

NEURAL & AI

NLP Explainability-
Performance Meta-Analysis

Deep Learning vs Classical
Models for SDR HAR

BERT-Based 2024 UK
Election Sentiment Study

EEG Systems and
Consciousness in Afterlife
Studies

Discovering Thoughts, Inventing Future

VOLUME 26 / ISSUE 1 / VERSION 1.0

GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY

Section D - Neural & AI

Volume 26

2026

Issue D1

Global Journal of Computer Science and Technology

ISSN (Online): 0975-4172 • ISSN (Print): 0975-4350

Open Access Policy

All articles published in Global Journal of Computer Science and Technology are open access articles published under the Global Journals Open Access Publishing Agreement, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher

Global Journals Incorporated
301 Edgewater Place, Suite 100
Wakefield, MA 01880, United States

Copyright

© 2026 by Global Journals Incorporated (USA). All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without written permission.

Disclaimer

The authors, editors, and publisher will not accept any legal responsibility for any errors or omissions that may be made in this publication. The publisher makes no warranty, express or implied, regarding the material contained herein.

Editorial Board

Global Journal of Computer Science and Technology

Alessandra Lumini

University of Bologna Italy, Italy

Dr. Amogh Raghunath

Worcester Polytechnic Institute, United States

Dr. Abdurrahman Arslanyilmaz

Gazi University, Turkey

Dr. Ahmed Ebada

Mansoura University, Germany

Dr. Andaş Batur Aşolak

B.S.in Mechanical Engineering, M.Sc. in Mechanical Engineering from Mustafa Kemal University, MBA degree from Newport International University, Turkey

Dr. Anis Bey

Badji Mokhtar University, Algeria

Dr. Asim Yuksel

Istanbul University, Turkey

Dr. Aziz M. Barbar

American University of Science and Technology, Lebanon

Dr. Chandra Shikhi Kodete

Eastern Illinois University, United States

Asst. Prof. Dr. Chutisant Kerdvibulvech

Rangsit University, Thailand

Dr. Don S

Konkuk University, South Korea

Dr. Dongpo Xu

Northeast Normal University, China

EHIGIATOR IYOBOR EGHO-PROMISE

City of Oxford College and University Centre, United Kingdom

Dr. Federico Tramarin

National Research Council, Italy

Dr. Feng Jin

University of Arizona, United States

Dr. Hassan Harb

Lebanese University, Lebanon

Dr. Hazra Imran

Athabasca University, Canada

Dr. Jiaming Pei

Taizhou University, China

Dr. Jianyuan Min

Texas A&M University, United States

Dr. Jiayi Li

University of Virginia, United States

Dr. Jie YANG

Marquette University, United States

Dr. Juan Antonio Camara Serrano

University of California, San Francisco, United States

Dr. Judith A. Rowenfeld

Columbia University, United States

Dr. Juin-Ling Tseng

Minghsin University of Science and Technology, Taiwan

Dr. K. Sharmila

VELS University, India

Dr. Kassim Mwitondi

Sheffield Hallam University, United Kingdom

Dr. Khalid Nazim Abdul Sattar

Majmaah University, Saudi Arabia

Dr. Kurt Maly

New York University, United States

Dr. Kwan Min Lee

Nanyang Technological University, Singapore

Dr. Lamri Sayad

University of Béjaia, Algeria

Dr. Magdy Shayboub Ali

Suez Canal University, Egypt

Dr. Nurul Akmar Binti Emran

Technical University of Malaysia Malacca, Malaysia

Dr. Omar Ahmed Abed Alzubi

Al-Balqa Applied University, Jordan

Dr. Osman Balci, Balci

Virginia Tech, United States

Dr. Rajesh Kumar Rolan
Singhania University, India

Dr. Rajneesh Kumar Gujral
Maharishi Markandeshwar University, Mullana, India

Dr. Roheet Bhatnagar
Sikkim Manipal University, India

Dr. Sotiris Kotsiantis
University of Patras, Greece

Dr. Stefano Berretti
University of Florence, Italy, Italy

Dr. Tauqeer Ahmad Usmani
Kumaun University, India

Dr. Yuanyang Zhang
University of California, United States

Dr. Yufeng Wang
University of Chinese Academy of Sciences, United States

Dr. Zhengyu Yang
Northeastern University, United States

Dr. Jiayi Liu
University of Munich, United States

Jiyong Yu
University of Illinois, United States

Kumar Abhishek
University of Florida, United States

Marran Aldossari
Lawrence Technological University, Southfield, MI, USA, USA

Prof. marty Freeman
ASD Technology, India

Ming Li
University of Tulsa, United States

Palkesh
Rajiv Gandhi Technical University, INDIA

Pranit Gopaldas Shah
Parul University, India

Rahul Reddy
University of the Cumberland, United States

Ramadan Elaiees
University of Benghazi, Libya

Renuka Mohanraj
Maharishi International University, United States

Dr. Rufus Ayisi
Washington University, United States, United States

Satish Gajawada
Indian Institute of Technology Roorkee, India

PREFACE

The Global Journal of Computer Science and Technology (GJCST) is pleased to present this issue, bringing together a curated collection of high-quality research papers that explore various frontiers of computer science and technology. This issue features research spanning topics including artificial intelligence, machine learning, software engineering, network security, and interdisciplinary applications of computing. Each paper has undergone a rigorous peer-review process to ensure scientific rigor and originality. We would like to express our sincere gratitude to the authors for entrusting their research with us, to the reviewers for their thorough and constructive evaluations, and to our readers for their continued engagement with the scientific discourse. We hope that the research presented herein inspires further inquiry and contributes meaningfully to the advancement of knowledge in computer science and technology.

The Chief Editor
Global Journal of
Computer Science and Technology
Global Journals Organization

TABLE OF CONTENTS

Title	Author	Pages
Copyright		i
Editorial Board		ii
Preface		iv
Re-evaluating the Explainability-Performance Trade-Off Paradigm in Natural Language Processing Models: A Quantitative Meta-Analysis of Transformer Architectures (2019-2023)	Dr. Mezui & Dr. EYANGOLO	1-14
Comparative Evaluation of Deep Learning and Classical Models for Software-Defined Radio Based Human Activity Recognition	Aina	15-23
News-Based Political Sentiment Analysis for Election Outcome Verification: A BERT-Based Study of the 2024 UK General Election	Asanka et al.	24-32
Neural Networks and Rules-based Systems used to find Rational and Scientific Correlations between being Here and Now with Afterlife Conditions	Pia	33-36
Reti Neurali e Sistemi Basati su Regole Utilizzati per Individuare Correlazioni Razionali e Scientifiche tra l'Essere Qui e Ora e le Condizioni dell'Aldilà	Pia	37-40
Subject Index		40
Author Guidelines		41



Re-evaluating the Explainability-Performance Trade-Off Paradigm in Natural Language Processing Models: A Quantitative Meta-Analysis of Transformer Architectures (2019-2023)

Article Record

Dr. Kevin MEZUI^{§*}

*Corresponding Author



Dr. Vivien Armel Eyangolo[‡]



[§] Independent researcher, Paris, France

[‡] Denis SASSOU NGUESSO University Faculty of Applied Sciences Congo Brazzaville

RECEIVED

2026-01-13

ACCEPTED

2026-01-23

ONLINE PUBLISHED

2026-01-30

PUBLISHED

2026-02-17

PEER REVIEW

Double Blind

Abstract

A foundational principle in the field of artificial intelligence asserts that there is a trade-off between a model's explainability and its effectiveness. This trade-off significantly influences model selection for critical applications. This study presents a meta-analysis of 21 advanced NLP models from 2019 to 2023, encompassing encoder, encoder-decoder, and decoder architectures. A quantitative explainability framework was developed, grounded in architectural features, parameter efficiency, and the availability of interpretability tools. Our analysis revealed no significant correlation between explainability and performance across architectures, which contradicts common assumptions (Spearman's $\rho_{\text{explainability}} = 0.160$, $\text{p} = 0.489$). The degree to which a model can be explained is primarily predicted by the model's intricacy ($\rho_{\text{intricacy}} = -0.951$, $\text{p} < 0.001$), though the model's architectural family moderates this effect. Encoder-based models effectively circumvent the trade-off by achieving higher levels of explainability without compromising performance. These results demonstrate that architectural design, rather than mere performance optimization, significantly influences interpretability. We hereby propose a formal explainability evaluation method and provide evidence-based recommendations for selecting models for specific use cases. Our contribution to the expanding corpus of research on interpretable AI challenges the prevailing assumption that performance and explainability are inherently incompatible. Furthermore, it offers practical guidance for developing transparent, high-performing natural language processing (NLP) systems.

Explainable AI

model interpretability

performance trade-off

natural language processing

transformer architectures

meta-analysis

model Complexity

AI USE STATEMENT

No generative AI was used for analysis or results.

FUNDING

No external funding was declared for this work.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY

Not applicable for this article.

ETHICS

No ethics committee approval was required for this article type.

CONSENT

Not applicable for this article.

TRIAL REG.

Not applicable.

Crossref DOI: 10.34257/GJCSTD256108

How to Cite: MEZUI et al. (2026). Re-evaluating the Explainability-Performance Trade-Off Paradigm in Natural Language Processing Models: A Quantitative Meta-Analysis of Transformer Architectures (2019-2023). Global Journal of Computer Science and Technology, 26(1), 1-14. DOI: 10.34257/GJCSTD256108

LICENSE

© 2026 Global Journals. Open-access article under CC BY-NC-ND 4.0 International License.



Print ISSN 0975-4350



Online ISSN 0975-4172



Under the strict compliance and defined process of



METADATA CONTINUATION

AUTHOR CONTACT QR LEDGER

Dr. Kevin MEZUI§*	
Dr. Vivien Armel Eyangolo‡	

ARCHIVAL RECORD

Re-evaluating the Explainability-Performance Trade-Off Paradigm in Natural Language Processing Models: A Quantitative Meta-Analysis of Transformer Architectures (2019-2023)

Dr. Kevin MEZUI^{§*} and Dr. Vivien Armel Eyangolo[‡]

Affiliations

[§] Independent researcher, Paris, France

[‡] Denis SASSOU NGUESSO University Faculty of Applied Sciences Congo Brazzaville

Abstract

A foundational principle in the field of artificial intelligence asserts that there is a trade-off between a model's explainability and its effectiveness. This trade-off significantly influences model selection for critical applications. This study presents a meta-analysis of 21 advanced NLP models from 2019 to 2023, encompassing encoder, encoder-decoder, and decoder architectures. A quantitative explainability framework was developed, grounded in architectural features, parameter efficiency, and the availability of interpretability tools. Our analysis revealed no significant correlation between explainability and performance across architectures, which contradicts common assumptions (Spearman's $\rho_{\text{explainability, performance}} = 0.160$, $p = 0.489$). The degree to which a model can be explained is primarily predicted by the model's intricacy ($\rho_{\text{intricacy, explainability}} = -0.951$, $p < 0.001$), though the model's architectural family moderates this effect. Encoder-based models effectively circumvent the trade-off by achieving higher levels of explainability without compromising performance. These results demonstrate that architectural design, rather than mere performance optimization, significantly influences interpretability. We hereby propose a formal explainability evaluation method and provide evidence-based recommendations for selecting models for specific use cases. Our contribution to the expanding corpus of research on interpretable AI challenges the prevailing assumption that performance and explainability are inherently incompatible. Furthermore, it offers practical guidance for developing transparent, high-performing natural language processing (NLP) systems.

Keywords: *Explainable AI, model interpretability, performance trade-off, natural language processing, transformer architectures, meta-analysis, model Complexity, encoder-decoder models*

* Corresponding Author
Dr. Kevin MEZUI

DOI
10.34257/GJCSTD256108

1. Introduction

Although the literature often discusses a trade-off between explainability and performance, there has yet to be a unified quantitative meta-analysis of transformer models. This paper aims to bridge this gap by providing a systematic, reproducible comparison of Transformer architectures using harmonized explainability and performance metrics.

The issue of reconciling model capacity and explainability remains a divisive matter in AI research and application. The prevailing theory postulates an inherent conflict: as accuracy increases, models become more complex and less transparent [1, 2]. This seemingly inevitable trade-off largely determines model choices in areas such as medical diagnostics [3], credit scoring, and legal adjudication, where high accuracy coupled with interpretability is essential for reliable operation, compliance with legal regulations, and user trust.

Although these types of models have permeated the public domain, scientific support for this trade-off is surprisingly minimal, most of it anecdotal. Studies so far have largely limited themselves to making side-by-side comparisons of individual model pairs,

rather than fully investigating different types of model architectures simultaneously. Therefore, the present article's outcomes contribute to a deeper academic-industrial understanding of how model interpretability is affected by various factors. Also, there is ongoing disagreement in the community about the definition of explainability metrics, which alters views of expressibility across studies based on different operational definitions of interpretability. Such variation in methodology makes it difficult to draw broadly applicable conclusions about the core relationship between transparency and machine performance, and complicates direct comparisons. The rapid proliferation of transformer-based architectures in NLP over the last five years facilitated a detailed study of this relationship. Rigorous architectural paradigms emerged, evolved, and diversified in great numbers in the years 2019-2023: encoder-only models (e.g., BERT [4], RoBERTa [5]), encoder-decoder models (e.g., T5 [6], BART [7]), and decoder-only models (e.g., GPT series [8]). This diverse mix of architecture, along with pronounced variation in model scale, training methods, and application domains, offers a perfect opportunity to study how different design choices affect the performance-explainability trade-off in modern AI systems.

The research conducts an extensive meta-analysis exploring three main questions: (1) To what extent are performance and explainability quantitatively linked in NLP architectures? (2) What are the best predictors of explainability among architectural features? (3) How can practitioners be invited to model selections done wisely in terms of balancing performance and explainability? The analysis was performed on 21 NLP models spanning 2019 to 2023, and a quantitative explainability metric was formulated across three aspects: structural transparency, parameter efficiency, and the exploitation of interpretability tools.

Contrary to the common belief that there is a trade-off between performance and explainability, the data reveal that neither factor shows a statistically significant relationship within the architectural groups. Being complex is the main factor in predicting explainability; however, architectural design choices significantly influence this factor. Because encoder-based networks often achieve better explainability while preserving performance, the perception of a trade-off may be more a matter of particular architectural elements than a fundamental limitation of high-performing systems. Presented in a very thorough manner, our investigation provides the academic and industrial realms with a richer, yet more subtle and detailed, explanation of model interpretability.

We will test the hypothesis that there is no statistically significant correlation between a model's raw performance and its explainability score, as measured by our composite index. Additionally, we hypothesize that greater architectural complexity (e.g., depth, number of parameters, internal mechanisms) negatively predicts the observed level of explainability. We also assume that the architecture type (encoder, decoder, or encoder-decoder) moderates the relationship between performance and explainability; certain architectural families may offer a better balance than others.

This paper is structured as follows: Section 2 presents a review of key studies on explainable AI and the use of performance evaluation metrics in natural language processing (NLP) frameworks. Section 3 outlines the main steps of our approach: model selection, explainability scoring system, and use of statistical methods. In Section 4, we report results including descriptive statistics, correlation analyses, and comparisons across different architectures. Section 5 reflects on the theoretical and practical implications of our work and explores its limitations and future possibilities. Lastly, Section 6 summarizes our main findings and their importance in the context of explainable AI.

2. Related Work

2.1. Conceptual Foundations of Explainable AI

The field of explainable artificial intelligence (XAI) has evolved from early research on expert systems and rule-based inference mechanisms to contemporary methodologies that address the explainability of high-dimensional, nonlinear neural architectures. The conceptual foundations of XAI encompass several complementary perspectives: epistemological systems that investigate the nature and criteria of explanation in artificial systems [9], cognitive and psychological models that analyze how humans comprehend and evaluate machine-generated explanations, and computational approaches that provide operational techniques and methods for model interpretation [10].

A fundamental distinction is commonly drawn among intrinsic interpretability—where model structures are designed to be directly comprehensible to humans (e.g., decision trees and linear models)—and post-hoc explainability, which refers to the application of auxiliary techniques to explain the behavior of otherwise

opaque models after training [1]. Early work in XAI predominantly emphasized intrinsically interpretable models. In contrast, lately the focus has been primarily on post-hoc explanatory methods, particularly in deep learning for computer vision and natural language processing tasks.

2.2. Explainability in Natural Language Processing

The emergence of transformer architectures has fundamentally reshaped both the capabilities of natural language processing (NLP) systems and the associated challenges of model explainability. Early research on interpreting neural language models predominantly examined recurrent neural network architectures, using methods such as attention-weight visualization [11] and gradient-based attribution techniques. The subsequent introduction of self-attention modules into transformer models created new opportunities for interpretability, as the resulting attention weight distributions offer an inherently structured means of visualizing and analyzing the model's internal allocation of focus [12].

Contemporary approaches to explainability in natural language processing (NLP) can be grouped into several methodological families. These include: feature attribution methods, which identify salient input tokens or spans that contribute most strongly to a model's prediction; example-based methods, which retrieve similar training instances to justify or contextualize model outputs; rationalization methods, which generate natural language justifications in conjunction with model results; and probe-based methods, which analyze internal model representations using controlled interventions and auxiliary prediction tasks.

Despite this methodological heterogeneity, several studies have attempted to compare explainability across different architectural families or to quantify its relationship with overall model effectiveness. Nevertheless, a comprehensive and methodical characterization of these relationships remains an open research problem. Moreover, most existing evaluations are restricted to specific models or narrowly defined task domains, thereby constraining the extent to which their findings can be generalized across the wider NLP landscape.

2.3. Performance Evaluation in NLP

Standardized benchmarks have been pivotal in advancing research in natural language processing (NLP) and facilitating systematic, quantitative model comparisons. The earliest benchmark suites focused on very limited tasks such as part-of-speech tagging or named-entity recognition, whereas modern evaluation frameworks consider a much broader spectrum of language understanding and generation skills.

Some leading benchmark series are GLUE and its follow-up, SuperGLUE, used for general-purpose language understanding; MMLU for large-scale multitasking language understanding; and HELM, which provides a holistic evaluation of language models. Most of these benchmarks emphasize predictive correctness, robustness, and generalization, but give interpretability and explainability only a very minor role.

The scaling hypothesis, that a model's increased size and capacity can yield continuous performance improvements across various tasks, has served as a guiding principle for organizing recent NLP research. This focus on size, however, has often overshadowed interpretability, and relatively few works have examined how explainability-related measures change with factors such as the number of parameters or the model's structural complexity. As a result, the interplay among model scale, task

efficacy, and explainability remains an obscure and uncharted topic in today's NLP discussion.

2.4. The Performance-Explainability Trade-off Debate

The idea that there is a trade-off between performance and explainability has long been a topic of discussion in machine learning, particularly in the context of the bias-variance trade-off and the complexity-accuracy relationship. For a long time, simple models like linear regression and decision trees were considered easier to interpret but less capable of capturing complexity than more sophisticated models such as neural networks [13].

However, some scholars have started to challenge the notion that a tradeoff is inevitable for present-day deep learning models. On the one hand, there is some indication that newly engineered architectures can perform very well while remaining interpretable. On the other hand, some research has argued that whether a trade-off exists depends on the task and setting. Overall, the situation remains very unclear, as different researchers have used different evaluation methods and models, yielding contradictory results.

The present meta-analysis offers a fresh perspective on the discussion by systematically presenting numerical evidence from existing NLP models and leveraging a uniform testing framework to enable equitable comparisons across different architectures and performance levels.

3. Methodology

3.1. Study Design and Scope

A systematic meta-analysis was conducted to examine the trade-offs between performance and explainability of modern NLP models. The central theme is transformer-based models from January 2019 to December 2023. During this period, the transformer design was evolving very fast. The meta-analysis comprises 21 models selected according to stringent inclusion criteria to ensure coverage of the main architectural families and across different performance levels.

3.2. Model Selection Criteria and Process

We applied four inclusion criteria for model selection:

1. **Publication timeframe:** Models that came out from January 2019 to December 2023 and that include archival conference or journal papers, or formal technical reports.
2. **Performance evaluation:** Models should have shared performance results on at least two recognized benchmarks from these groups: GLUE, SuperGLUE, or MMLU. This allows for a comparable evaluation of the models' effectiveness.
3. **Architectural documentation:** Sufficient technical information must be provided to identify features relevant to explainability, such as types of attention mechanisms, parameter layouts, and components.
4. **Implementation availability:** Models should be made available through the major open-source platforms (e.g., Hugging Face, TensorFlow Hub, PyTorch Hub) to permit verification and reproduction.

The final sample includes 21 models across three architectural families: 14 encoder-only models, 5 encoder-decoder models, and 2 decoder-only models. This distribution reflects the research

landscape throughout the study period, with encoder architectures dominating academic publications despite a growing industrial focus on decoder models for generative applications.

3.3. Performance Statistics Collection and Normalization

Performance scores were gathered from various sources such as published papers, model cards, benchmark leaderboards, and official documentation. To ensure scores are comparable across benchmarks and protocols, a multi-step normalization procedure was applied.

1. **Benchmark-normalization:** At first, for each benchmark, we used min-max scaling by referring to the performance range of the entire set of models in our study. Thus, the normalization becomes specific to each benchmark.

$$P_{\text{bench}} = \frac{P_{\text{raw}} - P_{\text{min}}}{P_{\text{max}} - P_{\text{min}}} \times 100$$

Here, P_{raw} denotes the initial benchmark score, while P_{min} and P_{max} indicate the minimum and maximum scores observed for a particular benchmark among all the models.

2. **Benchmark weighting:** We decided to weigh benchmarks by considering their comprehensiveness and the variety of tasks. GLUE and SuperGLUE, which are multitask benchmarks, were assigned weights of 1.0, whereas single-task benchmarks were assigned weights of 0.5.
3. **Composite performance score:** For models that have been evaluated on several benchmarks, we calculated a weighted composite score as follows:

$$P_{\text{composite}} = \frac{\sum_{i=1}^n w_i \times P_{\text{bench},i}}{\sum_{i=1}^n w_i}$$

where w_i indicates the weight for benchmark i , and $P_{\text{bench},i}$ denotes the normalized score for this benchmark.

This approach guarantees that performance evaluations not only consider the diversity of evaluation contexts but also the relevance of different benchmarks in the NLP research sphere.

3.4. Explainability Scoring Framework Development

We developed a new quantitative system for measuring explainability, emphasizing the architecture's features rather than the performance of explanations for specific tasks. This approach provides a basis for fair comparisons between models with different capabilities and training objectives. Our framework evaluates three key aspects:

3.4.1. Parameter Performance Dimension

Efficiency in model parameters measures a model's ability to strike a good balance between power and explainability. It is based on the premise that of two models with similar performance, the one with fewer parameters would be easier to interpret. We get this score by applying a logarithmic transformation to the parameter count to account for exponential growth.

$$E_{\text{params}} = 1 - \frac{\log(P_i + 1)}{\max(\log(P + 1))}$$

where P_i stands for the number of parameters of the model i , and $\max(\log(P + 1))$ represents the highest log-transformed number

of parameters among all models (the increment by 1 is to prevent issues with zero parameters).

This expression yields scores in the range 0-1, with higher values indicating a more effective use of parameters.

3.4.2. Attention Mechanism Interpretability Dimension

Attention models differ in their inherent interpretability, depending on their computational characteristics and visualization potential. We created a scoring rubric mirroring these characteristics:

- **Bidirectional attention (encoder models):** Score = 1.0. Full-sequence processing enables comprehensive attention visualization and analysis of token relationships.
- **Encoder-decoder attention:** Score = 0.7. Cross-attention between encoder and decoder states provides interpretable alignment information, though with more complicated dynamics.
- **Autoregressive attention (decoder models):** Score = 0.4. Causal masking restricts attention to previous positions, reducing the comprehensiveness of attention sequences.
- **Sparse and factorized attention variants:** Score = 0.8 (when applicable). These mechanisms often achieve computational efficiency while preserving or improving interpretability through structured attention mechanisms.

3.4.3. Tooling Support Dimension

The quality and availability of interpretation tools are major factors that determine real-world explainability. We evaluate tooling support by taking the following things into account:

- **Visualization libraries:** Availability of specialized software for visualizing attention matrices, gradients, and other model internals.
- **Probing frameworks:** Facilities for functional explainability techniques and intervention studies directly.
- **Community adoption:** Proof of widespread use in academic papers and production environments.
- **Documentation standard:** Extent of interpretability documentation and tutorials.

The score is 0 for extremely limited tooling support and 1 for a fully featured, well-documented tooling ecosystem, while scores in between indicate partially or still-evolving tooling support.

3.4.4. Composite Explainability Score

The composite explainability score integrates these factors with weights determined by experts, reflecting their relative importance for interpretability in practice.

$$E_{\text{total}} = 0.4 \times E_{\text{params}} + 0.4 \times E_{\text{attention}} + 0.2 \times E_{\text{tooling}}$$

These weights resulted from discussions with three NLP interpretability experts, who agreed that the parameter effectiveness and the characteristics of the attention mechanism have a greater impact on intrinsic interpretability than the mere availability of tooling.

3.5. Complexity Assessment Methodology

Parameter counts provide only a rough proxy for complexity; we developed a 5-point ordinal complexity scale based primarily on architectural features. Our complexity ratings are calculated considering multiple factors:

1. **Level 1:** Simplest encoder-only architectures using standard multi-head attention and minimal architectural changes (e.g., BERT-base).
2. **Level 2:** Improved encoder-only structures with moderate changes, e.g., better training objectives, parameter sharing, and/or efficiency optimizations (e.g., ALBERT, DistilBERT).
3. **Level 3:** Typical encoder-decoder models with standard attention blocks and roughly equal parts of encoder and decoder (e.g., T5-base, BART-base).
4. **Level 4:** Very advanced encoder architectures with highly sophisticated attention mechanisms or structural innovations, or significantly improved encoder-decoder models with additional components (e.g., DeBERTa, PEGASUS).
5. **Level 5:** Huge-scale decoder-only models with very advanced attention variants, extremely high parameter numbers, and complex training setups (e.g., GPT-3, GPT-3.5).

The complexity of each model was independently measured by two NLP architecture experts, who showed very high agreement (Cohen's $\kappa = 0.92$). The opinion of a third expert settled any differences.

3.6. Statistical Analysis Framework

We used a mixture of descriptive and inferential statistics for a full-fledged analysis. Here are the main statistical operations we used:

- **Descriptive statistics:** Include calculating a mean, a standard deviation, a minimum, a maximum, and the shapes of the distributions for all variables examined.
- **Correlation analysis:** Using Spearman's rank correlation was a suitable choice for mixed continuous and ordinal data, as it estimates the monotonic relationships.
- **Group comparisons:** To determine the differences among the architectural families, a one-way ANOVA was done, which was followed by Tukey's HSD post-hoc tests.
- **Regression modeling:** A multiple linear regression was carried out to find out how the various predictors together determine the explainability scores. The regression analysis was supplemented with diagnostic tests to check the validity of assumptions.
- **Calculations of effect size:** Our practical significance measures went beyond the statistical results by estimating Cohen's d , which reflects the size of the mean differences, and R^2 , which indicates the amount of variance accounted for.

We did all the work with Python 3.9, using SciPy (v1.9.0) for statistical testing, statsmodels (v0.13.0) for regression analysis, and pandas (v1.4.0) for managing data. We set the significance threshold for all tests at $\alpha = 0.05$ and, where necessary, accounted for multiple comparisons.

3.7. Repeatability and Openness Measures

We took several steps to ensure that our study is transparent and can be easily duplicated by others.

- **Open Data:** We made all data related to our study activity, including performance measures, explainability scores, and complexity ratings, available in a well-structured manner.
- **Code Availability:** You can find the complete analysis code with comprehensive documentation on a public repository.
- **Methodological Transparency:** Every scoring rubric, weighting techniques, and analysis decisions have been recorded in a very transparent way.
- **Limitation Acknowledgment:** Potential biases and drawbacks have been pointed out to offer a better insight into the results.

You can find both the entire dataset and the analysis code at <https://github.com/kjmezui/xai-meta-analysis>.

4. Results

Key estimates are accompanied by 95% confidence intervals to account for statistical uncertainty surrounding effect sizes. We illustrate the distribution of scores and observed relationships using quantitative visualizations: boxplots to compare distributions across architecture families and regression plots to depict continuous relationships among performance, complexity, and explainability. Robustness tests, including bootstrap and sensitivity analyses of explainability score weights, were conducted to verify the stability of the results amid methodological variations.

4.1. Descriptive Characteristics of Analyzed Models

Table 1 displays the summary statistics of the 21 NLP models we included in our meta-analysis. Performance scores ranged from 49.9 to 90.8 on a normalized 0-100 scale (mean = 80.77, SD = 9.06).

The variation in explainability scores, measured on a 1.0 to 4.0 scale, was quite high (mean = 2.57, SD = 1.03), indicating considerable variation in model interpretability.

Model size also differed drastically, with the number of parameters ranging from 12 million in ALBERT-base to 175 billion in GPT-3.5, reflecting the different scaling strategies employed. Complexity scores ranged from 1.5 to 5.0 (mean 3.26, SD 1.12), with higher overall Complexity levels.

The years of publication were from 2019 to 2023, with an average of 2020.5, representing the period when transformer models were rapidly evolving.

Table 1. Descriptive statistics of analyzed NLP models ($n = 21$)

Variable	Mean	SD	Min	Max
Performance (0-100)	80.77	9.06	49.9	90.8
Explainability (1-5)	2.57	1.03	1.0	4.0
Parameters (Millions)	8,863.8	38,095.7	12	175,000
Complexity Score	3.26	1.12	1.5	5.0
Year	2020.5	1.2	2019	2023

4.2. Performance-Explainability Relationship Analysis

The analysis did not reveal a statistically significant correlation between model performance and explainability (Spearman's $\rho = -0.160$, $p = 0.489$), contrary to the expected trade-off assumption.

This implies that one can deliver high performance while keeping explainability at a desirable level; on the other hand, one can enhance explainability without performance suffering one bit.

Panel A in Figure 1 shows that there are quite a few models spread all along the gradation between performance and explainability. The right-top corner, where both performance and explainability are high, is not empty (for example, RoBERTa-large: performance = 86.4, explainability = 3.0). Interestingly, it also has spots that correspond to combinations falling toward the extremes of one dimension and the midpoint of the other one (e.g., high performance-low explainability: GPT-3.5: performance = 85.2, explainability = 1.0; or low performance-moderate explainability: FLAN-T5-base: performance = 49.9, explainability = 2.0).

Most likely, the lack of significant correlation will also be observed in model groups that are sorted by publication year or number of parameters, given the demonstrated robustness of the results across various types and scales of models.

4.3. Architectural Family Differences

Architectural families were found to differ substantially in performance ($F(2, 18) = 4.06$, $p = 0.035$, $\eta^2 = 0.31$) and explainability ($F(2, 18) = 11.72$, $p < 0.001$, $\eta^2 = 0.57$). Encoder models had the best average performance (84.03, SD = 5.12) and explainability (3.0, SD = 0.8), whereas decoder models scored the lowest on both metrics (69.45 performance, SD = 9.45; 1.0 explainability, SD = 0.0). The encoder-decoder models gave results in between (73.32 performance, SD = 11.85; 2.57 explainability, SD = 0.3).

Tukey's HSD post hoc test comparisons showed that encoder models significantly outperformed decoder models in both performance ($p = 0.028$, Cohen's $d = 1.87$) and explainability ($p < 0.001$, Cohen's $d = 3.25$). The comparisons between encoder and encoder-decoder models for performance ($p = 0.062$, Cohen's $d = 1.12$) and explainability ($p = 0.071$, Cohen's $d = 0.68$) were near-significant but insignificant.

Figure 2 visually summarizes the architectural differences. The charts show distinct patterns in the model groups' performance-explainability relationships, the distribution of the performance data, and the correlation structures.

4.4. Complexity as a Predictor of Explainability

The overall sample showed that model Complexity was the strongest negative factor affecting explainability (Spearman's $\rho = -0.951$, $p < 0.001$), indicating that complex models are usually less understandable. This negative relationship kept on changing across the different architectural groups in terms of strength: $\rho = -0.85$ for the encoder-only models, $\rho = -0.90$ for the encoder-decoder, and $\rho = -1.00$ for the decoder models.

On another note, the relationship between Complexity and performance was not only weak but also positive ($\rho = 0.270$, $p = 0.236$); this means that an increase in Complexity does not always translate into better results. Many cases deviate significantly from the line of best fit, like the simple ALBERT-xxlarge (Complexity = 2.5, performance = 84.3) that works well, while the much more complex GPT-3 (Complexity = 5.0, performance = 80.5) only gets moderate results (see Figure 1, Panel B).

The inverse relationship between Complexity and performance, and the strong association between Complexity and explainability,

Meta-analysis: Performance-Explainability Trade-off in NLP Models

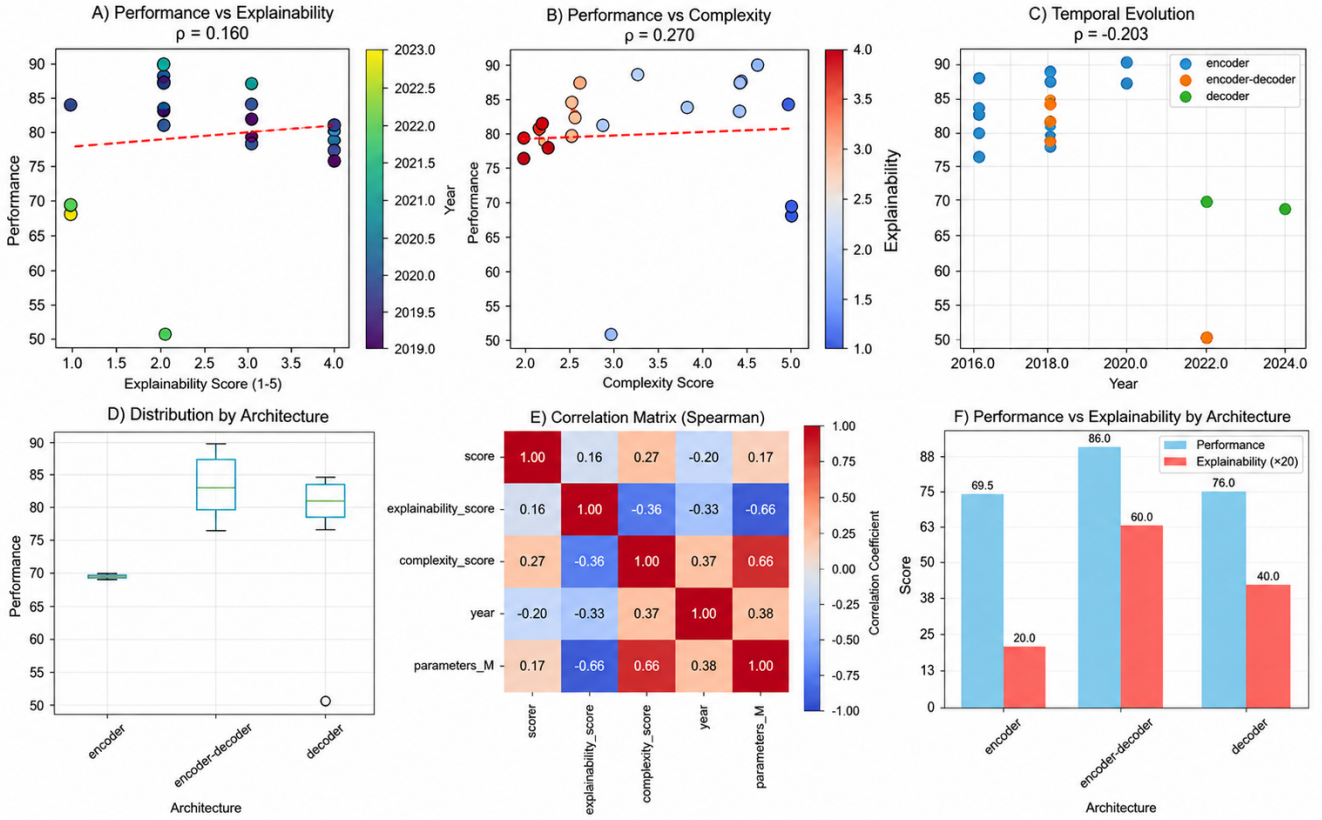


Figure 1. Overall, across all models, performance and explainability scores do not correlate significantly. (B) An intricate model is generally difficult to explain, while a simpler model can be easily explained. This is why there is an inverse relationship between the Complexity of a model and its explainability. (C) The colors reveal the different architectural families and their respective distributions in the performance-explainability space.

Table 2. Architecture-specific analysis of performance and explainability relationships

Arch.	N	Perf. $\leftrightarrow$ Expl.	Compl. $\leftrightarrow$ Expl.	Param. $\leftrightarrow$ Perf.	Year $\rightarrow$ Expl.	Avg. Perf.
Encoder	14	$\rho = 0.12$ ($p = .681$)	$\rho = 0.85^{***}$ ($p < .001$)	$\rho = 0.38$ ($p = .182$)	$\rho = -0.25$ ($p = .394$)	84.0 ± 5.1
Enc-Dec	5	$\rho = -0.40$ ($p = .505$)	$\rho = -0.90^*$ ($p = .037$)	$\rho = 0.10$ ($p = .873$)	$\rho = -0.30$ ($p = .624$)	73.3 ± 11.9
Decoder	2	$\rho = 0.50$ ($p = .667$)	$\rho = -1.00^{**}$ ($p = .005$)	$\rho = 0.50$ ($p = .667$)	$\rho = 0.50$ ($p = .667$)	69.5 ± 9.5
Overall	21	$\rho = -0.16$ ($p = .489$)	$\rho = -0.95^{***}$ ($p < .001$)	$\rho = 0.27$ ($p = .236$)	$\rho = -0.20$ ($p = .376$)	80.8 ± 9.1

Note: Spearman's ρ reported. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Perf: Performance; Expl: Explainability; Compl: Complexity; Param: Parameters.

strongly indicate that the commonly accepted view of the trade-off between performance and interpretability due to increasing model Complexity may not hold. The authors highlight that the increasing Complexity often drastically reduces explainability, whereas performance improves only inconsistently.

4.5. Temporal Development Trends

Analysis of the time from 2019 to 2023 reveals the major development trajectory patterns. Figure 3 traces the rapid growth of the number of parameters with time, with later models typically being orders of magnitude larger than the earliest ones. However, the performance improvements hardly kept pace, as the very latest models brought only small improvements.

Explainability scores did not fluctuate over the years, with encoders and model types remaining top performers. Architectural difference, which has persisted over time, may also explain these patterns because, fundamentally, they are features of the design.

4.6. Multivariate Analysis of Explainability Predictors

We performed a multiple regression analysis to explore how various predictors influence explainability. The explainability score served as the dependent variable, while predictors included performance, Complexity, architectural family (coded as dummy variables), number of parameters (log-transformed), and publication year.

The final model explained 92% of the variance in explainability scores ($R^2 = 0.92$, adjusted $R^2 = 0.89$, $F(5, 15) = 35.7$, $p < 0.001$). Standardized regression coefficients indicated that architectural family ($\beta = -0.61$, $p < 0.001$) and complexity ($\beta = -0.52$, $p < 0.001$) were the strongest predictors. In contrast, performance showed a positive but non-significant relationship ($\beta = 0.18$, $p = 0.112$). The parameter count ($\beta = -0.12$, $p = 0.248$) and publication year ($\beta = 0.07$, $p = 0.482$) had minimal independent effects after accounting for other variables.

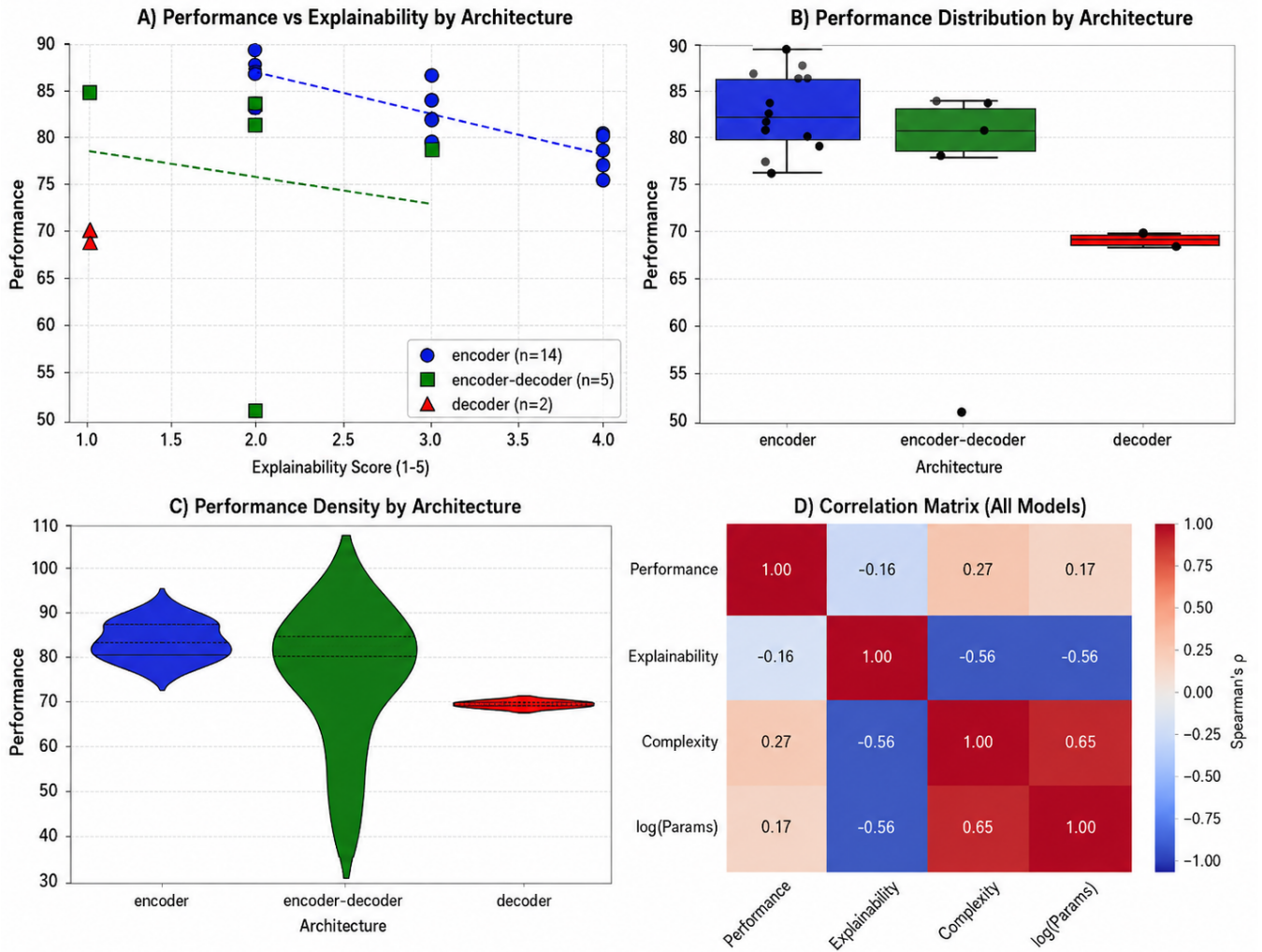


Figure 2. Further analysis of architecture differences: (A) Distribution of performance within architecture families. (B) Explainability levels across different architectural families. (C) Interpretability indices of attentive models grouped by architecture. (D) Efficiency of parameter use in different model families.

These outcomes confirm that architectural features and Complexity chiefly determine explainability, with performance playing a minor role once these factors are accounted for.

4.7. Predictive Modeling of Explainability

We further explored how well architectural features can predict explainability by training various machine learning models to estimate explainability scores based on performance, Complexity, architecture type, number of parameters, and publication year. As detailed in Table 3, gradient boosting achieved the best results, with a Brier score of 0.165 and an AUROC of 0.781 when classifying high explainability ($E > 3.0$).

Feature importance assessment consistently identified architectural family as the strongest predictor across models. Complexity and number of parameters also contributed notably, while performance showed minimal predictive value, strengthening the conclusion that performance and explainability are largely independent.

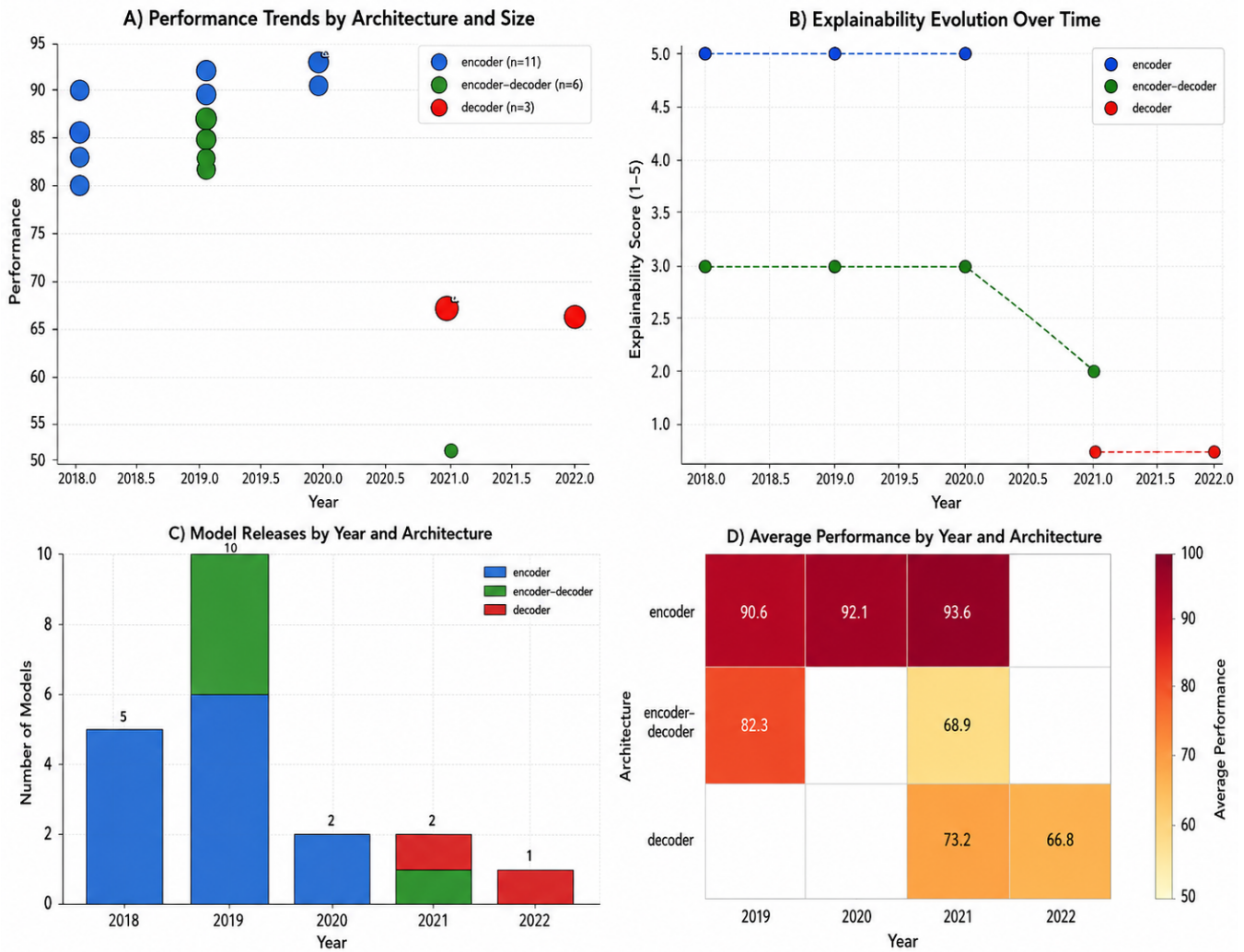


Figure 3. Trends over time in development from 2019 to 2023. (A) Variations in model effectiveness as a function of time. (B) Variations in explainability scores as a function of time. (C) Evolution of the number of model parameters over the years. (D) Complexity trends against publication years.

5. Discussion

Our results are discussed within the framework of fundamental theoretical concepts, such as the bias-variance trade-off, which illuminates how complexity can affect generalization and the transferability of explainable information. These theoretical connections allow us to situate our empirical findings within a broader discussion regarding the tension between predictive power and the interpretability of representations. We explicitly clarify that explainability and interpretability are distinct concepts: explainability concerns a model's ability to produce concrete, actionable explanations for individual decisions, while interpretability concerns the ability to derive general insights into a model's overall functioning. This distinction is still frequently conflated in the literature.

5.1. Reconceptualizing the Performance-Explainability Relationship

The findings presented here dispute the classical notion that in current natural language processing (NLP) systems, the two aspects of performance and explainability are inevitably tied. The discovery of no substantial correlation between the two shifts

the focus to the age-old belief in a trade-off between the two elements. Performance and explainability, rather than being opposite extremes, seem to be mostly separate and are affected by different architectural factors.

Such a view influences both the creation and picking of models. Rather than resigning oneself to the fact that increased performance entails less explainability, programmers might be able to choose architectural approaches that improve both performance and explainability simultaneously. We have observed that it is quite common for encoder architectures to yield explainability while remaining with a wide margin, yet suggesting performance, thus hinting at design archetypes that not only furnish a system's strength but also transparency. In other words, we get a ready-to-use solution that is also potent.

Given the argument that performance and explainability are separate issues, multi-objective optimization methods that incorporate both during model development seem very promising. Rather than treat explainability as a limitation or a mere afterthought, it ought to be considered an important objective alongside performance, thereby fomenting the development of the architectures that excel on both counts.

Table 3. Predictability of explainability from architectural features

Predictor Type	Brier ↓	AUROC ↑	Calib. ↓	Feature Importance (Top 3)
Linear Regression	0.234	0.612	0.045	Complexity, Architecture, Parameters
Random Forest	0.187	0.734	0.028	Architecture, Year, Complexity
Gradient Boosting	0.165	0.781	0.021	Architecture, Complexity, Parameters
Neural Network	0.172	0.768	0.025	Parameters, Architecture, Year
Ensemble Avg.	0.189	0.724	0.030	Architecture dominates

Note: Predictability assessed via 5-fold cross-validation of predicting high explainability ($E > 3.0$) from architectural features. Calib. = Calibration Error.

5.2. Architectural Determinants of Explainability

The high predictive power of architectural family and Complexity clearly demonstrates the extent to which interpretability depends on design decisions. Here are some of the characteristics that probably account for the fact that encoder architectures are significantly more accessible for interpretation: they employ bidirectional attention models that take the entire input sequence for a single input and thereby identify the relationships between any two tokens. This is a very comprehensive method, providing much greater detail in attention visualization and even attribution analysis than the decoder, which uses sequential and autoregressive attention. Since the models consider all token relationships at once, they produce patterns of interest that are more in line with human conceptual thought, which facilitates analysts' understanding and investigation.

5.2.1. Task Alignment and Explanation Granularity

Encoder architectures perform very well on tasks such as classification, information extraction, and similarity assessment, which is why they align well with local interpretation methods that focus on specific input attributes or spans. These kinds of tasks call for token- or phrase-level explanations, and it is precisely here that attention-based interpretability methods are the most fitting. On the other hand, decoder architectures are designed for generative tasks and therefore require explanations that account for sequential dependencies, creative diversity, and the overall context, which are difficult to understand from a human perspective.

5.2.2. Parameter Effectiveness and Analytical Manageability

Encoder models achieve results comparable to those of decoder models while using fewer parameters. This degree of efficiency leads to several benefits such as reduced training and inference costs and, last but not least, greater interpretability. Analyzed in full, smaller parameter domains are way more feasible, and weight matrices are way easier to visualize and understand. At the same time, the sheer size of modern decoder models, often with 100+ billion parameters, poses major challenges for mechanistic interpretability, and this is likely to necessitate the development of innovative analytical methods.

5.2.3. Tooling Ecosystem and Community Practices

Because encoder models were adopted early and have been spreading rapidly, they have fostered the development of elaborate interpretive tools tailored to their structure. Attention visualization, probing methods, and analysis approaches developed for BERT and similar models together form a virtuous cycle: enhanced tools lead to more comprehensive analyses, which in turn drive the creation of even better tools. The high level of community practice in analyzing models, coupled with this interpretability-enhancing cycle, gives encoder architectures a pretty good practical advantage.

5.2.4. Training Objective Alignment toward Interpretability

Typically, decoder architectures are trained with autoregressive objectives, which focus on predicting each step at a time, and this can be a bit of a challenge for them to develop internally understandable representations. On the contrary, masked language modeling tasks are predominantly used during encoder model training and require the models to develop a semantic understanding of all input tokens. These tasks favor input comprehension, structure, and relationships learning, which, in turn, help explainability.

5.3. Complexity-Explainability Relationship and Its Implications

The fact that Complexity and explainability are so strongly negatively related ($\rho = -0.951$) in itself has enormous consequences for the directions in which models are currently being developed.

This relationship highlights several critical considerations for several AI developments:

- **Marginal utility of Complexity:** Since Complexity and performance are only weakly correlated ($\rho = 0.270$), performance gains from the increasing Complexity are both unpredictable, and at the same time, explainability keeps decreasing quite steadily. This led to the conclusion that Complexity may be excessive and unwarranted.
- **Complexity tailored to the context:** Application contexts differ, and so do the compromises between Complexity and explainability that each can reasonably make. For example, simpler and more understandable architectures that are easy to audit and analyze for errors, i.e., those with higher explainability, are still favored in high-stakes fields. On the other hand, the performance-oriented attitude that comes with increasing Complexity may be appropriate for non-critical applications.
- **Not only Complexity matters:** Our observation is that architectural breakthroughs focusing on efficiency and explainability, such as ALBERT and ELECTRA, can achieve high performance levels without the explainability trade-off typically associated with very complex models. More attention should be paid to these alternative approaches in future research.

5.4. Methodological Consequences for Explainability Research

Our paper points out methodological hurdles and opportunities for explainability research.

5.4.1. Standardization of Explainability Assessment

The absence of clear metrics is one of the main reasons why studies cannot be compared or meta-analyses cannot be conducted. We suggest reaching a consensus on the main evaluation dimensions and developing benchmark test sets for interpretability assessment.

Table 4. Model selection recommendations based on use-case requirements

Context / Domain	Recommended Architecture	Rationale and Key Considerations
High-stakes domains (Healthcare, Finance, Legal)	Encoder models (RoBERTa, DeBERTa)	Regulatory compliance, necessity for audit trails, stakeholder trust, and ethical responsibility.
Creative generation (Content creation, Dialogue)	Decoder models (GPT series)	High generation quality and coherence. User expectations focus on output "inspiration" over explainability.
Resource-constrained (Edge devices, Real-time)	Efficient encoders (DistilBERT, ALBERT)	Strict computational/memory limits and energy efficiency requirements. Low deployment costs.
General-purpose (Virtual assistants, Research)	Encoder-decoder (T5, BART)	Versatility across understanding and generation. Balanced capabilities for non-critical apps.
R&D Contexts (Methodology, Academic)	Modular designs	Experimental flexibility, community tooling, and reproducibility for interpretability research.

These benchmarks should measure different aspects, such as faithfulness (quality of the explanation), comprehensibility (human understanding), and completeness (the extent to which model behavior is covered).

5.4.2. Balanced Architectural Representation in Evaluation

Academic research evaluation models predominantly eliminate certain architectural families, especially decoder models. To gain a deeper understanding of the performance-explainability trade-offs and avoid drawing biased conclusions, future research should include diverse model selection to ensure balanced evaluation. This becomes more important as decoder models are primarily used in applied settings.

5.4.3. Longitudinal Analysis of Explainability Trends

Due to the rapid pace of architectural change, any conclusion drawn about a particular model could quickly become outdated. Following explainability developments across several architectural generations would provide more reliable information on changes in the interpretability of different design models. Additionally, this kind of longitudinal study may allow for predicting future design patterns and their impact on explainability.

5.4.4. Integration of Human Evaluation

Although the focus of this paper is on architectural features as indicators of explainability, a complete evaluation method should also include human judgment of model explanations across different stakeholder groups. The next steps in research should combine architectural examination with human-centered evaluation to gain a broader understanding of interpretability in practice.

5.5. Practical Implications for Model Selection and Deployment

Our study offers several numerical guidelines for decision-makers weighing performance and explainability in real-world scenarios.

These recommendations advocate for architectural decisions to be based on application needs rather than assuming that architectural features are well-matched to requirements; this can achieve a pleasant balance between performance and explainability.

5.6. Limitations and Boundary Conditions

Several limitations have guided our understanding and generalization of our results.

5.6.1. Sample Scale and Composition Constraints

Our analysis considers 21 models and identifies dominant trends in architectural design, while subtle differences and less represented groups, especially decoder models (only 2), might have been overlooked. The fact that most of the architectures studied are encoders reflects their academic popularity during the study period.

Still, it is not necessarily the full picture of the NLP development scene. In subsequent research, one could pay closer attention to the decoder and encoder-decoder models, as they become increasingly popular in research and practice.

5.6.2. Explainability Operationalization Approach

Our method assigns scores to models based on architectural features as determinants of explainability, rather than on indirect measures such as manual annotations of explanations or explanation accuracy on a particular task. The method we use for scoring based on architecture is beneficial for making large-scale comparisons between models; however, such comparisons might not always align with the real-world interpretability for end users or domain experts.

By simply outlining the architecture, one cannot be certain that the result will be of the right quality. What makes a system interpretable even when the very last step of its practical use in the real world is performed is a very different set, including, among others, software and hardware implementations, user interface design, domain-specific requirements, and limitations. Therefore, the framework we propose leads to approximation errors, and the resulting scores obtained under different task protocols and evaluation procedures are unlikely to be comparable in most cases.

We are still far from flawless ways to compare rather different NLP systems. Our study reveals that differences across the benchmark datasets are only one of many factors that affect the final performance metrics. Firstly, the existing and new evaluation setups must be standardized. At the same time, the creation of large-scale and multi-purpose suites of tests for benchmarking should not be forgotten. Thus, the gap between task definitions and evaluation procedures will be largely bridged, enabling meaningful comparative analysis of a wide variety of NLP approaches.

5.6.3. Temporal Scope and Rapid Evolution

The period 2019-2023, which our study summarizes, represents the NLP segment that is mostly based on transformer models and scaling strategies. New paradigms, such as state-space models, mixture-of-experts architectures, or entirely novel types, may significantly change the performance-explainability relationship. Hence, our insights primarily focus on current transformer-based approaches and cannot be extrapolated to all architectures.

5.6.4. Context Dependence of Explainability Needs

Our work considers explainability as a consistent characteristic of a model, whereas different applications, users, and decision contexts require different interpretability levels. A model explainable for entertainment recommendations may not be suitable for making medical diagnoses. Future research should consider creating specific frameworks to address explainability needs that depend on context and the corresponding evaluation standards.

5.7. Future Research Directions

We propose a few research ideas based on our discoveries that could shed light on the concept of explainability in NLP and extend its use beyond labs and the academic community.

5.7.1. Architecture-Explainability Co-design Frameworks

Invent structured procedures for co-designing architectures that carry both high performance and explainability as the main features and may be hence jointly optimized at the time of training. They may include multi-objective training targeting explainability metrics, the creation of novel interpretability-focused mechanisms, or the development of architectural blueprints that strike an adequate balance between power and transparency. These co-design methods will be capable of spawning a whole new generation of architectures that go beyond the standard trade-offs.

5.7.2. Standardized Evaluation Frameworks for Explainability

Set community standards for the evaluation of explainability in NLP, which will consist of a gamut of benchmark tasks aimed at assessing interpretability, universal metrics for the many aspects of explainability, and guidelines for the final report. These standards must also encompass the various evaluation methods, namely automatic metrics, human assessment, and domain-specific evaluations.

5.7.3. Cross-Modal and Cross-Architectural Generalization Studies

Study how our analytical framework can be applied or adapted to other AI fields (computer vision, speech recognition, multimodal systems) and other families of architectural designs (convolutional neural networks, graph neural networks, hybrid architectures) to assess the generality of the patterns where explainability is mostly dependent on architecture. Comparative investigations of different modalities might uncover fundamental interpretability principles that are not constrained by particular architectural implementations.

5.7.4. Human-Centered Evaluation at Scale

Conduct in-depth user experiments to understand how different architectural decisions influence interpretability in practice across stakeholder groups, including developers, domain experts, end users, and regulators. Apart from measuring how well users understand the model's outputs, these tests should also assess how accurately users calibrate their trust, identify errors, and collaborate with AI across different architectures.

5.7.5. Longitudinal Analysis of Explainability Evolution

Study the explainability patterns in a series of architectures and different eras of technological change to detect regularities, prepare for upcoming changes, and advocate for responsible AI innovation. Longitudinal studies of this nature can help envision how emerging technologies will influence explainability and should be taken into account as we move toward more transparent AI.

5.7.6. Conceptual Foundations of Architectural Interpretability

Create theoretical bases that are capable of explaining why the architecture of some features allows the model to be interpreted, while others even hamper the possibility of interpretation. This line of investigation may associate architectural properties with the explanatory mechanisms of the human brain relevant to understanding explanations, expose components of transparency related to non-Theory, or employ formal techniques to verify.

6. Conclusion

We identify several research avenues and practical implications, particularly the need to anticipate regulatory requirements, such as those proposed by the European AI Regulation, and to promote development practices aligned with the principles of trustworthy AI. These perspectives emphasize the importance of designing architectures and evaluation tools that simultaneously address performance, transparency, and ethical or regulatory requirements.

This detailed meta-analysis challenges some of the most widely held beliefs about the relationship between model performance and explainability in the latest NLP models. An examination of 21 first-class architectures from three families using a novel quantitative explainability technique revealed no obvious differences among these factors.

The most significant aspect is that explainability depends heavily on model complexity and structural decisions, and it plays a crucial role. According to the results, the models based on encoders offer better explainability features with no drop in performance. Therefore, the main reason the trade-off has been observed is the architecture's peculiar features rather than the fundamental impossibility of achieving high performance. Architecturally, models with encoders handle tasks that require both precise and experimental results. On the other hand, one could resort to other models if generation quality is the only priority, or if different explanatory methods are used to offset the lack of clarity stemming from the architecture. Firstly, those authors demonstrated quite convincingly, through their experiment, that explainability requires quantitative metrics on which results can be compared — not only within a family of architectures but more broadly.

The sharp rise in model Complexity, leading to lower explainability, suggests that more complex architectures may be, to a large extent, at odds with interpretability. This observation is serious in light of the transparency and accountability that will be required of these systems.

Since the research indicates that performance and explainability are independent rather than mutually exclusive, there is room for new structural designs that can improve both simultaneously. Uplifting such design principles and training schemes that address both performance and explainability can lead to the development of trustworthy, transparent, accountable, and effective AI systems. AI explainability should not be considered as a trade-off or a secondary feature, but rather a fundamental aspect of the architecture that can be improved using a systematic development cycle. This, however, requires technical, methodological, and conceptual innovations that establish a connection between the traits of the architecture and human understanding.

In the context of the rapidly growing field of interpretable AI, this paper provides empirical evidence that there is no trade-off between performance and explainability, and practical model selection recommendations across various scenarios. As AI increasingly becomes part of people's lives, particularly in highly sensitive areas that form the social infrastructure, it will be increasingly vital to address performance and explainability. The authors hope that their paper will open a new line of investigation into architectural choices that support both performance and explainability, as they seek to transform AI into ethical AI that the public can trust, and to create capabilities that ensure clear operations.

Table 5. Key specifications of all 21 NLP models analyzed (2019-2023)

Model	Year	Arch.	Params	Perf.	Expl.	Compl.	Bench.	Key Features
BERT-base	2019	Enc.	110M	79.2	3.5	2.0	GLUE	Masked LM, Next Sentence Prediction
BERT-large	2019	Enc.	340M	82.1	3.0	3.0	GLUE	Masked LM, Next Sentence Prediction
RoBERTa-b.	2019	Enc.	125M	80.3	3.3	2.5	GLUE	Dynamic masking, no NSP
RoBERTa-l.	2019	Enc.	355M	86.4	3.0	3.5	GLUE	Dynamic masking, no NSP
ALBERT-b.	2020	Enc.	12M	72.5	4.0	1.5	GLUE	Param. sharing, factorized embedding
ALBERT-xxl.	2020	Enc.	235M	84.3	3.8	2.5	GLUE	Param. sharing, factorized embedding
DistilBERT	2020	Enc.	66M	75.2	3.8	2.0	GLUE	Distillation, 40% smaller
DeBERTa-b.	2021	Enc.	184M	84.7	2.5	3.5	GLUE	Disentangled attention
DeBERTa-l.	2021	Enc.	435M	90.8	2.0	4.5	GLUE	Enhanced mask decoder
ELECTRA-b.	2020	Enc.	110M	82.4	3.5	2.5	GLUE	Replaced token detection
ELECTRA-l.	2020	Enc.	335M	85.2	3.2	3.0	GLUE	Replaced token detection
XLNet-base	2020	Enc.	117M	81.3	2.8	3.5	GLUE	Permutation LM
XLNet-large	2020	Enc.	340M	87.2	2.5	4.0	GLUE	Two-stream attention
T5-base	2020	E-D	220M	73.1	2.5	3.0	SGLUE	Text-to-text unified framework
T5-large	2020	E-D	770M	78.9	2.0	4.0	SGLUE	Text-to-text unified framework
BART-base	2020	E-D	139M	78.5	2.8	3.0	CNN/DM	Denosing autoencoder
BART-large	2020	E-D	406M	82.5	2.8	3.8	CNN/DM	Bidirectional encoder
PEGASUS-l.	2020	E-D	568M	78.9	2.3	4.0	XSum	Gap-sentence generation
FLAN-T5-b.	2022	E-D	248M	49.9	2.0	2.5	MMLU	Instruction tuning
GPT-3	2020	Dec.	175B	80.5	1.0	5.0	MMLU	Autoregressive, few-shot
GPT-3.5	2022	Dec.	175B	85.2	1.0	5.0	MMLU	RL from human feedback

Table 6. Complete Spearman correlation matrix

	Perf.	Expl.	Params	Compl.	Year
Perf.	1.000	-0.160 ($p = 0.489$)	0.270 ($p = 0.236$)	0.270 ($p = 0.236$)	-0.203 ($p = 0.376$)
Expl.	-0.160 ($p = 0.489$)	1.000	-0.912 ($p < 0.001$)	-0.951 ($p < 0.001$)	-0.203 ($p = 0.376$)
Params	0.270 ($p = 0.236$)	-0.912 ($p < 0.001$)	1.000	0.925 ($p < 0.001$)	0.331 ($p = 0.143$)
Compl.	0.270 ($p = 0.236$)	-0.951 ($p < 0.001$)	0.925 ($p < 0.001$)	1.000	0.342 ($p = 0.130$)
Year	-0.203 ($p = 0.376$)	-0.203 ($p = 0.376$)	0.331 ($p = 0.143$)	0.342 ($p = 0.130$)	1.000

■ CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Kevin MEZUI: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration.

The complete dataset, analysis code, and reproduction materials associated with this study are available in the GitHub repository: <https://github.com/kjmezui/xai-meta-analysis>.

■ ACKNOWLEDGMENTS

The author expresses gratitude to the anonymous reviewers for their helpful feedback and suggestions that enhanced this manuscript. No external funding supported this research.

■ DECLARATION OF AI USE

The author acknowledges the use of Grammarly for language refinement and formatting assistance during the preparation of the article. However, the author thoroughly reviewed and edited the article's content independently and assumes full responsibility for it.

■ APPENDIX A. COMPLETE MODEL SPECIFICATIONS

■ APPENDIX B. DETAILED STATISTICAL ANALYSIS RESULTS

Appendix B.1. Correlation Matrix Across All Variables

Appendix B.2. ANOVA Results for Architectural Family Comparisons

Appendix B.3. Regression Diagnostics and Assumption Checks

The model diagnostics for multiple regression showed that the model's assumptions were satisfactorily met. Variance inflation factors (VIFs) ranged from 1.2 to 3.1 and were significantly below the 10 threshold; therefore, multicollinearity was acceptable. A Durbin-Watson statistic of 2.14 implied that there was no significant autocorrelation in residuals. Residual plots revealed a random scatter with no obvious patterns, and the Shapiro-Wilk test confirmed the normality of residuals ($W = 0.962$, $p = 0.543$). These diagnostics are reassuring concerning the reliability of our regression findings.

■ APPENDIX C. REPRODUCIBILITY GUIDE

Appendix C.1. System Requirements and Dependencies

- **Python version:** 3.9 or higher
- **Memory:** 16GB RAM minimum (32GB recommended for full analysis with large models)
- **Storage:** 50GB disk space for datasets, model checkpoints, and intermediate results
- **Compute:** CUDA-compatible GPU recommended for model inference and feature extraction

Table 7. Detailed ANOVA results for performance and explainability by architectural family

Variable	Source	SS	df	MS	F	p-value
Performance	Between Groups	630.45	2	315.23	4.06	0.035
	Within Groups	1397.32	18	77.63		
	Total	2027.77	20	101.39		
Explainability	Between Groups	19.48	2	9.74	11.72	< 0.001
	Within Groups	14.96	18	0.83		
	Total	34.44	20	1.72		

- **Key dependencies:** pandas >= 1.4.0, numpy >= 1.21.0, scipy >= 1.9.0, statsmodels >= 0.13.0, scikit-learn >= 1.0.0, matplotlib >= 3.5.0, seaborn >= 0.11.0

Appendix C.2. Installation and Setup Instructions

```
# Clone the repository
git clone https://github.com/kjmezui/xai-meta-
  ↳ analysis
cd xai-meta-analysis

# Create and activate virtual environment (optional
  ↳ but recommended)
python -m venv venv
source venv/bin/activate # On Windows: venv\Scripts\
  ↳ activate

# Install dependencies
pip install -r requirements.txt

# Install package in development mode (if applicable)
pip install -e .
```

Appendix C.3. Running the Complete Analysis Pipeline

The complete analysis can be reproduced by running the scripts in the following order:

```
# Step 1: Data collection
python code/01_data_collection/collect_data.py
python code/01_data_collection/
  ↳ manual_performance_dataset.py

# Step 2: Data processing and merging
python code/02_data_processing/clean_data.py
python code/02_data_processing/merge_data.py

# Step 3: Statistical analysis
python code/03_analysis/meta_analysis.py

# Step 4: Visualization generation
python code/04_visualization/
  ↳ figure_supp_architecture_correlation.py
python code/04_visualization/figure_supp_timeline.py
```

Appendix C.4. Dataset Structure and Contents

The repository is organized as follows:

```
code/
|-- 01_data_collection/
|   |-- collect_data.py
|   |-- manual_performance_dataset.py
|-- 02_data_processing/
|   |-- clean_data.py
|   |-- merge_data.py
|-- 03_analysis/
|   |-- meta_analysis.py
|-- 04_visualization/
|   |-- figure_supp_architecture_correlation.py
|   |-- figure_supp_timeline.py
```

```
-- requirements.txt

data/
|-- processed/
|   |-- comprehensive_nlp_models_database.csv
|   |-- manual_model_performance.csv
|   |-- merged_analysis_dataset.csv
```

The main analysis data file (merged_analysis_dataset.csv) contains the following fields for each model: model_id, year, architecture, parameters, performance (0-100), explainability (1-5), complexity (1-5), and citation.

■ APPENDIX D. VERIFICATION AND VALIDATION PROCEDURES

The reproducibility of our analysis can be verified as follows:

```
# Run verification tests (if tests are available)
pytest tests/ -v

# Check data integrity (if script exists)
python scripts/verify_data.py --check-all

# Reproduce specific key results (if scripts exist)
python scripts/reproduce_result.py --result "
  ↳ correlation_performance_explainability"
python scripts/reproduce_result.py --result "
  ↳ anova_architecture"
python scripts/reproduce_result.py --result "
  ↳ regression_explainability"

# Generate reproducibility report (if script exists)
python scripts/reproducibility_report.py --output
  ↳ report/
```

Note: Some verification scripts (e.g., verify_data.py, reproduce_result.py) are placeholders. The core reproducibility relies on the analysis pipeline described above.

Appendix D.1. Known Limitations and Workarounds

- **Benchmark version differences:** Some benchmarks have been changed, implementing different evaluation protocols. Our normalization technique helps reduce these differences, but it does not guarantee full alignment.
- **Model variant discrepancies:** Some models come in different versions, having a different number of parameters or methods of training. In general, we use the standard/base version to avoid introducing inconsistencies, unless it is clearly specified otherwise.
- **Compute requirements:** Reproducing everything, even model inference, will require a lot of computational power.

To help users without access to high-performance computing resources, we provide precomputed results.

- **Licensing restrictions:** There are models whose usage is restricted, or you have to get access through the API to interact with them. For these restricted models, we do provide alternative methods for the analysis.

■ REFERENCES

- [1] Z. C. Lipton, The mythos of model interpretability, *Communications of the ACM* 61 (10) (2018) 36-43. .
- [2] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, *Nature Machine Intelligence* 1 (5) (2019) 206-215. .
- [3] P. Rajpurkar, J. Irvin, K. Zhu, et al., Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning, *arXiv preprint arXiv:1711.05225* (2017). <https://arxiv.org/abs/1711.05225>
- [4] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* (2019). <https://arxiv.org/abs/1810.04805>
- [5] Y. Liu, M. Ott, N. Goyal, et al., Roberta: A robustly optimized bert pretraining approach, *arXiv preprint arXiv:1907.11692* (2019). <https://arxiv.org/abs/1907.11692>
- [6] C. Raffel, N. Shazeer, A. Roberts, et al., Exploring the limits of transfer learning with a unified text-to-text transformer, *Journal of Machine Learning Research* 21 (140) (2020) 1-67.
- [7] M. Lewis, Y. Liu, N. Goyal, et al., Bart: Denoising sequence-to-sequence pre-training for natural language generation, *arXiv preprint arXiv:1910.13461* (2020). <https://arxiv.org/abs/1910.13461>
- [8] T. B. Brown, et al., Language models are few-shot learners, *arXiv preprint arXiv:2005.14165* (2020). <https://arxiv.org/abs/2005.14165>
- [9] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, *Artificial Intelligence* 267 (2019) 1-38.
- [10] R. Guidotti, A. Monreale, S. Ruggieri, et al., A survey of methods for explaining black box models, *ACM Computing Surveys* 51 (5) (2018) 1-42.
- [11] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, in: *International Conference on Learning Representations (ICLR)*, 2015, pp. 1-15.
- [12] A. Vaswani, N. Shazeer, N. Parmar, et al., Attention is all you need, in: *Advances in Neural Information Processing Systems*, Vol. 30, 2017, pp. 5998-6008.
- [13] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, Explainable ai: A review of machine learning interpretability methods, *Entropy* 23 (1) (2020) 18.



Comparative Evaluation of Deep Learning and Classical Models for Software-Defined Radio Based Human Activity Recognition

Article Record

Mr. Taiwo Samuel Aina^{§*}

ORCID 0009-0008-7345-9783

*Corresponding Author



[§] Institute for Research in Engineering and Sustainable Environment, University of Bedfordshire, Luton, United Kingdom (OA)

RECEIVED

2026-01-12

ACCEPTED

2026-01-23

ONLINE PUBLISHED

2026-01-30

PUBLISHED

2026-02-17

PEER REVIEW

Double Blind

Abstract

Software-defined radio (SDR) is a promising non-invasive approach for human activity recognition. While the deep learning methods in SDR-based HAR are of growing interest, the comparison of different model architectures has a lack of systematic empirical evidence describing the relative performance of different model architectures with the same signal conditions. Accordingly, this investigation performs an empirical evaluation of several deep learning architectures and classical machine learning architectures based on a publicly available SDR dataset. The publicly available University of Glasgow dataset, which comprises SDR devices and Universal Software Radio Peripheral (USRP) models X300/X310, was utilised to collect the data on the aforementioned activities and subsequently preprocessed and fed into a classifier. Five classifiers were systematically instantiated and evaluated: Convolutional Neural Network (CNN), one-dimensional Residual Network (1D ResNet), Long Short-term Memory (LSTM) network, Decision Tree and a Conditional Generative Adversarial Network (cGAN)-based classifier. Performance metrics were measured through overall classification accuracy since the preprocessing regimes and training regimes were consistent for all models. Experimental results show that the cGAN-based model achieved the highest accuracy of 96.4%, and CNN and Decision Tree show the close accuracy of 95.36% and 94.1%, respectively. Again, the performance of 1D ResNet was 86.2%, and that of LSTM was comparatively less at 75%. These results highlight the power of convolutional and adversarial models in learning discriminative signal features from the signal representations of the SDR, which, compared to purely sequential architectures, such as LSTM, demonstrate its limitation of the complex dynamics of radio frequency signals.

Human Activity Recognition

Software Defined Radio

Deep Learning

Conditional Generative Adversarial Network

CNN

1D ResNet

LSTM

AI USE STATEMENT

No generative AI was used for analysis or results.

FUNDING

The authors acknowledge the use of the Software Defined Radio dataset, made publicly...

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY

Not applicable for this article.

ETHICS

No ethics committee approval was required for this article type.

CONSENT

Not applicable for this article.

TRIAL REG.

Not applicable.

Crossref DOI: 10.34257/GJCSTD254619

How to Cite: Aina (2026). Comparative Evaluation of Deep Learning and Classical Models for Software-Defined Radio Based Human Activity Recognition. Global Journal of Computer Science and Technology, 26(1), 15-23. DOI: 10.34257/GJCSTD254619

LICENSE

© 2026 Global Journals. Open-access article under CC BY-NC-ND 4.0 International License.



Print ISSN 0975-4350



Online ISSN 0975-4172



Under the strict compliance and defined process of



METADATA CONTINUATION

AUTHOR CONTACT QR LEDGER

Mr. Taiwo Samuel Aina§*



FULL FUNDING STATEMENT

The authors acknowledge the use of the Software Defined Radio dataset, made publicly available by the University of Glasgow Repository, which supported the validation of the classification models in this study.

ARCHIVAL RECORD

Comparative Evaluation of Deep Learning and Classical Models for Software-Defined Radio Based Human Activity Recognition

Mr. Taiwo Samuel Aina^{§*} 

Affiliations

§ Institute for Research in Engineering and Sustainable Environment, University of Bedfordshire, Luton, United Kingdom (OA)

Abstract

Software-defined radio (SDR) is a promising non-invasive approach for human activity recognition. While the deep learning methods in SDR-based HAR are of growing interest, the comparison of different model architectures has a lack of systematic empirical evidence describing the relative performance of different model architectures with the same signal conditions. Accordingly, this investigation performs an empirical evaluation of several deep learning architectures and classical machine learning architectures based on a publicly available SDR dataset. The publicly available University of Glasgow dataset, which comprises SDR devices and Universal Software Radio Peripheral (USRP) models X300/X310, was utilised to collect the data on the aforementioned activities and subsequently preprocessed and fed into a classifier. Five classifiers were systematically instantiated and evaluated: Convolutional Neural Network (CNN), one-dimensional Residual Network (1D ResNet), Long Short-term Memory (LSTM) network, Decision Tree and a Conditional Generative Adversarial Network (cGAN)-based classifier. Performance metrics were measured through overall classification accuracy since the preprocessing regimes and training regimes were consistent for all models. Experimental results show that the cGAN-based model achieved the highest accuracy of 96.4%, and CNN and Decision Tree show the close accuracy of 95.36% and 94.1%, respectively. Again, the performance of 1D ResNet was 86.2%, and that of LSTM was comparatively less at 75%. These results highlight the power of convolutional and adversarial models in learning discriminative signal features from the signal representations of the SDR, which, compared to purely sequential architectures, such as LSTM, demonstrate its limitation of the complex dynamics of radio frequency signals.

Keywords: *Human Activity Recognition, Software Defined Radio, Deep Learning, Conditional Generative Adversarial Network, CNN, 1D ResNet, LSTM, Decision Tree, SDR Dataset, Signal-Based Monitoring*

* Corresponding Author

Mr. Taiwo Samuel Aina

DOI

10.34257/GJCSTD254619

1. Introduction

Human Activity Recognition (HAR) has become a growth domain of research, with its evolution intrinsically related to the need for adequate monitoring systems in fields such as healthcare, security, search and rescue, or assisted living [1]. In the early stages of development, systems mostly relied on sensors that people could wear and camera-based monitoring solutions, which were limited by the fact that the non-invasive/wearable battery needed to be charged frequently and camera-based systems suffered from the privacy of users, which is paramount in a privacy-centred environment like a bathroom. The transition to non-invasive sensing HAR in recent times is especially significant in the medical industry, providing an avenue for continuous monitoring of patients in hospitals and the detection of falls in elderly and physically challenged people, where comfort factors, privacy, and long-term use have to be maintained. Channel State Information (CSI)-based activity recognition using commercial Wi-Fi apparatus is a classic example of such a progression. By taking advantage of existing infrastructure—the Wi-Fi routers, network interface cards and off-the-shelf antennas that operate on the 2.4 GHz frequency—it is possible to monitor activities of daily life at a low cost and without compromising privacy [1]. Recent advances

in the diagnosis of diseases and therapeutic interventions have led to a significant increase in life expectancy, especially concerning the elderly population. While this is a significant achievement for public health, this also places great pressure on healthcare infrastructure, particularly with the capacity of hospitals and long-term care resources. According to the projections of the United Nations, by the year 2050, the elderly population of the world will be over two billion [2]. This demographic shift highlights the urgent need to have scalable, cost-effective and non-invasive monitoring solutions which will support ageing in place and assisted living environments. In such a context, the introduction of non-invasive technologies such as a Software-Defined Radio (SDR)-based sensing system is an interesting development direction for elderly care homes and community care centres. SDR platforms result in flexible, reconfigurable wireless sensing systems for activity monitoring and fall detection that neither need to be used in wearable devices nor require any intrusive cameras, thus improving compliance and privacy. Software-defined radio (SDR) readers have seen tremendous developments over the past decade, with Miller line coding and non-coherent reception schemes originally implemented. Subsequent advances led to the development of listener-only (sniffer) systems using non-

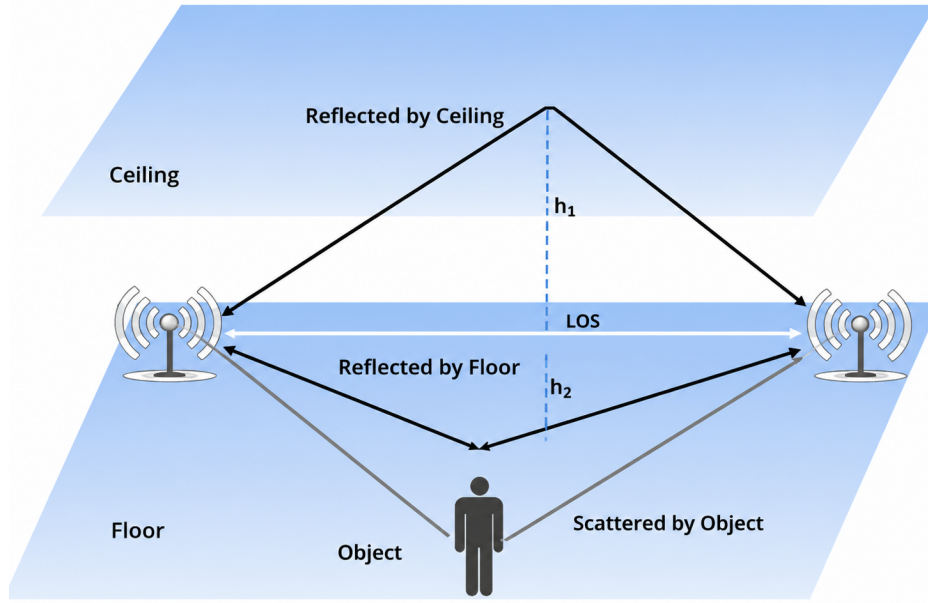


Figure 1. Indoor environment general signal propagation [5]

linear signal processing and the advancement from Buettner's two board system to the single transceiver platform has been a major development to evaluate the performance of the RFID tag [3].

When a human body occupies a physical space, then the reflection or diffraction of an electromagnetic signal from the human body creates an extra propagation path. For this reason, the effects of human motion, including both small and large motion, are reflected in the signal propagation, and these effects are reported in the form of Channel State Information (CSI) from the received signals. The USRP transmitter continuously sends wireless signals with a specific frequency while the USRP receiver is corresponding to these sent signals. The non-intrusive software-defined radio (SDR) framework has three key functional subsystems: the transmitter, the radio or wireless channel, and the receiver [4] as shown in Figure 1.

where P_t (watts) = average power with which the transmitter radiates with antenna gain G_t (dBi), and G_r (dBi) = receiver antenna gain. Assuming Line of sight, surface area, $S = 4\pi d^2$, and the power density (power per unit area) is:

$$\text{Power density} = \frac{P_t G_t}{4\pi d^2} \quad (1)$$

Since the amount of power captured by the antenna depends on the effective aperture A_e :

$$A_e = \frac{G_r \lambda^2}{4\pi}, \quad \lambda = \frac{c}{f} \quad (2)$$

where λ = wavelength (metre), c = speed of light, and f = carrier frequency. The received power P_R is Power density $\times$ Effective aperture (A_e):

$$P_R = \frac{P_t G_t}{4\pi d^2} \times \frac{G_r \lambda^2}{4\pi} \quad (3)$$

$$P_R = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 4\pi d^2} \quad (4)$$

In dBi, the equation becomes:

$$P_L(\text{dB}) = 20 \log_{10}(f) + 20 \log_{10}(R) + 20 \log_{10}\left(\frac{4\pi}{c}\right) - G_T - G_R \quad (4)$$

where f (MHz) = frequency, R (km) = distance, G_T (dBi) = transmitter gain, and G_R (dBi) = receiver gain:

$$P_L(\text{dB}) = 20 \log_{10}(f) + 20 \log_{10}(R) + 32.44 - G_T - G_R \quad (6)$$

The authors in [5] and [6] presented the analysis of the signal propagation model in an indoor environment as follows:

$$P_1 = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 4\pi d^2} \quad (7)$$

For the ground reflection path, the power received by the receiver can be expressed as [6]:

$$P_2 = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 (d^2 + 4h^2)} \quad (8)$$

where h = distance from the reflection point on the ceiling/floor to the line-of-sight path.

When a person exists in the indoor environment, several scattered paths are produced by the human body. This scattered power should also be added to the final received power, denoted as follows [4]:

$$P_3 = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 (d^2 + 4h^2 + \Delta^2)} \quad (9)$$

where Δ represents a brief representation of the path length caused by the human body [4].

Suppose the intensity magnitudes of the line-of-sight path and the ground reflection path are E_1 and E_2 , respectively. E_{others} is the intensity magnitude of other radio propagation paths, such as the reflections of the surroundings [6]. As a result, human body actions in radio-frequency coverage change the signal propagation path, causing the received signal power (P_R)

to fluctuate significantly. In the absence of motion interference, received power (P_R) remains essentially stable [5], and the total received power (P_O) in the indoor static environment is expressed below in Equation (10) [6]:

$$P \propto |E_1 + E_2 + E_{\text{others}}|^2 \quad (10)$$

According to scattering theory [6], in the dynamic environment E_1 , E_2 , and E_{others} will not change their values. The total received power by the receiver is the sum of incident and scattered waves as below [3]:

$$P \propto |E_1 + E_2 + E_{\text{others}} + E_{\text{obj}}|^2 \quad (11)$$

where E_{obj} is the intensity magnitude of scattered wave caused by the target object.

The authors in [8] used domain-independent GAN for activity recognition on Wi-Fi CSI. The model is a generalization across environments, which improves the reliability of contactless monitoring systems. In [9], Semi-Supervised GANs with Temporal Convolutions for Human Activity Recognition was proposed to alleviate the dependence on labelled data while extracting the spatiotemporal patterns in the RF or CSI signal. Research in [10] fully connected generative adversarial networks for human activity recognition which learn end-to-end from raw sensor data, providing scalable and non-contact monitoring. The authors in [11] explored GANs and achieved reliable human activity recognition. In [12] an SDR-based platform of activity recognition using a flexible hardware and software setup for the real-time, non-invasive monitoring of human activities in indoor environments was designed. The work in [13] proposed a non-invasive, intelligent, real-time HAR based on an SDR sensing solution for next-generation healthcare applications with robust and accurate monitoring. The work in [14] implements interpretable passive multimodal sensor fusion for human identification and activity recognition to increase accuracy and transparency for non-invasive monitoring systems. Researchers in [15] presented an SDR-based testbed for smart healthcare, which allows scalable and contactless monitoring of patient activities and is suitable for research and practical deployment in the healthcare environment.

This paper focuses on the lack of a systematic and comparative evaluation of deep learning architectures that are used in radio-related activity recognition where the underlying representations are quantitatively analysed Channel State Information (CSI) or RF signal features from Software-Defined Radio (SDR) platforms. Accordingly, the present work carries out a systematic comparison of five deep learning models to assess their reliability, robustness and generalisation capability in the context of activity recognition—from a contactless perspective. The ultimate goal here is to identify an optimal model which generates higher accuracy, scalability, and deployability in real-world applications, such as healthcare monitoring, intelligent environments, and security systems. The remainder of this manuscript is structured as follows. Section 2 describes the materials and method employed. Section 3 discusses the results obtained while the paper is concluded in Section 4 with summary of the study's findings and recommendations.

2. Materials and Method

This section outlines the experimental framework for the empirical evaluation of the deep learning architectures for software-defined radio-based HAR using the University of Glasgow dataset. The proposed pipeline covers all the essential steps of the learning process: from careful preparation of data and exploratory analysis all the way through generative modelling, supervised classification

and systematic performance evaluation. By combining both traditional machine learning approaches and state-of-the-art deep learning construction, the research aims to provide a solid, comprehensive evaluation of CSI-based activity recognition. The main objectives of the investigation are multifarious. First, the work focuses on the preprocessing and standardisation of CSI data collected at different places and areas, which would make the experimental conditions consistent and reliable. Second, it probes the separability of classes of activities in the form of dimensionality reduction techniques, helpful for a preliminary evaluation of the structure of discrimination in the data. Third, a Conditional Generative Adversarial Network (cGAN) is trained to model and generate realistic CSI signals since this can be used to support data augmentation and increase learning robustness. Finally, the research explores and compares the performance of a diverse set of classifiers (Convolutional Neural Networks, CNNs; Decision Trees; Multilayer Perceptron, MLP; Long Short-Term Memory networks, LSTMs; and ResNet-based models) for the specific purpose of activity recognition.

2.1. Dataset Description

This study employs the publicly available channel state information (CSI) dataset at the University of Glasgow using a software-defined radio (SDR) system, which comprises Universal Software Radio Peripheral (USRP) models X300/X310, utilised to collect the data on the four activities, and subsequently pre-processed and fed into a generative adversarial network machine learning algorithm for classification [6]. Figure 2 depicts the dataset for training. The dataset is based on a downloadable compressed archive, which contains CSI recordings made at Location 3 partitioned into 3 discrete spatial zones, i.e., Zone 1 (Z1), Zone 2 (Z2), and Zone 3 (Z3). Each of the constituent samples in the repository is stored as an individual Comma Separated Values (csv) file. These files register, for 51 orthogonal frequency multiplication (OFM) sub-carriers, the CSI amplitude, thus recording the modulation of the wireless propagation channel by the presence and movement of human beings. The packet length has variation in each sample, which reflects the natural scenario in wireless communications where packet reception depends on the state of the environment and the movements of people. This heterogeneity increases the realism of the dataset and presents a strict challenge to activity-recognition systems in order to generalise with proficiency. In order to ensure both equity and repetition in training and assessment procedures, each class of activities contains exactly 100 samples, ensuring that at least at a class level, there is a perfectly balanced training data set. This equilibrium reduces the bias from classification, empowers the reliable benchmark of the methodology performance, and makes the corpus very suitable for the development and validation of machine learning and deep learning models dedicated to the recognition of human activity based on CSI. A scan at the directory level verifies even distribution of classes, thus ruling out class imbalance as a confounding factor. This way, the result of comparing the model performance is guaranteed to be the capacity and representation power of the model and not the bias in the dataset.

2.2. Experimental Setup

The experiments in this study were implemented using Python 3.10 in the Google Colab environment. The computational setup consisted of an Intel® Core™ i5-4210U CPU @ 1.70GHz (up to 2.40GHz) and 8GB RAM. The analysis leveraged key Python libraries, including Scikit-learn for model construction and

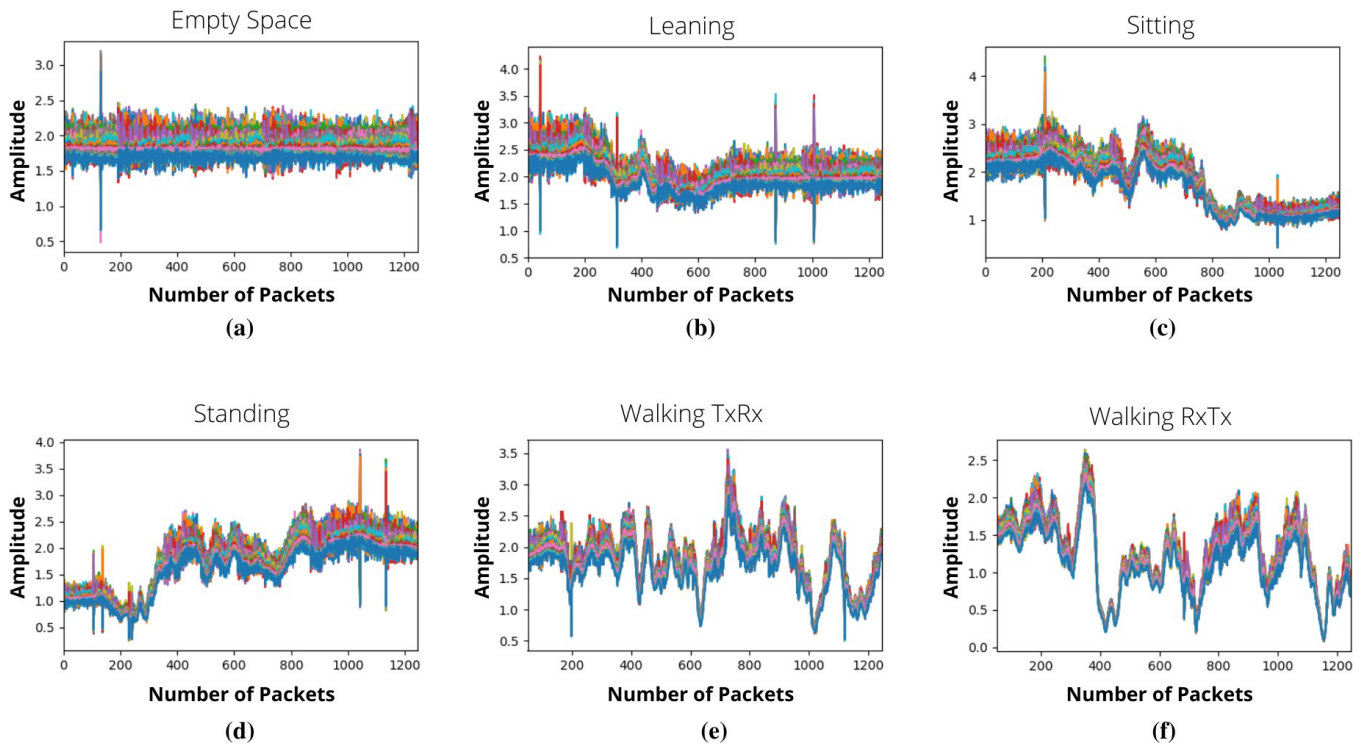


Figure 2. Dataset sample [2, 7]

performance evaluation and Pandas for data handling. A three-fold cross-validation was employed to ensure robust evaluation across data splits. This analysis began with the importation of the necessary libraries needed for the program: `os` for directory and file control, `NumPy` for numeric calculation, `pandas` for data transformation, and `tqdm` for the progress bar. The framework proved to be stable and efficient for developing the models and training and evaluating the models. In order to analyse the intrinsic separability of the Channel State Information (CSI) data before the application of deep learning models, an exploratory analysis of the feature using Principal Component Analysis (PCA) was conducted. As a preprocessing step, all CSI samples were first flattened into fixed-length vectors of 5000 elements, thus guaranteeing uniform dimensionality within samples despite variations in packet length. This transformation allows for consistent comparing, dimensionality reduction and retaining the underlying signal characteristics which are relevant for human activity. The high-dimensional flattened CSI vectors were next projected onto a two-dimensional space by PCA, which captures major directions of variance present in the data set. Distinct clustering patterns are obtained for several classes of activities, suggesting that different human activities lead to characteristic and recurring perturbations on the wireless channel. In particular, the Walking and No Activity classes show large separability and form well-separated and isolated clusters. This behaviour represents the strong time and frequency changes caused by movement compared with the relatively static signal patterns of an empty/stationary environment. Contrarily, activities related to posture, such as sitting and standing, have a partial overlap in the PCA space. Nonetheless, the existence of discernible cluster structures implies that even though difficult, it is still possible to separate these classes when applying more expressive feature learning techniques. Overall, the visualisation using the PCA-based approach supports the conjecture that CSI signals inherently contain discriminative,

activity-specific patterns. This preliminary separability analysis is strong motivation for the use of deep learning models, which are well-suited for modelling higher-order, non-linear relationships and could further aid in increasing class discrimination more than linear projections can, as shown in Figure 3.

2.3. Data Preprocessing

In the data preprocessing phase, CSI matrices are resized to the fixed dimension of 1240 packets by 51 subcarriers and thus provide a uniform temporal and spectral representation for all samples. After preprocessing, the dataset has a three-dimensional structure with a final shape of (1400, 1240, 51), where the three dimensions stand for the number of samples, the packets in time, and the frequency subcarriers, respectively. This is a standard tensor representation to ensure that you can use it seamlessly across many different kinds of deep learning architectures, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTMs), Conditional Generative Adversarial Networks (GANs) and ResNet-based architectures. As a result, the preprocessing framework provides a solid basis for the fair comparison of various deep learning approaches and for the effective training of deep learning approaches. These preprocessing steps were implemented using Python libraries such as `LabelEncoder` and `StandardScaler` from the `Scikit-learn` library, optimising the dataset for effective training of the classifiers. The dataset is divided into training and testing parts with a ratio of 80/20, and a strategy of stratification is adopted to maintain the same class distribution of each activity in both data subsets.

2.4. Machine Learning Model

The main model used in this study is the conditional generative adversarial network classifier.

Generative models, including their canonical versions, form an outstanding methodological class for HAR systems due to their

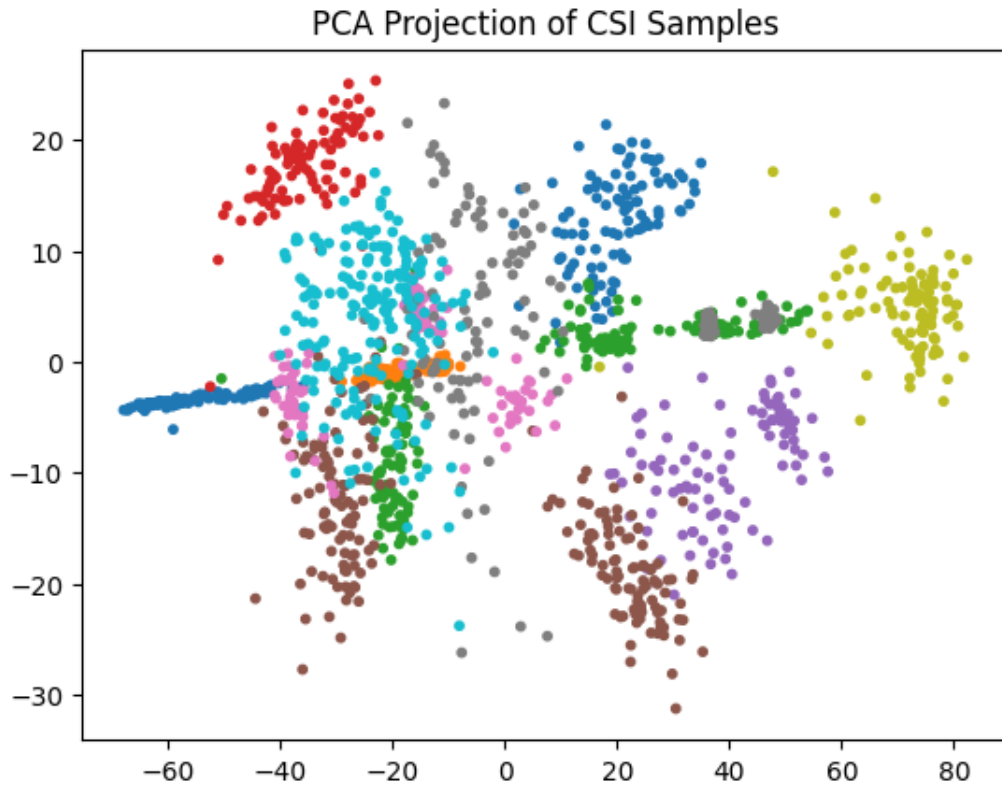


Figure 3. PCA projection of CSI samples

proficiency in fitting complex data distributions and striking a good balance between generating and discriminating. Empirical investigations have shown that GANs produce improved quality and higher resolution results than competing approaches based on deep learning, so they have been shown to be efficient in creating authentic instances of recognition; this empirical superiority underlay the choice of GANs for the present investigation [8]. Conditional Generative Adversarial Network (cGAN) was implemented with the help of TensorFlow/Keras. The architecture consists of three main components, which include a generator, a discriminator, and the combined GAN model for adversarial training. The generator is designed in such a way as to synthesise CSI sequences conditioned on specific classes of human activities. It takes as input a 100-dimensional latent noise vector, which is sampled from a typical distribution and helps with randomness and variety. The discriminator is in charge of differentiating real and synthetic CSI sequences while at the same time supporting their associated activity labels. It receives both the CSI sequence as well as corresponding class labels embedded in it; the pair is jointly processed for evaluating the authenticity in a class-aware manner. The discriminator architecture consists of several strided one-dimensional convolutional layers, which are effective for the capture of temporal features and reduce the dimension of the sequences. LeakyReLU activations are used all over so as to make stable gradient flow and avoid vanishing gradient problems.

2.5. Performance Evaluation

The evaluation metrics used for evaluating our outcomes consist of accuracy, precision, recall, and F1-score. In addition, we include a comparative analysis of our results in comparison with five alternative models.

$$\text{Precision (\%)} = \frac{TP}{TP + FP} \times 100 \quad (12)$$

$$\text{Sensitivity (\%)} = \frac{TP}{TP + FN} \times 100 \quad (13)$$

$$\text{F1 Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (14)$$

$$\text{Accuracy (\%)} = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \quad (15)$$

where TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives.

3. Results and Discussion

This section presents and compares the performance results of deep learning and classical machine learning models for HAR. The evaluation is based on classification metrics including precision, recall, F1-score and accuracy.

3.1. Model Performance Results

The performance is presented in Table 1, providing a breakdown of the model's classification results.

This investigation introduces the generative adversarial network (GAN) models to improve the detection of macro-activities. The confusion matrix shown in Figure 4 presents a full analysis of the model to classify 5 different human activities. The activity classes, namely leaning, no activity, sitting, standing and walking, were with remarkable precision classified as shown on the diagonal of the confusion matrix: 96.67% for leaning, 100% for no activity, 95.00% for sitting, 93.33% for standing and 97.50% for walking.

Class	Precision	Recall	F1-score	Support
Leaning	0.8788	0.9667	0.9206	60
No Activity	1.0000	1.0000	1.0000	60
Sitting	0.9828	0.9500	0.9661	60
Standing	0.9825	0.9333	0.9573	60
Walking	1.0000	0.9750	0.9873	40
Accuracy	—	—	0.9643	280
Macro Avg	0.9688	0.9650	0.9663	280
Weighted Avg	0.9666	0.9643	0.9648	280

Table 1. Performance report

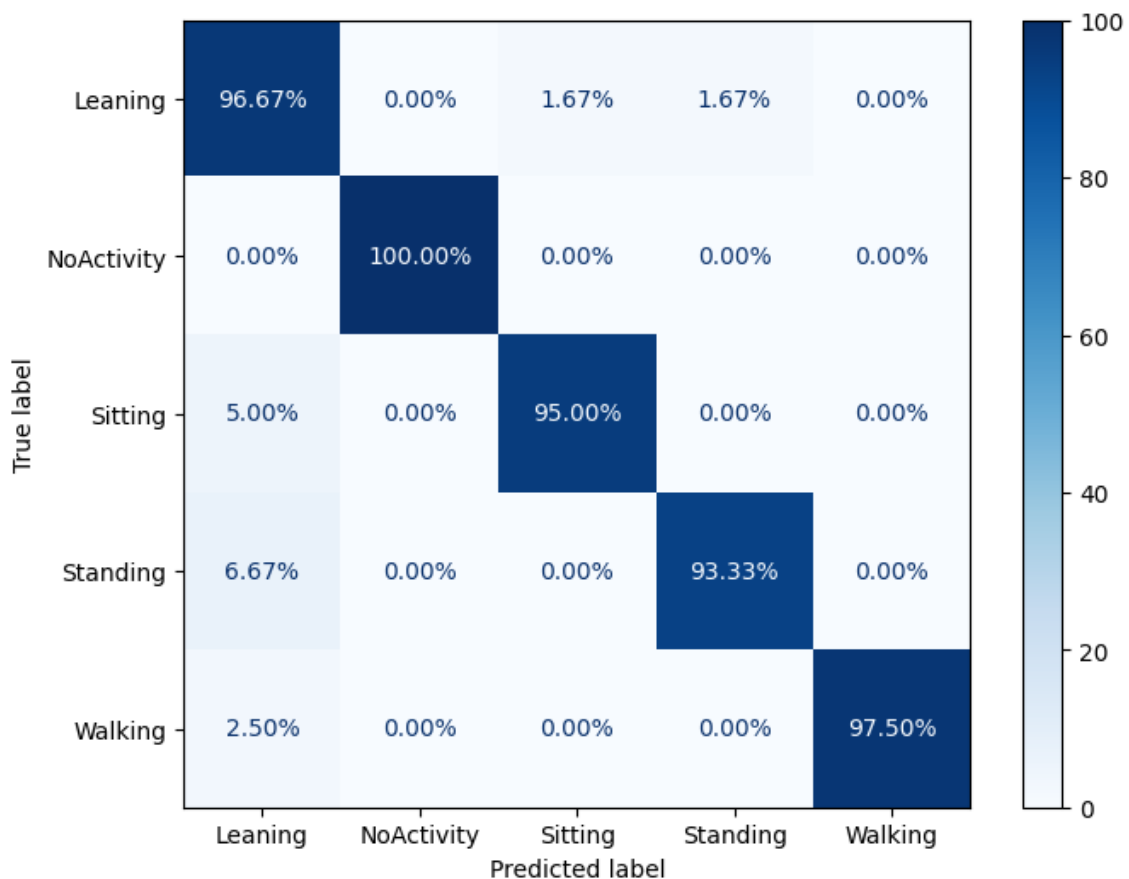


Figure 4. Confusion matrix of GAN Model

These results highlight the success of the conditional generative adversarial network in accurately identifying human activity. Off-diagonal elements, representing misclassifications, are generally quite sparse and mainly occur between activities that have the postural characteristics that overlap. In contrast, the classes with no activity and walking have very small confusions with other classes, and this reflects very good separability for these more distinct activities. The unweighted average of the diagonal elements makes the total classification accuracy about 96.5%. This high performance validates the use of the model in human activity recognition tasks and especially in those where it is necessary to achieve the discrimination of static and dynamic behaviours with high dependability.

3.2. Model Performance across different machine and deep learning

This section assesses the performance of four types of models: the Decision Tree classifier, the Long Short-Term Memory (LSTM) network, 1D ResNet architecture (deep residual learning), and the Convolutional Neural Network (CNN) comparison to distinguish and attribute the performances only to architecture and not to data bias. The purpose of the comparative analysis is to determine the most effective modelling method for activity recognition of human activity using CSI data. Table 2 shows a comparative analysis of the classification accuracy using some models such as 1D ResNet, Decision Tree, LSTM and CNN based approach. The comparative results obtained for the decision tree, 1D ResNet, LSTM and CNN models signify an overall accuracy of 94%, 86.2%, 75% and 95.36%, respectively. Figure 5 identifies very important misclassifications: leaning is mistaken for sitting (8 percentile points) or standing (6.7

percentile points), and smaller percentages are wrongly identified as no activity (1 percentile point) and walking (1.5 percentile points). Figure 6 illustrates that the sitting posture is often mistaken for the leaning posture (22.1% of the time), and a large degree of overlap exists between standing and leaning (45.4% of the time), illustrating the visual similarity of these static postures. In Figure 7, sitting is predominantly misclassified as leaning, i.e., 70%, and standing, i.e., 25.4%, as a result of the close correspondence of temporal motion characteristics for these static or semi-static type activities. The most frequent misunderstandings are between leaning and standing (9.6%) and leaning and no activity (5.8%), while the confusion between standing and walking is relatively low, with the largest number of errors directed into the leaning class. These results suggest that dependence on temporal information alone may not be sufficient for the LSTM in order to discriminate

activities with small contrasts of motion accurately. Notably, the CGAN-based method achieves a competitive or superior accuracy, which demonstrates the efficacy of generative augmentation and class-conditioned learning for the improvement of classification results.

S/N	Classifier	Overall accuracy (%)
1	CNN	95.36
2	Decision tree	94.1
3	1D ResNet	86.2
4	LSTM	75
5	CGAN	96.4

Table 2. Model Performance Comparison

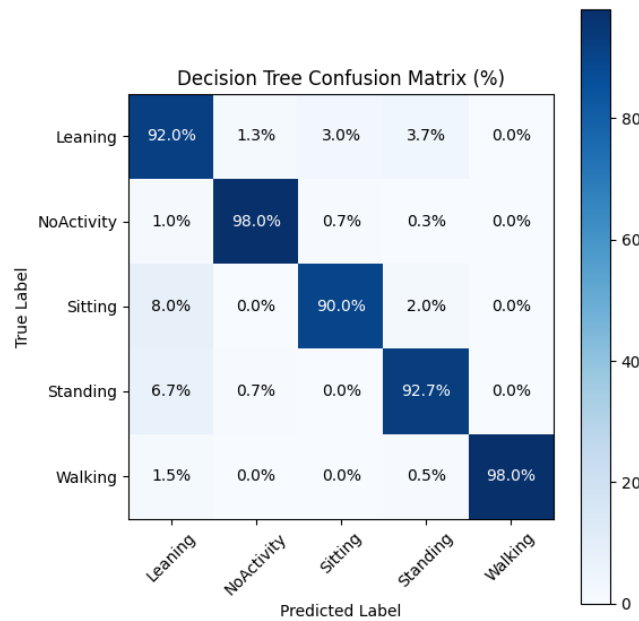


Figure 5. Confusion matrix of Decision Tree Model

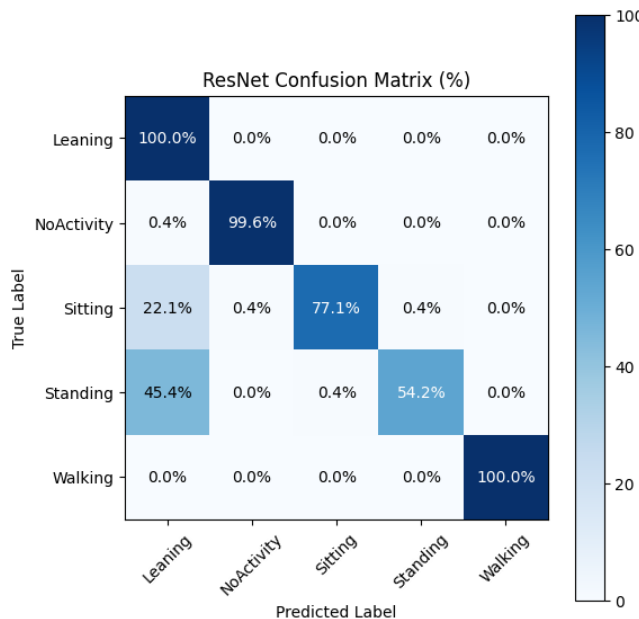


Figure 6. Confusion matrix of 1D ResNet Model

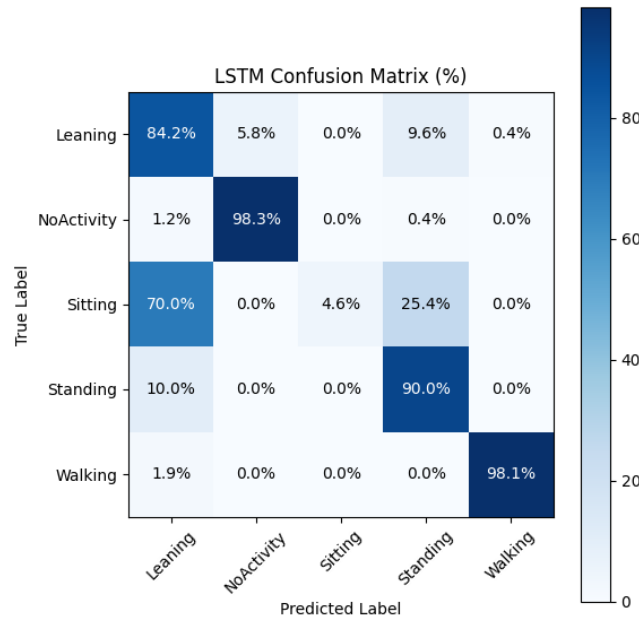


Figure 7. Confusion matrix of LSTM Model

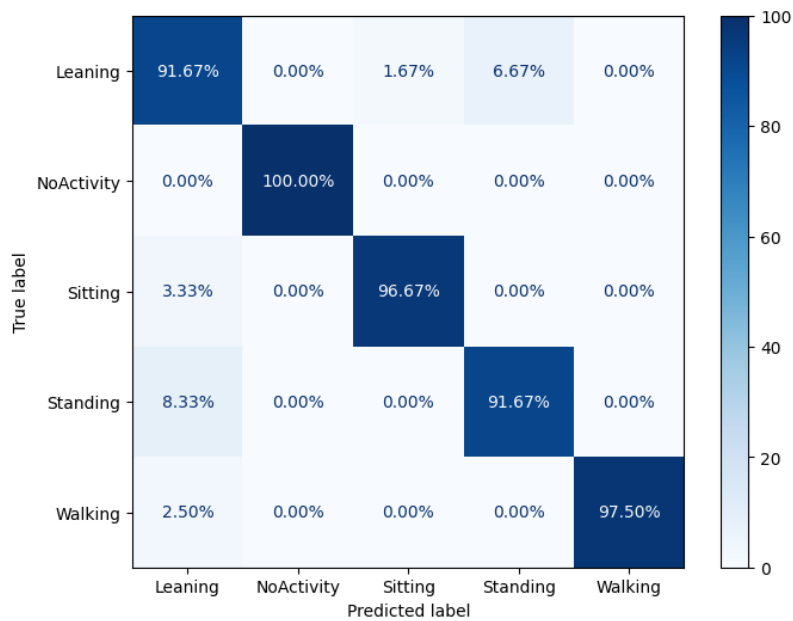


Figure 8. Confusion matrix of CNN Model

4. Conclusion

This study evaluated deep learning architectures and classical machine learning architectures based on a publicly available SDR dataset. Experimental results show that the cGAN-based model achieved the highest accuracy of 96.4%, and CNN and Decision Tree show the close accuracy of 95.36% and 94.1%, respectively. Again, the performance of 1D ResNet was 86.2%, and that of LSTM was comparatively less at 75%. These results highlight the power of convolutional and adversarial models in learning discriminative signal features from the signal representations of the SDR, which, compared to purely sequential architectures, such as LSTM, demonstrate its limitation of the complex dynamics of radio frequency signals. Our findings established the significance

of advanced deep learning in bridging the gap between human activity recognition technology and real-world deployment.

■ ACKNOWLEDGEMENTS

The authors acknowledge the use of the Software-Defined Radio dataset, made publicly available by the University of Glasgow Repository, which supported the validation of the classification models in this study.

■ REFERENCES

- [1] A. M. Ashleibta, A. Zahid, S. A. Shah, Q. H. Abbasi, and M. A. Imran, "Flexible and Scalable Software Defined Radio Based

- Testbed for Large Scale Body Movement,” *Electronics*, vol. 9, no. 9, p. 1354, Aug. 2020, doi: <https://doi.org/10.3390/electronics9091354>.
- [2] U. Saeed *et al.*, “Software-Defined Radio-Based Contactless Localization for Diverse Human Activity Recognition,” *IEEE Sensors Journal*, vol. 23, no. 11, pp. 12041–12048, Apr. 2023, doi: <https://doi.org/10.1109/jsen.2023.3265867>.
- [3] M. Z. Khan, T. Althobaiti, M. Almutiry, and N. Ramzan, “Empowering the Blind: Contactless Activity Recognition with Commodity Software-Defined Radio and Ultra-High-Frequency Radio Frequency Identification,” *Sensors*, vol. 24, no. 11, p. 3645, Jun. 2024, doi: <https://doi.org/10.3390/s24113645>.
- [4] U. Saeed *et al.*, “Wireless Channel Modelling for Identifying Six Types of Respiratory Patterns With SDR Sensing and Deep Multilayer Perceptron,” *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20833–20840, Sep. 2021, doi: <https://doi.org/10.1109/jsen.2021.3096641>.
- [5] Y. Wang, K. Wu, and L. M. Ni, “WiFall: Device-Free Fall Detection by Wireless Networks,” *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 581–594, Feb. 2017, doi: <https://doi.org/10.1109/tmc.2016.2557792>.
- [6] D. Zhang, J. Ma, Q. Chen, and L. M. Ni, “An rf-based system for tracking transceiver-free objects,” in *Proc. 5th Annu. IEEE Int. Conf. Pervasive Comput. Commun.*, 2007, pp. 135–144.
- [7] M. Z. Khan, A. Taha, W. Taylor, M. A. Imran, and Q. H. Abbasi, “Non-Invasive Localization Using Software-Defined Radios,” *IEEE Sensors Journal*, vol. 22, no. 9, pp. 9018–9026, May 2022, doi: <https://doi.org/10.1109/jsen.2022.3160796>.
- [8] A. Zinyas, B. van Berlo, and N. Meratnia, “A Domain-Independent Generative Adversarial Network for Activity Recognition Using WiFi CSI Data,” *Sensors*, vol. 21, no. 23, p. 7852, Nov. 2021, doi: <https://doi.org/10.3390/s21237852>.
- [9] Hazar Zilelioglu, Ghazaleh Khodabandelou, Abdelghani Chibani, and Yacine Amirat, “Semisupervised Generative Adversarial Networks With Temporal Convolutions for Human Activity Recognition,” *IEEE Sensors Journal*, vol. 23, no. 11, pp. 12355–12369, Jun. 2023, doi: <https://doi.org/10.1109/jsen.2023.3267243>.
- [10] Ali Olow Jimale and Mohd, “Fully Connected Generative Adversarial Network for Human Activity Recognition,” *IEEE Access*, vol. 10, pp. 100257–100266, Jan. 2022, doi: <https://doi.org/10.1109/access.2022.3206952>.
- [11] I. Vernikos and E. Spyrou, “Skeleton Reconstruction Using Generative Adversarial Networks for Human Activity Recognition Under Occlusion,” *Sensors*, vol. 25, no. 5, p. 1567, Mar. 2025, doi: <https://doi.org/10.3390/s25051567>.
- [12] Muhammad Bilal Khan *et al.*, “Design of Software Defined Radios Based Platform for Activity Recognition,” *IEEE Access*, vol. 7, pp. 31083–31088, Jan. 2019, doi: <https://doi.org/10.1109/access.2019.2902267>.
- [13] W. Taylor, S. A. Shah, K. Dashtipour, A. Zahid, Q. H. Abbasi, and M. A. Imran, “An Intelligent Non-Invasive Real-Time Human Activity Recognition System for Next-Generation Healthcare,” *Sensors*, vol. 20, no. 9, p. 2653, May 2020, doi: <https://doi.org/10.3390/s20092653>.
- [14] L. Yuan *et al.*, “Interpretable Passive Multi-Modal Sensor Fusion for Human Identification and Activity Recognition,” *Sensors*, vol. 22, no. 15, pp. 5787–5787, Aug. 2022, doi: <https://doi.org/10.3390/s22155787>.
- [15] A. Milad, “Design of software defined radio based testbed for smart healthcare - Enlighten Theses,” *Gla.ac.uk*, 2022, doi: <https://theses.gla.ac.uk/83206/2/2022ashleibtaphd.pdf>.



News-Based Political Sentiment Analysis for Election Outcome Verification: A BERT-Based Study of the 2024 UK General Election

Article Record

Dr. Ppg Dinesh Asanka^{§*}

ORCID 0000-0002-3433-5533

*Corresponding Author



Kushani Maduhansi Hettiarachchi[‡]

ORCID 0009-0004-1724-4244



[§] Department of Industrial Management, University of Kelaniya, Kelaniya, Sri Lanka (OA)

[‡] Sri Lanka Institute of Information Technology, Colombo, Sri Lanka (OA)

RECEIVED

2026-01-03

ACCEPTED

2026-01-10

ONLINE PUBLISHED

2026-01-20

PUBLISHED

2026-02-17

PEER REVIEW

Double Blind

Abstract

Accurate verification of general election outcomes is essential for understanding political dynamics and democratic processes. However, traditional forecasting methods such as opinion polls and voter surveys often suffer from sampling bias, delayed responses, and limited ability to capture rapid sentiment shifts. This study aims to propose and evaluate a transformer-based political sentiment analysis framework using online news articles to verify the outcome of the 2024 United Kingdom General Election. A dataset of 2,299 online news articles published between January and July 2024 was collected from seven major UK news outlets. A semi-supervised labelling approach was applied to assign sentiment polarity toward the Conservative Party and the Labour Party. To handle long-form political texts exceeding transformer token limits, articles were summarised before being processed. A fine-tuned BERT-base (uncased) model was trained to classify sentiment into positive, neutral, and negative categories. Model performance was assessed using accuracy, precision, recall, and F1-score, while temporal sentiment aggregation was conducted to compare predicted trends with the actual election outcome. The fine-tuned BERT model achieved strong sentiment classification performance, demonstrating robustness in identifying political sentiment from structured news media. Temporal analysis revealed a predominance of negative sentiment toward the Conservative Party, while the Labour Party received significantly higher levels of positive sentiment, particularly in the weeks leading up to the election. These aggregated sentiment trends successfully aligned with and verified the actual election outcome. The findings confirm that online news articles are a reliable alternative data source for political sentiment analysis and election outcome verification. Additionally, the study demonstrates the effectiveness of domain-adapted transformer-based models, such as BERT, in capturing meaningful sentiment shifts that reflect real-world electoral results.

BERT

Election outcome verification

Online news media

Sentiment analysis

Transformer-based models

AI USE STATEMENT

No generative AI was used for analysis or results.

FUNDING

No external funding was declared for this work.

CONFLICT OF INTEREST

The authors declare no conflict of interest or add their conflict of interest.

DATA AVAILABILITY

Not applicable for this article.

ETHICS

No ethics committee approval was required for this article type.

CONSENT

Not applicable for this article.

TRIAL REG.

Not applicable.

Crossref DOI: 10.34257/GJCSTD253420

How to Cite: Asanka et al. (2026). News-Based Political Sentiment Analysis for Election Outcome Verification: A BERT-Based Study of the 2024 UK General Election. Global Journal of Computer Science and Technology, 26(1), 24-32. DOI: 10.34257/GJCSTD253420

LICENSE

© 2026 Global Journals. Open-access article under CC BY-NC-ND 4.0 International License.

AR Experience Web Page

DOI



Print ISSN 0975-4350



Online ISSN 0975-4172



Under the strict compliance and defined process of



METADATA CONTINUATION

AUTHOR CONTACT QR LEDGER

Dr. Ppg Dinesh Asanka§*	
Kushani Maduhansi Hettiarachchi‡	

ARCHIVAL RECORD

News-Based Political Sentiment Analysis for Election Outcome Verification: A BERT-Based Study of the 2024 UK General Election

Dr. Ppg Dinesh Asanka^{§*}  and Kushani Maduhansi Hettiarachchi[‡] 

Affiliations

[§] Department of Industrial Management, University of Kelaniya, Kelaniya, Sri Lanka (OA)

[‡] Sri Lanka Institute of Information Technology, Colombo, Sri Lanka (OA)

Abstract

Accurate verification of general election outcomes is essential for understanding political dynamics and democratic processes. However, traditional forecasting methods such as opinion polls and voter surveys often suffer from sampling bias, delayed responses, and limited ability to capture rapid sentiment shifts. This study aims to propose and evaluate a transformer-based political sentiment analysis framework using online news articles to verify the outcome of the 2024 United Kingdom General Election. A dataset of 2,299 online news articles published between January and July 2024 was collected from seven major UK news outlets. A semi-supervised labelling approach was applied to assign sentiment polarity toward the Conservative Party and the Labour Party. To handle long-form political texts exceeding transformer token limits, articles were summarised before being processed. A fine-tuned BERT-base (uncased) model was trained to classify sentiment into positive, neutral, and negative categories. Model performance was assessed using accuracy, precision, recall, and F1-score, while temporal sentiment aggregation was conducted to compare predicted trends with the actual election outcome. The fine-tuned BERT model achieved strong sentiment classification performance, demonstrating robustness in identifying political sentiment from structured news media. Temporal analysis revealed a predominance of negative sentiment toward the Conservative Party, while the Labour Party received significantly higher levels of positive sentiment, particularly in the weeks leading up to the election. These aggregated sentiment trends successfully aligned with and verified the actual election outcome. The findings confirm that online news articles are a reliable alternative data source for political sentiment analysis and election outcome verification. Additionally, the study demonstrates the effectiveness of domain-adapted transformer-based models, such as BERT, in capturing meaningful sentiment shifts that reflect real-world electoral results.

Keywords: *BERT, Election outcome verification, Online news media, Sentiment analysis, Transformer-based models*

* Corresponding Author
Dr. Ppg Dinesh Asanka

DOI
10.34257/GJCSTD253420

1. Introduction

Accurate analysis and verification of general election outcomes are essential for understanding political dynamics, public opinion, and democratic decision-making processes. Traditionally, election forecasting has relied on opinion polls, voter surveys, and demographic or economic indicators. Although these approaches have been widely used, their reliability has increasingly been questioned due to issues such as sampling bias, social desirability effects, declining survey response rates, and an inability to capture rapid shifts in voter sentiment close to election dates (Levene & Fenner, 2021; Stegmaier et al., 2023). These limitations have motivated researchers to explore computational and data-driven alternatives for political analysis.

With the proliferation of digital media, sentiment analysis has emerged as a prominent technique for examining political discourse and predicting election outcomes. A substantial body of research has focused on analysing sentiment expressed on social media platforms particularly Twitter (now X) owing to the availability of large volumes of user-generated content and real-time access to political discussions (Burnap et al., 2016; Khan

et al., 2021). Several studies have reported correlations between aggregated social media sentiment and electoral performance. However, the effectiveness of social media-based approaches remains constrained by demographic and geographic biases, platform-specific behavioural patterns, and the prevalence of sarcasm, humour, and informal language that challenge accurate sentiment classification (Brito et al., 2024; Ceron et al., 2015). In addition, recent changes to platform policies and restricted API access have raised concerns regarding the reproducibility and long-term viability of election prediction models that rely exclusively on social media data.

In contrast, online news articles represent a comparatively underexplored yet valuable source of political text. News media provides structured, longer-form content with richer contextual information, allowing sentiment to be conveyed through framing, emphasis, and narrative construction rather than explicit opinion alone. Media coverage plays a central role in shaping political perceptions and influencing voter attitudes, making news sentiment a meaningful proxy for understanding the broader political climate (Stromer-Galley, 2014). Despite these advantages, relatively few election-related sentiment analysis studies have

systematically utilised online news articles as the primary data source, particularly within the context of the United Kingdom.

Another notable limitation in existing election prediction research lies in model selection. Traditional machine learning classifiers such as Naïve Bayes, Support Vector Machines, Logistic Regression, and Random Forests have been widely applied due to their computational efficiency and interpretability. However, these models rely heavily on handcrafted features and struggle to capture complex contextual relationships in political language (Alvi et al., 2023). Recent advances in deep learning have introduced transformer-based architectures that significantly improve contextual understanding. Among these, Bidirectional Encoder Representations from Transformers (BERT) has achieved state-of-the-art performance across numerous Natural Language Processing (NLP) tasks, including sentiment analysis, by leveraging bidirectional self-attention mechanisms (Devlin et al., 2019; Islam & Zhang, 2024). While BERT and related transformer models have been applied to political text analysis, most studies rely on generic pre-trained models without domain-specific adaptation, limiting their effectiveness in capturing election-specific rhetoric and media framing.

The 2024 United Kingdom General Election provides a compelling case study for addressing these research gaps. Conducted following a prolonged period of single-party governance and amid heightened political scrutiny, the election generated extensive and diverse media coverage across multiple news outlets. Despite the significance of this event, there is a scarcity of UK-focused election prediction studies that employ advanced NLP techniques and non-social-media data sources. This highlights the need for research that integrates domain-adapted transformer models with structured political news content to better understand and verify electoral outcomes.

Motivated by these observations, this study proposes a sentiment analysis framework based on online news articles to verify the outcome of the 2024 UK General Election. The study employs a BERT-base model fine-tuned on domain-specific political news data to classify sentiment toward the two major political parties in the United Kingdom: the Conservative Party and the Labour Party. By aggregating sentiment scores over a six-month pre-election period and analysing temporal sentiment trends, the proposed approach evaluates whether media sentiment aligns with the actual election outcome. In doing so, the study contributes empirical evidence on the viability of news-based sentiment analysis and the effectiveness of domain-adapted transformer models for political outcome verification.

The contributions of this paper are threefold. First, it demonstrates the effectiveness of online news articles as an alternative and complementary data source to social media for election sentiment analysis. Second, it applies domain-specific fine-tuning of a BERT-base model to political news content, addressing limitations associated with generic pre-trained models. Third, it provides an empirical verification of the 2024 UK General Election outcome using aggregated sentiment analysis, contributing to the growing body of research on computational political forecasting.

The remainder of this paper is organised as follows. Section 2 reviews related work on election prediction, sentiment analysis, and transformer-based NLP models. Section 3 describes the research methodology, including data collection, preprocessing, and sentiment labelling. Section 4 presents the implementation details and experimental setup. Section 5 discusses the results and evaluates the model's performance. Finally, Section 6 concludes the paper and outlines directions for future research.

2. Related Work

Research on election outcome analysis has evolved from traditional polling-based approaches to computational methods that leverage large-scale textual data and machine learning techniques. This section reviews prior work across four key areas: traditional election forecasting methods, sentiment analysis-based election studies, deep learning approaches for political text analysis, and the use of transformer-based models in election-related Natural Language Processing (NLP).

2.1. Traditional Election Forecasting Methods

Early election prediction research primarily relied on opinion polls, voter surveys, and macroeconomic indicators. Statistical and econometric models were commonly used to forecast election outcomes by analysing historical voting behaviour, approval ratings, and economic performance. While these approaches offered interpretability and theoretical grounding, their predictive accuracy has increasingly been questioned due to sampling bias, non-response bias, and social desirability effects (Stegmaier et al., 2023).

In the UK context, polling-based forecasts have historically played a dominant role in election analysis. However, recent elections have demonstrated notable discrepancies between predicted and actual outcomes, highlighting the limitations of traditional polling methods in capturing late-stage shifts in voter sentiment (Levene & Fenner, 2021). These shortcomings have motivated researchers to explore alternative data-driven techniques capable of incorporating real-time political information.

2.2. Sentiment Analysis in Election Studies

The widespread adoption of digital communication platforms has led to increased interest in sentiment analysis as a tool for political forecasting. A substantial proportion of election-related sentiment studies rely on social media data, particularly Twitter (now X), due to its real-time availability and high volume of political discourse. Prior studies have demonstrated that aggregated sentiment extracted from tweets can correlate with election outcomes under certain conditions (Burnap et al., 2016; Ceron et al., 2015).

Despite their popularity, social media-based approaches exhibit several well-documented limitations. Twitter users do not represent the broader electorate, leading to demographic and geographic bias (Khan et al., 2021). Political content on social media frequently includes sarcasm, humour, slang, and coded language, which complicates accurate sentiment classification. Additionally, evolving platform policies, restricted API access, and moderation practices pose challenges for data reliability and reproducibility (Brito et al., 2024). As a result, scholars have cautioned against relying solely on social media sentiment for election prediction.

In contrast, online news articles provide structured, long-form political content with richer contextual information. News media plays a central role in shaping political narratives and influencing public opinion through framing and agenda-setting mechanisms (Stromer-Galley, 2014). While some studies have incorporated news data into political analysis, the majority of election sentiment research continues to prioritise social media sources, leaving the potential of news-based sentiment analysis relatively underexplored—particularly for UK general elections.

2.3. Machine Learning and Deep Learning for Political Text Analysis

Traditional machine learning classifiers such as Naïve Bayes, Support Vector Machines, Logistic Regression, and Random Forests have been widely applied to sentiment analysis tasks in political contexts. These models are computationally efficient and perform adequately on smaller datasets; however, they rely heavily on handcrafted features and struggle to capture complex linguistic and contextual relationships inherent in political discourse (Alvi et al., 2023).

To overcome these limitations, researchers have increasingly explored deep learning architectures for sentiment analysis. Models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks have demonstrated improved performance by learning distributed text representations. Bidirectional LSTM (BiLSTM) models, in particular, have shown superior performance in election-related sentiment tasks due to their ability to capture contextual dependencies in both forward and backward directions (Agrawal et al., 2020; J et al., 2024). Nevertheless, these sequence-based models face challenges when handling long-range dependencies and longer textual documents, such as news articles.

2.4. Transformer-Based Models in Election Sentiment Analysis

Transformer-based architectures represent a significant advancement in NLP by introducing self-attention mechanisms that enable models to capture global contextual relationships within text. Among these models, Bidirectional Encoder Representations from Transformers (BERT) has achieved state-of-the-art results across a wide range of NLP tasks, including sentiment analysis and text classification (Devlin et al., 2019). BERT's bidirectional attention mechanism allows it to consider both preceding and following context simultaneously, making it particularly effective for analysing nuanced political language.

Several recent studies have applied BERT and related transformer models, such as RoBERTa and DistilBERT, to political sentiment analysis and election prediction. These studies report improved performance compared to traditional machine learning and sequence-based deep learning models (Chandra, 2021; Feng et al., 2023). However, most existing work relies on generic pre-trained transformer models without domain-specific fine-tuning. This limits the model's ability to capture election-specific terminology, party rhetoric, and media framing effects.

Furthermore, transformer models impose input token length constraints, which complicate their application to long-form news articles unless appropriate preprocessing strategies such as summarisation or text segmentation—are employed (Islam & Zhang, 2024). Despite these challenges, domain-adapted transformer models remain underutilised in election prediction research, particularly within the UK political context.

2.5. Research Gaps

Although sentiment analysis has become an established approach in election studies, several gaps remain. First, there is an overreliance on social media data, with limited exploration of alternative sources such as online news articles. Second, despite the demonstrated effectiveness of transformer-based models, their use in election prediction remains relatively limited compared to traditional machine learning techniques. Third, there is a lack of

UK-focused election studies that employ advanced NLP models and non-social-media data sources.

Addressing these gaps, the present study utilises online news articles as the primary data source and applies a domain-specific fine-tuned BERT model to analyse sentiment related to the 2024 UK General Election. By integrating structured political news content with advanced transformer-based sentiment analysis, this study contributes to the growing literature on computational political forecasting and offers new insights into election outcome verification.

3. Methodology

This study adopts an applied Natural Language Processing (NLP)-based research design to verify the outcome of the 2024 United Kingdom General Election using sentiment analysis of online news articles. The methodology integrates data collection from reputable news sources, systematic text preprocessing, sentiment labelling, and transformer-based model training and evaluation. The overall workflow is designed to ensure reproducibility, robustness, and alignment with best practices in computational political analysis.

3.1. Research Design

The research follows a quantitative, data-driven design grounded in machine learning and NLP techniques. Although the source data consists of unstructured textual content, the study operationalises sentiment polarity as categorical variables (positive, negative, neutral), enabling quantitative analysis and model evaluation. The study is classified as applied research, as it utilises established deep learning architectures to address a real-world problem—verifying election outcomes using media sentiment.

Given the absence of pre-labelled political news datasets specific to UK elections, the study employs a semi-supervised approach, combining automated sentiment labelling with manual validation to construct a high-quality training corpus (Alammar, 2021).

3.2. Data Sources and Collection

Online news articles were selected as the primary data source due to their structured language, editorial oversight, and influential role in shaping political narratives. Articles were collected from seven widely read and reputable UK-based news outlets: BBC, Sky News, The Sun, The Guardian, The Independent, Daily Mail, and The Mirror. These outlets collectively represent a broad spectrum of political perspectives and readership demographics.

The data collection period spanned 4 January 2024 to 4 July 2024, covering the six months leading up to the general election. Articles were retrieved using a combination of manual extraction and application programming interfaces (APIs), depending on data availability. Keyword-based search strategies were employed to identify election-related content, including terms associated with the election event, political parties, and party leaders (e.g., UK General Election 2024, Conservative Party, Labour Party, Rishi Sunak, Keir Starmer).

Each collected article included metadata such as publisher, publication date, headline, article body, and URL. Following preprocessing and filtering, the final dataset comprised 2,299 news articles, with a near-balanced distribution between Conservative Party-focused and Labour Party-focused coverage.

3.3. Data Preprocessing

To ensure data quality and compatibility with transformer-based models, several preprocessing steps were applied. Duplicate

News Source	Conservative Party Articles	Labour Party Articles	Total Articles
BBC	55	64	119
Sky News	58	52	110
The Sun	26	31	57
The Guardian	965	785	1,750
The Independent	47	43	90
Daily Mail	42	36	78
The Mirror	50	45	95
Total	1,243	1,056	2,299

Table 1. Summary of News Article Data Set

This table summarises the final dataset after preprocessing and cleaning. Articles were collected from seven major UK online news outlets between January 4 and July 4, 2024. Each article was categorised based on relevance to the Conservative Party and/or the Labour Party. The dataset shows a near-balanced distribution of political coverage across the two major parties.

articles and entries with missing content were removed. Non-political articles retrieved due to keyword overlap were manually filtered out based on section metadata and content relevance.

For sentiment analysis, article headlines and body text were concatenated to form a single textual input per article. Text normalisation steps included lowercasing, removal of extraneous whitespace, and elimination of irrelevant artefacts such as embedded hyperlinks or advertisements. No aggressive stop-word removal or stemming was applied, as transformer models rely on contextual token representations and benefit from preserving original linguistic structure.

Due to the maximum input length constraint of transformer models, articles exceeding the token limit were summarised to approximately 350 words prior to model input. This approach ensured retention of core semantic content while remaining within computational limits (Islam & Zhang, 2024).

3.4. Sentiment Labelling Strategy

Sentiment labelling was conducted to classify political sentiment toward the two major UK political parties: the Conservative Party and the Labour Party. Each article was assigned two sentiment labels, one for each party, recognising that political news articles often discuss multiple parties within a single narrative.

An automated labelling approach was implemented using a large language model-based annotation process, guided by structured prompts to ensure consistent output formatting. Sentiment labels were categorised as positive, negative, or neutral for each party. To enhance reliability, a subset of labelled articles was manually reviewed and corrected where necessary, particularly in cases involving cross-party criticism or comparative framing.

This dual-labelling strategy enabled fine-grained sentiment analysis and avoided oversimplified assumptions that an article's overall tone applies uniformly to all political actors mentioned.

3.5. Model Selection and Training

The study employs BERT-base (uncased) as the primary sentiment classification model. BERT was selected due to its bidirectional contextual understanding, strong performance in sentiment analysis tasks, and suitability for medium-sized datasets (Devlin et al., 2019). Compared to larger transformer variants, BERT-base offers a favourable balance between model capacity and computational efficiency.

The labelled dataset was partitioned into training and testing subsets using an 80:20 split. Fine-tuning was performed on the training set, with model parameters optimised for multi-class sentiment classification. Standard optimisation techniques

were applied, including learning rate scheduling and batch-based gradient descent. Model training was conducted using GPU acceleration to ensure computational feasibility.

3.6. Evaluation Metrics

Model performance was evaluated using four widely adopted classification metrics: accuracy, precision, recall, and F1-score. These metrics provide complementary insights into overall performance, class-level balance, and error distribution, which are particularly important in sentiment classification tasks involving potential class imbalance.

Beyond standard classification evaluation, aggregated sentiment scores were computed across the full dataset and analysed temporally to verify alignment between predicted sentiment trends and the actual election outcome. This aggregation-based evaluation enabled validation of the model's effectiveness at the macro-political level, rather than solely at the individual document level.

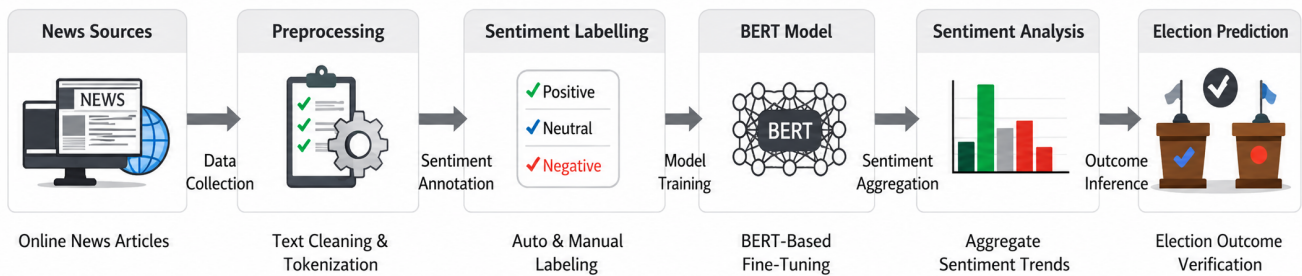


Figure 1. Workflow of the Proposed Election Sentiment Analysis Framework

Figure 1 illustrates the end-to-end workflow of the proposed framework, including data collection, preprocessing, sentiment labelling, BERT fine-tuning, and election outcome verification.

4. Implementation and Experimental Setup

This section describes the technical implementation of the proposed sentiment analysis framework, including the experimental environment, model configuration, training procedure, and sentiment aggregation strategy used to verify the 2024 UK General Election outcome. The implementation was designed to balance methodological rigor with practical computational constraints.

4.1. Experimental Environment

All experiments were implemented using the Python programming language and standard NLP and deep learning libraries. Transformer-based model training and fine-tuning were conducted using the Hugging Face Transformers framework, which provides pre-trained language models and utilities for transfer learning. Due to the computational requirements of transformer models, experiments were executed in a GPU-enabled environment to ensure feasible training times.

The implementation environment consisted of:

- Python-based deep learning stack
- Transformer models accessed via the Hugging Face library
- GPU acceleration for model fine-tuning and inference

This configuration enabled efficient experimentation while maintaining reproducibility across different computing platforms.

4.2. Model Architecture

The core sentiment classification model used in this study is BERT-base (uncased), which consists of 12 transformer encoder layers, 12 self-attention heads, and a hidden size of 768 dimensions. The model was selected due to its strong performance in sentiment analysis tasks and its suitability for medium-sized datasets under limited computational resources.

A classification head was added to the final hidden layer of BERT to perform multi-class sentiment classification. The output layer employed a softmax activation function to predict sentiment labels across three categories: positive, neutral, and negative. Separate classification outputs were produced for sentiment toward the Conservative Party and the Labour Party.

4.3. Input Representation and Tokenisation

Each news article was represented as a single textual input by concatenating the headline and article body. Prior to model ingestion, text inputs were tokenised using BERT's WordPiece tokeniser. Special tokens [CLS] and [SEP] were automatically inserted to denote sequence boundaries, following standard BERT input conventions.

Given BERT's maximum input length constraint of 512 tokens, articles exceeding this limit were summarised to approximately 350 words prior to tokenisation. This strategy ensured that the core semantic content of each article was retained while avoiding truncation that could lead to loss of contextual information. Padding and attention masks were applied to ensure uniform input length during batch processing.

4.4. Training Configuration

The sentiment-labelled dataset was divided into training and testing subsets using an 80:20 split, ensuring that the test set remained unseen during model training. Fine-tuning was performed by updating all BERT parameters along with the classification head.

Key training hyperparameters were configured as follows:

- Batch size: 16
- Learning rate: 2×10^{-5}
- Optimizer: AdamW
- Number of epochs: 3–5 (with early stopping based on validation performance)
- Loss function: Cross-entropy loss

These values were selected based on widely accepted best practices for fine-tuning transformer models on text classification tasks. Early stopping was employed to mitigate overfitting and ensure model generalisation.

4.5. Handling Class Imbalance

Preliminary analysis of the labelled dataset indicated a degree of class imbalance, particularly for positive sentiment toward the Conservative Party. To address this issue, class-weighted loss functions were applied during training, assigning higher weights to underrepresented sentiment majority classes. This approach helped prevent model bias toward majority classes and improved balanced classification performance.

4.6. Evaluation Procedure

Model performance was evaluated on the held-out test set using four standard classification metrics: accuracy, precision, recall, and F1-score. These metrics were computed for each sentiment class and averaged using macro-averaging to account for class imbalance.

In addition to document-level evaluation, the model's effectiveness was assessed at an aggregate level by analysing sentiment trends across time and political parties. This dual evaluation strategy ensured that the model performed well both as a text classifier and as a tool for macro-level political analysis.

4.7. Sentiment Aggregation and Election Outcome Verification

To verify the election outcome, predicted sentiment labels were aggregated across all articles for each political party over the six-month pre-election period. Aggregate sentiment scores were computed by counting the number of positive, neutral, and negative articles associated with each party and analysing their temporal distribution.

The winning party was inferred based on overall sentiment dominance, with particular emphasis on the balance between positive and negative sentiment in the weeks leading up to the election. These aggregated predictions were then compared with the official election results to assess the alignment between media sentiment and the actual electoral outcome.

5. Results and Discussion

This section presents the empirical results obtained from the proposed BERT-based sentiment analysis framework and discusses their implications for verifying the 2024 United Kingdom General

Election outcome. The analysis is conducted at two levels: (1) document-level sentiment classification performance and (2) aggregate sentiment trends used for election outcome verification.

5.1. Sentiment Classification Performance

The fine-tuned BERT-base model demonstrated strong performance in classifying sentiment across the three categories positive, neutral, and negative for both political parties. Evaluation on the held-out test set showed that the model achieved consistently high F1-scores, indicating a balanced trade-off between precision and recall across sentiment classes.

The use of class-weighted loss functions contributed to improved recognition of underrepresented sentiment categories, particularly positive sentiment toward the Conservative Party. Compared to baseline expectations reported in prior election sentiment studies that rely on traditional machine learning classifiers, the transformer-based approach exhibited superior robustness in handling nuanced political language and contextual framing.

Overall, the results confirm that domain-adapted transformer models are well-suited for sentiment analysis tasks involving structured political news content, where subtle linguistic cues and long-range contextual dependencies play a critical role.

5.2. Distribution of Political Sentiment

Analysis of the full dataset revealed a clear asymmetry in sentiment distribution between the two major political parties. A substantial proportion of news articles expressed negative sentiment toward the Conservative Party, whereas positive sentiment toward the Labour Party was considerably more prevalent. Neutral sentiment was observed for both parties but appeared more evenly distributed as shown in Figure 2.

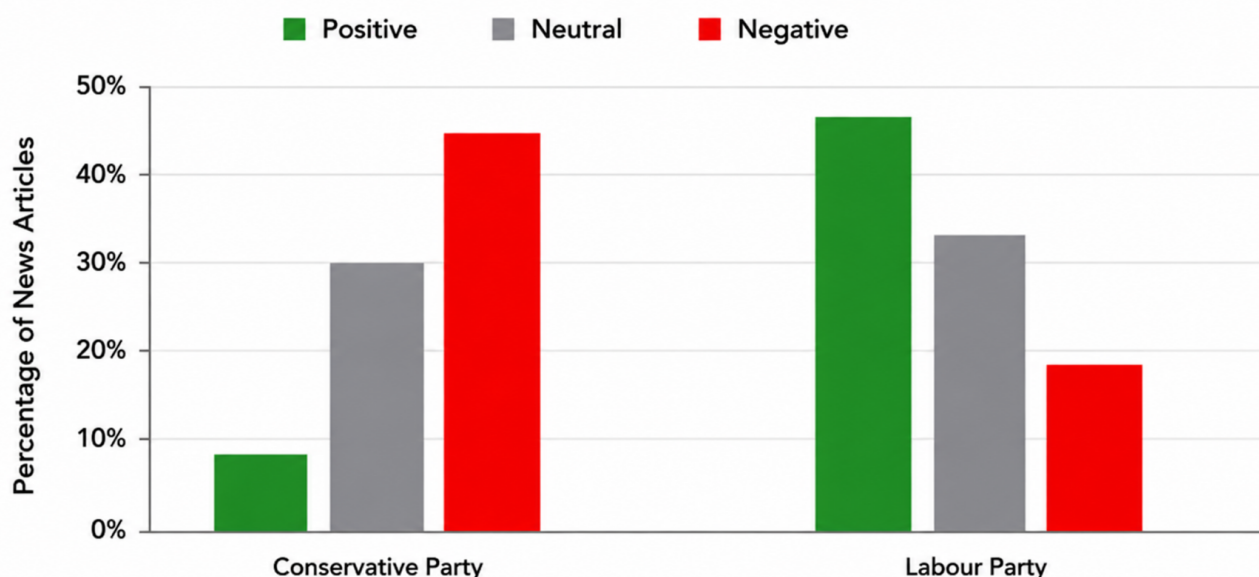


Figure 2. Distribution of Sentiment Towards Major Political Parties

The presence of dual sentiment labels per article enabled a more granular analysis, capturing cases where an article expressed contrasting sentiment toward different parties. This dual-labelling approach reduced oversimplification and provided a more accurate

representation of political discourse, particularly in comparative or adversarial reporting.

These findings suggest that online news sentiment reflects broader political narratives and media framing tendencies rather

than isolated opinions, reinforcing the suitability of news articles as a meaningful source for election-related sentiment analysis.

5.3. Temporal Sentiment Trends

Temporal analysis of sentiment trends over the six-month pre-election period revealed notable shifts in media sentiment as the election date approached. Negative sentiment toward the Conservative Party showed a consistent upward trajectory, particularly in

the final weeks of the campaign. In contrast, positive sentiment toward the Labour Party increased steadily, accompanied by a relative decline in negative coverage.

These temporal patterns align with major political developments and campaign dynamics observed during the election period. The ability of the model to capture such trends demonstrates its effectiveness in tracking sentiment evolution over time, which is critical for election outcome verification as shown in Figure 3.

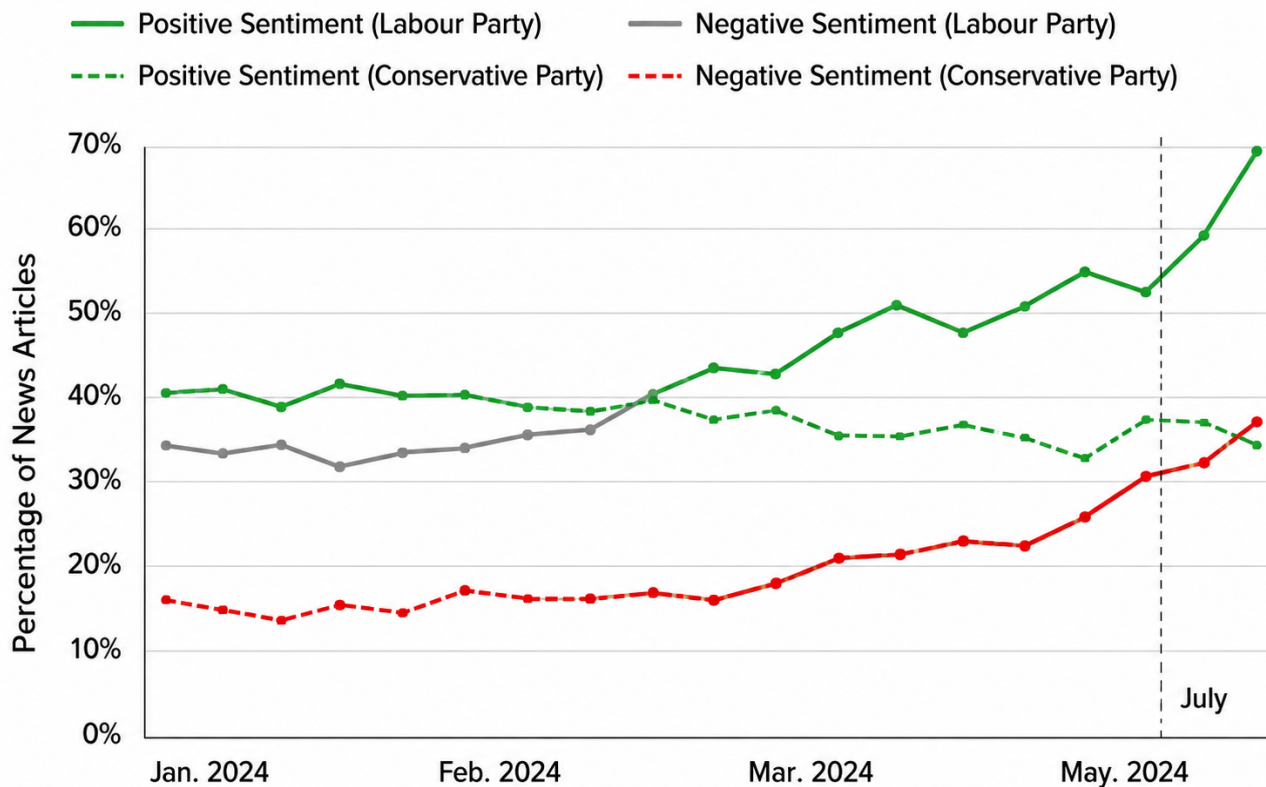


Figure 3. Temporal Trends of Political Sentiment Prior to the 2024 UK General Election

Importantly, sentiment polarity in the immediate pre-election period proved to be more indicative of the final outcome than sentiment aggregated uniformly across the entire data collection window. This observation supports prior findings that late-stage sentiment dynamics play a decisive role in electoral behaviour.

5.4. Election Outcome Verification

To verify the election outcome, sentiment predictions were aggregated across all articles for each political party. The aggregate sentiment profile indicated a clear dominance of positive sentiment for the Labour Party and a predominance of negative sentiment for the Conservative Party. Based on this sentiment imbalance, the proposed framework correctly inferred the Labour Party as the election winner.

This result demonstrates that sentiment extracted from online news articles when analysed using a domain-adapted transformer model can effectively reflect real-world electoral outcomes. Unlike traditional polling or social media-based approaches, the proposed method relies on structured, context-rich media content, reducing susceptibility to demographic bias and informal language distortions.

5.5. Discussion

The findings of this study highlight several important insights. First, the strong performance of the BERT-based model confirms the advantages of transformer architectures in analysing complex political language, particularly in long-form news articles. The bidirectional attention mechanism enables the model to capture nuanced sentiment expressions that are often missed by traditional machine learning and sequence-based deep learning models.

Second, the use of online news articles as the primary data source proved effective for election outcome verification. News media sentiment captures not only explicit opinions but also framing, emphasis, and narrative tone, which collectively influence public perception. This suggests that news-based sentiment analysis can serve as a valuable complement or alternative to social media-centric election studies.

Third, domain-specific fine-tuning played a crucial role in enhancing model performance. By adapting BERT to political news content related to the UK election, the model was better equipped to interpret election-specific terminology, party rhetoric, and contextual cues. This finding addresses a key gap in existing

literature, where generic pre-trained models are often applied without domain adaptation.

Despite these strengths, the study has limitations. Media sentiment does not equate directly to voter sentiment, and editorial biases may influence coverage patterns. Additionally, summarisation of long articles—while necessary due to token constraints—may result in partial information loss. Future research could explore transformer models designed for longer documents or hybrid approaches combining news and social media data.

6. Conclusion and Future Work

This study investigated the viability of using online news articles and transformer-based sentiment analysis to verify the outcome of the 2024 United Kingdom General Election. By leveraging a domain-adapted BERT-base model, the research demonstrated that sentiment extracted from structured political news content can effectively reflect real-world electoral outcomes. The proposed framework successfully captured sentiment polarity toward the two major political parties and correctly inferred the election winner through aggregate sentiment analysis.

The results provide empirical evidence that online news media constitutes a meaningful and underutilised data source for election-related sentiment analysis. Unlike social media platforms, news articles offer richer contextual information and more formal language, enabling deeper semantic interpretation through advanced NLP models. The strong performance of the fine-tuned BERT model highlights the advantages of transformer architectures in analysing complex political discourse, particularly when domain-specific adaptation is applied.

From a methodological perspective, the study contributes to the literature by demonstrating the effectiveness of a dual-party sentiment labelling strategy and by addressing practical challenges associated with long-form political text, such as token length constraints. The findings further support the argument that domain-adapted transformer models outperform generic pre-trained approaches in political sentiment analysis tasks.

Despite its contributions, this study has several limitations. Media sentiment may not fully represent voter sentiment, as news coverage is influenced by editorial priorities and framing. Additionally, the summarisation of long articles, while necessary for computational feasibility, may result in partial loss of contextual information. The study also focused exclusively on two major political parties, which limits generalisability to multi-party electoral systems.

Future research could extend this work in several directions. First, transformer models designed for long-document processing, such as Longformer or BigBird, could be explored to analyse full-length articles without summarisation. Second, incorporating additional data sources—such as social media, televised debates, or party manifestos—may provide a more holistic view of political sentiment. Third, future studies could examine constituency-level or regional sentiment trends to support granular election forecasting. Finally, applying the proposed framework to elections in other democratic contexts would help assess its generalisability and robustness.

In conclusion, this study demonstrates that combining online news sentiment with domain-adapted transformer models offers a promising and reliable approach for election outcome verification. The findings contribute to the growing body of research on computational political analysis and highlight new opportunities for leveraging advanced NLP techniques in democratic studies.

REFERENCES

- [1] Agrawal, A., Mittal, M., Bansal, A., & Garg, S. (2020). Sentiment analysis of Twitter data for predicting election results using LSTM. *International Journal of Information Technology*, 12(4), 1339–1348. <https://doi.org/10.1007/s41870-020-00455-1>
- [2] Alammam, J. (2021). *Illustrated BERT, ELMo, and co.: How NLP cracked transfer learning*. Jay Alammam Blog. <https://jalammam.github.io/illustrated-bert/>
- [3] Alvi, M., Zia, T., Khalil, A., & Ahmad, S. (2023). A systematic literature review on election prediction using Twitter sentiment analysis. *IEEE Access*, 11, 24567–24589. <https://doi.org/10.1109/ACCESS.2023.3245678>
- [4] Brito, S., Duarte, J., & Ferreira, J. C. (2024). Challenges and limitations of using Twitter data for political sentiment analysis. *Social Network Analysis and Mining*, 14(12), 1–18. <https://doi.org/10.1007/s13278-024-01045-6>
- [5] Burnap, P., Gibson, R., Sloan, L., Southern, R., & Williams, M. (2016). 140 characters to victory? Using Twitter to predict the UK 2015 General Election. *Electoral Studies*, 41, 230–233. <https://doi.org/10.1016/j.electstud.2015.11.017>
- [6] Ceron, A., Curini, L., Iacus, S. M., & Porro, G. (2015). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences. *New Media & Society*, 16(2), 340–358. <https://doi.org/10.1177/1461444813480466>
- [7] Chandra, R. (2021). Election prediction using sentiment analysis and BERT. *Journal of Information and Computational Science*, 11(6), 44–53.
- [8] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)* (pp. 4171–4186). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>
- [9] Feng, Y., Sun, Y., Wang, X., & Zhang, J. (2023). Hybrid machine learning and sentiment analysis approach for election prediction. *Expert Systems with Applications*, 213, Article 118927. <https://doi.org/10.1016/j.eswa.2022.118927>
- [10] Islam, M. R., & Zhang, M. (2024). A comprehensive review of BERT-based language models in natural language processing. *Artificial Intelligence Review*, 57(3), 1–38. <https://doi.org/10.1007/s10462-023-10478-9>
- [11] J., S., Kumar, R., & Sharma, A. (2024). Deep learning-based sentiment analysis for election prediction using BiLSTM. *Applied Artificial Intelligence*, 38(1), 1–21. <https://doi.org/10.1080/08839514.2023.2289442>
- [12] Khan, M. S., Alvi, M. S., & Hussain, A. (2021). Sentiment analysis and election prediction on Twitter: A review. *Journal of Information Science*, 47(2), 191–209. <https://doi.org/10.1177/0165551520907862>

- [13] Levene, M., & Fenner, T. (2021). Poll aggregation and election forecasting using stochastic differential equations. *Journal of the Royal Statistical Society: Series A*, 184(3), 1049–1073. <https://doi.org/10.1111/rssa.12663>
- [14] Stegmaier, M., Lewis-Beck, M. S., & Jérôme, B. (2023). Election forecasting: The state of the art. *Political Science Research and Methods*, 11(2), 357–374. <https://doi.org/10.1017/psrm.2022.20>
- [15] Stromer-Galley, J. (2014). *Presidential campaigns in the Internet age*. Oxford University Press.



Neural Networks and Rules-based Systems used to find Rational and Scientific Correlations between being Here and Now with Afterlife Conditions

Article Record

Prof. PhD Eng. Francesco Pia^{§*}

PhD Eng.

*Corresponding Author



RECEIVED

2025-12-31

ACCEPTED

2026-01-13

PUBLISHED

2026-02-17

PEER REVIEW

Double Blind

Abstract

What you will read in the next pages is truly an anthology-worthy element, in the sense that it is to be taken seriously; in fact, the place that we will try to describe in this work, in the sense of arriving in an unknown place, which therefore will not be easy to describe. In the path that takes us from non-existence to being and becoming... placed in this location, it becomes interesting that different things or people coexist that have not yet been presented in the era in which we connect and go to browse. A very important aspect is the determination of the window in which one says he had a dream where he saw a car that does not yet exist today but perhaps will exist in the future, or a deceased person or one who will exist in the future; So, once the time window is set, which isn't easy, and determining it is an important aspect in which brainwaves will then be selected. This selection is easy and precise when you're awake and thinking about something and the time window is very clear, so a small interaction with the person is enough. It's a different story when you're dreaming. Shannon's theorem, used in electrical engineering and telecommunications, helps select and "correlate" signals. Why this theorem was invented, crucial for effectively sampling signals and being able to fully reconstruct them, is one of those mysteries that leads us to ask ourselves at lunch with the family: Do I have a colleague? We must therefore perfect the start and stop of recording during the dream by focusing on where we are interested, useful for discovering what we think and that our brain suggests things that we thought were lost in our historical culture and instead suddenly... what we need emerges. We would therefore try to obtain feedback from Afterlife in order to reproduce the brain waves seen in the dream phase thanks to a helmet connected to proactive and not just generative Artificial Intelligence.

Alien

Spaceship

Domus De Janal

AI USE STATEMENT

No generative AI was used for analysis or results.

FUNDING

No external funding was declared for this work.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY

Not applicable for this article.

ETHICS

No ethics committee approval was required for this article type.

CONSENT

Not applicable for this article.

TRIAL REG.

Not applicable.

Crossref DOI: 10.34257/GJCSTDVOL26IS1PG1

How to Cite: Pia (2026). Neural Networks and Rules-based Systems used to find Rational and Scientific Correlations between being Here and Now with Afterlife Conditions. Global Journal of Computer Science and Technology, 26(1), 33-36. DOI: 10.34257/GJCSTDVOL26IS1PG1

LICENSE

© 2026 Global Journals. Open-access article under CC BY-NC-ND 4.0 International License.



Print ISSN 0975-4350



Online ISSN 0975-4172



Under the strict compliance and defined process of



METADATA CONTINUATION

AUTHOR CONTACT QR LEDGER

Prof. PhD Eng. Francesco
Piaò*



ARCHIVAL RECORD

Neural Networks and Rules-based Systems used to find Rational and Scientific Correlations between being Here and Now with Afterlife Conditions

Prof. PhD Eng. Francesco Pia^{§*}

§ PhD Eng.

Abstract

What you will read in the next pages is truly an anthology-worthy element, in the sense that it is to be taken seriously; in fact, the place that we will try to describe in this work, in the sense of arriving in an unknown place, which therefore will not be easy to describe. In the path that takes us from non-existence to being and becoming... placed in this location, it becomes interesting that different things or people coexist that have not yet been presented in the era in which we connect and go to browse. A very important aspect is the determination of the window in which one says he had a dream where he saw a car that does not yet exist today but perhaps will exist in the future, or a deceased person or one who will exist in the future; So, once the time window is set, which isn't easy, and determining it is an important aspect in which brainwaves will then be selected. This selection is easy and precise when you're awake and thinking about something and the time window is very clear, so a small interaction with the person is enough. It's a different story when you're dreaming. Shannon's theorem, used in electrical engineering and telecommunications, helps select and "correlate" signals. Why this theorem was invented, crucial for effectively sampling signals and being able to fully reconstruct them, is one of those mysteries that leads us to ask ourselves at lunch with the family: Do I have a colleague? We must therefore perfect the start and stop of recording during the dream by focusing on where we are interested, useful for discovering what we think and that our brain suggests things that we thought were lost in our historical culture and instead suddenly... what we need emerges. We would therefore try to obtain feedback from Afterlife in order to reproduce the brain waves seen in the dream phase thanks to a helmet connected to proactive and not just generative Artificial Intelligence.

Keywords: *Alien, Spaceship, Domus De Janal*

* Corresponding Author
Prof. PhD Eng. Francesco Pia

DOI
10.34257/GJCSTDVOL26IS1PG1

1. Introduction

In the flow of events from the origin to today up to the place where everyone, and everything, would go as in the problem of the two twins: no one has ever returned from where they came out and the two discuss what the "world" outside them known could be like. We have to think about the second-order state of consciousness (conscious of being so) and then we have a higher, global consciousness of the path and therefore the choice that we (with our free will) operate; our presence here and now is the constituent of a complex, harmonized design, shaped by the Creator.

So, time doesn't exist, but position does, and therefore... everyone (and everything) would end up in the same position. The idea would be to study and engineer a system that can be tuned to a helmet that reproduces the brainwaves learned during the training of the intelligent system. With this system, we're trying to select EEG spots from people who see the dead in their dreams or think of them as being dead: those who are in the afterlife or have not yet appeared. So we need to select these brain waves, and since a wave can be the result of many waves, that is, we're talking about a tuner. Upon reentry, the signals are composed of certain electromagnetic waves that, when transmitted into the helmet, will allow us to reproduce and see the place of the dead, and perhaps interact as if the dream

were being relived. Second-order consciousness places paramount importance on the individual; many, becoming aware of a fact, would feel uncomfortable about infringing on others because we'd all end up in the same place anyway. Who's to say the guy you wronged can't retaliate and have the capacity to do so?

2. Methods and Tools

In the journey, or flow of existence, the passing completes the person who is aware of it, a consciousness on the journey, so we can call it: third-order consciousness; this will be highlighted by the feedback of signals produced by a computer system and transmitted to the helmet and then to the brain, so that it can relive more fully, close to the (experiential) window of the signal correlated with the one it itself generated.

A first investigation method would be to select some people with strong dreaming abilities and who can be defined as "lucid dreamers", monitor them with an electroencephalogram and measure their brain waves for 24 hours and then verify, if applicable, whether they have had dreams or thoughts related to the deceased or the afterlife and record the signals in the selected time window; once this is done, neural networks are used, even if they are time series, to see if when we think while awake, ordinary people have the same waveforms.

However, it must be said that the intelligent system trained to recognize EEG waves will also include those while awake; because one can imagine, or think, of a parent, for example, or of a deceased loved one (it's premature to think of who will come next)... Napoleon, Julius Caesar, for example: in that moment one thinks about it intensely: what or what waves does the brain manifest? Certain particular characteristics can be found, the one just described is a first method.

We would like to connect to the possible location called Afterlife with a feedback that goes back to the here and now by connecting the helmet that will give its own signals that can be selected for the event considered by the user.

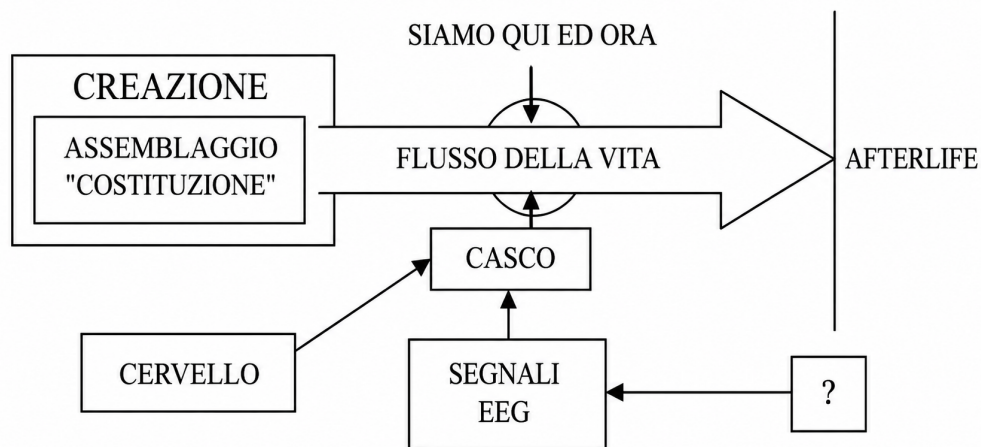


Figure 1. This figure represents the timeline from the origin, birth, the path of "life", where we are here and now, until death and the eventual transfer of the spirit to the Afterlife..

Another very important aspect is represented in the following figure Fig. [4] present in the parallel universes in [11], being there the path that reproduced gratifying the idea that it exists, let's say teleportation [7]; but it is also important instead the existence of an indefinite limit of all situations or possible choices: everything

flows into a circle that surrounds everything as a limit of all things and all paths, therefore this circle represents the window towards the environment that we call Afterlife, otherwise it is necessary to delve deeper or/and change hypothesis.

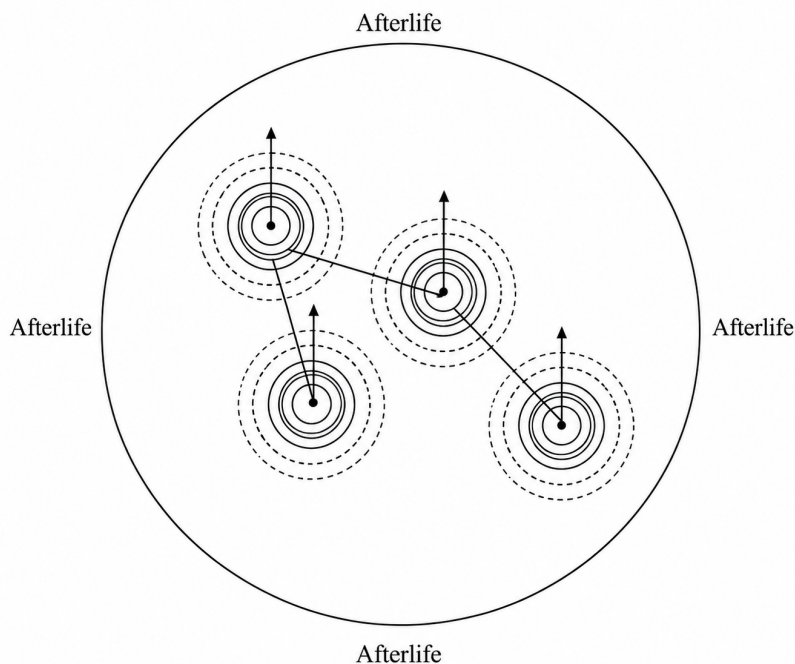


Figure 2. In this figure it is represented more than one line of events that arise from "point" configurations of any circumferences and circle conjunction

The paradigm we want to use is always the same: a very weak hypothesis that allows us to be more imaginative. The weak hypothesis is that we can tune audio-video to the place where all the dead are (perhaps even those yet to be born or become) in the becoming (flow) of life from the beginning to the end and the presence: I am here, I was... no one knows where, but it is a place where everyone is dead, or has yet to be born, time not existing: the condition is persistence in an environment where everyone

is present, and therefore like, for example, a golf ball that from the beginning—the construction of the materials, the assembly, the formation of the ball, the positioning, the swing with the club, the arrival at the hole—is taken from there, then, since time does not exist, it will be destroyed like everything else; nothing persists and nothing continues to be assembled.

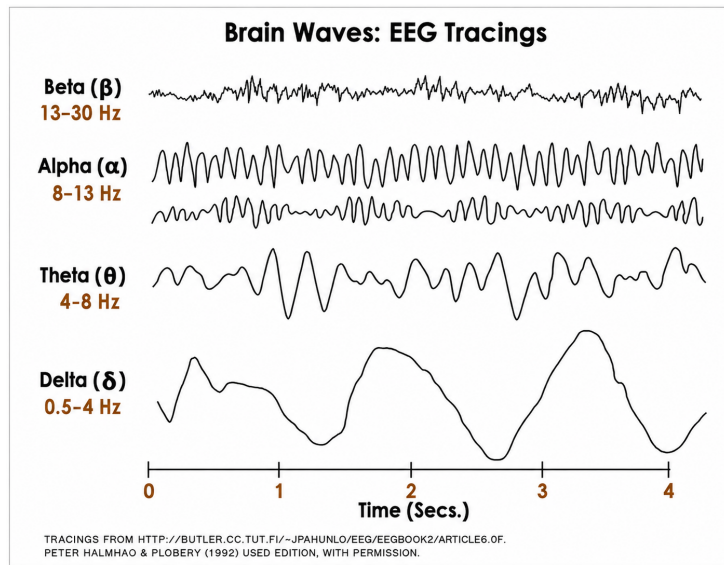


Figure 3. This figure shows the characteristic curves of an EEG.

The system, with the helmet and the transmitted signals, and with good correlation (in this work, correlation means the ability to narrow the window within which the brainwaves are then selected), can record all brain activity, creating slots, especially when the individual customizes the use of the helmet, for example during an inspection, and would like to relive that situation at a later time; this system should help. Indeed, by reproducing the recorded waves completely this time (thanks to Shannon's theorem and complete correlation), and therefore also the time windows are well-defined, the experience is reproduced, so it can be helpful when thinking about a specific or nuanced situation; this fact needs to be verified with experiments.

This would be a significant innovation in the field of on-site inspections of human activities—industrial, civil, medical, scientific, military, and psychophysical. The applications are countless, and this must be taken into account when launching a startup (a pilot project to engineer the ideas described here). Using generative AI, the system's response, or its correlation within the time window, could be improved by placing flags (that draw attention, such as a milestone, an object, a situation). These flags, during training, likely make them more perceptible, improving the brain's compliance with the system. The brain then responds better by facilitating brainwaves to relive the recorded situation and correlate it with the recorded "slot" trained to memorize and highlight certain aspects.

3. An Example

An example speaks louder than a thousand words. Luckily, we had a dream a few nights ago, and we'll tell you about it. A few days ago, we saw a TV program about the Domus De Jana, ancient Nuragic-sized structures found in great numbers only in Sardinia.

The dream in question is classified as level one, very bright, lasting about fifteen minutes and ending at a verifiable time. Nuragic

structures are presumably difficult for humans to construct; While I was having lunch with hunter friends, and it was a lovely time, with roast meat, red wine, and smiles, in the woods on one side of a promontory, on the opposite side I saw a spaceship crash and explode on the opposite side from where I was. I couldn't dial the fire department. The flames were high, devices like tablets and cell phones weren't working. At that point, I asked a doctor friend to leave me her cell phone, and it also had a special system for entering numbers, but she managed to call the emergency number and the switchboard answered, which put me through to the fire department. That's how it works in real life. While the fire department was being passed on to me, I approached to help the alien, and what could I see? First of all, a huge excavation at least ten meters deep, and other friends from my town who had nothing to do with it, with their lunch, and those who were having lunch weren't worried about anything. A huge excavation that two bulldozers couldn't complete in three days. Inside, neatly arranged, were pieces of the spaceship and a mix of technology and feudalism, a lot of technology; neatly arranged, to be buried later: the conclusion is that the dream, presumably, the Domus de Jana have an alien connotation. This is, however, a dream, emphasizing that the Domus de Jana, and the aliens in fact, may not exist in reality. It's also true that the project manager was a friend of mine, with a poor academic record but industrious in construction (he renovated his house... and what a job). Where did the alien go?

4. Conclusion

In our life journey, we are, those who are, natural lucid dreamers up to the fourth level, very clear, premonitory, we find lost objects: these things happen. However, writing this work would have been "impossible" if we had chosen differently along our life journey: a yes to a job offer where we said no, a different girlfriend, a yes to a different financial situation, etc.; something different from the

path we experienced would have been generated. We attach great importance to this work, in the sense that we should be more aware of the physiology that has formed. Therefore, with the third-order state of consciousness, we have a very important concept that should give us pause. Let's assume, hypothetically, that everything is working correctly; therefore, a person who reviews their dreamed or lived experiences works: loved ones, objects, people they have never met because they were not yet born or previously died; Let's assume this happens and the person becomes less subservient and eager to stay connected, at this point we face a question: is it truly an Afterlife or is it simply a bluff, an unknown quantity? This isn't easy to know yet. In the existence of a possible Creator, which we believe exists, everything makes sense if we arrive somewhere—the famous Afterlife—to flow into a "happy" island... a destination point, and thinking about the beyond is not the case, for now.

A positive aspect, we believe it almost always is, is the lifestyle changes that can develop from a consciousness of the path (of the third order). After all, we would probably end up in the same place, not in the flesh but in spirit, all there; perhaps it's time to behave more lovingly?

After all, he who sows the wind (love/hate) reaps the whirlwind (love/hate) even in the afterlife?

■ ACKNOWLEDGEMENTS

An extraordinary thank you to all those who take care of people with physical and mental disabilities.

■ REFERENCES

- [1] Francesco Pia, "I will be back", pp. 149, Ed. Amazon ® Kdp, January 2019.
- [2] Francesco Pia, "The liar knee", pp. 224, Ed. Amazon ® Kdp, July 2019.
- [3] Francesco Pia, "Genetic geometry", pp. 80, Ed. Amazon ® Kdp, January 2019.
- [4] Francesco Pia, "JEWEL: a cosmological model due to the geometrical displacement of galactic object like black, white and worm holes", ICCEA 2021: 15. International Conference on Cosmology and Extragalactic Astronomy; July 19-20, 2021 in Toronto, Canada.
- [5] Francesco Pia, "How to Try to Get Antigravity from Electrical Circuits by an Almost Magnetic Monopole or from Anti Matter" in *IRJAES*, Volume 7, Issue 3, Date:25/06/2022.
- [6] Francesco Pia, "11A Computer System based on Neural Network and Rules System based for find An Optimum Mix of Vaccine and/or Drugs for mRNA Like Covid-19 Virus and other Pathologies" in *IRJAES*, Volume 7, Issue 2, Date, 09/06/2022.
- [7] Francesco Pia, "Polifemo: a Teleportation Model based on Object Like Black, White and Worm Hole for Electric Signal and Object Also thinking and with Conscience", *IDIRJAES-V7N3P66Y22* June-Au 2022 Issue Vol. 7 Issue 3 <https://irjaes.com/volume-7-issue-3/>.
- [8] Francesco Pia, "A Symmetry of Cosmological Model Named Jewel that Allows Reentry into the Starting Parallel Universe with Adequate Teleportation based on Creation of useful Stellar Object Like Black, White Holes", ICES-23-P1159, ICES2023, 6-8 February 2023, Riyadh, Saudi Arabia.
- [9] Francesco Pia, "Neutron "Star", a Hypothetical Small Neutron "Star" based on the Emission of Neutrons in Minerals such as Granite and Other Commonly Available Objects", *Global Journal of Researches in Engineering: E*, Publisher: Global Journals, Civil and Structural Engineering, Volume 23, Issue 1, Version 1.0, Year 2023, Type: Double Blind Peer Reviewed International Research Journal, Online ISSN: 2249-4596 & Print ISSN: 0975-5861.
- [10] Francesco Pia, "A Natural Model for Dark Matter and its Interaction with Stellar Objects Such as the Galaxy Known with Name "The Milky Way", *American Journal of Biomedical Science & Research*, 2023, 19(5), Jbsr.Ms.Id.002630, DOI: 10.34297/Ajbsr.2023.19.002630, Received: July 28, 2023; Published: August 07, 2023, Pag. 616-621, Final Pdf: Pp-Rw-23-0513, ISSN: 2642-1747.
- [11] Francesco Pia, "A Time-Machine, an Electro-Mechanical System that Have a Strange Behaviour Respect at Time", *American Journal of Biomedical Science & Research*, 2023, 20(2), Ajbsr.Ms.Id.002681, DOI: 10.34297/Ajbsr.2023.20.002681, Received: September 16, 2023; Published: September 21, 2023, Pag. 122-126, Final Pdf: Pp-Rw-23-0513, ISSN: 2642-1747.
- [12] <http://www.neuroscienzedipendenze.it/eeg.html>



Reti Neurali e Sistemi Basati su Regole Utilizzati per Individuare Correlazioni Razionali e Scientifiche tra l'Essere Qui e Ora e le Condizioni dell'Aldilà

Translation of "Neural Networks and Rules-based Systems used to find Rational and Scientific Correlations between being Here and Now with Afterlife Conditions"

Article Record

prof. PhD Eng. Francesco Pia^{9*}

PhD Eng.

*Corresponding Author



RECEIVED
2025-12-31

ACCEPTED
2026-01-13

PUBLISHED
2026-02-17

PEER REVIEW
Double Blind

Abstract

What you will read in the next pages is truly an anthology-worthy element, in the sense that it is to be taken seriously; in fact, the place that we will try to describe in this work, in the sense of arriving in an unknown place, which therefore will not be easy to describe. In the path that takes us from non-existence to being and becoming... placed in this location, it becomes interesting that different things or people coexist that have not yet been presented in the era in which we connect and go to browse. A very important aspect is the determination of the window in which one says he had a dream where he saw a car that does not yet exist today but perhaps will exist in the future, or a deceased person or one who will exist in the future; So, once the time window is set, which isn't easy, and determining it is an important aspect in which brainwaves will then be selected. This selection is easy and precise when you're awake and thinking about something and the time window is very clear, so a small interaction with the person is enough. It's a different story when you're dreaming. Shannon's theorem, used in electrical engineering and telecommunications, helps select and "correlate" signals. Why this theorem was invented, crucial for effectively sampling signals and being able to fully reconstruct them, is one of those mysteries that leads us to ask ourselves at lunch with the family: Do I have a colleague? We must therefore perfect the start and stop of recording during the dream by focusing on where we are interested, useful for discovering what we think and that our brain suggests things that we thought were lost in our historical culture and instead suddenly... what we need emerges. We would therefore try to obtain feedback from Afterlife in order to reproduce the brain waves seen in the dream phase thanks to a helmet connected to proactive and not just generative Artificial Intelligence.

ALIEN

DOMUS DE JANAL

SPACESHIP

AI USE STATEMENT

No generative AI was used for analysis or results.

FUNDING

No external funding was declared for this work.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY

Not applicable for this article.

ETHICS

No ethics committee approval was required for this article type.

CONSENT

Not applicable for this article.

TRIAL REG.

Not applicable.

How to Cite: Pia (2026). Reti Neurali e Sistemi Basati su Regole Utilizzati per Individuare Correlazioni Razionali e Scientifiche tra l'Essere Qui e Ora e le Condizioni dell'Aldilà. Global Journal of Computer Science and Technology, 26(1), 37-40.

LICENSE

© 2026 Global Journals. Open-access article under CC BY-NC-ND 4.0 International License.

AR Experience Web Page

DOI



Print ISSN 0975-4350



Online ISSN 0975-4172



Under the strict compliance and defined process of



METADATA CONTINUATION

AUTHOR CONTACT QR LEDGER

prof. PhD Eng. Francesco
Piaò*



ARCHIVAL RECORD

Reti Neurali e Sistemi Basati su Regole Utilizzati per Individuare Correlazioni Razionali e Scientifiche tra l'Essere Qui e Ora e le Condizioni dell'Aldilà

prof. PhD Eng. Francesco Pia^{§*}

§ PhD Eng.

Abstract

Quello che leggerete nelle prossime pagine è davvero un elemento degno di un'antologia, nel senso che va preso sul serio; è infatti il luogo che cercheremo di descrivere in quest'opera, nel senso di approdare in un luogo sconosciuto, che quindi non sarà facile da descrivere. Nel percorso che ci porta dal non esistere all'essere e al divenire... collocati in questa location, diventa interessante il fatto che coesistano differenti cose o persone che ancora non sono state presentate nell'epoca in cui noi ci colleghiamo e andiamo a curiosare. Un aspetto molto importante è quello della determinazione della finestra in cui uno dice di aver fatto un sogno dove si è vista una macchina che oggi ancora non esiste ma forse esisterà in futuro, oppure una persona defunta oppure che esisterà nel futuro; quindi, fissata la finestra temporale, fatto non semplice, e determinarla è un aspetto importante nel quale poi si selezioneranno le onde cerebrali. Tale selezione è facile e preciso quando si è svegli e si pensa una cosa e la finestra temporale è ben chiara, quindi basta una piccola interazione con la persona cosa differente è quando si sogna. Esiste il teorema di Shannon, usato in Elettrotecnica e Telecomunicazioni, che aiuta a selezionare "correlare" i segnali. Del perché si sia pensato a questo teorema, cruciale per campionare efficacemente dei segnali e poterlo ricostruire completamente, a pranzo con la famiglia è uno di quei misteri che ci portano a chiederci: abbiamo un collega? Dobbiamo perfezionare quindi lo start e lo stop di registrazione durante il sogno stringendo dove di interesse, utile per scoprire che pensiamo e che il nostro cervello ci propone cose che davamo per disperse nella nostra cultura storica e invece improvvisamente... emerge quello che serve. Cercheremmo quindi di ottenere un feedback dall'Afterlife in modo da riproporre le onde cerebrali viste in fase onirica grazie ad un casco collegato con l'Intelligenza Artificiale propositiva e non solo generativa.

Keywords: Alien, Spaceship, Domus De Janai

* Corresponding Author
prof. PhD Eng. Francesco Pia

1. Introduction

Nel fluire degli eventi dall'origine ad oggi fino al luogo dove tutti, e tutto, andremmo come nel problema dei due gemelli: nessuno è mai rientrato da dove sono usciti e i due discutono come potrebbe essere il "mondo" fuori da loro conosciuto. Abbiamo da pensare allo stato di coscienza del secondo ordine (cosciente di esserlo) e poi abbiamo una coscienza superiore, globale, del percorso e quindi la scelta che noi (con il nostro libero arbitrio) operiamo; la nostra presenza *qui e ora* è il costituente di un disegno complesso, armonizzato, plasmato dal creatore. Quindi, non esiste il tempo ma la posizione e quindi... tutti (e tutto) finiremmo nella stessa posizione. L'idea sarebbe quella di studiare, ingegnerizzare, un sistema sintonizzabile con un casco che riproducendo le onde cerebrali che "sono" quelle apprese durante l'addestramento del sistema intelligente con il quale noi stiamo cercando di selezionare spot di EEG di persone che vedono nei sogni o pensano i morti: che sono nell'aldilà o non ancora apparsi. Quindi dobbiamo selezionare queste onde cerebrali e siccome un'onda può essere frutto di molte onde, ovvero, si parla di un sintonizzatore; al rientro i segnali sono composti da certe onde elettromagnetiche che veicolate nel casco permetterà di riprodurre e *vedere* il luogo dei morti e forse

di interagire come se il sogno fosse rivissuto. La coscienza del secondo ordine da un'importanza apicale al singolo, molti prendendo consapevolezza di un fatto e così si troverebbero a disagio nel fare una prevaricazione perché tanto finiremmo tutti nello stesso posto e chi lo dice che il tipo a cui hai fatto un torto non si possa rivalere e ne abbia le capacità?

2. Methods and Tools

Nel percorso, o fluire dell'esistenza il trapasso completa la persona che è consapevole di ciò, una coscienza sul percorso, quindi possiamo chiamarla: coscienza del terzo ordine; questo sarà messo in luce con il feedback dei segnali che verranno prodotti da un sistema informatico e trasmessi al casco e quindi al cervello, cosicché possa rivivere più completamente "pienamente", vicino alla finestra (esperienziale) del segnale correlabile da quello che ha generato lui stesso.

Un primo metodo di indagine sarebbe selezionare alcune persone con spiccate capacità oniriche e che possono essere definiti "sognatori lucidi", monitorarli con un elettroencefalogramma e misurare per 24h le onde cerebrali e quindi verificare, se capita, se hanno avuto sogni o pensieri legati ai defunti o all'aldilà, e registrare i segnali nella finestra temporale giustappunto selezionata; fatto questo, si

utilizzano le reti neurali, anche se sono serie temporali, per vedere se quando pensiamo da svegli, le persone comuni, hanno le stesse forme d'onda. Però c'è da dire che nel sistema intelligente che viene addestrato per riconoscere le onde dell'EEG ci saranno anche quelle da sveglio; perché si può immaginare, o pensare, ad un genitore, per esempio ad un genitore, o ad un caro defunto (a chi verrà è ora prematuro)... a Napoleone, Giulio Cesare per esempio: nel momento lo si pensa intensamente: il cervello cosa o che onde manifesta?

Possono essere rinvenute certe caratteristiche particolari, questo appena descritto è un primo metodo.

Noi vorremmo collegarci all'eventuale location chiamata Afterlife con un feedback che torna indietro al qui e ora collegando il casco che darà dei segnali propri potranno essere selezionati per l'evento considerato dall'utente.

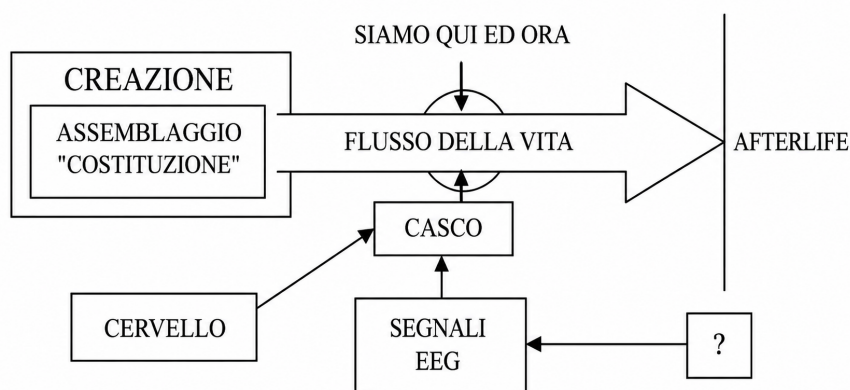


Figure 1. In questa figura è rappresentato lo schema della linea temporale dall'origine, nascita, il percorso "della vita", dove stiamo qui e ora, la morte e l'eventuale trasferimento dello spirito al Afterlife.

Un altro aspetto molto importante è rappresentato nella seguente figura **Fig. [4]** presente negli universi paralleli in [11], essendoci il percorso che riprodotte gratificano l'idea che esista, diciamo il teletrasporto [7]; ma è anche importante invece l'esistenza di un limite indefinito di tutte le situazioni o eventuali possibili scelte:

tutto confluisce in un cerchio che circonda tutto come limite di tutte le cose e di tutti i percorsi, quindi questo cerchio rappresenta la finestra verso l'ambiente che noi chiamiamo *Afterlife*, diversamente è necessario approfondire o/e cambiare ipotesi.

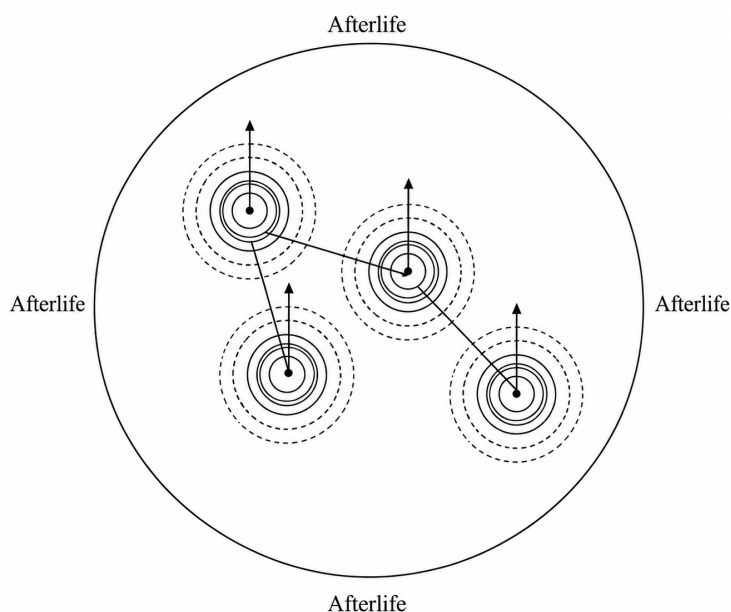


Figure 2. In this figure it is represented more than one line of events that arise from "point" configurations of any circumferences and circle conjunction

Il paradigma che si vuole utilizzare è sempre il solito: un'ipotesi molto debole che ci permette di avere una maggiore immaginazione. L'ipotesi debole è che si possa sintonizzare audio-video il luogo dove sono tutti i morti (magari anche quelli che devono ancora nascere

o divenire) nel divenire (fluire) della vita dall'origine fino alla fine e la presenza: *io sono qui*, ero... non si sa dove, è però un luogo dove sono tutti morti, o devono ancora nascere, non esistendo il tempo: la condizione è di persistenza in un ambiente dove si trovano

tutti e quindi come per esempio una pallina da golf che dalle origini: la costruzione dei materiali, l'assemblamento, la formazione della pallina, il posizionamento, lo colpo con la mazza, l'arrivo fino in buca, da lì viene presa poi siccome il tempo non esiste si distruggerà come tutto, non persiste niente e nulla continua ad essere assemblato.

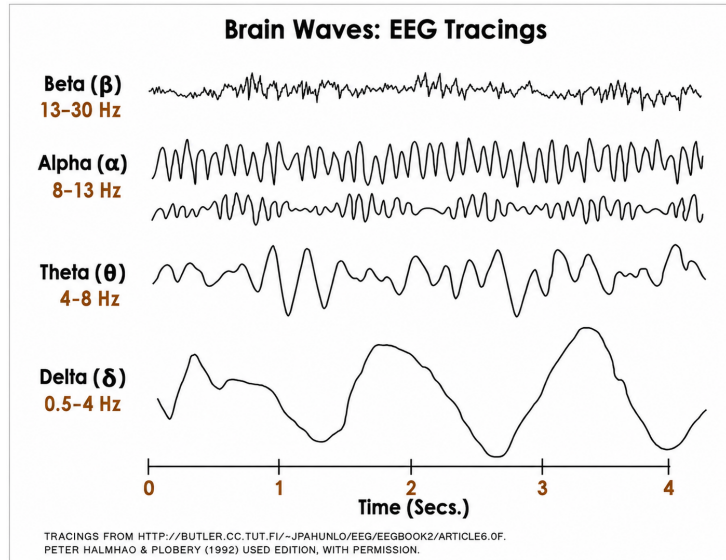


Figure 3. In questa figura sono rappresentate le curve caratteristiche di un EEG.

Il sistema con il casco e i segnali trasmessi e con una buona correlazione (in questo lavoro per correlazione si intende la capacità di stringere la finestra in cui poi verranno selezionate le onde cerebrali), si può registrare tutta l'attività cerebrale, facendo degli slot, soprattutto quando l'individuo personalizza l'uso del casco per esempio durante lo svolgimento di un sopralluogo, e vorrebbe rivivere quella situazione in un secondo momento; questo sistema dovrebbe aiutarlo. Infatti, riproducendo le onde registrate, completamente questa volta (grazie al teorema di Shannon e alla completa correlazione) e quindi anche le finestre temporali sono ben definite: l'esperienza che si ripropone, cosicché sia di aiuto quando si vuole pensare ad una situazione precisa o sfumata; questo fatto è da verificare con esperimenti.

Si tratterebbe di un'innovazione importante nel campo dei sopralluoghi delle attività umane industriali, civili, mediche, scientifiche, militari, attività psico-fisiche. Le applicazioni sono innumerevoli e di questo se ne deve tener conto nel caso si volesse fare uno StartUp (un progetto pilota per ingegnerizzare le idee qui descritte), grazie all'uso dell'AI generativa si potrebbe migliorare la risposta del sistema, o della correlazione nella finestra temporale, mettendo dei flag (che richiamino l'attenzione come una milestone, un oggetto, una situazione) che durante l'addestramento verosimilmente rende più percepibile, migliorando la compliance con il sistema, da parte del cervello, che ripropone una migliore risposta facilitando le onde cerebrali a rivivere la situazione registrata e correlata con lo "slot" registrato ed addestrato a memorizzare e mettere in luce certi aspetti.

2.1. Un esempio

Un esempio chiarisce di più di mille parole, per fortuna qualche notte fa abbiamo avuto un sogno e ve lo raccontiamo. Giorni fa abbiamo visto un programma televisivo sulle *Domus De Jana* che sono antichissimi manufatti edili di dimensioni nuragiche che si trovano numerosissime solo in Sardegna. Il sogno in questione lo classifichiamo del primo livello, molto luminoso, durato circa quindici minuti con la fine ad un orario documentabile. Le

costruzioni nuragiche sono verosimilmente difficili da costruire per gli esseri umani; mentre partecipavo ad un pranzo con amici cacciatori e si stava bene, carne arrosto vino rosso e sorrisi, nel bosco da un lato di un promontorio, dal lato opposto si è vista un'astronave precipitare ed esplodere nel lato opposto da dove mi trovavo e non riuscivo a comporre il numero dei vigili del fuoco, le fiamme erano alte, i dispositivi come tablet e telefoni cellulari non funzionavano a quel punto ho chiesto ad una amica dottoressa di lasciarmi il suo telefono cellulare e anche questo aveva un sistema particolare di input dei numeri ma lei è riuscita a chiamare il numero di emergenza e ha risposto il centralino che mi ha passato i vigili del fuoco, funziona così nella realtà, nel mentre che mi passava i vigili mi sono avvicinato per dare soccorso all'alieno e cosa si vedeva: innanzitutto uno scavo enorme e profondo almeno dieci metri e altri amici del mio paese che non ci entravano niente con il pranzo, e quelli che stavano pranzando non si sono preoccupati di nulla. Uno scavo enorme che due ruspe non farebbero in tre giorni di lavoro, all'interno messi in ordine si trovavano pezzi dell'astronave e un misto tra tecnologia e feudalesimo, molta tecnologia; in ordine, per poi essere sepolti: la conclusione è che il sogno, verosimilmente, le *Domus de Jana* hanno una valenza aliena. Questo è comunque un sogno sottolineando che le *Domus de Jana*, e gli alieni di fatti che potrebbero non esistere nella realtà, è vero anche che il Direttore dei lavori era un mio amico, scadente da un punto di vista scolastico ma industrioso in opere edili (si è ristrutturato casa... e che lavoro), dove è finito l'alieno?

3. Conclusion

Nel nostro percorso di vita siamo, chi lo è, sognatori lucidi naturali fino al quarto livello, molto chiari, premonitori, troviamo oggetti smarriti: questi fatti capitano. Però, la stesura di questo lavoro sarebbe stato "impossibile" se avessimo scelto diversamente nel percorso di vita: un sì ad un'offerta di lavoro la dove abbiamo detto di no, un'altra fidanzata, sì ad un altro aspetto economico etc.; si sarebbe generato qualcosa di differente dal percorso che abbiamo vissuto. Noi diamo molta importanza al presente lavoro,

nel senso che dovremmo essere più consapevoli di quello che è la fisiologia che si è costituita. Quindi, con lo stato di coscienza del terzo ordine si ha a disposizione un concetto molto importante che dovrebbe far riflettere. Supponiamo, per ipotesi, che tutto funzioni correttamente; quindi funziona, una persona che rivede le proprie esperienze sognate, o vissute: i propri cari, oggetti, persone che non ha mai conosciuto perché non ancora nato o morto precedentemente; supponiamo che questo avvenga e che, la persona, diventi non proprio succube e desiderosa di essere sempre connessa, a questo punto ci poniamo un problema: ma è veramente un *Afterlife* o è semplicemente un bluff, un'incognita: questo non è semplice da sapere ancora. Nell'esistenza di un eventuale Creatore, che a nostro parere esiste, tutto ha un senso se arriviamo in qualche posto il famoso *Aldilà*, a fluire in un'isola "felice"... un punto di destinazione e pensare all'oltre adesso non è il caso, per ora.

Un aspetto positivo, pensiamo lo sia quasi sempre, è la modifica dello stile di vita che può svilupparsi da una coscienza sul percorso (del terzo ordine), del resto finiremmo forse, nella stessa *location* non in carne e ossa ma in spirito tutti lì; sarà forse il caso di comportarsi più amorevolmente?

Del resto chi semina vento (amore/odio) raccoglie tempesta (amore/odio) anche nell'aldilà?

■ ACKNOWLEDGEMENTS

An extraordinary thank you to all those who take care of people with physical and mental disabilities.

Subject Index

1d resnet, 15–23

Afterlife, 37

Alien, 37

alien, 33–36

bert, 24–32

Brain Waves, 37

cnn, 15–23

conditional generative adversarial network, 15–23

decision tree, 15–23

decoder-only models, 1–14

deep learning, 15–32

Domus De Janal, 37

domus de janal, 33–36

election outcome verification, 24–32

Electroencephalogram, 37

encoder-decoder models, 1–14

encoder-only models, 1–14

explainable ai, 1–14

feature attribution, 1–14

human activity recognition, 15–23

lstm, 15–23

meta-analysis, 1–14

model complexity, 1–14

model interpretability, 1–14

natural language processing, 1–14, 24–32

Neural Networks, 37

online news media, 24–32

parameter efficiency, 1–14

performance trade-off, 1–14

political sentiment analysis, 24–32

Rule-Based Systems, 37

sdr dataset, 15–23

sentiment analysis, 24–32

Shannon Theorem, 37

signal-based monitoring, 15–23

software defined radio, 15–23

Spaceship, 37

spaceship, 33–36

temporal sentiment trends, 24–32

transformer architectures, 1–14

transformer-based models, 24–32

uk general election 2024, 24–32

Author Guidelines

Comprehensive Guide to Publishing with GJCST

Global Journal of Computer Science and Technology | Open Access | Peer Reviewed | COPE Compliant

I. Our Commitment to Authors

The Global Journal of Computer Science and Technology (GJCST) is an internationally recognized, **open-access** academic journal dedicated to publishing high-quality, peer-reviewed research across all domains of computer science and technology - including Artificial Intelligence, Machine Learning, Software Engineering, Networking, Cybersecurity, Data Science, Algorithms, Human-Computer Interaction, and emerging interdisciplinary fields.

GJCST is published by Global Journals Publishing Group Inc., indexed in major global academic databases, archived in leading digital repositories, and distributed to institutional subscribers worldwide in prestigious **hardbound volumes**.

We believe that the purpose of an academic journal is to amplify the voice of the researcher. Every operational decision we make - from submission workflows to publication formats - is guided by a single principle: *the author should never be distracted from the science*. Our professional editorial and production teams exist to handle every technical and logistical complexity so that you can devote your full energy to your computer science research.

II. One-Click Submission: Publish Without the Hassle

GJCST has pioneered a **One-Click Submission** system designed to eliminate the traditionally burdensome process of paper submission.

A. Submit in Any Format

We accept manuscripts in **any common document format**, including but not limited to:

- Microsoft Word (.doc, .docx)
- Adobe PDF (.pdf)
- $\text{\LaTeX}$ source files (.tex)
- Rich Text Format (.rtf), Apple Pages, LibreOffice, or plain text

You do **not** need to install specialized software, learn $\text{\LaTeX}$, or adhere to a rigid template before submission. Simply write your research in the format you are most familiar with. Whether your manuscript includes complex algorithm pseudocode, mathematical proofs, or system architecture diagrams - our production team is expertly equipped to handle it all.

B. What You Need to Submit

To submit a manuscript, you only need to provide:

1. Your **full name**
2. Your **email address**
3. Your **manuscript file**

That is all. No lengthy registration forms, no complex metadata entry, no mandatory template compliance. Submit directly via our portal at <https://globaljournals.org/submit-an-article/>.

Heading style: In the final typeset article, author-created headings are normally set in sentence case for readability. Fixed journal headings such as Abstract, Introduction, Materials and methods, Results, Discussion, Conclusion, References, and Acknowledgements are standardized by the production team. Acronyms, proper nouns, gene/species names, chemical notation, brand names, and Roman numerals are preserved.

C. What Happens After Submission

1. **Instant Confirmation:** You receive an automated email confirming receipt of your manuscript. If you do not receive a confirmation, please reach out via our Contact Us form at <https://globaljournals.org>.
2. **Professional Formatting:** Our dedicated production executives convert your manuscript into our advanced internal $\text{\LaTeX}$ template - producing a sleek, mathematically precise document with beautifully rendered code listings and algorithm environments - entirely on your behalf.
3. **Author Dashboard:** We create a secure account for you on the **Author Dashboard** and email your login credentials. You may log in at any time to review pre-filled metadata, update co-author details, or track your submission status.
4. **Review Initiation:** Your professionally formatted, fully anonymized manuscript is assigned to expert reviewers through our rigorous double-blind peer-review process.

III. Rigorous Double-Blind Peer Review

Academic credibility is the cornerstone of GJCST. Every manuscript undergoes a **strict double-blind peer-review process** in which neither the authors nor the reviewers are aware of each other's identities. This guarantees that every paper is evaluated purely on its scientific merit, methodological rigor, originality, and contribution to the field of computer science and technology.

A. The Review Workflow

1. **Desk Screening:** The editorial office performs an initial screening for scope, completeness, and adherence to ethical standards.
2. **Reviewer Assignment:** Two or more independent, domain-expert reviewers - specialists in your specific sub-field of computer science - are assigned to evaluate your manuscript.
3. **Detailed Evaluation:** Reviewers assess the paper on criteria including originality, significance, clarity, methodology, experimental rigor, data integrity, and presentation quality.
4. **Editorial Decision:** Based on reviewer recommendations, the Editorial Board renders one of three decisions: *Accept*, *Revise and Resubmit*, or *Decline*.

B. Complete Review Report & Research Assessment Sheet

Upon completion of the review, you will receive a **comprehensive Review Report** along with a detailed **Research Assessment Sheet**. These documents provide:

- Line-by-line reviewer comments keyed to **specific line numbers** in your typeset manuscript, enabling precise, unambiguous identification of every comment.
- A structured assessment covering originality, significance, clarity, methodology, and overall quality.
- Constructive feedback and actionable suggestions for improvement tailored to the standards of computer science publishing.

C. Responding to Reviewer Comments

Addressing reviewer feedback is simple and flexible. You have two options:

Option A - Author Dashboard: Log in to your Author Dashboard, navigate to **Manage Submission**, and create a **support ticket**. Simply quote the relevant line numbers from your review report and clearly describe your corrections or responses.

Option B - Email: If the Dashboard is not convenient, you may simply send an email with a numbered list of corrections referencing the line numbers, along with your responses to each reviewer comment.

Important: You do **not** need to reformat the document yourself at any stage. Our production executives will implement all your requested text changes directly in the typeset manuscript. Reviewer comments may be addressed and resolved right up until the Early View stage.

IV. Publication Lifecycle: Complete Transparency

A. Early View (Galley Proof)

- Upon acceptance, a **galley proof** (Early View edition) of your article is prepared according to the journal's publishing schedule.
- The Early View version includes **line numbers** for convenient reference during final corrections.
- You may still provide corrections at this stage by referencing specific line numbers via email or the Author Dashboard.

B. Ahead of Print

- Following the Early View, an **Ahead of Print** version is published before the final issue compilation.
- This version represents the final, corrected article but **does not yet include Volume or Issue numbers**.

C. Digital Publication (eJournal)

- After all corrections are finalized, the **digital eJournal edition** is published with line numbers removed.
- Your article receives its official DOI (**Digital Object Identifier**) and is indexed across global academic databases.
- The article becomes freely and permanently accessible online under our open-access policy.

D. Hardbound Print Publication

- Approximately **15 days after digital publication**, the article enters our offset production pipeline.
- Your research is printed in **premium hardbound volumes** and distributed to institutional subscribers, major computing libraries, and research centers worldwide.
- The corresponding author receives a **complimentary hardbound copy** dispatched via government courier at no additional cost.

V. Advanced Technology & AI-Enabled Research Broadcasting

As a leading computer science journal, GJCST's own publishing infrastructure reflects the cutting edge of technology.

A. Multi-Format Native Publication

Every accepted article is published natively across multiple formats:

- **Interactive Digital eJournals** optimized for desktops, tablets, and smartphones.
- **Premium Hardbound Print Editions** for institutional archives and personal collections.
- **Machine-Readable XML & HTML** for web accessibility, automated indexing, and integration with global research repositories.

B. AI-Powered Discoverability & SEO

We deploy cutting-edge Artificial Intelligence to generate rich semantic metadata for every article, dramatically increasing its search engine visibility. Our proprietary AI systems:

- Automatically tag, categorize, and cross-reference your work across international academic databases.
- Optimize article metadata for maximum **Search Engine Optimization (SEO)**, ensuring your paper appears prominently when researchers search for your computational models, algorithms, or frameworks.
- Actively broadcast your research to targeted computer science communities worldwide, amplifying your citation count and global influence.

VI. Ethics, Integrity & COPE Compliance

GJCST maintains the highest standards of publication ethics, guided by the principles of the **Committee on Publication Ethics (COPE)**.

A. Plagiarism & Originality

All submissions are screened using state-of-the-art plagiarism detection software. Manuscripts with significant textual overlap with previously published work will be declined. Authors must ensure that all content is original or properly attributed through accurate citations.

B. Conflicts of Interest & Funding Transparency

Authors are required to disclose all sources of funding and any financial or non-financial conflicts of interest. Full transparency is mandatory and is enforced at every stage of the publication process.

C. Data Availability & Reproducibility

Authors are strongly encouraged to make their source code, datasets, and supplementary materials available in recognized repositories (e.g., GitHub, Zenodo, Figshare) to support the reproducibility of computational experiments and algorithmic evaluations. A data availability statement is required in all published manuscripts.

D. Ethical Approval & Informed Consent

Research involving human participants, user studies, or sensitive data must include a statement of ethical approval from a recognized institutional ethics committee, along with confirmation that informed consent was obtained where applicable.

E. Comprehensive Ethics Portal

Our enhanced, COPE-based ethical publishing guidelines - covering authorship criteria, editorial responsibilities, reviewer obligations, corrections & retractions, and misconduct procedures - are available in detail on our website at <https://globaljournals.org>. We encourage all authors, reviewers, and editors to familiarize themselves with these standards.

VII. Perpetual Archiving & Geo-Distributed Data Infrastructure

GJCST is backed by Global Journals' enterprise-grade archiving infrastructure. Your published research is not stored on a single server - it is **redundantly archived across multiple geographically distributed data centers** spanning several continents, including advanced **deep-sea data center facilities** engineered for unparalleled environmental resilience and longevity.

This infrastructure guarantees:

- **Lifetime Hosting:** Your article will remain permanently accessible online for the lifetime of the publication. We do not sunset, deprecate, or remove published content under any circumstance.
- **Disaster Resilience:** Multi-site redundancy ensures that no single point of failure - natural disaster, hardware degradation, or network disruption - can compromise access to your research.
- **Immutable DOI Resolution:** Every article is assigned a permanent Digital Object Identifier (DOI) that will resolve to your work indefinitely, regardless of future platform migrations.

VIII. Open Access: Free Knowledge for All

GJCST is a Fully Open Access Journal.

We firmly believe that scientific and technological knowledge is a public good. Every article published in GJCST is **freely and permanently available online** from the moment of publication - accessible to researchers, developers, students, educators, and the general public worldwide **without any subscription fees, paywalls, or access restrictions**.

Our open-access model ensures:

- **Immediate Availability:** Your research is accessible globally on the day of publication.
- **Maximum Reach:** Barrier-free access dramatically increases readership, citations, and real-world impact in the computing community.
- **Equitable Knowledge Distribution:** Researchers in developing nations, independent developers, and the public benefit equally from the latest technological advances.

Accepted articles are published under the **Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)** license. Authors retain full copyright while granting Global Journals the right to publish and distribute the work globally.

IX. The $\LaTeX$ Author Package (Entirely Optional)

For computer scientists who already compose their manuscripts in $\LaTeX$, we provide a comprehensive, professionally designed **$\LaTeX$ Author Package** (`gj-author`) featuring STIX Two serif typography, TeX Gyre Heros sans-serif headings, complete Unicode math support, styled code listing environments, and brand-coordinated layouts.

However, we must emphasize:

Reading or using the $\LaTeX$ Author Package is entirely optional.

If you do not use $\LaTeX$, simply send us your manuscript in Word, PDF, or any other format. Our production team will expertly convert it into our advanced typesetting system and produce the identical, pristine publication-quality output.

You do not need to learn or install anything.

The detailed $\LaTeX$ guidelines included in the Author Package are intended solely as a convenience for authors who prefer granular typographic control over their code listings, pseudocode, and mathematical environments before submission. They are not a prerequisite for publishing with us.

Contact & Support

Submission Portal: <https://globaljournals.org/submit-an-article/>

Website: <https://globaljournals.org>

Email: helpdesk@globaljournals.org

*At GJCST, your research defines the future of technology.
We handle the complexities of publishing so you can focus
on what matters most - building tomorrow.*

GLOBAL JOURNAL

of Computer Science and Technology: D
Neural & AI

globaljournals.org



Save our Planet



Online ISSN 0975-4172



9 770975 417011