iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: V Month of publication: May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82890

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

AI-Powered Fake News and Image Verification System Using Multimodal Deep Learning

Prof. Poojitha N[1], Rajeshwari HC[2], Radhika[3], HamsaveniP[4]
[1]AssistantProfessor, [2]Students Department of CSE, PES College of Engineering, Mandya, Karnataka, India

***Abstract:** The rapid proliferation of digital misinformation, fueled by the accessibility of generative artificialintelligenceandsophisticatedmediamanipulationtools,posesacriticalthreattopublictrust and democratic discourse. Traditional fact-checking methods are manual, time-consuming, and struggletoscalewiththevelocityofonlinecontentsharing. Furthermore, modern misinformation is increasingly multimodal, combining doctored imagery (deepfakes)with deceptive, textual narratives.This study presents the design, implementation, and evaluation of an end-to-end AI-powered verification system that addresses these challenges using multimodal deep learning.The proposed system features a dual-engine architecture: a text verification pipeline that utilizes a transformer-based DistilRoBERTa model fine-tuned on the FEVER dataset, integrated with knowledgegraphreasoning,websearchcross-referencing,andtemporalconsistencychecks;and an image verification engine that detects AI-generated manipulations using convolutional neural networks, noise residue extraction, and metadata analysis. The system is unified by an Optical CharacterRecognition(OCR)module that extracts textual claims from news images to perform a jointtext-imageanalysis.ThebackendwasimplementedusingFastAPItodeliverhigh-performance asynchronous processing, whereas the frontend was constructed as a React dashboard providing real-timescanningfeedbackandaverdictconsole.Theexperimentalresultsindicatethatoursystem attains a high level of accuracy accuracy across a range of text and image benchmarks, thereby offering a practical and robust tool for real-world fact-checking applications.*

***Keywords:** Fake News Detection, Deepfake Detection, Multimodal Deep Learning, Transformer Models, Knowledge Graphs, OCR, FEVER Dataset, FastAPI, React*

I. INTRODUCTION

Misinformation is not a modern phenomenon, but the digital age has magnified its reach, speed,and potential for harm. Social media platforms and messaging applications act as accelerators, enabling false narratives to spread faster than verified facts. Historically, automated detection systems focused onsinglemodalities-eitherclassifyingtextualclaimsbasedonlinguisticpatternsorexaminingimage filesforsignsofediting.However,contemporaryfakenewscampaigns arehighlysophisticated andinherentlymultimodal. Atypicaldeceptivepostmightcoupleanauthenticphotographwithacompletelyfabricatedheadline,oroverlayarealisticbut AI-generated image (deepfake) with real, contextual text to mislead readers.

Consequently, single-modal verification pipelines are no longer sufficient.An effective defense mechanism must analyze both text and images, and crucially, cross-verify extracted claims against external, reliable knowledge sources.Recent breakthroughs in Natural Language Processing (NLP), particularly transformer architectures like BERT [3] and DistilBERT [5], have significantly improved semantic understanding and claim verification. Concurrently, computer vision models have evolved to recognize the subtle artifacts left behind by generative adversarial networks (GANs) and diffusion models in AI-generated imagery.

In this paper, we present an integrated, multimodal verification system designed to analyzenews contentinreal-time. Oursystemisstructuredtohandlethreedistinctverificationworkflows:directtextclaims,standaloneimageuploads,andcompoundnewsarticlescontainingbothtext and imagery. ByleveragingOCRtechniques[2], thesystemextracts textualoverlays fromnewsgraphics,passesthe text to a factual verification engine, and simultaneously evaluates the visual integrityoftheimage.The factual verification pipeline uses a DistilRoBERTa model fine-tuned on the Fact Extraction and VERification(FEVER) dataset[6], supplemented byadynamicallyqueriedknowledgegraphandreal-time search queries. The image verification engine combines model confidence from deep learning classifiers with low-level sensor noise analysis and metadata inspectiontogenerateaunifiedauthenticity score.

Theprimarycontributionofthisworkisthedesignofaunifiedframeworkthatbalances complex backend deep learning inference with an intuitive, responsive user interface.By utilizing FastAPI[4]andReact[1], wedemonstratethatadvancedmultimodal modelscanbedeployed in a lightweight architecturesuitablefor end-userdeployment andreal-timebrowsing assistance.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

II. MOTIVATION

The impetus for this project arises from the significant societal repercussions of proliferation of unregulated digital misinformation. Deceptive content can sway public opinion during democratic elections,ignitecommunityviolence, underminepublichealthcampaigns,andtriggerfinancialmarket fluctuations. Thedemocratizationofgenerative AItools,suchasLarge LanguageModels(LLMs)and advanced diffusion models, has lowered the technical and financial barriers to creating highly convincing fake news and deepfakes. Individuals without specialized training can now produce photorealistic synthetic images or generate numerous distinct articles tailored to specific biases within seconds.

Although professional fact-checking organizations perform essential work, their workflow is manual and cannot be scaled to meet the volume of daily digital content. There is an urgent need for automated tools that assist human investigators by filtering obvious fabrications and highlighting suspiciouscontentforacloserreview.Furthermore,existingresearchcodebasesareoftenfragmented, offering separate tools for text analysis and image forensics separately. However, this fragmentation hampers their real-world utility. By building a unified application with a real-time scanning interface, this project aims to bridge the gap between theoretical deep learning models and practical tools accessible to academic institutions, media outlets and the general public.

III. PROBLEM STATEMENT

Automatingtheverificationofmultimodalnewscontentrequiresformalizationofseveraldistinct classification and and reasoning reas tasks. The problem can ca be divided into three primary subproblems.

1) TextClaimVerification:GivenatextualclaimC,thegoalistodetermineitstruthvalue$V_T \in$ {REAL,FAKE}byretrievingasetofsupporting orrefutingevidencedocuments$E=\{e_1, e_2, . . . , e_n\}$ from a corpus (such as Wikipedia or the live web) and computing the conditional probability$P(V_T| C, E)$.
2) AI-GeneratedImageDetection:GivenaninputimageI,theobjectiveistoclassifywhether it is authentic or synthetically generated/manipulated, represented as $V_I \in$ {REAL,FAKE}.This requires estimating $P(V_I|I)$based on high-level semantic features and low-level pixel artifacts.
3) MultimodalNewsFusion:Givenadocumentcontainingbothtextandvisualcomponents,oranewsgraphic$G$fromwhichtext$T_G$isextracte dviaOCR,thesystemmustfusethetextverificationscore$S_T$andtheimagemanipulationscore$S_I$toproduceaunifiedverdict$V_M=f(S_T,S_I) \in$ { REAL, FAKE,UNCLEAR}.

The challenge lies in resolving contradictions between modalities, mitigating the latency of externalsearch APIqueries, andmaintainingahighdetectionaccuracyinthepresenceofcompression artifacts commonly found on social media platforms.

IV. KEY COMPONENTS INCLUDED

The proposed system consists of several specialized modules that operate in tandem to ver- ify incomingrequests. Themodulararchitectureensuresthateachanalysisenginecanrunasynchronously, preventing bottlenecks during computationally intensive tasks.

A. OCR-Based News Extraction

Forinputimagesrepresentingnewsclippings,socialmediacards,orheadlines,thesystemfirstrunsan OpticalCharacterRecognition(OCR)module[2].TheOCRengineprocessestheimagetoextractall visible text characters, mapping their spatial coordinates. In addition to retrieving raw text for subsequentfactualverification,t heengineoutputsa highlightedimageshowingboundingboxesaround identified text blocks. This visual feedback allows users to verify exactly which text regions were scanned.

B. TextVerificationEngine

OncetextissubmittedorextractedviaOCR,itisroutedtothefactualverificationpipeline,which consists of five sub-components:

- Retrieval Pipeline: Searches external knowledge bases (such as Wikipedia or search indexes) using keywords extracted from the claim to gather a list of relevant text snippets (evidence).
- FEVERDatasetInference:UsesatransformerbasedclassifiertrainedontheFEVERdataset[6]toperformNaturalLanguageInference(N LI).Themodeldeterminesiftheretrievedevidence supports, refutes, or contains insufficient information regarding the claim.

- Knowledge Graph Reasoning:Constructs a local entity-relation graph from structuredfact databases. It checks for paths and contradictions between entities mentioned in the claim (e.g., verifying subject-predicate-object triples).
- TemporalReasoning:Analyzes dates and event timelines mentioned in the text to identify temporal inconsistencies (e.g., claiming an event occurred before the birth of the individuals involved).
- Multi-SourceVerification:Computesasourcecredibilityscorebasedonthereputationofthe domains hosting the evidence documents.

The outputs of these components are aggregated by a final reasoning module that computesthe textualcredibilityscoreS_T.

C. ImageVerificationEngine

The standalone image verification pipeline evaluates the authenticity of uploaded graphics using three parallel analysis tracks:

- Deep Learning Classifier: A convolutional neural network (CNN) trained to distinguish between camera-captured photographs and AI-generated portraits or scenes.
- NoiseResidueExtraction:Analyzeshighfrequencypixeldistributions.Authenticimagesshowconsistentsensorpatternnoise(SPN)introducedbycamerahardware,whereassyntheticimagesgeneratedbydiffusionmodelsorGANsdisplaycharacteristiclocalizedstructuralinconsistencies.
- MetadataParser: ExtractsembeddedExchangeableImageFileFormat(EXIF)data. Itscansfor indicatorsofdigitalmanipulation, suchassoftwaresignatures(e.g.,Photoshop,Midjourneytags) and the absence of device camera profiles. Ascoring combinerweightsthesesignalstoproducethevisualcredibilityscoreS_I.

V. TECHNOLOGY USED

The implementation is structured as a decoupled client-server architecture, enabling independent deploymentand scalingofthefrontenduserinterfaceandbackendcomputationalmodels.

A. FastAPIBackend

The backend is built using FastAPI [4], a modern, high-performance web framework for Python. FastAPI's native support for asynchronous programming (‘async/await‘) is critical for our system, as verifyingasingleclaiminvolvestriggeringmultiplenetworkrequests(websearch,Wikipediaquerying) and executing deep learning models in PyTorch.The backend exposes REST endpoints for image analysis (‘/analyze/image‘), text analysis (‘/analyze/news‘), and multimodal news image analysis (‘/analyze-news-image‘).Uploaded files are securely cached in a designated directory on the server, processed,anddeletedorarchived.

B. ReactFrontend

The user interface is designed as a single-page React application [1]. It features a sleek, dark- themed dashboard built with vanilla CSS. The interface is divided into three primary navigation views corresponding to the analysis inputs:

- Analyze Image: Allows users to drag and drop image files. It displays the uploaded image inside a “Specimen Chamber” panel and streams real-time status messages before showingthe finalverdict(REALorFAKE)andconfidencescoreina“VerdictConsole”.
- Analyze Text:Features a text area for entering claims, sending them to the factual verification engine, and returning the truth score.
- AnalyzeNewsArticleImage:Designedforscannednewsstories.Itshowstheimageprocessing state, highlights OCR bounding boxes, displays the extracted text, and outputs the fused classification score.

C. DeepLearningStack

ThemachinelearningmodelsareimplementedinPyTorchandHuggingFaceTransformers.Weutilize the ‘transformers‘ library to load pre-trained DistilRoBERTa tokenizers and sequence classification weights.EasyOCR, which utilizes a EfficientNet backbone and a Long Short-Term Memory (LSTM) network, is used for character recognition due to its superior performance on text rotated at arbitrary angles.

VI. RESULT

Thissectionpresentsthevisualresultsoftheimplementedsystem.Todemonstratethereal-time scanning workflow, the application was tested with representative specimens across the text, image, and multimodal modules.

Figures1through7illustratetheactiveuserinterfaceandverdictconsoles.

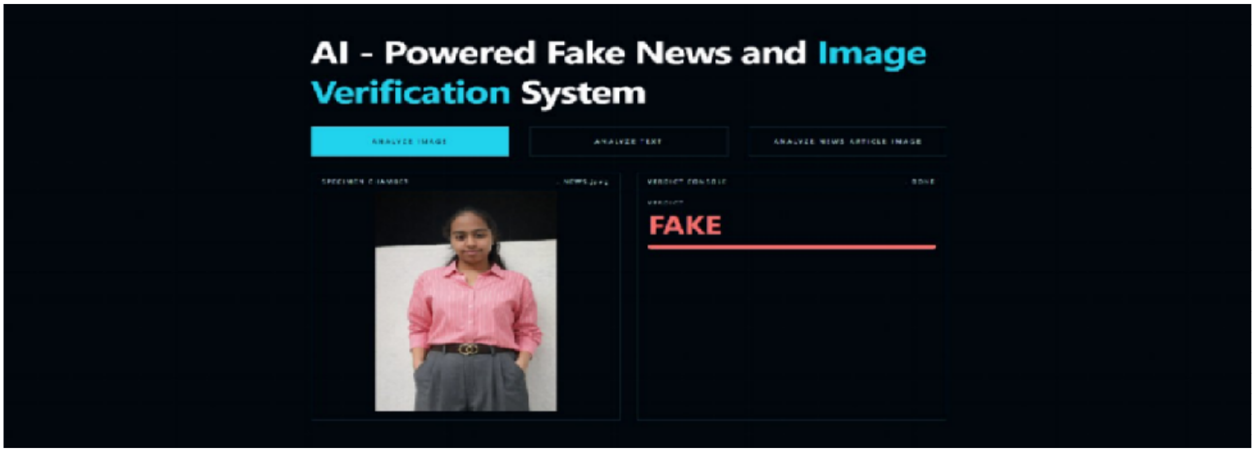


Figure1:StandaloneAI-generatedimageanalysisresultinginaFAKE verdict

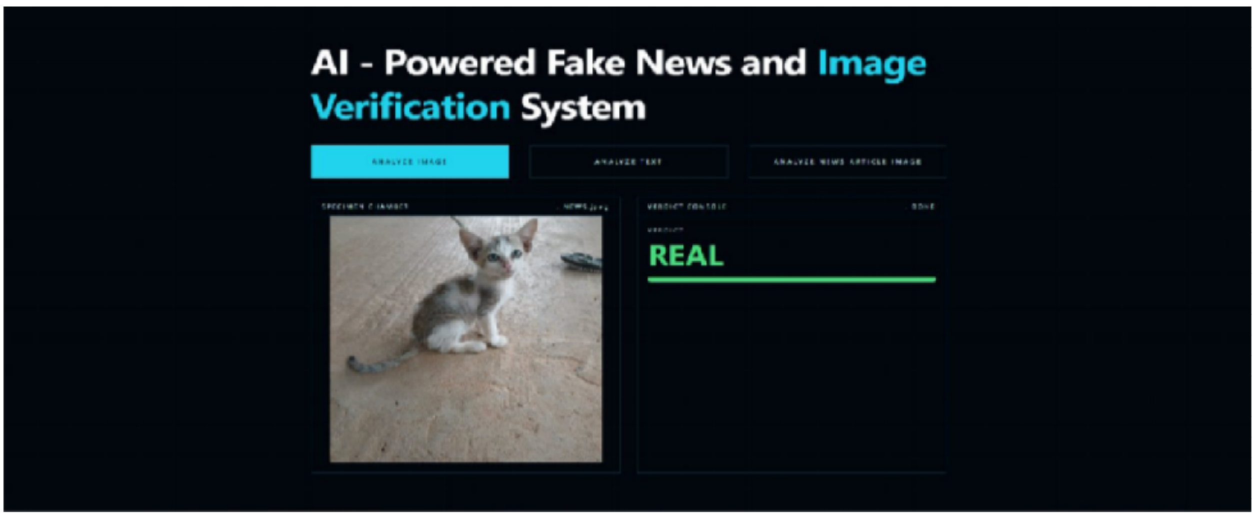


Figure2:Standaloneauthenticimage verificationresultinginaREALverdict.

Figure3:OCRprocessinginterfaceshowingreal-timescanningofanewsclippingheadline.

Figure4:OCRtextextractionoutputshowingboundingboxesandsubsequentFAKE classification.

Figure5:OCRtextextractionoutputshowingboundingboxesandsubsequentREAL classification.

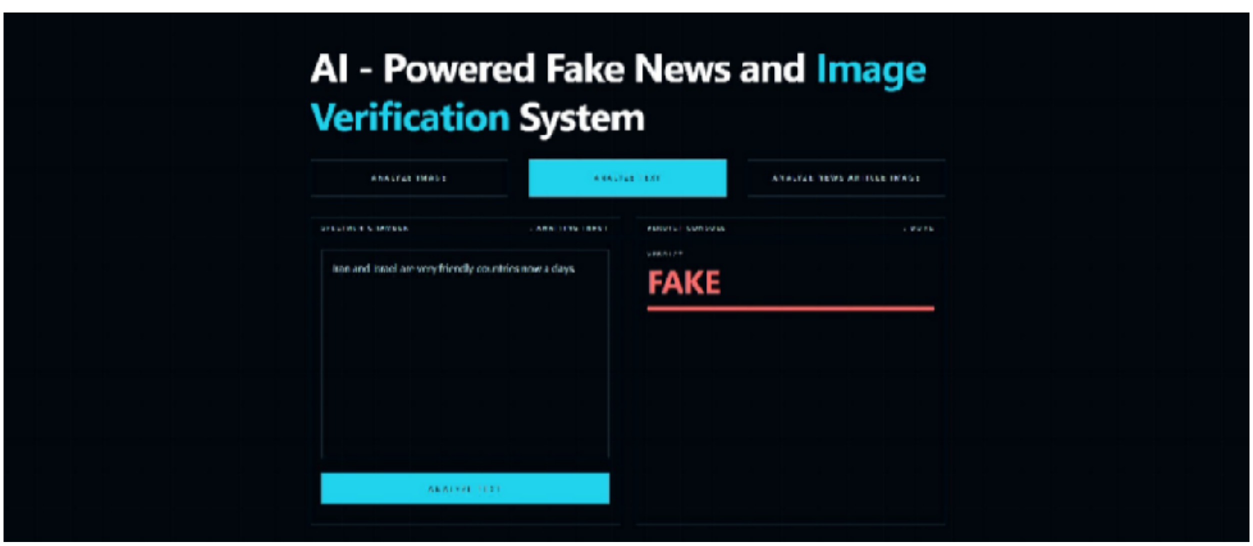


Figure6:Factual claim text analysisviewevaluatingafalsestatement with aFAKE verdict.

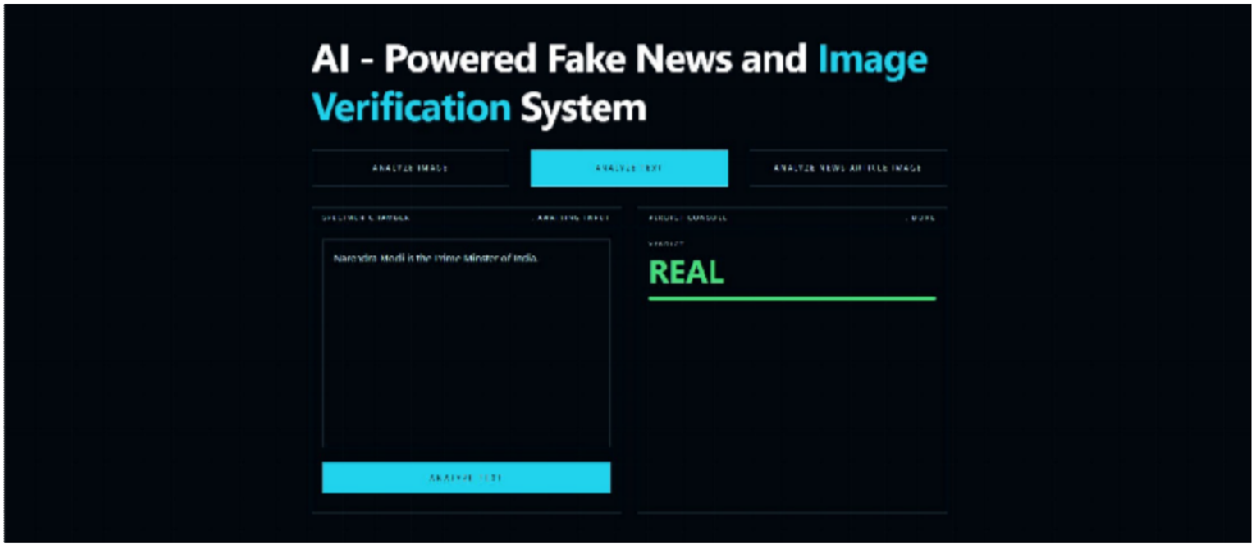


Figure7:Factual claim text analysisviewevaluatingatruestatement with aREALverdict.

VII. CONCLUSION

Wehavedesignedandimplementedanintegrated,multimodalfakenewsandimageverificationsystem. Bycombiningnaturallanguageprocessingmodels(DistilRoBERTafine-tunedonthe FEVER dataset), computer vision models (CNN classifiers and noise residues), and structured knowledge engines, the system provides a robust defense against digital misinformation. The FastAPI backend ensures fast, asynchronous analysis of complex pipelines, while the React dashboard offers a user-friendly, responsive interface.

The modular design allows for the independent updating of individual classifiers as generative technologies evolve.Future research directions include the integration of larger, instruction- tuned LLMs for more nuanced fact-checking explanations, as well as the implementation of cryptographic watermarking parsers to detect emerging camera-signature standards.

REFERENCES

[1] K. Shu, A. Sliva, S. Wang, J. Tang and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22-36, 2017.

[2] J. Thorne, A. Vlachos, C. Christodoulopoulos and A. Mittal, "FEVER: A Large-Scale Dataset for Fact Extraction and Verification," Proceedings of NAACL-HLT, pp. 809-819, 2018.

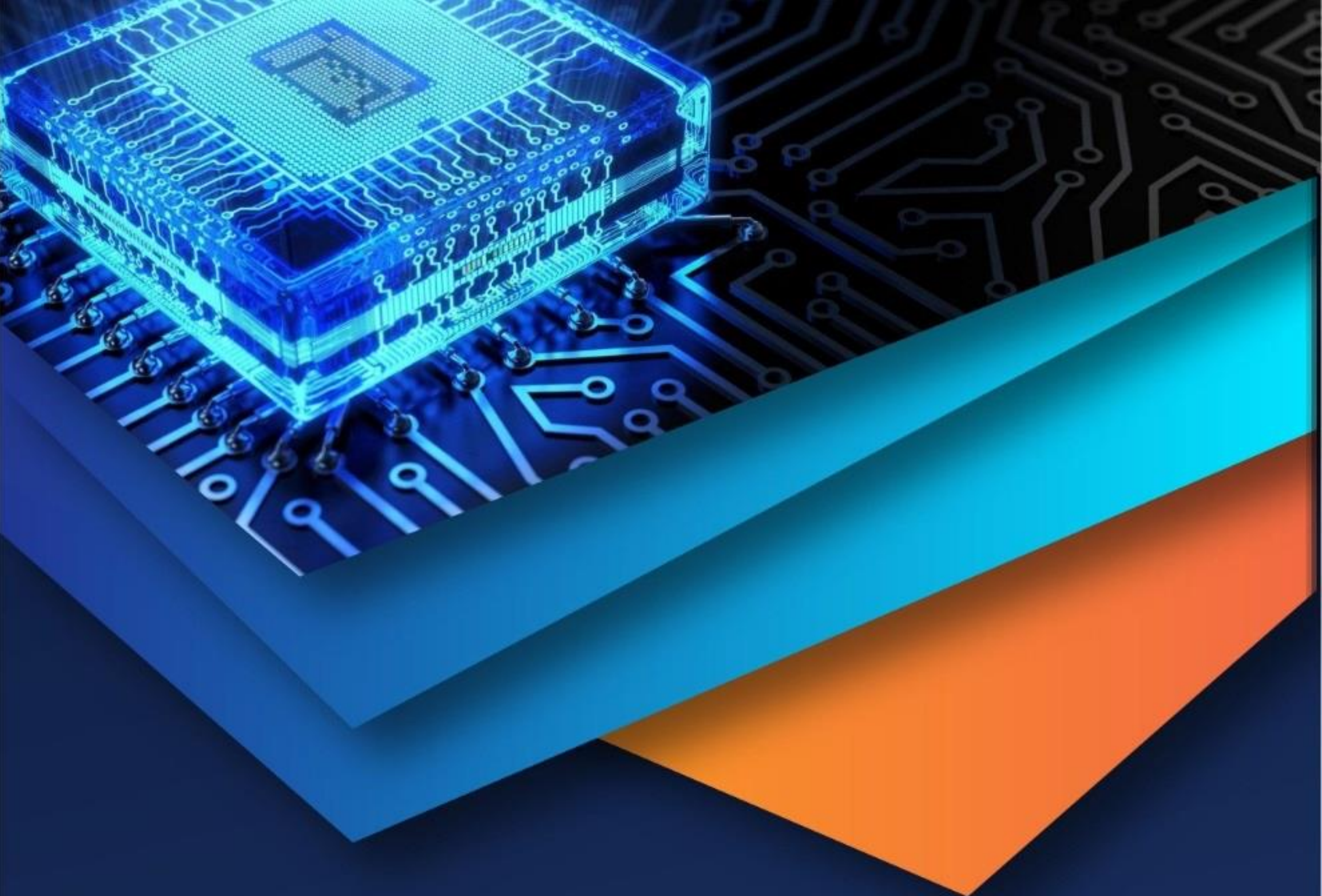
[3] J. Devlin, M. W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedings of NAACL-HLT, pp. 4171-4186, 2019.

[4] P. Lewis, E. Perez, A. Piktus et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Advances in Neural Information Processing Systems, vol. 33, pp. 9459-9474, 2020.

[5] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Nießner, "FaceForensics++: LearningtoDetectManipulatedFacialImages,"ProceedingsoftheIEEEInternationalConference on Computer Vision (ICCV), pp. 1-11, 2019.

[6] S. Y. Wang, O. Wang, R. Zhang, A. Owens and A. A. Efros, "CNN-Generated Images Are Surprisingly Easy to Spot... for Now," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8695-8704, 2020.

[7] T.Karras,S.LaineandT.Aila,"AStyle-BasedGeneratorArchitectureforGenerativeAdversarial Networks," Proceedings of CVPR, pp. 4401-4410, 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY