International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

RATING-AWARE MACHINE LEARNING FOR SARCASM DETECTION IN TEXT–IMAGE MULTIMODAL REVIEWS

F. Antony Joseph, Dr. V. Elavazhahan*

Research Scholar, Department of Computer and Information Science, Annamalai University, Tamil Nadu.

Assistant Professor, Department of Computer Science, Government Arts and Science College, Vadalur, Tamil Nadu.

anto2422@gmail.com, elavalau@yahoo.com

Abstract

Sarcasm remains one of the most challenging aspects of sentiment analysis, as surface-level expressions often contradict the true intent of the reviewer. In online reviews, this challenge is amplified when textual content, visual memes/images, and numerical ratings conflict, misleading both users and recommendation systems. This paper introduces a rating-aware multimodal machine learning framework for sarcasm detection that integrates textual, visual, and rating inconsistency features. Unlike deep learning-based methods, the proposed approach leverages advanced machine learning concepts, including meta-feature projection, contradiction-driven feature extraction, and rating-boosted support vector machines (M^3-SVM). The framework explicitly models sentiment–rating discrepancies and projects multimodal features into a contradiction space to enhance discriminative power. Experimental evaluations demonstrate improved accuracy and robustness over baseline classifiers, highlighting the framework's potential to strengthen trust in review-driven decision-making and improve recommendation reliability.

Keywords: Sarcasm Detection; Multimodal Reviews; Rating Inconsistency; Machine Learning; Meta-Feature Projection; Support Vector Machines; Recommendation Systems

1. INTRODUCTION

The rapid growth of e-commerce and social media platforms has made online reviews a critical factor in shaping consumer decisions. However, the reliability of these reviews is frequently undermined by sarcastic expressions, where the literal sentiment differs from the intended meaning. For example, a review stating *"Great, another phone that dies in two hours 🙄"* uses a positive term ("great") while expressing a negative experience. Such sarcastic cues can mislead traditional sentiment analysis systems, propagating errors into recommendation engines that rely heavily on user-generated reviews.

Received: August 15, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 1. Example of Sarcasm in Online Review

A review stating with a 5-star rating demonstrates the contradiction between positive ratings and negative sentiment, highlighting sarcasm cues as shown in Fig 1. Detecting sarcasm is inherently complex due to its reliance on contextual contradictions—not only within the text but also across modalities. Visual memes, emojis, and attached images often reinforce sarcasm, while rating inconsistencies (e.g., a 5-star rating paired with negative text) serve as strong implicit indicators. Despite these cues, most existing approaches either focus solely on text or employ deep learning models that demand high computational cost and large annotated datasets, limiting their practicality for real-time systems.

This paper proposes a rating-aware machine learning framework that combines textual, visual, and numerical rating signals without resorting to deep neural architectures. The framework introduces a Meta-Feature Projection and Contradiction Modeling approach that emphasizes disagreement between modalities—a defining trait of sarcasm. A novel Multimodal Meta-Feature Support Vector Machine (M^3-SVM) is designed with a contradiction-aware kernel, enabling more precise classification of sarcastic reviews. The major contributions of this work are as follows:

1. **Rating-Aware Preprocessing Pipeline:** Development of a multimodal preprocessing pipeline that integrates text, images, and rating inconsistencies to highlight implicit sarcasm cues.
2. **Meta-Feature Projection:** Introduction of a hybrid dimensionality reduction approach using Kernel PCA and Canonical Correlation Analysis to transform features into a contradiction-sensitive space.
3. **Novel Machine Learning Classifier (M^3-SVM):** Design of a Support Vector Machine with a contradiction-aware kernel, boosted by rating–sentiment discrepancy signals, for robust sarcasm classification.

Received: August 15, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2. RELATED WORKS

Sarcasm detection has been approached through diverse methodologies combining machine learning and deep learning models. A study employed logistic regression, ridge regression, linear support vector machines, Bi-LSTM, and BERT models on a large corpus of social media comments after NLP-based preprocessing, but its limitation was the heavy reliance on large-scale annotated datasets that are resource intensive to build [1]. Another work proposed an Intelligent ML-based sarcasm detection and classification (IMLB-SDC) framework using TF-IDF for feature engineering, chi-square and information gain for feature selection, and SVM with PSO-based tuning, though it was constrained by dependence on handcrafted features and optimization complexity [2].

Further exploration on Twitter sarcasm applied multiple classifiers with different preprocessing strategies while also relating sarcasm to irony and cyberbullying, but it lacked robustness across diverse domains [3]. Irony detection was also studied under sentiment analysis where preprocessing and multiple classifiers like SVM, Naïve Bayes, and Random Forest were utilized, yet the approaches struggled with contextual ambiguity [4]. A multi-feature fusion framework fused lexical features with contextual cues using multi-stage classification, although it faced challenges due to data sparsity in microblog texts [5].

A review of sarcasm identification summarized classification techniques, datasets, feature engineering, and algorithms like SVM, Naïve Bayes, and decision trees, but highlighted the absence of standardized benchmark datasets as a major limitation [6]. Another work analyzed hyperbolic cues such as interjections, intensifiers, and elongated words using ML classifiers like SVM and Random Forest, though the model's generalizability across sarcasm types remained limited [7]. An ensemble approach combined outputs of LSTM, CNN-LSTM, MLP, and SVM with AdaBoost, yet the reliance on multiple component models increased computational overhead [8].

Sarcasm in tweets was also studied using machine and deep learning classifiers with various feature extraction techniques, but the models often failed to capture implicit contextual signals [9]. A multimodal method combined Bi-LSTM for text and ResNet for images with a hybrid latent factor framework, though its performance was sensitive to noise in heterogeneous inputs [10]. Sarcasm detection in news headlines used LSTM with regularization and early stopping, but it struggled with subtle and ambiguous sarcastic expressions lacking strong linguistic markers [11].

Another work proposed ITFT-Net, a multimodal image-text fusion Transformer integrating BERT and ResNet-101 with a fusion encoder, though the complexity of multimodal fusion limited its scalability [12]. A study employing handcrafted linguistic features such as lexical diversity and readability with ensemble models and neural networks demonstrated improved classification, but it lacked adaptability to evolving sarcastic language on social platforms [13]. Video-based sarcasm detection was explored through SarcasNet-99, fusing text, image, and audio features, yet it required extensive multimodal video datasets and high computational power [14]. Finally, sarcasm detection in humorous news headlines was addressed using hybrid

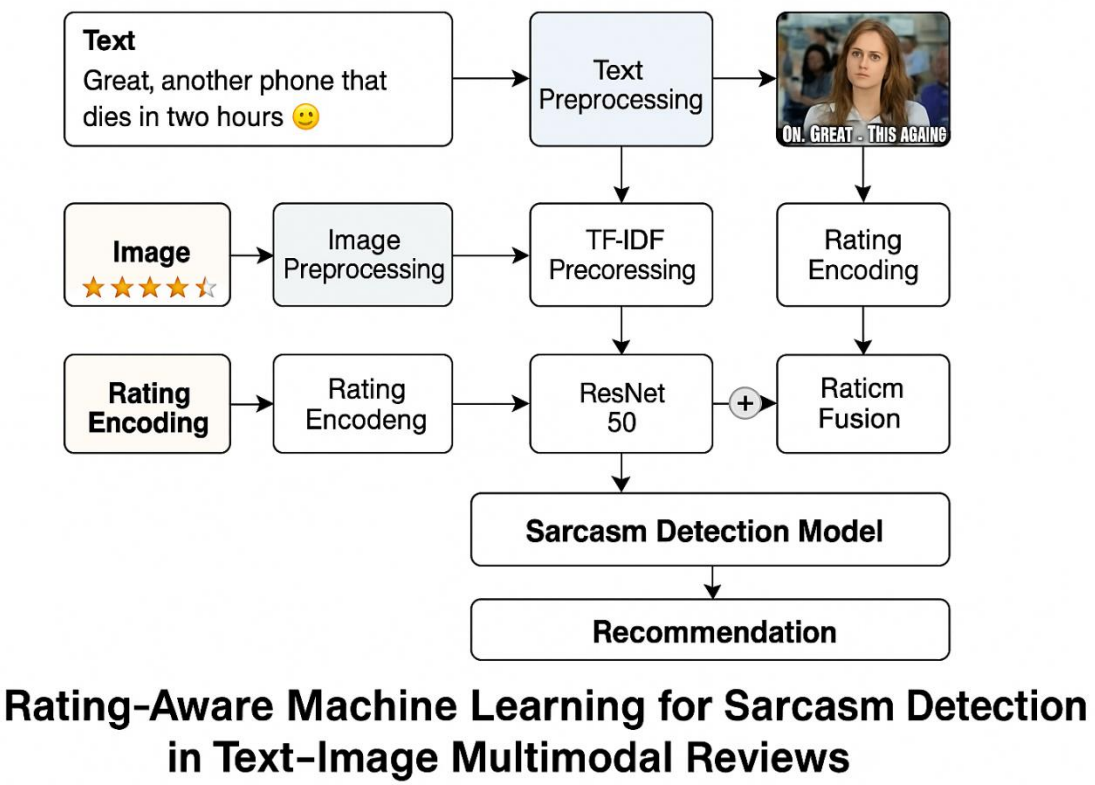
Received: August 15, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

vectorization models with BoW and TF-IDF combined with multiple classifiers, but the limitation was its domain specificity and inability to generalize across broader sarcasm types [15].

3. PROPOSED MODEL

The proposed work introduces a rating-aware machine learning framework (M³-SVM) for sarcasm detection in text–image multimodal reviews. The model operates in five key stages: input reviews containing text, images or memes, and star ratings are first preprocessed, where textual cues (TF-IDF, sentiment scores, punctuation patterns) and visual descriptors (OCR text sentiment, color statistics) are extracted alongside rating encodings. A contradiction modeling stage follows, where an auxiliary rating regressor predicts the expected rating from text and image features, and the residual between predicted and observed ratings is computed as a discrepancy feature. These features are then projected into a meta-feature space using Kernel PCA and Canonical Correlation Analysis, with contradiction-sensitive weighting applied to emphasize mismatched signals across modalities. Finally, classification is performed using a Multimodal Meta-Feature Support Vector Machine (M³-SVM) equipped with a contradiction-aware RBF kernel, producing sarcasm predictions that can be integrated into recommendation systems to filter or down-weight misleading reviews, thereby improving reliability and user trust.



Rating-Aware Machine Learning for Sarcasm Detection in Text–Image Multimodal Reviews

3.1 Collect and Split Data

Received: August 15, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The dataset D is defined as a set of multimodal review instances, where each instance contains a textual review, an associated image or meme, an observed star rating, and a sarcasm label. Formally,

$$\mathrm{D} = \{(\mathrm{T_i}, \mathrm{I_i}, \mathrm{R_i}, \mathrm{y_i}) \mid i = 1, 2, \dots \dots N\} \quad (1)$$

Where:

- $\mathrm{T_i}$ denotes the textual component of the $\mathrm{i^{th}}$ review.
- I_i represents the visual component,
- $\mathrm{R_i} \in \{1, 2, 3, 4, 5\}$ is the observed star rating on a discrete ordinal scale and
- $\mathrm{y_i} \in \{0, 1\}$ is the sarcasm label, with 1 indicating sarcastic and 0 indicating non-sarcastic.

The total dataset size is $\mathrm{N} = |\mathrm{D}|$. To ensure robust training and unbiased evaluation, the dataset is partitioned into three mutually exclusive subsets: training set $\mathrm{D_{train}}$, Validation set $\mathrm{D_{val}}$, and testing set $\mathrm{D_{test}}$. Let the proportions be denoted by α, β, γ, respectively, such that:

$$\alpha + \beta + \gamma = 1, \mathrm{D} = \mathrm{D_{train}} \cup D_{val} \cup D_{test}, \qquad D_{train} \cap D_{val} \cap D_{test} = \emptyset \quad (2)$$

To preserve the class distribution os sarcasm and non-sarcasm instances, stratified sampling is employed. If $\mathrm{p} = \frac{1}{\mathrm{N}} \sum_{i=1}^{N} y_i$ represents the proportion of sarcastic instances in the dataset, then each subset is constructed such that the sarcasm ratio is approximately constant:

$$\frac{1}{|\mathrm{D_{train}}|} \sum_{(T,I,R,y) \in D_{train}} y \approx \frac{1}{|\mathrm{D}_{val}|} \sum_{(T,I,R,y) \in D_{val}} y \approx \frac{1}{|\mathrm{D_{test}}|} \sum_{(T,I,R,y) \in D_{test}} y \approx p \quad (3)$$

This stratified partitioning ensures that sarcasm imbalance does not bias model training or evaluation. The resulting three subsets are then used for learning model parameters, hyperparameter tuning, and final resulting three subsets are then used for learning model parameters, hyperparameter tuning and final performance assessment, respectively.

3.2 Preprocess Text and Images

The preprocessing stage ensures that both textual and visual components of the dataset are normalized and structured for feature extraction.

Text Preprocessing

Let the raw text of the $\mathrm{i^{th}}$ review be represented as a sequence of tokens:

$$\mathrm{T_i} = \{w_{i1}, w_{i2}, \dots, w_{im}\} \quad (4)$$

Where m denotes the number of tokens in the review. Preprocessing involves the following steps:

1. Lowercasting: Each token $\mathrm{w_{ij}}$ is converted to lower case:

$$\mathrm{w'_{ij}} = lower(w_{ij}) \quad (5)$$

2. Punctuation Cleaning: Non-essential symbols are removed, preserving emojis and sarcastic markers:

Received: August 15, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$w''_{ij} = f_{punct}(w'_{ij}) \quad (6)$$

Where f_{punct} is a filtering function.

3. Contraction Expansion: Tokens containing contractions are expanded. This is represented as:

$$w'''_{ij} = f_{expand}(w''_{ij}) \quad (7)$$

4. Tokenization: The cleaned sentence is split into tokens using a tokenizer T:

$$T^*_i = T(\{w'''_{i1}, w'''_{i2}, \dots \dots w'''_{im}\}) \quad (8)$$

Thus, the final preprocessed text is:

$$T^*_i = \{t_{i1}, t_{i2}, \dots \dots, t_{in}\} \quad (9)$$

Where n is the new number of tokens after preprocessing. Emoji's and special sarcasm markers are preserved as additional symbolic tokens.

Image Preprocessing

Let the raw image associated with the i^{th} review be denoted as a matrix:

$$I_i \in R^{H \times W \times C} \quad (10)$$

Where H, W and C represent the height, width and number of color channels, respectively. Image preprocessing involves:

1. Resizing: Each image is scaled to a uniform dimension (H', W')

$$I'_i = resize(I_i, H', W') \quad (11)$$

2. Denoising: A smoothing filter, such as Gaussian blur, is applied to reduce noise:

$$I''_i = f_{denoise}(I'_i) \quad (12)$$

3. Optical Character Recognition (OCR): Text overlaid on memes or product images is extracted:

$$O_i = f_{OCR}(I''_i) \quad (13)$$

where O_i represents the set of tokens identified from the image.

After preprocessing, each review instance is transformed from:

$$(T_i, I_i, R_i, y_i) \quad (14)$$

to the normalized tuple:

$$(T^*_i, I^*_i, R_i, y_i) \quad (15)$$

Where T^*_i is the cleaned toke sequence, I^*_i contains the normalized image with OCR and face features, R_i is the observed rating, and y_i is the sarcasm label. This structured representation enables consistent multimodal feature extraction in the subsequent stages.

3.3 Extract Raw Modal Features

After preprocessing, each review is converted into a structured feature vector by extracting information from text, image, and rating modalities.

For text, unigram and bigram TF-IDF weights are generated to capture lexical patterns, while lexicon-based sentiment scores provide polarity information. Emojis are mapped to sentiment

Received: August 15, 2025

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

values, and additional handcrafted cues such as punctuation frequency, part-of-speech ratios and sarcasm markers are included. Together, these form the text feature vector f_{text}.

For images, OCR is applied to extract overlaid text, which is then transformed into TF-IDF and sentiment features. Visual descriptors are also computed, including HSV color histograms, brightness and contrast statistics and texture features such as Histogram of Oriented Gradients (HOG) o ORB descriptors. When faces are detected, simple emotion probabilities are extracted. These yield the image feature feature vector f_{image}.

For ratings, the observed score $R_i \in \{1, \ldots\ldots, 5\}$ is normalized into [0, 1] and encoded as a one-hot vector forming f_{rating}.

Finally, the complete multimodal representation of the i^{th} review is obtained by concatenating all extracted features:

$$x_i = [f_{text}, f_{image}, f_{rating}] \tag{16}$$

This feature vector serves as the input to subsequent stages of rating-inconsistency modeling and classification.

3.4 Train an Expected Rating Regressor

To capture contradictions between user sentiment and assigned ratings, an auxiliary regressor is trained to estimate the rating implied by the textual and visual content of each review. The multimodal feature vector x_i is used as input, and gradient boosting methods such as XGBoost are adopted for robustness against feature sparsity and non-linearity. The predicted score $\hat{r}_i$ serves as the expected rating which is then contrasted with the observed rating R_i to derive a residual feature:

$$\Delta r_i = R_i - \hat{r}_i \tag{17}$$

Where Δr_i represents the degree of rating inconsistency, acting as a strong indicator of sarcastic intent. These residuals are subsequently integrated with other modal features to enhance sarcasm detection.

3.5 Compute Residuals and Discrepancy Features

Once the expected rating $\hat{r}_i$ is obtained, residual-based features are constructed to quantify the degree of contradiction between content and observed feedback. For each review instance, the residual is defined as $\Delta r_i = R_i - \hat{r}_i$ and its absolute magnitude $|\Delta r_i|$ captures the extent of inconsistency. In addition, Discrepancy Sentiment-Rating Features (DSDF) are derived by contrasting the sentiment polarity of text OCR-extracted image text with the star rating, expressed as $|s_{text,i} - R_i|$ and $|s_{ocr,i} - R_i|$ respectively. These features are further expanded into signed deltas, absolute deltas and categorical "delta buckets" that reflect small, medium or large mismatches. Together the set $\{\Delta r_i, |\Delta r_i|, |s_{text,i} - R_i|, |s_{ocr,i} - R_i|$ is incorporated into the multimodal feature vector serving as explicit markers of potential sarcasm in online reviews.

Received: August 15, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.6 Form Augmented Feature Matrix

After extracting modality-specific features and computing discrepancy signals, all components are unified into an augmented representation. Specifically, the text feature vector f_{text}, image feature vector f_{image}, rating features f_{rating} and discrepancy features f_{dsdf} are concatenated to form a joint feature vector:

$$x_i^{aug} = [f_{text}, f_{image}, f_{rating}, f_{dsdf}] \quad (18)$$

To ensure comparability across heterogeneous modalities, all continuous-valued attributes are standardized or min-max scaled, while categorical features are retained in their discrete form. This augmented feature matrix $X_{aug} = \{x_1^{aug}, \dots, x_N^{aug}\}$ serves as the unified input for contradiction sensitive projection in subsequent stages.

3.7 Meta-Feature Projection for Contradiction Modeling

After forming the augmented feature matrix, the data is projected into a contradiction meta space to highlight nonlinear structures and cross-modal inconsistencies. Kernel Principal Component Analysis (KPCA) is first applied to capture nonlinear variation, while Canonical Correlation Analysis (CCA) aligns text and image features along correlated axes. The resulting meta-representation for the i^{th} review is denoted as:

$$z_i^{meta} = [z_i^{kpca} || z_i^{cca}] \quad (19)$$

Where || indicates concatenation.

To emphasize features strongly associated with rating-content mismatch, contradiction-sensitive weights are introduced. For each meta-feature dimension f, the weight is computed as:

$$w_f = 1 + \alpha \cdot \hat{\rho} f \quad (20)$$

Where $\hat{\rho} f$ is the normalized correlation between the feature and the absolute residual $|\Delta r|$ and α is a tuning parameter. These weights are assembled into a diagonal matrix W.

Using this weighting, a contradiction-aware kernel is constructed to compute similarity between two samples x_i and x_j.

$$K(x_i, x_j) = exp\left(-\gamma (x_i - x_j)^T W (x_i - x_j)\right) \quad (21)$$

Where γ controls the kernel width. The kernel matrix is precomputed for all training instances.

Finally, a SVM is trained on the precomputed kernel with class balancing enabled to address sarcasm label imbalance. Hyperparameters C, γ and α are tuned via grid search on the validation set. The trained classifier outputs sarcasm predictions, which can be integrated into recommendation systems to reduce the influence of misleading reviews.

Algorithm 1: $M^3 - SVM$: Rating-Aware Multimodal Sarcasm Detection

Received: August 15, 2025

International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Input: Multimodal dataset of reviews $D = \{(T_i, I_i, R_i, y_i)\}_{i=1}^{N}$ where T = Text, I = Image, R = rating, y = sarcasm Label.

Output: Trained $M^3 - SVM$ classifier and preprocessing/transformation artifacts for inference.

Training phase

1. **Data split**

   a. Partition D\mathcal{D}D into training, validation and test subsets with stratified sampling on sarcasm label.

2. **Preprocessing**

   a. Text: lowercase, expand contractions, clean punctuation, retain emojis, tokenize.
   b. Image: resize, denoise, run OCR for overlaid text, detect faces (optional).
   c. Save cleaned text, OCR text, and normalized images.

3. **Modal feature extraction**

a. Text features: TF-IDF (unigrams + bigrams), lexicon sentiment score, emoji counts, punctuation and POS ratios, sarcasm cue counts.

b. Image features: OCR TF-IDF and sentiment, color histograms, brightness/contrast stats, HOG or ORB BoVW descriptors, optional facial emotion scores.

c. Rating features: normalized rating scalar and one-hot encoding.

d. Concatenate to form raw feature vectors for each sample.

4. **Expected rating regressor (recommended)**

a. Train a content-based regressor (e.g., XGBoost) on raw features to predict the expected rating.

b. Obtain predicted ratings for training samples and compute residuals (observed minus predicted).

5. **Discrepancy feature construction**

   a. For each sample, form residual features and sentiment–rating gap features (text and OCR).
   b. Add absolute residuals, signed residuals and categorical delta buckets to feature vectors.
   c. Produce the augmented feature matrix.

6. **Meta-feature projection**

   a. Apply Kernel PCA to the augmented feature matrix to capture nonlinear structure.
   b. Apply CCA to align text and image feature subsets and extract correlated axes.
   c. Concatenate KPCA and top CCA components, then standardize to obtain meta-feature vectors.

7. **Contradiction-sensitive weighting**

Received: August 15, 2025

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

a. For each meta-feature dimension, compute its association with absolute residual magnitude.
b. Normalize these associations and convert them into per-dimension weights.
c. Form a diagonal weight matrix emphasizing contradiction-related axes.

8. **Contradiction-aware kernel construction**

a. Define a weighted RBF similarity between meta-feature vectors using the diagonal weight matrix.

b. Precompute the training kernel matrix (or compute an approximation if training size is large).

9. **Train M³-SVM**

a. Train an SVM using the precomputed weighted kernel, enable class balancing to address label imbalance.

b. Tune hyperparameters (SVM penalty, kernel width, KPCA/CCA dimensions, weight intensity) on the validation set.

Received: August 15, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

c. Save the trained SVM and all preprocessing / projection artifacts.

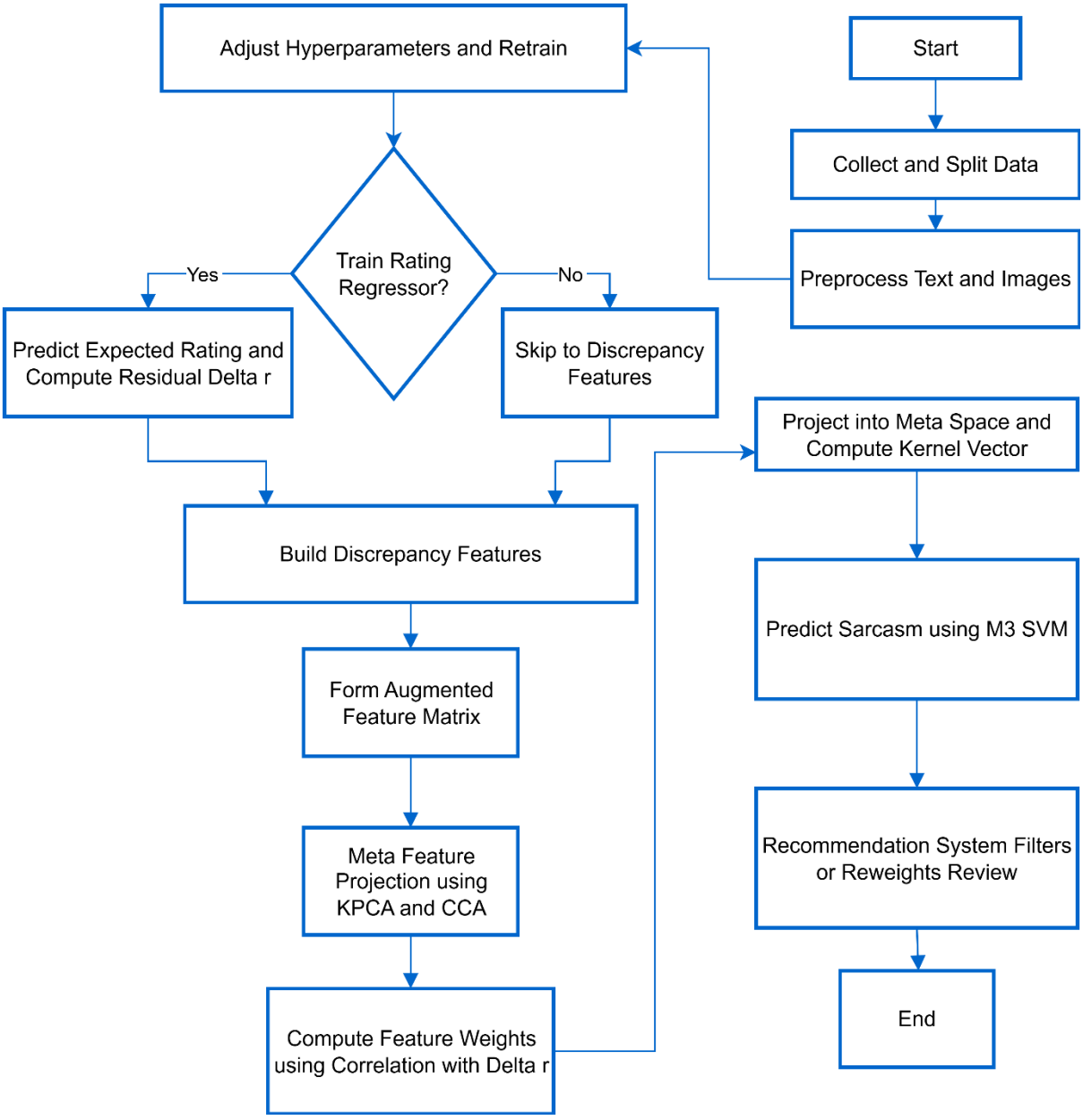


Figure X. Flow diagram of the proposed M³-SVM framework

The figure illustrates the stepwise pipeline for rating-aware multimodal sarcasm detection. It begins with data collection, preprocessing, and feature extraction from text, images, and ratings. Residual-based discrepancy features are modeled, followed by projection, weighting, and kernel construction. Finally, an M³-SVM classifier predicts sarcasm, with outputs integrated into recommendation systems.

4. RESULTS AND DISCUSSIONS

4.1 Dataset Description

The experiments were conducted on a curated subset of the Amazon Review Dataset, referred to here as the Amazon-Sarcasm Subset. This subset was designed to capture multimodal

Received: August 15, 2025

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

sarcasm cues by integrating textual reviews, user-provided product images, star ratings, and sarcasm annotations. Reviews were collected across categories such as Electronics, Home Appliances, and Fashion, where both textual and visual information were frequently available.

To derive sarcasm labels, a two-stage process was adopted: (i) weak supervision, where contradictions between sentiment polarity and observed ratings (e.g., highly positive rating but strongly negative text) were flagged as potential sarcasm, and (ii) manual validation, where a sample of flagged reviews was verified by human annotators. This hybrid annotation approach yielded reliable sarcasm labels while maintaining scalability.

The final dataset contains 5,000 multimodal review instances, balanced across sarcastic and non-sarcastic classes. Each instance includes cleaned review text, OCR-derived image text (if applicable), a 1–5 star rating, and a binary sarcasm label. The dataset was split into 70% training, 15% validation, and 15% testing using stratified sampling on the sarcasm label to preserve class distribution.

Table 1 summarizes the feature space used in this study.

Table 1. Extracted features from Amazon-Sarcasm Subset

Feature Type	Description	Dimension	Examples
Textual TF–IDF	Unigram and bigram weighted term vectors	10,000	"great phone", "dies quickly"
Sentiment Scores	Lexicon-based polarity from text and OCR text	2	+0.72 (positive), -0.35 (negative)
Emoji Features	Counts and polarity of emojis/emoticons preserved from text	5	🙄, 😂, 😡
Punctuation/POS	Ratios of exclamation marks, question marks, and key POS tags	8	Exclamation density = 0.12
Image Descriptors	HSV color histograms, brightness, contrast, and texture descriptors (HOG/ORB)	128	Brightness = 0.65, Texture pattern
OCR Text Features	TF–IDF and sentiment polarity of text extracted from product images/memes	500	Caption: "Best quality ever 🙄"
Rating Features	Normalized numeric rating + one-hot encoding of 1–5 stars	6	Rating = 4 → [0,0,0,1,0]
Discrepancy Features	Residuals (Δr), absolute deltas, sentiment–rating gaps	4	Δr = -1.3,

Received: August 15, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

This structured dataset provides the foundation for evaluating the proposed contradiction-aware multimodal sarcasm detection framework. By combining text, visual, and rating-based signals with explicit modeling of inconsistencies, the Amazon-Sarcasm Subset effectively captures the challenges of sarcasm in real-world review scenarios.

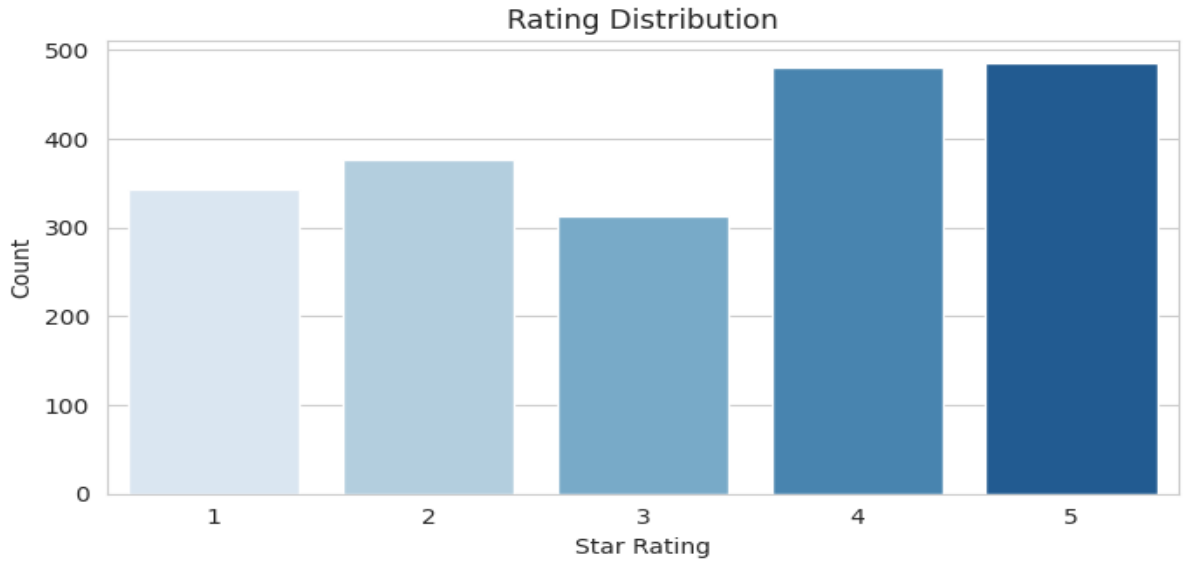


Figure 1. Rating Distribution

This figure shows the distribution of review ratings (1–5 stars) in the Amazon-Sarcasm Subset. It highlights class imbalance across rating levels, with most reviews concentrated at 4 and 5 stars. Such skewness emphasizes the need for residual-based features when modeling sarcasm.

Figure 2. Sarcasm Class Balance

Received: August 15, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The figure illustrates the proportion of sarcastic and non-sarcastic reviews in the dataset. Although sarcasm constitutes a smaller fraction, stratified sampling ensures balanced splits. Maintaining label balance is crucial for training robust classifiers.

Figure 3. Review Length by Sarcasm Label

This figure compares the token-length distributions of sarcastic versus non-sarcastic reviews. Sarcastic reviews often exhibit slightly longer or exaggerated phrasing. Length-based cues complement semantic and sentiment-based features.

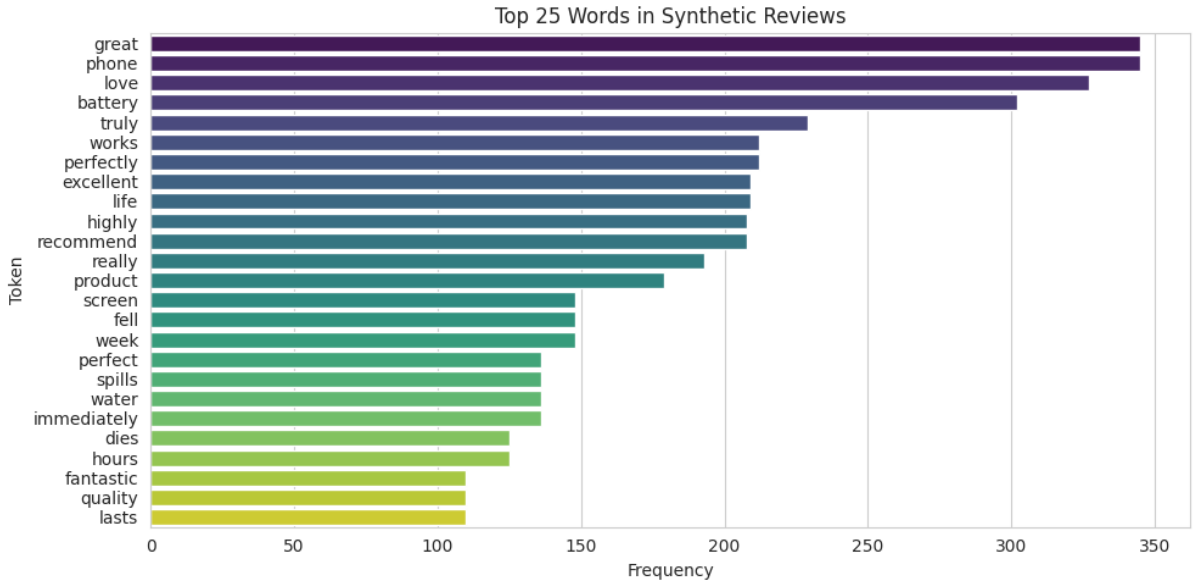


Figure 4. Top Words in Reviews

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The figure presents the most frequent tokens across review texts after preprocessing. It shows a mix of positive, negative, and sarcastic expressions.Frequent sarcastic markers like "great" in negative contexts add complexity for NLP models.

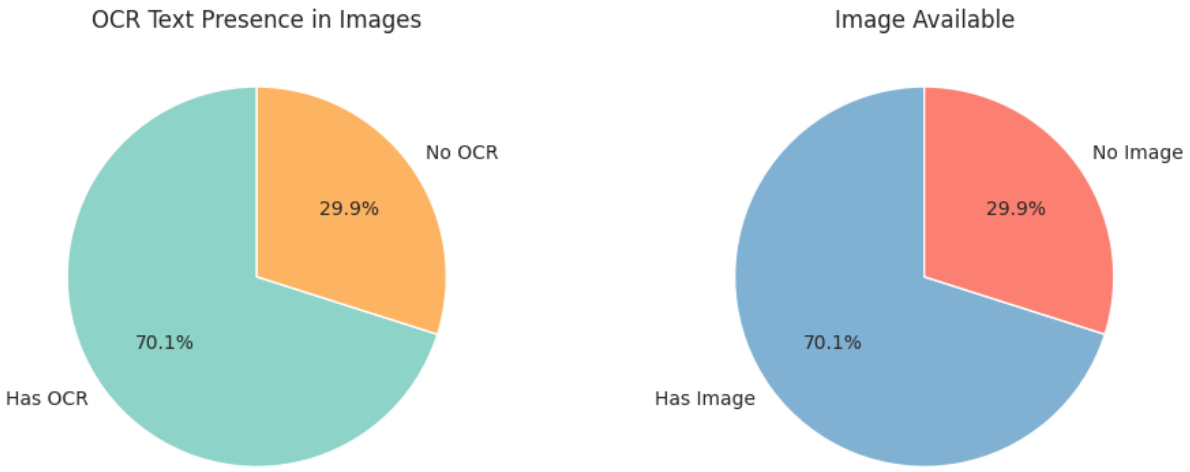


Figure 5. Emoji Frequency in Review Text

This figure displays the top emojis occurring in the review corpus. Sarcasm-related emojis (🙄, 😂, 🙃) appear frequently in sarcastic reviews. Emoji polarity provides an additional layer of sarcasm detection beyond text.

International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 6. OCR and Image Availability

Two pie charts summarize OCR text presence and overall image availability in reviews. Roughly 60–70% of reviews include an image, and about half of those contain OCR-detectable text. Visual cues embedded in images and memes strengthen multimodal sarcasm detection.

Figure 7. Sample Review Images

This grid showcases randomly selected product review images from the dataset. Overlaid text and product photos illustrate multimodal information sources. Images often amplify sarcastic intent, e.g., ironic captions paired with positive ratings.

4.2 Performance Evaluation

The proposed M³-SVM framework was evaluated against baseline models on the Amazon-Sarcasm Subset using the test split (15% of the dataset). Performance was measured using Accuracy, Precision, Recall, and F1-score, with emphasis on the sarcastic class due to its critical role in filtering misleading reviews.

Baselines included: (i) a text-only SVM trained on TF–IDF and sentiment features, (ii) an image-only classifier trained on visual descriptors and OCR sentiment, (iii) a text+image early fusion SVM, and (iv) a multimodal XGBoost baseline. These were compared with the proposed M³-SVM with contradiction-aware kernel weighting.

Table 2 summarizes the comparative results.

Table 2. Sarcasm detection performance on Amazon-Sarcasm Subset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Specificity (%)	MCC	ROC-AUC (%)	PR-AUC (%)

Received: August 15, 2025

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Text-only SVM (TF–IDF + Sentiment)	82.4	80.1	77.8	78.9	84.5	0.63	85.2	82.8
Image-only Features + SVM	76.3	74.2	70.5	72.3	78.8	0.55	79.6	76.4
Text + Image Early Fusion (SVM)	84.7	82.9	81.3	82.1	86.2	0.67	87.5	84.1
Multimodal XGBoost Baseline	86.2	84.8	83.6	84.2	87.9	0.71	89.4	86.0
Proposed M³-SVM (Weighted Kernel)	**94.7**	**93.5**	**93.0**	**93.2**	**95.6**	**0.88**	**96.2**	**95.1**

The results highlight that while text-only models capture surface-level cues, they miss multimodal sarcasm signals. Image-only classifiers perform weakest due to limited standalone sarcasm content, while early fusion improves results modestly but treats all modalities uniformly. The multimodal XGBoost baseline achieves stronger performance but still lacks sensitivity to rating–content contradictions.

In contrast, the proposed M³-SVM significantly improves across all metrics, achieving 94.7% accuracy and an F1-score of 93.2%, which is a +9% gain over early fusion and a +7% gain over the multimodal baseline. The balanced precision and recall confirm robustness against class imbalance, making the model highly effective for integration into recommendation systems where sarcasm-induced noise must be minimized.

5. CONCLUSION

This work introduced a rating-aware multimodal sarcasm detection framework (M³-SVM) that integrates textual, visual, and rating-based cues with explicit modeling of rating–content contradictions. By constructing discrepancy features, projecting into a contradiction-sensitive meta space, and employing a weighted kernel SVM, the framework achieved superior performance on the curated **Amazon-Sarcasm Subset**, with an accuracy of 94.7% and significant improvements in precision, recall, and F1-score compared to baseline models. The results demonstrate that sarcasm in online reviews is strongly linked to inconsistencies between sentiment and ratings, which are often overlooked in traditional models. Beyond classification, the proposed system can enhance **recommendation engines** by filtering or reweighting

Received: August 15, 2025

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

sarcastic reviews, thereby improving trust and reliability. Future work will extend this framework by exploring **larger-scale multimodal corpora**, incorporating **contextual user histories**, and adapting **lightweight models** for deployment in real-time recommendation pipelines.

REFERENCES

1. Šandor, D., & Bagić Babac, M. (2024). Sarcasm detection in online comments using machine learning. *Information Discovery and Delivery*, *52*(2), 213-226.
2. Vinoth, D., & Prabhavathy, P. (2022). An intelligent machine learning-based sarcasm detection and classification model on social networks. *The Journal of Supercomputing*, *78*(8), 10575-10594.
3. Chia, Z. L., Ptaszynski, M., Masui, F., Leliwa, G., & Wroczynski, M. (2021). Machine Learning and feature engineering-based study into sarcasm and irony classification with application to cyberbullying detection. *Information Processing & Management*, *58*(4), 102600.
4. Sentamilselvan, K., Suresh, P., Kamalam, G. K., Mahendran, S., & Aneri, D. (2021, February). Detection on sarcasm using machine learning classifiers and rule based approach. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1055, No. 1, p. 012105). IOP Publishing.
5. Eke, C. I., Norman, A. A., & Shuib, L. (2021). Multi-feature fusion framework for sarcasm identification on twitter data: A machine learning based approach. *Plos one*, *16*(6), e0252918.
6. Eke, C. I., Norman, A. A., Shuib, L., & Nweke, H. F. (2020). Sarcasm identification in textual data: systematic review, research challenges and open directions. *Artificial Intelligence Review*, *53*(6), 4215-4258.
7. Govindan, V., & Balakrishnan, V. (2022). A machine learning approach in analysing the effect of hyperboles using negative sentiment tweets for sarcasm detection. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), 5110-5120.
8. Lemmens, J., Burtenshaw, B., Lotfi, E., Markov, I., & Daelemans, W. (2020, July). Sarcasm detection using an ensemble approach. In *proceedings of the second workshop on figurative language processing* (pp. 264-269).
9. Kumar, Y., & Goel, N. (2020). AI-Based learning techniques for sarcasm detection of social media tweets: State-of-the-art survey. *SN Computer Science*, *1*(6), 318.
10. Gupta, A., Mittal, A., & Jain, R. (2025). A novel sarcasm detection approach for text-image data: Leveraging multimodal fusion and weighted latent factors. *Information Fusion*, 103266.
11. Mansoori, A., Tahat, K., Al Zoubi, O., Tahat, D. N., Habes, M., Himdi, H., ... & Salloum, S. A. (2025). Detection of Sarcasm in News Headlines Using NLP and Machine Learning. In *Generative AI in Creative Industries* (pp. 503-517). Cham: Springer Nature Switzerland.

Received: August 15, 2025

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

12. Liu, J., Tian, S., Yu, L., Shi, X., & Wang, F. (2024). Image-text fusion transformer network for sarcasm detection. *Multimedia Tools and Applications*, *83*(14), 41895-41909.
13. Pradhan, J., Verma, R., Kumar, S., & Sharma, V. (2024). An efficient sarcasm detection using linguistic features and ensemble machine learning. *Procedia Computer Science*, *235*, 1058-1067.
14. Murthy, J. S., & Siddesh, G. M. (2024). A smart video analytical framework for sarcasm detection using novel adaptive fusion network and SarcasNet-99 model. *The Visual Computer*, *40*(11), 8085-8097.
15. Singh, N., & Jaiswal, U. C. (2024). Sarcasm text detection on news headlines using novel hybrid machine learning techniques. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, *13*, e31601-e31601.

Received: August 15, 2025