AutoML-Based Fake Review Detection: A Unified Textual and Multimodal Learning Framework

Tayasan M. H. Gedara
National cybersecurity program
IMT School for Advanced Studies
Lucca, Italy
tayasanmilinda.hewadaundagedara@imtlucca.it

Rocco Loffredo
dept. Industrial Engineering
University of Salerno
Salerno, Italy
rloffredo@unisa.it

Nadeeka M.K. Arachchige
dept. Industrial Engineering
University of Salerno
Salerno, Italy
nkiringodaarachchige@unisa.it

***Abstract*—Fake reviews undermine the credibility of online platforms, yet most existing detection methods rely on manually designed pipelines with extensive feature engineering and heuristic model selection, limiting robustness and cross-domain generalization. To address these challenges, this paper proposes a unified AutoML-based framework for fake review detection that automates model selection, feature optimization, and hyperparameter tuning in an end-to-end manner, evaluated in both text-only and multimodal settings. The framework is evaluated across multiple AutoML systems and five benchmark datasets, including Amazon, FRD, Yelp-NYC, TripAdvisor, and a synthetic E-Commerce dataset, and integrates textual, numerical, and visual information during the multimodal stage. Experimental results show that multimodal AutoML consistently outperforms traditional machine learning models and remains competitive with state-of-the-art deep learning approaches, achieving up to 10% accuracy increment and 9.1% F1-score, demonstrating AutoML's effectiveness.**

***Index Terms*—Fake Review Detection, AutoML, Multimodal Learning,**

I. Introduction

In this era of the internet, e-commerce customers can easily purchase goods on online platforms and share their evaluations, which will rate customer satisfaction in the process. These reviews benefit both organizations and potential customers by providing insight into products and services before purchase decisions [1], [2]. Reviews significantly influence product reputation, sales rankings, and consumer trust, and even small changes in ratings can lead to substantial financial gains or losses for businesses. Driven by profit-oriented incentives, marketing managers and dishonest sellers increasingly fabricate fake reviews to manipulate product reputation, making online review systems prime targets for abuse [3]. The rapid and widespread propagation of such deceptive reviews misleads consumers' purchasing decisions and threatens trust and the sustainable development of the e-commerce ecosystem.

Although considerable progress has been made in fake review detection, most current methods rely on manually constructed learning pipelines in Deep Learning (DL) Architectures [4]. Model selection is generally heuristic, with researchers assessing only a limited range of classifiers or neural architectures, while feature engineering and selection are predominantly handcrafted and tailored to specific domains. Hyperparameter tuning is frequently conducted manually or through constrained grid searches, and comparative evaluations typically emphasize fixed models rather than a comprehensive exploration of the model space. As a result, these approaches often exhibit limited robustness, poor reproducibility, and weak adaptability across various platforms, domains, or languages. While Automated Machine Learning (AutoML) has shown significant promise in minimizing human bias and optimizing end-to-end learning pipelines in other fields [6], its application to fake review detection remains largely uninvestigated.

To address this gap, this paper proposes an AutoML-based framework for fake review detection that automates model selection, feature optimization, and hyperparameter tuning in an end-to-end manner. The proposed framework integrates textual, sentiment-based, and review-related features within a unified automated optimization process, enabling adaptive and data-driven learning across diverse datasets and domains. Extensive experiments against representative traditional machine learning and deep learning baselines demonstrate that the proposed AutoML approach achieves competitive or superior performance with improved robustness and reproducibility [7]. By minimizing manual intervention and tuning effort, the framework offers a scalable and deployment-ready solution for modern e-commerce platforms facing increasingly sophisticated fake review generation techniques.

II. Related Work

Online reviews significantly influence consumer decision-making on e-commerce platforms, making them a prime target for manipulation through fake or deceptive reviews [8]. Fake review detection has therefore attracted substantial research attention, with approaches evolving from heuristic rules to advanced machine learning and deep learning models.

Early studies primarily relied on handcrafted features and traditional machine learning classifiers, such as Support Vector Machines and Random Forests [10], [11]. These approaches typically use lexical, syntactic, and sentiment-based features

This work was supported by the University of Salerno under the project "Metodologie di soft computing per affrontare il problema della disinformazione nel settore dei beni culturali."

extracted from review text, along with reviewer behavior indicators such as rating deviation and review frequency. Although effective to some extent, such methods require extensive feature engineering and often fail to generalize across domains and datasets .

To improve detection performance, sentiment-aware methods have been widely explored. Aspect-based sentiment analysis has shown that fake reviews often exhibit abnormal sentiment distributions across product aspects, leading to improved detection accurcy across heterogeneous product categories [12]. Ontology-driven sentiment analysis further enhances interpretability by integrating linguistic and semantic knowledge, particularly in scenarios with limited labeled data [13] . However, these approaches remain sensitive to domain-specific feature design.

Recent research has increasingly adopted deep learning models which enhanced detection performance. For example Li et al. built a fake review detector based on CNN [14] and another approach CNN [15] with the user reviewing behavior and social relations embedding method used fruad review detection. More effectively, an LSTM-based approach proposed by Wang et al.[24] around accuracy 89% and Aslam et al.[32]used BiLSTM approaches for Fake review detection. Moreover, Aspect-aware deep neural networks [17]further reduce noise by focusing on salient opinion targets and sentiments, achieving superior performance on benchmark datasets. Transformer-based models such as BERT [18]and RoBERTa (Robustly Optimized BERT Approach) [4]have also achieved state-of-the-art results but require careful hyperparameter tuning and substantial computational resources.

A. *Multimodal for fake review detection*

Moreover, recent research introduce the effective fake review detection, resulting in an increased interest in multimodal techniques that use many sources of information. Multimodal fake review detection often integrates textual features with reviewer behavior, temporal patterns, statistical metadata, and, in certain circumstances, visual or contextual clues to better detect fraudulent signals.Chen et al. [5] introduced "PKU Real20M", a large-scale e-commerce benchmark designed for cross-domain retrieval between products and proposed a two-stage framework: a multimodal pretraining pipeline incorporating standard vision–language objectives and a Query-driven Cross-Domain retrieval framework that aligns query representations with fused, visual, and textual features via contrastive learning while using reconstruction losses to prevent text modality dominance.In another research, Li and Chen [19]introduced a deep neural architecture that combines BERT-based textual representations with handmade reviewer and temporal characteristics, demonstrating that multimodal feature fusion considerably increases detection accuracy over text-only baselines. Similarly, Jian et al. [20] developed a metadata-aware multimodal system that combines transformer-based text embeddings with rich review metadata via attention-based fusion, resulting in high robustness and generalization performance on restaurant review datasets. Beyond text–metadata fusion, Hou et al. [21] extended multimodality by incorporating visual context and review intent, highlighting that multimodal representations can capture nuanced deceptive intentions overlooked by binary text classifiers . These findings collectively indicate that multimodal learning enables complementary feature interactions, reduces ambiguity in deceptive content, and substantially enhances detection performance, thereby establishing multimodal fusion as a critical direction in contemporary fake review detection research. More recently, Gedara et al. [22] introduced a fuzzy-based multimodal framework that jointly analyzes text, images, and their semantic consistency using LSTM, ANFIS, and a fuzzy inference system, enabling both improved detection performance and interpretability by explicitly reasoning over cross-modal inconsistencies. Collectively, these studies demonstrate that multimodal fusion not only enhances detection accuracy by leveraging complementary modalities but also improves robustness and explainability, positioning multimodal learning as a critical paradigm in advanced fake review and deceptive content detection research.

B. *Automated Machine Learning*

While multimodal deep learning models have demonstrated strong performance in detecting fake reviews and misinformation, they often require extensive manual effort in model selection, feature engineering, and hyperparameter tuning. To address these limitations, recent studies have increasingly explored Automated Machine Learning (AutoML) as a means to streamline and optimize the machine learning pipeline for fake information detection. AutoML aims to automate critical stages of model development, particularly algorithm selection and hyperparameter optimization, thereby reducing human bias and improving reproducibility. This objective is formally captured by the Combined Algorithm Selection and Hyperparameter (CASH) optimization problem, which jointly determines the optimal learning algorithm and its corresponding hyperparameters, rather than treating these steps sequentially as in traditional machine learning workflows [23].Recent AutoML frameworks have been extended to support multimodal learning by automatically integrating heterogeneous data sources such as text, images, and structured features. AutoGluon adopts an ensemble-driven strategy that combines modality-specific models through stacking and late fusion, enabling robust multimodal learning without exhaustive hyperparameter tuning [24], [25]. AutoKeras focuses on deep-learning-based AutoML and supports multimodal inputs through automated neural architecture search over text and image modalities [26]. H2O AutoML emphasizes scalable ensemble construction over diverse learners and provides limited multimodal capability through automatic feature encoding and stacking [27], [28]. In contrast, TPOT employs evolutionary algorithms to optimize complete pipelines and has been explored for unimodal settings by evolving heterogeneous feature-processing workflows [29]. These techniques demonstrate that AutoML based unimodal and multimodal learning can effectively balance accuracy, robustness, and computational efficiency.

In the context of fake information detection, AutoML has shown particular promise in handling heterogeneous and high-dimensional data, such as tabular metadata, visual features, and derived statistical attributes. Gedara et al. [6] proposed an explainable multimodal fake news detection framework that integrates AutoML with prompt-based large language models, where AutoML is employed in the image-based pipeline to automatically select and ensemble classifiers over extracted visual features and another approch TPOT used to detect forgery images using adaptive neuro-fuzzy inference system [30]. These study demonstrates that AutoML-based pipelines can achieve competitive or superior performance compared to manually designed deep learning models. Similar observations have been reported by Liuliakov et al. [31] demonstrate that ensemble and evolutionary based AutoML frameworks can effectively handle high-dimensional and sparse feature spaces, achieving competitive or superior performance to manually tuned models while improving robustness and generalization in complex detection tasks.In summery, it worth to try investigate AutoML approaches for fake review detection whether AutoML approach able increase accuracy on detection

III. Methodology

This section presents the proposed methodology for fake review detection, which is designed in two progressive stages. The first stage focuses on text-only fake review detection using AutoML with tabular representations, while the second stage extends the framework to a multimodal AutoML-based architecture that integrates textual, visual, and numerical metadata. The overall workflow of both approaches is illustrated in Figures 1 and 2.

A. *Approach I: Text-Based Fake Review Detection Using AutoML*

In the first approach, as illustrated in Figure 1, only textual reviews are employed, and an AutoML-based tabular learning strategy is used for fake review detection. Each review is initially subjected to standard text preprocessing steps, including tokenization, lowercasing, stop-word removal, and lemmatization. The cleaned text is then transformed into numerical representations using word-encoding techniques and embedding-based vectorisation, yielding fixed-length (covering 90% of the data) feature vectors suitable for tabular learning.

The resulting tabular feature matrix is provided as input to multiple AutoML frameworks, including AutoKeras, AutoGluon, H2O AutoML, and TPOT. These frameworks automatically explore a appropriator ML model from the following Table I, preprocessing strategies, and ensemble configurations without manual intervention. Unlike traditional machine learning pipelines that require handcrafted model selection and hyperparameter tuning, the AutoML systems autonomously identify optimal model configurations by leveraging ensemble learning, stacking, and evolutionary search strategies.

At the end of the approach, the output of this stage is a binary classification indicating whether a review is fake or genuine, based solely on textual content. This text-only

TABLE I
Classification Approaches Used by AutoML Frameworks for Tabular Data

AutoML Framework	Classification Approaches
TPOT	Linear (LogReg, SVM); Distance-based (KNN); Tree-based (DT, RF, ET); Boosting (GB, AdaBoost)
AutoGluon	GBDT (LightGBM, CatBoost, XGBoost); Bagging (RF, ET); Linear models; KNN; Neural networks (MLP)
AutoKeras	Neural networks (MLP); Embedding-based DL; Neural architecture search
H2O AutoML	GBM; XGBoost; Distributed RF; GLM; Deep learning (MLP); RuleFit

AutoML pipeline serves as a strong baseline and allows us to assess the effectiveness of automated model discovery before incorporating additional modalities.

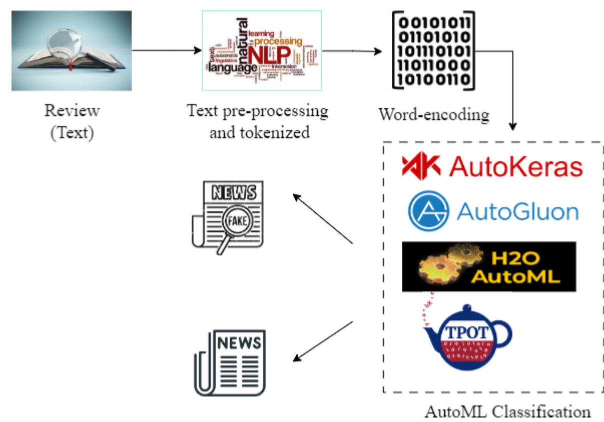

Fig. 1. Proposed methodology for text-based fake reviews detection

B. *Approach II: Multimodal Fake Review Detection Using AutoML*

The second approach extends the text-based framework by incorporating multimodal information, including review text, review-related numerical metadata, and associated images. This design aims to capture complementary signals that are often overlooked by unimodal models.As shown in Figure 2, the review data are decomposed into three modalities:

Text & Numerial data Modality: Review text is preprocessed and encoded following the same pipeline described in Approach I.For Numerical metadata includes structured attributes such as rating scores, posting time, reviewer activity statistics, and other numerical indicators are normalized and included as tabular features.

Image Modality: Review images undergo image preprocessing operations, including resizing and normalization. Visual features are extracted either through pre-trained convolutional networks or automated feature extractors supported by AutoML frameworks.

In the multimodal learning and classification, image and tabular inputs are feed the to AutoML approches which are supported for multimodal classifcation, AutoGloun and AutoKeras. In particular, AutoGluon AutoML automatically identifies the input modalities, selects suitable pretrained encoders for each modality, and learns a joint representation through a late-fusion neural architecture, where the outputs of the modality-specific encoders are combined using a fusion module such as a multilayer perceptron or transformer. The fused representation is then utilized by a classification head to predict the review label. In contrast, AutoKeras adopts a fully neural approach in which separate input branches are constructed for each modality, and their learned feature representations are merged through feature-level fusion as part of a neural architecture search process. Both frameworks enable end-to-end multimodal learning by automating model selection, representation fusion, and optimization, thereby facilitating the effective exploitation of complementary textual, structural, and visual cues in multimodal fake review detection.

Finally, the multimodal AutoML system outputs a binary decision indicating whether the review is fake or genuine. By leveraging heterogeneous data sources and automated optimization, this approach enhances robustness, generalization, and detection accuracy compared to the text-only baseline.

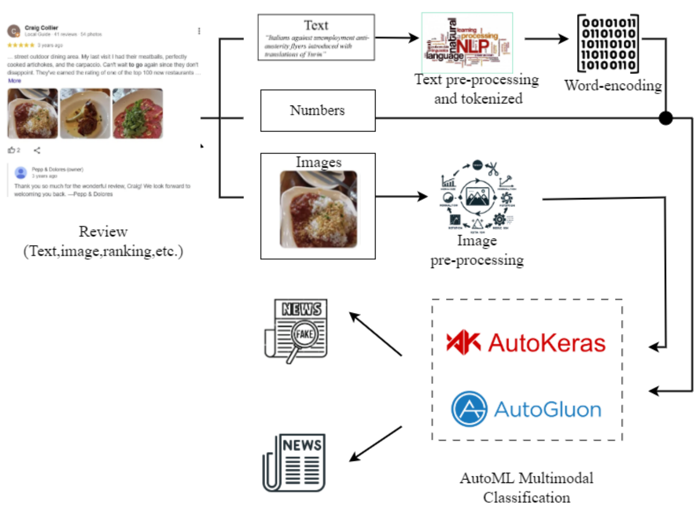

Fig. 2. Proposed AutoML-based multimodal approach for multimedia (text and image) fake reviews detection

IV. NUMERICAL EXPERIMENTS

This section presents the experimental evaluation of the proposed approach for forgery image detection. The experiments were conducted on four publicly available datasets, as detailed in the next subsection.

A. Experimental Setup

Hardware choice and resource specifications: All experiments were implemented in Python 3.10 and executed on a system equipped with a 13th Gen Intel® Core™ i3-1360P 2.10 GHz CPU and an NVIDIA GeForce RTX 4070 GPU.

Data Partitioning: The dataset was randomly shuffled and split into 80% for training and 20% for testing. Each experiment was conducted ten times, and the average performance was reported. Results from 5-fold cross-validation, which yielded comparable outcomes, are omitted for conciseness.

B. Baselines

Our proposed approach is compared against the following state-of-the-art methods:

- **Hierarchical Fusion Attention Network (HFAN)**learns joint user–product representations through multiattention mechanisms and orthogonal decomposition, modeling reviews as user–product relations and encoding them using the TransH knowledge graph embedding model [34].
- **Fine-tuned RoBERTa (fakeRoBERTa)**: A transformer-based fake review detection model which trained on a balanced Amazon review dataset consisting of 20,000 original (human-written) reviews and 20,000 GPT-2–generated fake reviews, distinguish computer-generated fake reviews from genuine ones [35].
- **Aspect-Rating Local Outlier Factor (AR-LOF)**: The approach computes aspect-level sentiment ratings using a lexicon-based method and applies tensor factorization to address data sparsity [28].
- **Joint-bEhavior-and-Social-relaTion-infERable (JESTER)**:A unified latent space embeds user behavior, user–item social relations, and review text by modeling co-occurrence-based behaviors via likelihood maximization, capturing social context through random-walk-based user–item graphs, and encoding reviews with a CNN using CBOW embeddings, enabling effective detection under sparse user histories [15].
- **PKU Real20M**: Multimodal aprroach uses a shared vision-text backbone , a Frame-level Temporal Transformer to aggregate video frames, and a fusion encoder to integrate visual and textual features archiving 70.71% accuracy [5].
- **BERT-based Sentiment Classification(BERT-SC)**: A sentiment analysis framework applied to Amazon Electronics product reviews using multiple machine learning classifiers. Among all evaluated models, the BERT classifier achieved the best performance with an accuracy of 0.8893 on the Amazon Electronics dataset, outperforming traditional machine learning baselines [16].

C. Datasets

The proposed method was evaluated using multiple publicly available datasets widely adopted in fake review detection research. These datasets cover real-world, synthetic, and computer-generated reviews, enabling a comprehensive assessment across diverse review generation scenarios.

- **Amazon Electronics Reviews (AER)**: This dataset is derived from the Amazon 5-core product review collection in the Electronics category, spanning from May 1996 to July 2014. It contains 1,689,188 reviews, where each reviewer and each product has at least five associated reviews. The dataset includes rich metadata such as product identifiers, reviewer information, ratings, review text, timestamps, and helpfulness scores. It is commonly used for sentiment analysis, review usefulness prediction, and fake review detection tasks. The dataset was obtained from the Amazon Product Data repository curated by McAuley et al. [32], [33].

- **Fake Reviews Dataset (FRD)**: This dataset consists of 40,000 reviews, evenly split between 20,000 genuine reviews and 20,000 fake reviews. The real reviews were sampled from the Amazon Review Data (2018) k-core subset, ensuring dense user–item interactions. The fake reviews were generated using large language models, with GPT-2 selected due to its superior performance compared to ULMFiT in terms of training loss, validation loss, perplexity, and training efficiency. This dataset enables controlled experimentation on computer-generated fake reviews and has been widely used in recent fake review detection studies [35].
- **Yelp NYC Restaurant Reviews (Yelp NYC)**: This dataset contains reviews for 44,023 restaurants located in New York City. Each restaurant entry includes review text, metadata, and contextual information such as restaurant names, locations, and visual references. The dataset is frequently used in recommendation systems and opinion analysis research and provides a realistic large-scale benchmark for detecting deceptive restaurant reviews [38].
- **TripAdvisor Restaurant Reviews (Trip)**: This dataset is based on the benchmark collection introduced by Ott et al. [37]. Genuine reviews were collected from the 20 most popular hotels in Chicago on TripAdvisor, while deceptive reviews were artificially generated by crowdworkers recruited via Amazon Mechanical Turk. Only review texts and class labels are provided, without reviewer metadata or temporal information, making this dataset suitable for text-based fake review detection.
- **Synthetic E-Commerce Relational Dataset**: This dataset consists of simulated transactional and behavioral data designed to mimic real-world e-commerce environments. It includes customer demographics, transaction records, product categories, and purchasing behavior patterns. The dataset supports advanced modeling of consumer behavior and complements text-based review datasets by enabling experiments on structured and relational data in fake review detection research [39].

D. Performance measures

The proposed approach's performance has been assessed using accuracy, Precision, Recall and F1-score:

E. Impact of AutoML only Text-based review

To evaluate the effectiveness of the proposed text-based AutoML framework described in Approach I, the TPOT AutoML framework was employed as the primary model selection and hyperparameter optimization mechanism for fake review detection.In this setup, only tabular representations derived from textual reviews were used as input, without incorporating any visual or metadata features. Analysis of the pipelines generated by TPOT shows that the *Multilayer Perceptron (MLPClassifier)* was the most frequently selected best-performing model across datasets. In addition to model selection, TPOT automatically optimized the associated hyperparameters of the MLPClassifier to maximize cross-validation performance. The representative hyperparameter configurations selected by TPOT are summarized in Table II.

TABLE II
TPOT-Selected Hyperparameters of the MLPClassifier

Hyperparameter	**Value**
Hidden layer size	(100,) or (100, 50)
Activation function	ReLU
Solver	Adam
Learning rate	Adaptive
Initial learning rate	0.001
Regularization (α)	0.0001
Maximum iterations	300–500
Early stopping	True
Validation split	0.1
Random state	42

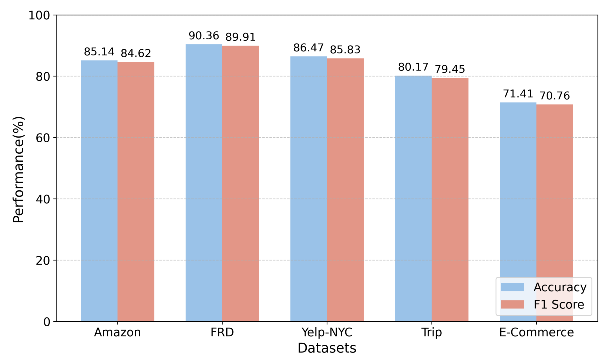

Fig. 3. Performance comparison of the TPOT-selected models across different datasets in terms of Accuracy and F1-score.

Figure 3 presents the classification performance of the TPOT-based approach across five benchmark datasets. The highest accuracy and F1-score are achieved on the FRD dataset, indicating TPOT's strong effectiveness in distinguishing machine-generated fake reviews from genuine ones. Comparable performance is observed on the Amazon and Yelp-NYC datasets, demonstrating stable generalization on large-scale real-world reviews. In contrast, lower performance on the TripAdvisor and Synthetic E-Commerce datasets reflects the increased difficulty caused by limited contextual cues and higher data complexity. Overall, the close correspondence between accuracy and F1-score across all datasets confirms that the TPOT-selected models maintain a balanced trade-off between precision and recall.

Following the same procedure in addition to TPOT, other AutoML frameworks were also evaluated under the same text-only experimental setting, and the corresponding results are illustrated in Figure 4. Among the evaluated frameworks, AutoGluon consistently demonstrates strong performance across all datasets, achieving higher accuracy and F1 Scores, particularly on the FRD and Yelp-NYC datasets. The best performing models selected by each AutoML framework are summarized in Table III. According to the best performance model, all frameworks suggest neural network–based architectures are predominantly selected as optimal models, indicating their effectiveness in capturing semantic and contextual patterns in review text and aligning with current state-of-the-art approaches in fake review detection. However, a consistent performance gap is observed in larger variant reviews and more human-related review datasets, such as Amazon and Trip, where unimodal text-only models exhibit relatively lower performance despite the application of deep learning techniques. This trend suggests inherent limitations of text-only representations and motivates the integration of additional modalities, thereby justifying the proposed multimodal AutoML approach in subsequent experiments.

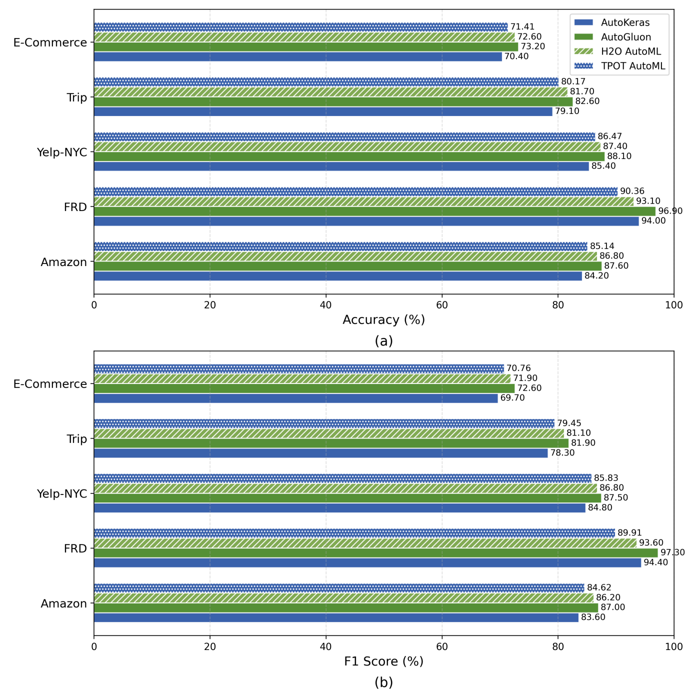

Fig. 4. Performance comparison of different AutoML frameworks under the text-only fake review detection setting across five datasets. (a) reports classification accuracy, (b) presents the corresponding F1-score.

TABLE III
BEST-PERFORMING MODELS SELECTED BY DIFFERENT AUTOML FRAMEWORKS IN THE TEXT-ONLY SETTING

AutoML Framework	Best Model
TPOT	MLPClassifier
AutoGluon	Multilayer Perceptron (MLP)
H2O AutoML	XGBoost
AutoKeras	Text Embedding + Dense NN

F. Impact of Multimodal AutoML only Metadata review

As explained in the second approach, review data is considered for additional data. In this approach, the AutoML

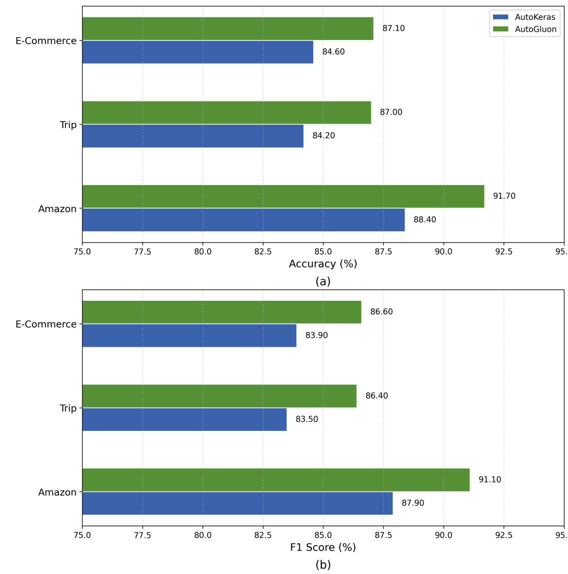

Fig. 5. Performance comparison of multimodal AutoML frameworks for fake review detection. (a) Classification accuracy and (b) F1-score achieved by AutoKeras and AutoGluon.

frame with a multimodal approach, AutoGluon, and AutoKeras are considered, and the Trip, E-commerce, and Amazon datasets are considered due to the fact that only these datasets contain metadata regarding the reviews. Figure5(a) and Figure5(b) illustrate the classification accuracy and F1-score achieved by AutoKeras and AutoGluon, respectively. Across all datasets and evaluation metrics, AutoGluon consistently outperforms AutoKeras, demonstrating superior robustness and generalization capability in multimodal fake review detection. The best performance gain from the AutoGloun Multimodal configuration, where the text encoder uses the HuggingFace model, and a ResNet-based approach is used for images. On the Amazon dataset, AutoGluon achieves the highest performance with an accuracy of 91.7% and an F1-score of 91.1%, surpassing AutoKeras by margins of 3.3% in accuracy and 3.2% in F1-score. This improvement indicates that AutoGluon is particularly effective at leveraging the rich multimodal information in Amazon reviews, including text, images, and extensive metadata. Similar performance trends are observed on the TripAdvisor and E-Commerce datasets. On TripAdvisor, AutoGluon achieves 87.0% accuracy and 86.4% F1-score, outperforming AutoKeras by approximately 2.8% in both metrics. On the E-Commerce dataset, AutoGluon achieves 87.1% accuracy and an F1-score of 86.6%, again demonstrating consistent gains over AutoKeras. Moreover, resutls of unimodal over mutlimodal across three datasets is compared in Table IV. The results clearly demonstrate that incorporating multimodal information leads to significant and consistent performance improvements over unimodal learning. For the Amazon dataset, the multimodal model improves accuracy from 87.6% to 91.7% and F1-score from 0.870 to 0.911, indicating that the combination of text, images, and metadata captures richer behavioral and contextual cues than

TABLE IV
PERFORMANCE COMPARISON OF UNIMODAL AND MULTIMODAL AUTOML APPROACHES

Dataset	Unimodal (Tabular)				Multimodal			
	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
Amazon	0.876	0.870	0.870	0.870	**0.917**	**0.911**	**0.911**	**0.911**
Trip	0.826	0.819	0.819	0.819	**0.870**	**0.864**	**0.864**	**0.864**
E-Commerce	0.732	0.726	0.726	0.726	**0.871**	**0.866**	**0.866**	**0.866**

tabular features alone.

The performance gains are even more pronounced for the E-Commerce dataset, where accuracy increases from 73.2% in the unimodal setting to 87.1% in the multimodal configuration. This substantial improvement suggests that unimodal tabular features are insufficient to model complex deception patterns in heterogeneous e-commerce environments, whereas multimodal representations enable more discriminative and robust fake review detection. Similar trends are observed for the TripAdvisor dataset, where multimodal learning improves accuracy by over 4% and F1-score by 4.5% compared to the unimodal.Overall, these results confirm that multimodal AutoML significantly outperforms unimodal approaches, particularly in datasets with diverse and information-rich review metadata. The findings emphasize the importance of jointly modeling textual, visual, and structured signals and demonstrate the effectiveness of AutoGluon AutoML as a scalable and automated solution for real-world fake review detection systems.

G. Baseline Comparison

Finally, the proposed AutoML-based fake review detection framework is compared with representative state-of-the-art baseline approaches, including transformer-based models, sentiment-driven methods, behavior-aware frameworks, and multimodal learning techniques. Table V summarizes the performance of these baselines across commonly used fake review detection datasets. The comparison highlights the effectiveness of automated model selection and multimodal feature integration in improving detection accuracy and robustness across diverse review domains. Overall, the results in Table V demonstrate that the proposed AutoML-based framework consistently outperforms traditional sentiment-based, behavior-driven, and unsupervised fake review detection approaches across multiple datasets. While transformer-based models such as fine-tuned RoBERTa achieve high accuracy on controlled or synthetic datasets, their performance is highly dependent on data distribution and task-specific fine-tuning. In contrast, the proposed AutoML framework generalizes well across real-world datasets by automatically selecting optimal model architectures and hyperparameters. The strong performance of AutoGluon particularly on large-scale datasets such as FRD and Yelp-NYC, confirms the effectiveness of automated multimodal learning for scalable and robust fake review detection.

TABLE V
PERFORMANCE COMPARISON WITH STATE-OF-THE-ART FAKE REVIEW DETECTION METHODS

Dataset	Approach	Accuracy (%)
Amazon	BERT-SCR [16]	88.93
	PKU Real20M [5]	70.71
	AutoML-Text (AutoGloun)	87.60
	AutoML-Multimodal (AutoGloun)	**91.70**
Trip	AR-LOF [28]	79.00
	User-Product Attention Model [36]	84.78
	AutoML-Text (AutoGloun)	82.60
	AutoML-Multimodal (AutoGloun)	**87.00**
Yelp-NYC	HFAN [34]	85.40
	JESTER [15]	85.00
	AutoML-Text (AutoGloun)	**88.10**
FRD	fakeRoBERTa [35]	97.00
	AutoML-Text (AutoGloun)	**96.90**
E-Commerce	PKU Real20M [28]	70.71
	AutoML-Text (AutoGloun)	73.20
	AutoML-Multimodal (AutoGloun)	**87.10**

V. CONCLUSION AND FUTURE WORK

An AutoML-based framework offers an effective, scalable solution for fake review detection by eliminating the need for manual model design, feature engineering, and hyperparameter tuning. Automated model selection and optimization enable the proposed approach to achieve competitive and robust performance across diverse datasets. The AutoML framework is evaluated in two stages: a text-only pipeline and an extended multimodal setting that integrates textual, numerical metadata, and visual information. Experiments on benchmark datasets show that AutoML frameworks identify effective model architectures, and improve performance, increasing accuracy from 87.6 to 91.7 on the Amazon dataset and from 73.2 to 87.1 on the E-Commerce dataset. Compared with transformer-based, sentiment-driven, and behavior-aware, the proposed multimodal AutoML framework demonstrates strong performance in fake review detection, especially in content multimedia form. The findings suggest a further deep-down multimodal approach in future work and investigate multilingual extensions, large language model integration, and explainable AutoML for adversarial review settings.

ACKNOWLEDGMENT

This work was supported by the University of Salerno under the project “Metodologie di soft computing per affrontare il problema della disinformazione nel settore dei beni culturali.”

REFERENCES

[1] R Xu,Y Xia, K Wong, and W Li. ”Opinion Annotation in On-line Chinese Product Reviews.” In LREC, vol. 8, pp. 26-30. 2008.

[2] A. K Samha, Y. Li, and J. Zhang, “Aspect-based opinion extraction from customer reviews,” 2014, arXiv:1404.1982. [Online]. Available: http://arxiv.org/abs/1404.1982.

[3] S Vosoughi, D Roy , S Aral,” The spread of true and false news online”,in science. 2018 Mar 9,359(6380):1146-51.

[4] R. Mohawesh,X Shuxiang ,N.T Son ,O Robert , S Matthew ,J Yaser , and M Sumbal , ”Fake Reviews Detection: A Survey,” in IEEE Access, vol. 9, pp. 65771-65802, 2021, doi: 10.1109/ACCESS.2021.3075573.

[5] Y. Chen, H. Zhong, X. He, Y. Peng, and L. Cheng,"Real20M: A large-scale e-commerce dataset for cross-domain retrieval,"in *Proc. 31st ACM Int. Conf. Multimedia (MM '23)*, Ottawa, ON, Canada, Oct. 2023, pp. 4939-4948, doi: 10.1145/3581783.3612408.

[6] T. M. H. Gedara, R. Gaeta, S. Tomasiello and R. El Shawi, "Toward Explainable Multimodal Fake News Detection: A Fusion of Prompt Learning and Automated Machine Learning," in IEEE Access, vol. 13, pp. 188150-188169, 2025, doi: 10.1109/ACCESS.2025.3627504.

[7] Reddy, Tummalapalli Sudhamsh. "Ethical Dimensions of AutoML: Addressing Bias, Transparency, and Responsible Development." Journal of Computer Science and Technology Studies 7, no. 5 (2025): 981-988.

[8] Paul, Himangshu, and N Alexander . "Fake review detection on online E-commerce platforms: a systematic literature review." Data Mining and Knowledge Discovery 35, no. 5 (2021): 1830-1881.

[9] Q. Li, Q. Wu, C. Zhu, J. Zhang, and W. Zhao, "An inferable representation learning for fraud review detection with cold-start problem," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2019, pp. 1–8.

[10] E. F. Cardoso, R. M. Silva, and T. A. Almeida, "Towards automatic filtering of fake reviews," Neurocomputing, vol. 309, pp. 106-116, Oct. 2018.

[11] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in Proc. 49th Annu.Meeting Assoc. Comput. Linguistics: Human Lang. Technol., vol. 1, 2011, pp. 309-319.

[12] Z Shunxiang, A Zhu, G Zhu , Z Wei , and K Li, "Building fake review detection model based on sentiment intensity and PU learning." IEEE Transactions on Neural Networks and Learning Systems 34, no. 10 (2023): 6926-6939.

[13] M Jaffar, U J Muhammad , T J Muhammad , K H Muhammad ,R Zeeshan , M A Shafqat , and A H J Rao , "Ontology-based sentiment analysis for real-time product reputation modeling", Spectrum of Engineering Sciences (2025): 648-667.

[14] Li, L., Ren, W., Qin, B., & Liu, T. (2015). Learning Document Representation for Deceptive Opinion Spam Detection. In M. Sun, Z. Liu, M. Zhang, & Y. Liu (Eds.), Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data (pp. 393–404). Springer International Publishing. https://doi.org/10.1007/978-3-319-25816-4_32.

[15] Q. Li, Q. Wu, C. Zhu, J. Zhang and W. Zhao, "An Inferable Representation Learning for Fraud Review Detection with Cold-start Problem," 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 2019, pp. 1-8, doi: 10.1109/IJCNN.2019.8852437.

[16] A. Hamza, K. B. Majeed, M. Rashad, and A. Jaffar, "An integrated approach for Amazon electronic products reviews by using sentiment analysis,"*Bull. Bus. Econ.*, vol. 13, no. 2, pp. 142–153, 2024.

[17] G. Bathla, P. Singh, R. K. Singh, E. Cambria, and R. Tiwari, "Intelligent fake reviews detection based on aspect extraction and analysis using deep learning," Neural Computing and Applications, vol. 34, no. 22, pp. 20213–20229, Nov. 2022, doi: 10.1007/s00521-022-07531-8.

[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, arXiv:1810.04805. [Online]. Available: http://arxiv.org/abs/1810.04805.

[19] X. Li and L. Chen, "Fake Review Detection Using Deep Neural Networks with Multimodal Feature Fusion Method," in *Proceedings of the IEEE Conference on Big Data and Smart Computing*, Inner Mongolia University, China, 2022, pp. 1–8.

[20] Y. Jian, X. Chen, X. Wang, Y. Liu, X. Chen, X. Lan, W. Wang, and H. Wang, "A Metadata-Aware Detection Model for Fake Restaurant Reviews Based on Multimodal Fusion," *Neural Computing and Applications*, vol. 37, pp. 475–498, 2025, doi: 10.1007/s00521-024-10647-8.

[21] J. Hou, Z. Tan, S. Zhang, Q. Hu, and P. Wang, "Detecting Fake Review Intentions in the Review Context: A Multimodal Deep Learning Approach," *Electronic Commerce Research and Applications*, vol. 70, Art. no. 101485, 2025, doi: 10.1016/j.elerap.2025.101485.

[22] T. M. H. Gedara, V. Loia, and S. Tomasiello, "A fuzzy-based multimodal approach for interpretable fake news detection," *Applied Soft Computing*, vol. 179, Art. no. 113277, 2025, doi: 10.1016/j.asoc.2025.113277.

[23] Thornton, C., Hutter, F., Hoos, H.H. and Leyton-Brown, K., 2013, August. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 847-855).

[24] N. Erickson, J. Mueller, A. J. Smola, and M. Li,"AutoGluon: A Toolkit for Automated Machine Learning,"arXiv:2003.06505, 2020.

[25] X. Shi, J. Mueller, N. Erickson, M. Li, and A. J. Smola,"Benchmarking Multimodal AutoML for Tabular Data with Text Fields,"in *Advances in Neural Information Processing Systems*, 2021.

[26] H. Jin, F. Chollet, Q. Song, and X. Hu,"AutoKeras: An AutoML Library for Deep Learning,"*J. Mach. Learn. Res.*, vol. 24, pp. 1–6, 2023.

[27] E. LeDell and S. Poirier,"H2O AutoML: Scalable Automatic Machine Learning,"in *Proc. ICML Workshop on Automated Machine Learning*, 2020.

[28] L. You, Q. Peng, Z. Xiong, D. He, M. Qiu, and X. Zhang, "Integrating aspect analysis and local outlier factor for intelligent review spam detection," *Future Generation Computer Systems*, vol. 102, pp. 163–172, 2020.

[29] R. S. Olson and J. H. Moore,"TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning," in *Proc. ICML Workshop on Automated Machine Learning*, 2016.

[30] T. M. H Gedara, V. Loia and S. Tomasiello, "Detecting Fake Images Using Neuro-Fuzzy Inference Systems: A Brief Comparative Analysis," 2024 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Yokohama, Japan, 2024, pp. 1-7, doi: 10.1109/FUZZ-IEEE60900.2024.10612033.

[31] A. Liuliakov, L. Hermes, and B. Hammer,"AutoML technologies for the identification of sparse classification and outlier detection models,"*Applied Soft Computing*, vol. 133, Art. no. 109942, 2023,doi: 10.1016/j.asoc.2022.109942.

[32] S. AlZu'bi, A. Alsmadiv, S. AlQatawneh, M. Al-Ayyoub, B. Hawashin and Y. Jararweh, "A Brief Analysis of Amazon Online Reviews," 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), Granada, Spain, 2019, pp. 555-560, doi: 10.1109/SNAMS.2019.8931816

[33] Y. Hou, J. Li, Z. He, A. Yan, X. Chen, and J. McAuley,"Bridging language and items for retrieval and recommendation,"*arXiv preprint arXiv:2403.03952*, 2024.

[34] C. Yuan, W. Zhou, Q. Ma, S. Lv, J. Han, and S. Hu, "Learning review representations from user and product level information for spam detection," in Proc. IEEE Int. Conf. Data Mining (ICDM), Nov. 2019, pp. 1444–1449.

[35] J. Salminen, C. Kandpal, A. M. Kamel, S. Jung, and B. J. Jansen,"Creating and detecting fake reviews of online products,"*J. Retailing Consum. Serv.*, vol. 64, Art. no. 102771, 2022, doi:10.1016/j.jretconser.2021.102771.

[36] Y. Tian, M. Mirzabagheri, P. Tirandazi, and S. M. H. Bamakan, "A nonconvex semi-supervised approach to opinion spam detection by rampone class SVM," Inf. Process. Manage., vol. 57, no. 6, Nov. 2020, Art. no. 102381.

[37] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics: Human Language Technologies (ACL-HLT)*, Portland, OR, USA, 2011, vol. 1, pp. 309–319.

[38] Circiu, Andreea (2021). yelp. figshare. Dataset. https://doi.org/10.6084/m9.figshare.14888283.v1

[39] M. R. Islam, M. Hossain, M. Alam, M. M. Khan, M. M. K. Rabbi, M. F. Rabby, K. K. Bishnu, B. C. Das, and M. T. R. Tarafder,"Leveraging machine learning for insights and predictions in synthetic e-commerce data in the USA,"*Journal of Ecohumanism*, vol. 4, no. 2, pp. 2394–2420, 2025, doi: 10.62754/joe.v4i2.6635.