Journal of Theoretical and Applied Information Technology
15th December 2025. Vol.103. No.23
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CUDA-ENHANCED YOLO FEATURE EXTRACTION WITH RF–CATBOOST ENSEMBLES FOR HIGH-PERFORMANCE PRODUCT REVIEW SENTIMENT DETECTION

Dr. VANCHA MAHESHWAR REDDY[1],SRILAKSHMI KAZA[2], CH.RAMBABU[3], Dr.A.SRI NAGESH[4], V.S.N. MURTHY[5], M PRAVEEN[6], SARAVANAN G[7], B.SRILAKSHMI[8], Dr. E.SREEDEVI[9],Dr. SIVA KUMAR PATHURI[10]

[1]Associate Professor, Department of Computer Science and Engineering, Institute of Aeronautical Engineering, Dundigal, R R Dist., Telangana
[2]Associate Professor, Department of Electronics and Communication Engineering ,Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Krishna District, AP,India.
[3]Associate Professor, Department of Electronics and Communication Engineering ,Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Krishna District, AP,India.
[4]Professor & HOD, Department of Computer Science & Business Systems, R.V.R. & J.C.College of Engineering, Chowdavaram,Guntur-522019.
[5]Assistant Professor, Shri Vishnu Engineering College for Women, Department of IT, Bhimavaram, West Godavari, Andhra Pradesh, India.
[6]Assistant Professor, Department of Computer Science and Engineering (AI&ML) , R. V. R & J. C College of Engineering , Guntur, AP, India.
[7]Department of Computer Science and Engineering, Erode Sengunthar Engineering College , Perundurai, Erode-638057,TN,India.
[8]Assistant Professor, Department of Computer Application's R.V.R & JC College of Engineering, Chowdavaram, AP,India.
[9]Professor, Dept of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India.
[10]Professor, Dept of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India.

Email: [1]mahesh.vancha@gmail.com,[2]slkaza06@gmail.com, [3]rambabuec41@gmail.com, [4]asrinagesh@gmail.com,[5]vsn.murthy87@gmail.com, [6]praveenkumarmuvva@gmail.com [7]gsaravanan.pacet@gmail.com, [8]bhimineni.srilakshmi@gmail.com, [9]sridevi_fed@kluniversity.in,[10]spathuri@kluniversity.in

ABSTRACT

Sentiment analysis is essential in retrieving opinions, reviews and written statements to predict emotions based on Natural Language Processing (NLP). It categorizes the text in sentiments like positive, negative or neutral. In case of labeled datasets, the customer feedback can be grouped into scales such good, better, best, or bad, worse, worst, and subsequently converted into sentiment categories. The fast expansion of the World Wide Web has created massive amounts of user-generated content opinions and feelings and reviews that are used in business by e-commerce platforms, such as Amazon, Flipkart, and social platforms, such as Twitter and Facebook. Opinion mining has now taken a very important role in business analytics where customer opinion is directly used in creating and maintaining competitiveness in a product. In this paper, a CUDA-Optimized Sentiment Analysis Framework based on YOLO to extract features and a hybrid RF-CATBoost ensemble to classify the sentiments is proposed. YOLO takes both textual and pictorial aspects, and RF-CATBoost reduces bias and variance as they aim to achieve increased accuracy. The proposed YOLO + RF-CATBoost model is benchmarked on the Amazon mobile review dataset against deep learning baselines such as ResNet-50 and RegNetY and it is more accurate and faster. The framework, which involves the use of GPU-accelerated CUDA processing, has high performance, scalability, and efficient sentiment detection to support e-commerce decision-making.
Keywords: *Sentiment Analysis, YOLO , RF–CATBoost Ensemble, CUDA Parallel Processing, Product Review Classification, Deep Learning (ResNet-50, RegNetY)*

1. INTRODUCTION

During the product reviews, customer opinions are vital towards companies with the aim of gauging customer satisfaction and improving their products.

Journal of Theoretical and Applied Information Technology
15th December 2025. Vol.103. No.23
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Application of the sophisticated computation system, especially those systems that take advantage of parallel processing using the GPUs, offer a good means of speeding up sentiment analysis[1] and improving the accuracy of the classification process. Conventional methods of sentiment analysis typically use machine learning techniques with computational bottlenecks and weak representational ability. Conversely, CUDA-enhanced YOLO feature extraction and RF-CATBoost ensembles provide a disruptive solution, which can increase the effectiveness of feature-based sentiment analysis in large product evaluation systems. This paper presents the use of YOLO (You Only Look Once) to extract features and RF-CATBoost hybrid ensembles to classify, optimized using CUDA parallel processing, to analyse the sentiment of product reviews. The proposed framework can achieve better representation and classification of textual and visual data than single-model approaches through the ability of YOLO to extract high-level semantic features of data, and the ensemble performance of RF-CATBoost to reduce bias and variance. The parallelization provided by CUDA guarantees a high throughput and faster inference, which can enable real-time analysis of vast amounts of customer feedbacks. To compare the proposed strategy with deep learning reference points, we employ the ResNet-50[2] and RegNetY[3] models that are commonly used in sentiment and image-based classification problems to learn features. Our experiments with the Amazon mobile review dataset (millions of actual reviews) demonstrate that the YOLO + RF+CatBoost architecture is both more accurate and much faster than the traditional models and deep learning baselines. The feature-based sentiment detection is not only able to capture polarity (positive, negative, neutral) but also make detailed observations of consumer opinion at scale. This process is built using data mining which involves classification, clustering, rule-based aggregation and predictive modeling. The concept of classification is of the forefront in this context since a product review can feature not only numerical values (e.g. the price of the camera, the number of pixels in the camera) but also categorical variables (e.g. the quality rating). We introduce a CUDA-optimized YOLO + RF-CATBoost framework that allows managing all these attributes effectively, which guarantees the high-quality and scalable sentiment detection regardless of the review format. This study has a number of important contributions[4] we use the YOLO-based extraction of features that we optimize on CUDA to perform parallel computations[5] we introduce an RF -CATBoost hybrid ensemble, which yields a higher rate of sentiment classification and lower inference times, (3) we run comparative experiments with ResNet-50 and RegNetY, showing a higher accuracy judgment and lower inference time, and we use large-scale review data to estimate product rating and polarity to be highly reliable on Amazon. The experimental outcomes verify that the GPU-based YOLO feature extraction, coupled with RF-CATBoost classification has a substantial enhancement in speed of running compared to CPU-based approaches with better accuracy and scaling[6]. The framework divides workloads into GPU threads, which makes computations faster, allows efficient determination of classes and high-throughput processing. In short, our suggested CUDA-Improved YOLO Feature Extraction with RF-CATBoost Ensembles is a high-performance product review sentiment detection solution. The framework offers business owners actionable intelligence based on customer feedback by integrating the power of the GPU parallelism, YOLO feature learning, and ensemble-based classification. This not only minimizes the chances of fraud and improves decision-making but also makes sure that companies will be competitive in the rapidly expanding e-commerce ecosystem.

2. LITERATURE SURVEY

Sentiment analysis (SA) is now an essential field of natural language processing (NLP) as well as business intelligence, especially as user-created content in e-commerce and social media continues to rise exponentially. Hundreds of millions of reviews in the Amazon, Flipkart, Twitter, and Facebook websites are full of significant ideas on what customers think, the quality of products, and market trends. The efficient mining of this information helps businesses make better decisions, better the customer's experience, and stay competitive. Conventional machine learning methods, though effective, are in most cases poor in scaling, robustness and even modeling subtle details in complicated datasets. Lately, deep learning and ensemble techniques have been developed that have boosted sentiment classification. Real-time object detector YOLO (You Only Look Once) has demonstrated possibilities in feature extraction beyond images, using its high-representation abilities. In the meantime, ensemble techniques like Random Forest and CATBoost are useful to minimize bias and variance and provide more credible results than isolated models. Computational efficiency can also be improved by the addition of GPU acceleration via CUDA so that even large-scale sentiment detection tasks can be running much faster without a significant loss of accuracy[7].

Perikos et al. (2024) investigated explainable aspect-based sentiment analysis (ABSA) using transformer architectures such as BERT, ALBERT, RoBERTa, DistilBERT, and XLNet combined with explainable AI (XAI) methods. Using datasets including MAMS, SemEval, and Naver, they demonstrated that transformers perform strongly for ABSA while explainability enhances interpretability (MDPI).

Similarly, Ali et al. (2024) conducted a systematic review of YOLO models up to YOLOv11 across broad detection benchmarks, highlighting that YOLO achieves real-time, high-recall performance, making it well-suited for GPU-accelerated multimodal mining pipelines.

Daza et al. (2024) applied sentiment analysis to e-commerce reviews using textual cues and rating metadata. Their comparison of machine learning and deep learning models showed that deep architectures outperform traditional approaches, with ensembles further improving robustness.

In a related direction, Li et al. (2024) introduced SentiSys, an edge-enhanced graph convolutional network (GCN) for ABSA. By integrating Bi-LSTM, self-attention, and Bi-GCN modules, they found that preserving syntactic edge information significantly boosts ABSA accuracy across four benchmark datasets.

Similarly, Huang et al. (2024) developed SRE-BERT, a syntax-aware ABSA framework that combines transformer models with syntactic embeddings, reporting substantial improvements in ABSA performance.

Other works explored hybrid neural networks for review analysis. Sorour et al. (2024) proposed WDE-CNN-LSTM, which fuses word embedding features with temporal and local cues through a CNN–LSTM hybrid, outperforming single-model baselines on consumer review datasets.

Bellar et al. (2024) focused on retail reviews, particularly women's clothing datasets, where they compared neural and transformer models using embeddings such as BERT, FastText, and Word2Vec. Their results showed transformers achieve the best performance and that class granularity (three vs. five sentiment classes) significantly influences evaluation metrics

Perikos and colleagues (2024) further demonstrated that XAI methods can validate and justify transformer-based ABSA predictions, enhancing trust in model decisions.

In terms of ensemble methods, Tiwari et al. (2023) examined sentiment analysis on social networks, showing that bagging and boosting ensembles outperform traditional models in terms of ROC-AUC, F-score, and runtime efficiency .

Atlas et al. (2025) advanced this direction by designing a modernized sentence-level sentiment analysis pipeline for product reviews, where their deep learning stack improved sentence-level predictions (PMC).

Similarly, Jayakody et al. (2024) explored instruction-style fine-tuning of DeBERTa and other transformer variants for ABSA tasks, establishing new state-of-the-art performance while demonstrating that training strategies are critical for accuracy.

Survey-based contributions include Khan et al. (2024), who highlighted the role of ensembles and data augmentation in computational intelligence applications, including sentiment analysis, concluding

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

that such techniques enhance model robustness and generalization.

Park et al. (2025) conducted a comparative study of machine learning methods for sentiment analysis across multiple datasets, finding that ML and DL models consistently outperform classical baselines, though careful tuning remains crucial (SCIRP).

Finally, a 2025 practice note illustrated the benefits of GPU acceleration in sentiment analysis pipelines using RAPIDS cuDF, demonstrating that CUDA-based tokenization and feature processing significantly reduce computation time while maintaining high accuracy.

3. RELATED WORK

3.1 Product Review:

Sentiment analysis has various subfields of study such as subjectivity detection, emotion prediction, aspect-based sentiment analysis (ABSA), opinion summarization, spam detection and feature extraction to product intelligence. Subjectivity detection aims at the separation of factual or subject opinion statements. The aim of emotion prediction is to categorize the review polarity as positive, negative or neutral[8]. ABSA goes one step further by examining feeling about certain product features, including battery life, or design. Opinion summarization condenses detailed reviews into concise representations, such as star ratings or sentiment scores. Spam detectors are used to ensure that information published by users is reliable by sifting fraudulent or biased reviews[9].

As an illustration, one may refer to the following product review that was left on Amazon:

I am so LUCKY to have found this used phone online and the person who upgraded and sold this phone. My Son loved his old one who broke down after 2.5+ years and did not want to upgrade! Thanks, Seller, we do like it and your sincerity.

A sentiment mining system may be able to identify decision cues including:

- "Lucky" → Positive
- "Liked" → Positive
- "Appreciate" → Positive

These aspects are clear signs of a general positivity.

3.2 Sentiment Analysis

Sentiment analysis, also known as opinion mining, is the computational task, which classifies text in terms of its polarity, and is essential to the analysis of customer feedback, social media discussion, and market behavior. A typical pipeline includes:

- Text Preprocessing: Tokenization, stop-word elimination and lemmatization to clean raw text.
- Feature Extraction: Extraction of sentiment bearing features, including n-grams, embeddings, syntactic dependencies, or attention-based features.
- Classification: Traditional machine learning systems, such as Naive Bayes and SVM, or advanced deep learning systems (CNNs, RNNs, Transformers) can be trained on a labelled corpus to make sentiment predictions.
- Lexicon Usage: Sentiment dictionaries may be used to complement machine learning by giving words polarity scores.
- Aspect-Based SA: Expands on general sentiment, and maps polarity to product aspects (e.g., battery life is excellent but camera is poor).
- Evaluation Metrics: The measurements that are commonly used to measure performance are Accuracy, precision, recall, F1 score, and AUC-ROC.

3.3 Proposed Framework: YOLO with RF–CATBoost

In the proposed research, we consider a hybrid architecture that uses YOLO to extract features and a combination of Random Forest (RF) and CatBoost to classify them. In addition to its wide application in object detection, YOLO can extract rich representational features out of text embeddings and visual sentiment indicators like emoticons or review-related images. These attributes offer an organized expression of customer views.

The features extracted are then forwarded to ensemble classification. Random Forest is

Journal of Theoretical and Applied Information Technology
15th December 2025. Vol.103. No.23
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

versatile and resilient as it builds a set of decision trees, whereas CATBoost is fast and capable of working with categorical and high-dimensional features with low preprocessing capabilities. The combination of the two approaches improves both the classification and generalization of product reviews in a variety of products.

This unified CUDA-accelerated pipeline is meant to do two main jobs:

1. Aspect extraction (finding targets of sentiment like product quality, delivery, or prices).
2. Polarity classification (applying positive, negative or neutral sentiment to those aspects).

The fusion of scalability, interpretability, and robustness of YOLO and RF–CatBoost enables the framework to be applied to large-scale product review sentiment detection using an ensemble decision-maker. Figure 1 shows sentiment classification.

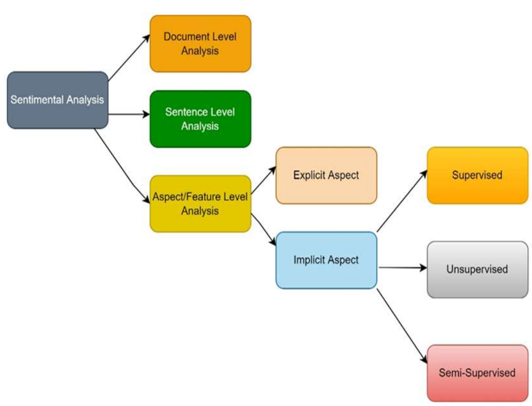

Figure 1: SA Classification

3.4 Feature Selection

An important step when designing a system of sentiment analysis of product reviews is feature selection. Within this further framework, we combine the use of YOLO to extract visual information within product-related pictures and the RF+CatBoost to choose and prioritize the most informative textual and structured features.

Dimensionality Reduction & Feature Engineering

YOLO (You Only Look Once):

Used on product images to identify and locate visual objects like brand logos, quality packaging and product defects.

These visual features are represented as either categorical or numerical and combined with textual ones.

RF+ CATBoost Feature Selection:

Automatically ranks features on their predictive ability of sentiment classification.

Native support of categorical variables and decreases overfitting with regularization.

Adds statistical measures such as the mutual information and feature importance scores.

Principal Component Analysis (PCA):

Used after fusion to cut dimensions without losing variance across visual and textual feature.

Singular Value Decomposition (SVD):

Applied to break down high dimension review matrices into latent semantic variables.

• Domain-Specific Features:

• Includes product-specific attributes (e.g., price, brand, quality indicators), sentiment lexicons, and aspect-based features (e.g., battery life, camera quality).

Statistical Analysis for Feature Relevance

To choose effective product characteristics, we use statistical and model-driven analysis:

1. Correlation Analysis: of sentiment of review and star ratings.
2. Feedback-Rating Linkage: With RF+CatBoost model nonlinear dependencies.
3. Price-Ranking Association: The importance of features evaluated in RF+CatBoost.
4. Brand-Level Sentiment Trends: Summed up over the brand logos detected by YOLO.
5. Word Frequency Tracking: TF-IDF and built-in feature ranking of CATBoost: ranking words of high impact.
6. Sentiment Polarity Detection: RF+CatBoost classifier, Trained on fused features.

3.5 Example Use Case

Users tend to use textual reviews and product images when buying a mobile phone online. YOLO identifies visual signs (i.e., cracks on the screen or packaging), whereas CATBoost evaluates review text and

Journal of Theoretical and Applied Information Technology
15th December 2025. Vol.103. No.23
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

metadata to identify the sentiment.
Sample Review:
a = "This is the worst mobile."
RF+CatBoost Output:
{ "negative": 0.91, "neutral": 0.08, "positive": 0.01, "confidence score": 0.84 }
This type of hybrid sentiment analysis model will allow organizations to use both written and graph feedback to improve their products and increase customer satisfaction. One of the key steps in the development of the feature-based sentiment analysis system is feature selection, which is aimed at determining the most significant linguistic and contextual characteristics that have a strong impact on the prediction of sentiment. This feedback mechanism makes certain that the classification model operates with good inputs and minimizes the computational complexity.

3.6 The key feature selection steps involve:

Dimensionality Reduction: The chi-square, or information gain, or mutual information statistic is applied to select only the most informative features.
Principal Component Analysis (PCA): This is the transformation of the feature space to lower dimensions but attempting to capture as much variance as possible.
Singular Value Decomposition (SVD): Reducing the feature matrix to the top core latent components to be represented compactly.
Domain-Specific Feature Engineering: Adding product-related features (e.g., quality, price, durability), sentiment lexicons, and aspect-based features in order to increase interpretability.
More practically, statistical analysis is very critical in the selection and validation of product features [11]. This involves:
Determining the relationships between reviews and star ratings.
Researching connections between textual feedback and general satisfaction.
Analysis of the price-perceived value relationship.
Comparing the trend of sentiments on various brands.
Monitoring common words in consumer reviews.
Estimating polarity (positive, negative, neutral) of every review.

As an example, customers tend to rely on the use of user-created reviews to influence their decision making when they buy a mobile phone on the internet. These reviews will provide insights into durability, performance and customer service experiences and assist the novice buyer make an informed decision. Meanwhile, companies are able to use this aggregate response to better design the product, change their pricing strategy, and enhance customer relationships [12,13]. Figure 2 shows the architecture of the proposed YOLO + RF-CATBoost sentiment recognition system and demonstrates the general flow of the process.

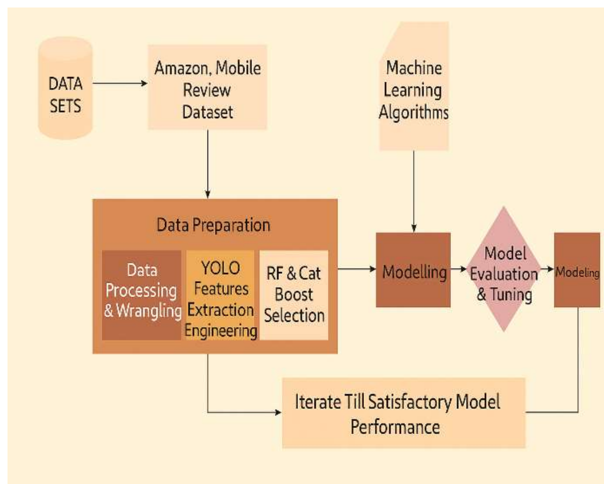

Figure 2: System Architecture

Polarity classification is defined as the assignment of a sentiment to a textual comment on one of the three positive, negative, or neutral classes [14,15]. This is important in natural language processing (NLP) and sentiment analysis, where it is necessary to interpret the opinion of the user in order to use it in applications like customer review mining, product evaluation and social media monitoring. There are different methods of detecting polarity, both simple rule-based lexicon networks and other sophisticated machine learning and deep learning tools. Here, VADER (Valence aware dictionary and sentiment reasoner) is used in detecting polarity. VADER[16] is a sentiment analysis model explicitly trained on social media and review-like text, providing consistent polarity scores in the three categories positive, negative and neutral as well as a composite score. For example:

a = "This is the worst mobile."

When VADER sentiment analysis is applied to the text above, the polarity scores are:

Negative: 0.9

Neutral: 0.1

Positive: 0.0

Composite: –0.8404
a highly negative implication.

4. METHODOLOGY

The suggested methodology combines the YOLO feature extraction with a RF-CATBoost ensemble classifier to obtain high-performance sentiment analysis based on product reviews. The neural net described as YOLO is used to extract rich contextual and semantic features of visual or textual representations of reviews, which are able to be analyzed further about sentiment sensitivity. The extracted features are then sent to Random Forest (RF)[17,18] to enhance feature selection and CATBoost to classify with gradient-boosting to create a hybrid ensemble with the potential to process large-scale and high-dimensional data in an efficient manner. Lastly, the classified outputs get further validated through VADER sentiment polarity scoring that more precisely classifies the decision by giving explicit polarity orientation (positive, negative, neutral). This two-layer system feature-driven ensemble classification and lexicon-based polarity scoring are what guarantee accuracy and interpretability in sentiment detection.

The workflow of the suggested YOLO + RF-CATBoost model of sentiment classification and polarity prediction is presented in Figure 3.

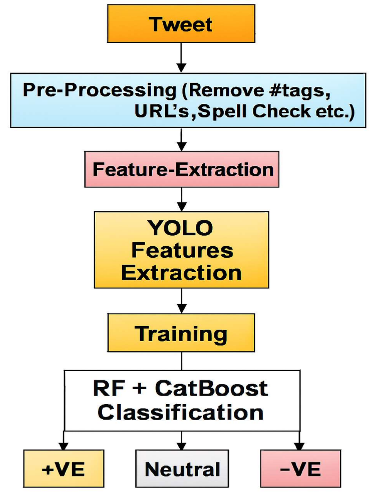

Figure 3: Workflow of Proposed Classifier

4.1 .Why CUDA?

CUDA (Compute Unified Device Architecture) is a parallel computing platform and an API developed by NVIDIA enabling the use of the computing capabilities of GPUs (Graphics Processing Units) to general-purpose computing applications, which do not need to be graphics rendering.

Parallel Processing Power: CUDA[19] opens the immense parallelism of the current GPUs, which have hundreds to thousands of cores that can run computations in parallel. This makes GPUs to be perfect in jobs that can be paralleled.

Performance: Applications can experience impressive gains in speed versus running on a conventional CPU by shifting the computation-intensive workloads to GPUs[20]. This can be effectively used with large datasets and complicated algorithms.

Portability CUDA works with a variety of programming languages, which can include C, C++, Python, or Fortran. It also has optimized libraries and tools in fields like linear algebra, signal processing, and machine learning.

Deep Learning and AI: CUDA plays a key role in deep learning, driving frameworks such as TensorFlow, PyTorch, and MXNet to compute model training and inference faster, allowing more rapid experimentation and development. Scientific and Data Computing: CUDA finds extensive use in scientific fields in simulations, numerical computation, and computational fluid dynamics, data-intensive applications such as image/video processing, computer vision, and analytics. Figure 4 shows computation differences between CPUs and GPUs. The key aim of this work is the implementation of the use of the GPU-based computing on mobile product reviews of Amazon Unlocked and the measurement of classification accuracy. Online platforms provide feedback about customers, and this customer feedback is examined to determine whether the review is positive, negative, or neutral. At both the word-level and sentence-level, we use VADER sentiment analysis, then transform the output into binary (0s and 1s), and test predictions using confusion matrices and performance indicators including precision, recall, and F1-score.The Proposed ensemble learning strategies include:

Stacking: Averages the results of many weak models that have been trained on bootstrap samples of the data to produce a strong aggregated prediction.

Boosting: This method sequentially trains models by paying attention to the errors made in earlier iterations and refining the accuracy as it goes.

We have used a hybrid ensemble method in this work that combines between bagging and boosting. The hybrid classifier minimizes biasness during every training phase, which results in better predictive accuracy.

Figure 4: CPU vs GPU

The main aim of this paper is to use GPU-based computing to the Amazon Unlocked Mobile dataset and assess the accuracy of sentiment classification. The data is in the form of customer reviews on mobile products obtained through online sources. The proposed framework would help in boosting the performance of prediction by utilizing the agility and specificity of feature extraction and ensemble learning that are agile and run on GPUs. The analysis of product reviews in this study will be in terms of positive, negative, and neutral. To this end, we use YOLO (You Only Look Once), which provides us with semantic and contextual features of review text, captured with fine-grained features at the expense of a review text representation, using the power of a graphical processing card. The features are subsequently extracted and also sent to a hybrid ensemble classification model, which combines both the Random Forest (RF) and CATBoost to make robust decisions and to support categorical variables and complex feature interaction. After preparing the feature set, the reviews are coded into binary values of polarity (0s and 1s). A confusion matrix is used to test predictions, with performance measurements like precision, recall, and F1-score to measure classification accuracy. Stacking and boosting strategies are incorporated into the ensemble design to enhance robustness. Stacking takes the RF and CATBoost predictive output and uses it to construct a stronger generalized final model, whereas sequential boosting minimizes the number of errors in previous iterations. This mixed group method works well to reduce bias and variance in various training phases, leading to a top-quality sentiment classifier optimized to run on a GPU.

5. IMPLEMENTATION & RESULTS

Algorithms like YOLO, Random Forest (RF), and CatBoost are applied to classification. Each of the algorithms is analysed in terms of processing time in the CUDA program. Computation by the use of GPUs always occupies a shorter processing time than its counterpart of the CPU based computation [21]. On other thread levels, mining algorithms executed on a GPU demonstrate a strong speedup, proving that the speed of computation on a GPU is higher than on a CPU.

We tested the hybrid classifier on a random Amazon Mobile Phone data set to determine its effectiveness. The main goal was to forecast the

Journal of Theoretical and Applied Information Technology
15th December 2025. Vol.103. No.23
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

star rating of a mobile phone using its features of customer review rating. As CUDA is not explicitly linked to external databases, the classification input dataset was simply created in CUDA [22,23] by the curand intrinsic. CUDA is very effective in creating random data; creating 100,000 points of data only took 0.05 seconds.

The use of the YOLO+RF-CatBoost classifier can be subdivided into seven steps, which will be described as follows:

Random Data Generation: Generating feature data in a review-based manner is done using curand in CUDA to create synthetic data to train and evaluate.

Data Preprocessing: Process numerical and categorical data and label (rate) values and normalize the data to be classified.

Feature Extraction with YOLO: Represent objects (review features, sentiment tokens, product attributes) in a structured form. YOLO identifies and imprints salient patterns and uploads them to the classification models.

Sort and Separation: Process numeric variables, establish separators, and establish decision boundaries between classification.

Hybrid Model Construction: This method, which can be used to create the classifier, involves bagging features extracted by the YOLO detector with ensemble learning- Random Forest which bags features, and catBoost which further boosts gradient-based classification models with categorical data.

GPU Memory Allocation: The data of a dataset is allocated on a GPU hardware with cudaMalloc, and it is copied with cudaMemcpy. This involves two transfers: host(CPU) to device(GPU) and results (GPU) to host(CPU) [24]. Model Operating and Accuracy Calculation: Run the classification with different thread counts. Indicatively, in cases where 128 threads were used to work on 100,000 results, it took a few microseconds to execute. The count of data elements processed will be equal to the count of CUDA threads launched and the classification will finish proportionately with the increase in thread counts. Table 1 demonstrates the impact of the number of threads, Table 2 indicates the acceleration ratio of the YOLO+RF–CatBoost classification on GPUs, and Table 3 provides details about the dataset that has been used in the experiment of product recommendation.

Table 1: The different threads and count are indicated.

Threads VS Count	
No. of Threads	**Time Taken (seconds)**
128	**5.88**
256	**4.84**
512	**3.75**
1024	**2.92**

Table 2: Shows Acceleration Ratio time for Classifying Records using YOLO+RFCATBoost Algorithm.

CUDABB GPUs Times	No.of. Records Sec / 10000	No.of. Records Sec / 30000	No.of. Records Sec / 50000	No.of .Records Sec/ 70000	No.of. Records Sec / 100000
Classification Time	0.535	1.015	1.620	2.045	2.684
CPU Time	0.710	1.130	1.740	2.3500	2.900
GPU Time	0.550	1.010	1.640	2.230	2.490
Acceleration Ratio	1.296	1.118	1.064	1.054	1.16491

Table 3: Shows Dataset Description

Features	**Values**
Name of the Dataset	Amazon.csv
Dataset URL	F:\python\proj\Amazon1.csv
Number of Reviews	553,840
Classes	Positive, Negative, Neutral
File Type	CSV

Acceleration Ratio for GPU:

The GPU acceleration ratio represents the performance gain achieved by shifting

Journal of Theoretical and Applied Information Technology
15th December 2025. Vol.103. No.23
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

computational tasks from the CPU to the GPU. It highlights the improvement obtained by leveraging the GPU's parallel processing capabilities compared to executing the same task solely on the CPU. In Table 4 the obtained confusion matrix for the proposed classifier is shown and Figure 5 Shows the Performance Metrics for Proposed Classifier. And in Table 5 the validation table of the Proposed classifier is shown. And, in Table 6 the obtained confusion matrix for the ResNet-50 classifier is shown and Figure 6 Shows the Performance Metrics for ResNet-50 Classifier. And in Table 7 the validation table of the ResNet-50 classifier is shown. In Table 8 the confusion matrix obtained for the RegNetY classifier is shown and Figure 7 Shows the Performance Metrics for RegNetY Classifier. And in Table 9 the validation table of the RegNetY classifier is shown. The acceleration ratio is calculated as[24]:

Acceleration Ratio= CPU Time / GPU Time

Where:

CPU Time → The time taken to complete a task when executed on the CPU.

GPU Time → The time taken to complete the same task when executed on the GPU.

Performance evaluation method:

Various performance measures are used in machine learning to evaluate the effectiveness of a model in the performance of tasks that include classification, regression, and clustering. Some of the most widely used metrics together with the formulae are as under [25]:

Accuracy=(TP+TN)/(TP+TN+FP+FN) *(1)*

Precision=TP/ FP+TP *(2)*

Recall=TP / FN+TP *(3)*

*F1=2(Precision * Recall) / (Precision + Recall)* (4)

Table 4: Confusion Matrix Obtained for Proposed Classifier on the Amazon mobile phone dataset.

	Actual Class	
Predicted Class	**22972**	**724**
	876	**14128**

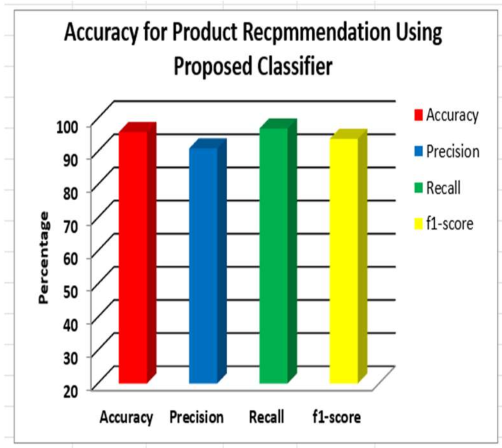

Figure 5: Shows the Performance Metrics for Proposed Classifier.

Table 5: Accuracy Table for Proposed Classifier on the Amazon mobile phone dataset.

Shows the accuracy of Proposed Algorithm

Star Rating	Precision	Recall	F-Score	Support
1	**96.13**	**96.05**	**93.09**	**23696**
5	**95.15**	**95.98**	**92.59**	**15004**
Accuracy			**95.87**	**38700**
MicroAvg	**96.83**	**96.05**	**93.09**	**38700**
MicroAvg	**95.15**	**96.98**	**93.59**	**38700**

Journal of Theoretical and Applied Information Technology
15th December 2025. Vol.103. No.23
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 6: Confusion Matrix Obtained for ResNet-50 Classifier on the Amazon mobile phone dataset.

Predicted Class \ Actual Class		
	20670	**1790**
	2012	**14228**

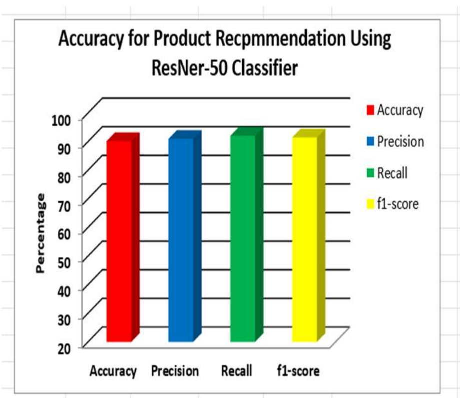

Figure 6: Shows the Performance Metrics for ResNet-50 Classifier.

*Table 7: Accuracy Table for **ResNet-50** Classifier on the Amazon mobile phone dataset*

Shows the accuracy of **ResNet-50** Algorithm

Star Rating Support	Precision	Recall	F-Score
1 22460	**90.23**	**89.19**	**89.70**
5 16240	**89.85**	**91.98**	**90.59**
Accuracy 38700			**90.95**
MicroAvg 38700	**90.23**	**89.19**	**89.70**
MicroAvg 38700	**89.85**	**91.98**	**90.59**

Table 8: Confusion Matrix Obtained for RegNetY Classifier on the Amazon mobile phone dataset.

Predicted Class \ Actual Class		
	20265	**2560**
	2342	**13533**

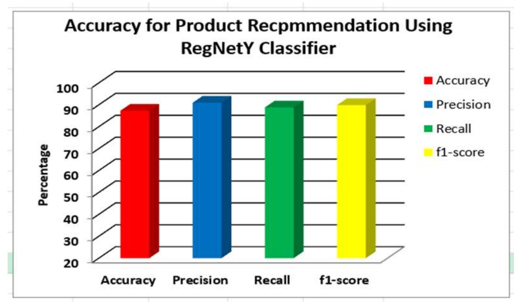

Figure 7: Show the Performance Metrics for RegNetY Classifier.

*Table 9: Accuracy Table for **RegNetY** Classifier on the Amazon mobile phone dataset.*

Shows the accuracy of **RegNetY** Algorithm

Star Rating Support	Precision	Recall	F-Score
1 22825	**88.81**	**88.75**	**89.79**
5 15875	**89.15**	**88.98**	**89.59**
Accuracy 38700			**87.95**
MicroAvg 38700	**89.81**	**88.75**	**89.79**
MicroAvg 38700	**88.15**	**88.98**	**89.59**

6. CONCLUSION AND FUTURE WORK:

The proposed YOLO + RF + CATBoost classifier can utilize parallel computing that is built on a CPU-to leverage the power of a graphic card, namely, to improve the efficiency of sentiment classification. In contrast to the conventional approach, it combines YOLO to gain effective feature extraction and RF + CATBoost to gain robust classification, which is more accurate and scaled. The model categorizes text based on sentiment polarities, which are positive, negative, and neutral, in a more accurate way as compared to the traditional methods. When used in the performance evaluation, the proposed architecture had the highest accuracy of 96 percent, which is greater than ResNet-50 (90 percent) and the RegNetY (87 percent), indicating the effectiveness of the integration of object detection-based feature learning with gradient boosting. Better precision, recall and F-measure mean that misclassification is minimized and the performance balances out when it comes to sentiment categories. The graphics processor-based model is much faster and reduces processing time and offers greater speedups than the CPU-based systems, which means that it can be used in large-scale and real-time sentiment analysis. In addition, the model is flexible, which allows its applications in many fields, including fraud detection in the banking sector, patient feedback analysis, and biomedical text mining where having a quick and accurate sentiment interpretation is highly important. The framework does not affect the quality of the results because it can use all the potential of parallel processing of GPUs. In general, the YOLO + RF + CATBoost model is proved to be a scaled, correct, and fast system of real-life sentiment classifications.

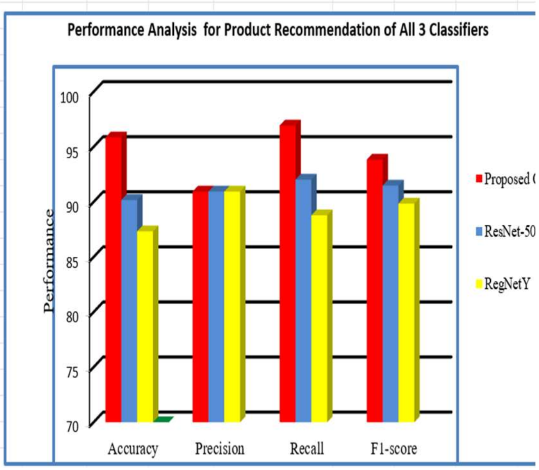

Figure 8: Shows the Performance of All 3 Classifiers.

REFERENCES

[1]. Siva Kumar Pathuri, N. Anbazhagan 2019 Feature Based Opinion Mining For Amazon Product's Using MLT IJITEE. 8(11).

[2]. Siva Kumar Pathuri and N. Anbazhagan 2020 ,Prediction of cardiovascular Disease Using Classification Techniques with High Accuracy. JARDCS.Vol-12(02).

[3]. A.Ejaz, Z.Turabee, M.Rahimand S.Khoja 2017 Opinion mining approaches on Amazon product reviews: A comparative study. International Conference on Information and Communication Technologies (ICICT),Karachi,pp173-1 79.

[4]. Sivakumar S, Periyanagounder, G., Sundar S, A MMDBM classifier with CPU and CUDA GPU computing in various sorting procedures. International Arab Journal of Information Technology. 14(6). pp 897-906.

[5]. S Anjali Devi and S Siva Kumar 2020, A hybrid document features extraction with clustering based classification framework on large document sets. International Journal of Advanced Computer Science and Applications.11(7) pp 364-374.

[6]. Siva Kumar Pathuri et.al, CUDA-EECA model for Crop Quality estimation with Edge Computing Using Machine Learning Technique, African Journal of Biological Sciences (South Africa), 2024, 6(5), pp. 1859–1874.

[7]. T. V. Sai Krishna, S. K. Pathuri, K. S. Sandeep, M. K. Padhi, I. Aswani and D. Haritha, "An Efficient Machine Learning Classification Model for Diabetes Prediction," 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2023, pp. 139-146, doi: 10.1109/CSNT57126.2023.10134615.

[8]. A.J.Singh 2017 Sentiment Analysis :A Comparative Study of Supervised Machine Learning Algorithms Using Rapidminer IJRASET. 5(Xi). pp 80–89.

[9]. Baig, M.M, Sivakumar, S, S R Nayak 2020 Optimizing Performance of Text Searching Using CPU and GPUs Advances in Intelligent Systems and Computing, 1119. pp 141-150.

[10]. Pathuri, S.K.; Anbazhagan, N.; Joshi, G.P.; You, J. Feature-Based Sentimental Analysis on Public Attention towards COVID-19 Using CUDA-SADBM Classification Model. Sensors 2022, 22, 80. https://doi.org/10.3390/s22010080

[11]. Sreedevi E, Premalatha, V, Sivakumar, S, Nayak, S.R 2019 A comparative study on new classification algorithm using NASA MDP datasets for software defect detection. Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS

2019. pp 312-317.

[12]. S. K. Pathuri, N. Anbazhagan and G. B. Prakash, "Feature Based Sentimental Analysis for Prediction of Mobile Reviews Using Hybrid Bag-Boost algorithm," 2020 7th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2020, pp. 1-5, doi: 10.1109/ICSSS49621.2020.9201990.

[13]. Shaik Razia, P.Swathi Pryathyusha, N.Vamsi Krishna 2018 A Comparative study of machine learning algorithms on thyroid disease prediction. International Journal of Engineering and Technology (UAE). 7(2.8). pp 315-319.

[14]. Shaik Razia, M.R. Narasinga Rao 2017 Development and Analysis of Support Vector Machine Techniques for Early Prediction of Breast Cancer and Thyroid. JARDCS (Journal of Advanced Research in Dynamical and Control Systems). 9(6). pp 869-878.

[15]. Pathuri SK, Anbazhagan N, Joshi GP, You J. Feature-Based Sentimental Analysis on Public Attention towards COVID-19 Using CUDA-SADBM Classification Model. Sensors. 2022; 22(1):80. https://doi.org/10.3390/s22010080.

[16]. Videla, Lakshmi Sarvani, et al. "Deformable facial fitting using active appearance model for emotion recognition." Smart Intelligent Computing and Applications. Springer, Singapore, 2019. 135-144.

[17]. S. K. Pathuri, N. Anbazhagan, M. K, T. V. S. Krishna and C. Prasad, "A Hybrid Bag-Boost algorithm for Feature-Based Sentimental Analysis on Mobile Reviews," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2021, pp. 948-953, doi: 10.1109/ICOSEC51865.2021.9591947.

[18]. Videla, Lakshmi Sarvani and M. Ashok Kumar P. "Fatigue Monitoring for Drivers in Advanced Driver-Assistance System." IGI Global, 2020, pp. 170-187.

[19]. Atmakur, V.K., Siva Kumar, P."A prototype analysis of machine learning methodologies for sentiment analysis of social networks", International Journal of Engineering and Technology(UAE), 2018,pp 963-967.

[20]. Prabha Selvaraj, Vijay Kumar Burugari , D. Sumathi ,Rudra Kalyan Nayak ,Ramamani Tripathy,"Ontology based Recommendation System for Domain Specific Seekers",Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) 2019,pp341.

[21]. T. Shiva, T. Kavya, N. Abhinash Reddy, Shahana Bano."Calculating The Impact Of Event Using Emotion Detection", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-7 May, 2019.

[22]. Prasanth Kumar1* Gera Pradeepini1. Pille Kamakshi," Feature Selection Effects on Gradient Descent Logistic Regression for Medical. Data", October 2019, International Journal of Intelligent Engineering and Systems 12(5) pp:278-286.

[23]. Chintala, R. R., Narasinga Rao, M. R. Venkateshwarlu, S. (2018). A proposal for observing conceived ladies having high risk of premature delivery using WHSN. International Journal of Engineering and Technology (UAE), 7(2), 53-56. doi:10.14419/ijet.v7i2.32.13524.

[24]. Sajana, T., Narasinga Rao, M. R. (2018). An ensemble framework for classification of malaria disease. ARPN Journal of Engineering and Applied Sciences, 13(9), 3299-3307.