucinations. This issue becomes especially critical in high-stakes domains such as scientific research, healthcare, and policy-making, where accuracy and reliability are essential.

Retrieval-Augmented Generation (RAG) has emerged as a practical approach to improve the grounding of LLM outputs by incorporating external knowledge during inference. By retrieving relevant document segments and conditioning generation on this information, RAG reduces dependence on the model's internal parametric knowledge. However, most RAG pipelines rely on a single-pass retrieval process and assume that the retrieved content is both accurate and sufficient. In reality, retrieved information may be incomplete, noisy, or even misleading, which can still lead to hallucinations despite the presence of external knowledge sources.

A major limitation of existing RAG systems is the lack of explicit verification mechanisms. Retrieved documents are often treated as reliable without validating whether the generated statements are actually supported by evidence. Additionally, the tight coupling between retrieval and generation prevents independent verification of individual claims. This absence of claim-level validation limits the system's ability to identify incorrect assertions, ultimately contributing to hallucinations.

To address these challenges, this paper introduces ARA, a claim-centric verification architecture designed to enhance factual reliability in RAG pipelines. The proposed approach decomposes documents into atomic, self-contained claims and evaluates each claim through a structured multi-stage verification process. The pipeline integrates claim extraction, query reformulation, multi-source evidence retrieval, and evidence-based validation. By separating claim generation from verification, the architecture reduces reliance on internal model knowledge and enables more robust decision-making.

The effectiveness of the proposed architecture is evaluated on a curated dataset of 496 claims derived from scientific documents. Experimental results show notable improvements in accuracy and macro F1-score, along with a significant reduction in hallucination rates compared to a standard RAG baseline. These findings suggest that structured claim-level verification offers a reliable approach for improving the trustworthiness of generative AI systems.

The main contributions of this paper are as follows:

1) A claim-centric verification architecture that decomposes documents into atomic, self-contained, and verifiable units.

2) A multi-stage evidence retrieval and validation pipeline that leverages multi-source external knowledge, reducing dependence on single-pass retrieval.

3) An empirical evaluation demonstrating significant improvements in accuracy and a more than fivefold reduction in hallucination rates compared to a standard RAG baseline.

The remainder of the paper is organized as follows. Section II reviews related work. Section III presents the proposed methodology. Section IV describes the experimental setup. Section V discusses the results and analysis. Section VI provides discussion and limitations, and Section VII concludes the paper.

II. Related Work

A. Retrieval-Augmented Generation

RAG has become a widely utilized approach for grounding LLMs in external knowledge bases during inference. The foundational RAG framework introduced by Lewis et al.[1]combined retrieval with sequence generation which enabled models to ground outputs on retrieved documents. Subsequent work has lead to improvement in retrieval quality which includes better embedding models, hybrid retrieval strategies and re ranking mechanisms. Apart from these advances, mostly RAG systems rely on a single pass retrieval assume that the retrieved context is accuracy. This means trust excessive trust in documents lead to propagation of errors which results in persistent hallucinations.

B. Fact Verification and Claim Checking

Fact verification has been widely studied and implemented in natural language processing and natural language understanding, particularly in the context of claim checking. Early approaches for verification such as FEVER and verification pipelines breakdown of claims and retrieve supporting evidence from large pool of pre defined documented claims. Recent work has used transformer based architectures for evidence retrieval and claim classification, improving performance on benchmark datasets which demonstrated significant improvement over the past few years. Many of these systems are designed for small, isolated claims and mostly depend on structured datasets with evidence source with no real mechanism for gathering real time data. In contrast, long form documents or research papers require search mechanisms for extracting and verifying multiple claims across external web sources. Existing verification systems rely on limited retrieval strategies which again means there will be issues related to reliability and single pass.

C. Multi-step Reasoning and Agentic LLM Systems

Recent research has explored multi step reasoning frameworks that uses LLMs with multiple reasoning steps and tool use. Approaches such as ReAct agents [3] reasoning and action steps, allows models to query external tools during inference but ReAct agents often restrict tool usage. Similarly, multi agent systems have been proposed to break down complex tasks into smaller tasks handled by specialized agents. While these improve reasoning capabilities and task decomposition, they are not designed for factual verification of the query. In particular, they lack structured mechanisms for claim level validation and often do not enforce strict evidence grounding, leaving them prone to yet again hallucinations.

D. Summary and Research Gap

Even though significant progress has been made in retrieval-augmented generation, fact verification, and multi agent systems, existing approaches lack a combined framework that uses claim level decomposition with multi source external evidence grounding and explicit verification. RAG systems do not check validity of generated content, while fact checking systems are not optimized for document processing for automated extraction of claims. Agentic frameworks focus on reasoning and task execution and not systematic verification. This gap motivates the need for a structured and claim first verification architecture that integrates multi source external web retrieval with evidence based validation, as experimented in this research work.

III. Methodology

A. System Overview

Our proposed Agentic Research Assistant architecture is designed to improve factual reliability, claim extraction and verification in large language models outputs through structured, claim verification with external knowledgebase. Unlike conventional retrieval-augmented generation pipelines, which rely only on generation, Our system ARA introduces an explicit verification layer that operates at the level of atomic claims. Our overall ARA pipeline consists of five main stages: claim extraction, multi query generation, external evidence retrieval, evidence combining, and claim verification. Our architecture enforces a clear separation between generation and validation.

B. Claim Extraction

The architecture entry point is an input document or generated response, the first part is to break the content into set of atomic self contained claims. Each claim is unique and verifiable, avoiding pronouns or contextual dependencies which could make evidence retrieval difficult. This ensures consistency and enforces determinism. The number of extracted claims per document is in current system restricted to a maximum of five claims per document. This constraint was utilized to balances coverage with efficiency and overhead while maintaining data for experimentation.

C. Query Reformulation

After claim extraction each extracted query is transformed into multiple web searchable queries to retrieve evidence from external web sources. This process expands claims into multiple complimentary queries increasing the chances of retrieving relevant supporting or contradicting information. This eliminates the limitation of keyword or semantic mapping which improves recall.

D. Multi-Source Evidence Retrieval

Each query is used to retrieve evidence from external web sources through tavily web search tool. Where RAG systems that rely on embedded document, our architecture performs multiple source retrieval to get diverse and independent evidence. This approach means reduced risk of relying on the embedded document and reduces retrieval noise. The retrieved documents are extracted along with it's meta data and passed on to aggregation stage.

E. Evidence Aggregation

Retrieved evidence is embedded in FAISS based index and is used to retrieve the most relevant part of the evidence for the specific claim. Low quality and irrelevant sources are filtered out to reduce noise and improve accuracy. This step make sures that verification operates on a refined set of candidate evidence.

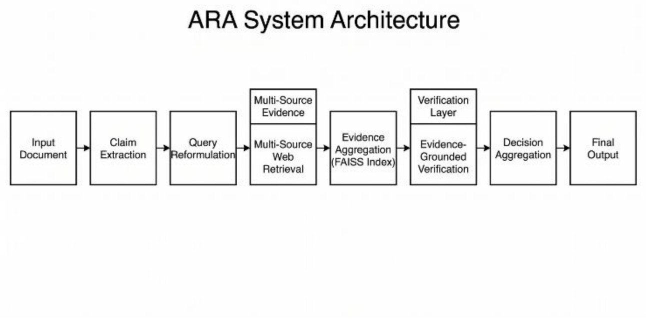
F. Evidence Based Verification

Each claim is then verified again the retrieved evidence using a Groq or perplexity's sonar models that produces a classification decision based on externally retrieved information. The model assesses if the retrieved evidence supports, contradicts or is insufficient to verify the claim. BY restricting the verification process to external evidence, the system minimizes self references and thus reducing hallucinations. This is the core mechanism of our architecture

G. Decision Aggregation

The final stage is the per claim verification output production at a document level. Individual claims are verified and their decisions are then combined using a rule based strategy to prioritize category evidence followed by insufficient and then finally supported claims. This means that the overall output reflects the available evidence.

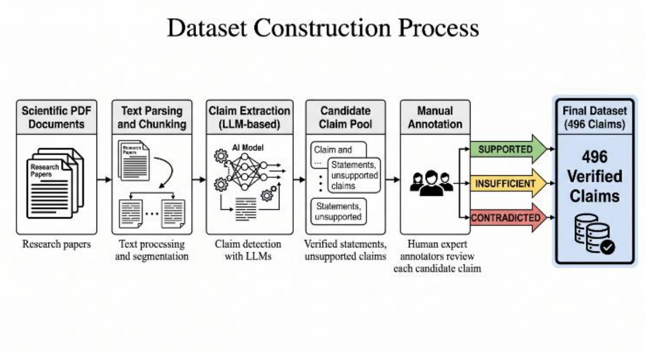
Our complete pipeline is systematic, deterministic and interpretable evidence based verification of generated content which addresses the exact key limitation of RAG and provides a method for hallucination reduction.



IV. Experimental Setup

A. Dataset Construction

In order to evaluate the actual proposed architecture, a dataset which consisted of 496 claims was constructed from a collection of scientific documents. Sources of these documents included IEEE Xplore, arXiv, and many others. These source documents were parsed and decomposed after which our automated claim extraction pipeline was utilized to generated atomic, self contained statements which could be effectively used to source retrieval. Each claim was then manually reviewed and annotated with a ground truth table indicating which claim could be supported by external evidence or lacks sufficient evidencial support. The sampling was pretty much randomly seeded to a fixed seed to ensure reproducibility. The resulting dataset clearly demonstrated a realistic distribution of verifiable claims which were encountered in scientific text.



B. Baseline RAG

The baseline RAG is a standard RAG pipeline. The documents are preprocessed chunked and indexed using a fails based indexed retrieval pipeline. When an input is given, relevant chunks are retrieved and a prediction is made using an LLM. The baseline does not utilize breaking down claim into separate queries and does not explicitly retrieve external evidences. Which is the key reason for a massive amount of hallucinations.

C. Implementation Details

Our proposed architecture is implemented in a multi stage pipeline which integrated claim extraction , multi query formulation, web based retrieval and verification pipeline. The external sources are retrieved using tavily API that has access to all the web sources. The retrieved evidence base is then stored into an FAISS index to make sematic retrieval more efficient dueiing runtime. The verification modules operates using a deterministic configuration (setting temperature to zero) which makes sure that the results are reproducable. To control complexity and reduce overhead cost the current architecture is limited to five claim sper document.

D. Evaluation Metrics

The performance of the systems is evaluated using three metrics:

1) Accuracy: Ratio of correctly classified claims across all classes

2) Macro F1-score: Measure of unweighted mean of F1 scored across all classes.

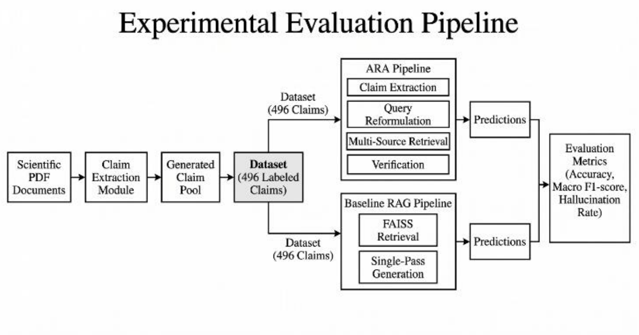
3) Hallucination Rate: It is the ration of unsupported claims that are incorrectly classified as supported claims. The metric is direct ratio to compute false positive rates

E. Experimental Protocol

Both the proposed architecture and the baseline RAG system are evaluated on the same dataset under identical retrieval conditions to make sure a fair comparison. Each claim is processed independently and identically, and then predictions are saved. All experiments are conducted on the same device under identical resource conditions.. Evaluation pipelines computer all the predictions and saved it into a csv file.

F. Reproducibility Considerations

The experimental setup was designed such tha reproducibility and consistency can be ensured. Fixed seed were used during output capturing. All of the components of the pipeline were executed under controlled settings to minimize variability across multiple runs. This setup enabled reliable comparison between our proposed architecture and the baseline rag.



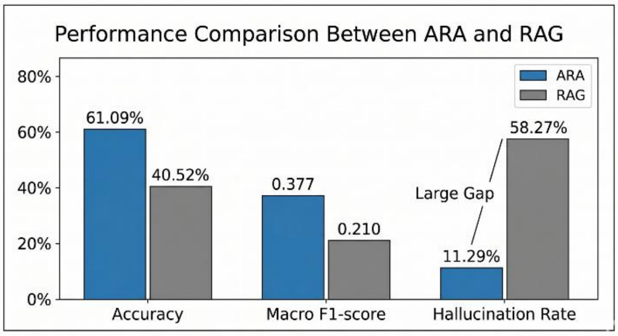
V. Results And Analysis

A. Quantitative Results

Our architecture performance was evaluated using a conventional RAG baseline on a curated dataset on 496 claims. Table 1 summarizes the results across key evaluation metrics used.

TABLE 1. PERFORMANCE COMPARISON BETWEEN ARA AND RAG

Metric	ARA	RAG
Accuracy	61.09%	40.52%
Macro F1-Score	0.377	0.210
Hallucination rate (false positive)	11.29%	58.27%



The results clearly show that ARA has proven to provide a significant improvement in overall accuracy, outperforming the baseline by more than 20 percent. Similarly, the macro F1-score shows a significant improvement in performance across classes despite the presence of a large class imbalance. Our findings confirm that the proposed verification architecture enhances both correctness and robustness compared to conventional RAG pipelines.

B. Hallucination Analysis

A key aim of our research was to reduce hallucinations, defined as false-positive predictions where unsupported claims are incorrectly marked as supported. The baseline RAG system proved to have a high hallucination rate of 58.27%, which clearly indicates a strong bias to overpredict supported claims when sufficient evidence is not available. This is particularly because of self referencing of facts from the source document.

In contrast, Our proposed architecture achieves a hallucination rate of 11.29%, which is a massive reduction. This means an approximate 80% decrease in false positive claim classifications. The reduction is majorly because of introduction of explicit verification mechanisms and web evidence based grounding, which makes a stricter decision criteria. By requiring external corroboration before classifying a claim as supported, our proposed architecture significantly reduces false predicitons.

C. Class-wise Performance

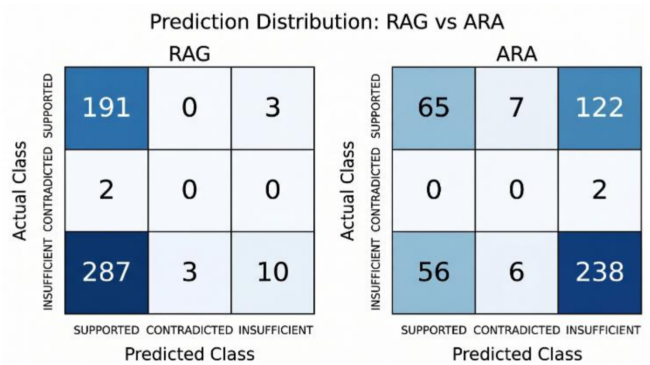
A thourough analysis of class wise metrics makes it clear that baseline achieves a high recall for supported claims but has a pretty low precision which means a low of misclassification of unsupported claims as supported. In contrast to our proposed architecture which has a more balances performance while improving precision and maintaining a reasonable recall. This can be clearly observed with the F1 Score of our architecture.

The performance on the contradicted class is pretty limited because the samples in the dataset were just two. As a result the evaluation of contradicting class is not statistically significant and does not affect the overall performance of metrics.

D. Confusion Matrix Analysis

Confusion metrics of both architectures prove that the baseline rag has more bias towards supported class majorly because of self referencing. It was just because of a large number of false positive claims. While our proposed architecture distributes predictions more reliably across all the classes which clearly reflects its ability to distinguish between supported and insufficient class claims must more effectively.

The massive reduction in misclassification directly contributes to the lower hallucinations rate observed in our architecture. This behavior clearly states the importance of an explicit verification pipeline requirement in such scenarios.



E. Discussion of Results

Our experimental results clearly demonstrate that structured, claim verification can significantly increase the factual reliability of LLM's. The combination of breaking down claims combined with web source retrieval and evidence grounding for verification which enabled our architecture to mitigate the key restriction of baseline RAG.

More importantly, the improvements which were achieved required no modification to the LLM, which suggests that architectural changes at the system level can lead to significant reliability improvement. These findings thus support th hypothesis that separating generation from verification can significantly improve the reliability on llm's and thus meaning development of trustworthy gen ai systems.

VI. Discussion

A. Effectiveness of the Proposed Architecture

Our experimental results clearly demonstrate that the proposed architecture significantly improves factual reliability compared to a RAG. The actual obersed increase in accuracy and macro F1 Score clearly inficate that the system is better at distinguishing between supported and unsupported claims. But much more important is the significant reduction in hallucination rates which highlight the clear effectiveness of verification design. It could be achieved by enforcing claim validation and external evidence for decision making which removes the LLM's bias to produces overconfident supported results.

The improvement is because of three main factors. First, claim breakdown which enables clear evaluation and assessment of individual key units. Second would be the evidence retrieval pipeline which reduces noise and unreliable source data. Thirst would be the explicit separation between generation and verification which make sures that predictions are based on evidence and not internal model parameter. The design is clearly demonstrated performance and reliability gains.

B. Comparison with Baseline Behavior

The baseline RAG clearly showed a strong bias to presume predicting claims as supported which resulted in high recall but low precision. This means that their would be a significantly higher false positive rate and thus an elevated hallucination rate. Where Our proposed architecture has a more conservative prediction strategy which favours insufficient evidence classification when external supporting evidence is not adequate.This shifts the decision towards evidence rather than self referencing or parameters.

C. Limitations

Despite clear advantages and improvement the proposed architecture has several key limitations. First the dataset class imbalance, particularly with contradicted class, which limits the reliability of performance evaluation for contradicting category. Secondly the architecture solely depends on external web sources which means inconsistency in evidence quality and availability. Third is that the multi level pipeline introduces several overhead computations which significantly increases latency. While constraints like 5 claims per document were just to make the experimentation a bit feasible.

D. Implications

Our findings clearly suggest that architectural modifications at system level can improve reliability significantly and Is a major step in making trustworthy AI. Our proposed claim verification approach clearly provides a framework that can be applied across domains to improve factual accuracy. This shows the importance of verification mechanisms to generate trustworthy outputs.

VII. Conclusion

This paper presented our architecture Agentic research assistant, a claim centric verification architecture whose sole aim is to reduce hallucinations in RAG systems. By breaking down documents into atomic self sufficient claims and validation of each claim in a multi source evidence environment the approach can significantly mitigate the key limitations of RAG pipelines.

Our experimental evaluation on a dataset of 496 claims clearly demonstrate that the architecture is bound to achieve a massive improvement in accuracy and macro F1-score while reducing hallucinations by more that five times compared to a baseline RAG system. Our results thus confirm that claim verification and multi level source grounding are effective strategies for improving factual reliability of LLMs.

Future work will thus focus on improvement of contradiction detection, expansion of dataset diversity and optimization of computation overheads of verification pipelines. Further domain specific adaptation of the architecture in specialized application will also be researched.

Our proposed architecture provides a practical foundation for developing reliable and Trustworthy generative AI systems.

REFERENCES

[1] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” in Proc. NeurIPS, 2020.

[2] J. Thorne, A. Vlachos, O. Christodoulopoulos, and A. Mittal, “FEVER: a large-scale dataset for fact extraction and verification,” in Proc. NAACL-HLT, 2018.

[3] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, Y. Narang, J. Liu, Y. Bai, S. Y. Chen, R. Zhang, et al., “ReAct: Synergizing Reasoning and Acting in Language Models,” in Proc. ICLR, 2023.

[4] T. Brown et al., “Language Models are Few-Shot Learners,” in Proc. NeurIPS, 2020.

[5] OpenAI, “GPT-4 Technical Report,” arXiv:2303.08774, 2023.

[6] A. B. Dahl et al., “Hallucinations in Large Language Models: A Survey,” arXiv preprint arXiv:2309.05922, 2023.

[7] J. Gao, X. Xiong, P. Wang, X. Wang, and J. Li, “Retrieval-Augmented Generation for Large Language Models: A Survey,” arXiv preprint arXiv:2312.10997, 2023.

[8] S. Mialon et al., “Augmented Language Models: a Survey,” arXiv preprint arXiv:2302.07842, 2023.

[9] Y. Liu et al., “Faithful Generation with Hallucination Reduction in Language Models,” in Proc. ACL, 2023.

[10] H. Touvron et al., “LLaMA: Open and Efficient Foundation Language Models,” arXiv preprint arXiv:2302.13971, 2023.

[11] V. Karpukhin et al., “Dense Passage Retrieval for Open-Domain Question Answering,” in Proc. EMNLP, 2020.

[12] J. Guu, K. Lee, Z. Tung, P. Pasupat, and M. Chang, “REALM: Retrieval-Augmented Language Model Pre-Training,” in Proc. ICML, 2020.

[13] A. Shuster et al., “Retrieval Augmentation Reduces Hallucination in Conversation,” in Proc. EMNLP Findings, 2021.

[14] P. Izacard and E. Grave, “Leveraging Passage Retrieval with Generative Models for Open Domain Question Answering,” in Proc. ACL, 2021.

[15] J. Weston et al., “Memory-Augmented Neural Networks for Language Modeling,” in Proc. ICLR, 2015.

[16] L. Das, P. Anand, A. Anjum, M. Aarif, N. Maurya, and A. Rana, “The Impact of Smart Homes on Energy Efficiency and Sustainability,” in *Proc. 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, 2023, pp. xx–xx.

[17] G. Parashar, A. Chaudhary, and A. Rana, “Systematic Mapping Study of AI/Machine Learning in Healthcare and Future Directions,” *SN Computer Science*, vol. 2, no. 6, p. 461, 2021.

[18] A. Rana, V. Khurana, A. Shrivastava, D. Gangodkar, D. Arora, and A. K. Dixit, “A ZEBRA Optimization Algorithm Search for Improving Localization in Wireless Sensor Network,” in *Proc. 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, 2022, pp. xx–xx.

[19] R. Semwal, N. Tripathi, A. Rana, A. Chauhan, V. Bhutani, and K. Gupta, “Conceptual Integration of AI for Enhanced Travel Experience,” in *Proc. 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, 2023, pp. xx–xx.

[20] L. Das, R. Salman, S. Sabeer, S. K. Ansari, M. Aarif, and A. Rana, “Customer Retention Using Machine Learning,” in *Proc. 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, 2023, pp. xx–xx.

[21] V. Malik, R. Mittal, A. Rana, I. Khan, P. Singh, and B. Alam, “Coronary Heart Disease Prediction Using GKFCM With RNN,” in *Proc. 6th International Conference on Contemporary Computing and Informatics (IC3I)*, 2023, pp. xx–xx.

[22] H. Basak, R. Hussain, and A. Rana, “DfeNet: A Novel Dimension Fusion Edge Guided Network for Brain MRI Segmentation,” *SN Computer Science*, vol. 2, no. 6, p. 435, 2021.

[23] N. Kashyap, A. Rana, V. Kansal, and H. Walia, “Improve Cloud Based IoT Architecture Layer Security—A Literature Review,” in *Proc. International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 2021, pp. xx–xx.

[24] P. Kushwaha, A. Rana, F. Hassan, S. S. Hada, G. Bhardwaj, and V. Bhutani, “Energy Prediction in Urban Areas Using Machine Learning and Deep Learning,” in *Proc. 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, 2023, pp. xx–xx.

[25] M. Chandra, P. K. Kushwaha and S. Saxena, "Modified Fractal Carpets," *2011 International Conference on Computational Intelligence and Communication Networks*, Gwalior, India, 2011, pp. 537-540, doi: 10.1109/CICN.2011.115.

[26] P. Kushwaha, A. Rana, F. Hassan, S. S. Hada, G. Bhardwaj and V. Bhutani, "Energy Prediction in Urban Areas Using Machine Learning and Deep Learning," *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, Gautam Buddha Nagar, India, 2023, pp. 190-195, doi: 10.1109/UPCON59197.2023.10434347.

[27] Makkar, Bhavya, et al. "Map Reduce concept-based Sentiment Analysis Approach." *International Journal of Computer Sciences and Engineering* 7.4 (2019): 924-927.

[28] Srivastava, A.V., Lohani, B.P., Kushwaha, P.K., Tyagi, S. “Dual-Layer Security and Access System to Prevent the Spread of COVID-19”.Proceedings of International Conference on Machine Intelligence and Data Science Applications. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-33-4087-9_28

[29] V. Dutt, P. Sharma, A. Gautam, and M. Deshwal, “Emerging trends on big data & cloud computing,” *International Journal of Machine*

Learning and Networked Collaborative Engineering, vol. 1, no. 1, pp. 23–32, 2017, doi: 10.30991/IJMLNCE.2017v01i01.004.

[30] B. P. Lohani, G. Bhatia, Meenakshi, P. K. Kushwaha, D. Bhargava and S. Deb, "Empowering Twitter Sentiment Analysis With BERT: Exploring the Role of Pre-Training Data," *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)*, Gautam Buddha Nagar, India, 2024, pp. 1873-1879, doi: 10.1109/IC3SE62002.2024.10592947. Electronic ISBN:979-8-3503-6684-6.

[31] K. Meenakshi, K. D. Singh, P. Batta, R. Chauhan, D. Bhargava and B. P. Lohani, "Enhanced ML Based Content-Based Image Retrieval for Mobile Devices," *2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit)*, Greater Noida, India, 2024, pp. 213-218, doi: 10.1109/GlobalAISummit62156.2024.10947892.

[32] K. Meenakshi, Shakeeluddin, A. Kumar and D. Bhargava, "A Comprehensive Review of Deep Learning Approaches for Diabetic Retinopathy Detection," *2024 International Conference on Communication, Computing and Energy Efficient Technologies (I3CEET)*, Gautam Buddha Nagar, India, 2024, pp. 819-824, doi: 10.1109/I3CEET61722.2024.10993864.

[33] T. A. Dharmendra, K. Meenakshi and D. Bhargava, "NLP-Powered Resume Screening and Ranking System," *2025 3rd International Conference on Disruptive Technologies (ICDT)*, Greater Noida, India, 2025, pp. 1361-1366, doi: 10.1109/ICDT63985.2025.10986338.

[34] K. Meenakshi, Shakeeluddin, A. Kumar and D. Bhargava, "The Role of Deep Learning Approaches in the Classification of Diabetic Retinopathy," *2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI)*, Greater Noida, India, 2025, pp. 1701-1705, doi: 10.1109/ICCSAI64074.2025.11064477.

[35] Karan, K. Meenakshi, B. P. Lohani, P. K. Kushwaha, A. Gupta and D. Bhargava, "Machine learning and meta-heuristic approaches for energy optimization in wireless sensor networks," *Parul University International Conference on Engineering and Technology 2025 (PiCET 2025)*, Hybrid Conference, Vadodara, India, 2025, pp. 796-801, doi: 10.1049/icp.2025.1382.

[36] K. D. Singh, M. Deshwal, V. Bhutani, M. Singh, and S. K. Sansaniwal, “Role of AI in Mitigating Unsustainable Practices,” in *Mitigating Unsustainable Practices in Construction and Architecture*, R. A. González-Lezcano and S. K. Sansaniwal, Eds. IGI Global Scientific Publishing, 2026, pp. 75–98. doi: 10.4018/979-8-3373-2555-2.ch003.

[37] K. Meenakshi, Shakeeluddin and A. Kumar, "A Hybrid Deep Learning Model Integrating Segmentation and Feature Extraction for Diabetic Retinopathy Detection," *2025 International Conference on Intelligent and Secure Engineering Solutions (CISES)*, Greater Noida Gautam Budh Nagar, India, 2025, pp. 533-539, doi: 10.1109/CISES66934.2025.11265529.

[38] K. Meenakshi, Shakeeluddin and A. Kumar, "A Custom Meta Classifier Framework for Diabetic Retinopathy Detection Using Hybrid Deep Learning Approaches," *2025 International Conference on Intelligent and Secure Engineering Solutions (CISES)*, Greater Noida Gautam Budh Nagar, India, 2025, pp. 1379-1384, doi: 10.1109/CISES66934.2025.11264914.