International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

AI-Powered Brand Identity Modeling through Intelligent Name Generation and Market Evaluation

Balne Ramya, T. Srilekha*

Vaagdevi Engineering College, Warangal, 506005, Telangana, India.

*Correspondence: T. Srilekha

Abstract

Establishing a strong and distinctive brand identity is a key factor in gaining a competitive edge, but creating original, meaningful, and market-relevant brand names is often a complex and time-consuming task. Conventional approaches primarily depend on human imagination, brainstorming sessions, and rule-based strategies, which can be subjective, inconsistent, and limited in their ability to recognize complex semantic relationships or adapt to evolving market trends. To address these challenges, this work introduces an AI-powered framework that combines natural language processing (NLP) with machine learning to automate the generation and assessment of brand names. The proposed methodology starts with comprehensive text preprocessing, including data cleaning, tokenization, lemmatization, and stop-word elimination to produce high-quality textual inputs. It then employs transformer-based language models, such as DistilRoBERTa, to generate contextual embeddings that effectively capture semantic and linguistic information from words and phrases. For classification, several machine learning techniques, including Logistic Regression (LR), Random Forest Classifier (RFC), and Support Vector Machine (SVM), are trained using balanced datasets, where the Synthetic Minority Over-sampling Technique (SMOTE) is utilized to mitigate class imbalance. The developed system supports both individual and bulk brand name analysis, providing outputs such as predicted brand categories, relevance scores, and evaluation metrics through an intuitive web-based interface. By integrating advanced NLP capabilities with reliable machine learning algorithms, the proposed solution enhances scalability, efficiency, and prediction accuracy while minimizing manual effort, reducing bias, and accelerating the creation of effective, distinctive, and market-oriented brand names.

Keywords: Brand name generation, Natural Language Processing (NLP), transformer embeddings, DistilRoBERTa, semantic analysis, brand identity

1. Introduction

In today's fast-changing business landscape, characterized by the rapid growth of startups and continuous technological innovation, creating a unique and memorable brand name has become a crucial yet challenging aspect of business strategy. A brand name represents one of the first points of interaction between a business and its customers, playing a significant role in shaping brand perception, strengthening market identity, and improving long-term customer recognition. However, the increasing number of emerging businesses and online enterprises has created intense competition, making it more difficult to develop names that are original, legally compliant, and aligned with market expectations. Traditional brand naming methods, which mainly rely on manual ideation, brainstorming activities, or basic linguistic patterns such as single-word names, often suffer from limitations including trademark availability issues, insufficient creativity, and weak contextual relevance. These shortcomings highlight the need for

Received: 30-05-2026 | Accepted: 01-07-2026 | Published: 08-07-2026

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

more intelligent, scalable, and automated approaches to brand name creation [1,2,3].

Recent advances in artificial intelligence have introduced a transformative approach to branding, fundamentally changing the way organizations develop and manage their brand identity. AI-powered technologies, including ChatGPT, have significantly improved creative workflows by enabling automated content generation, supporting innovative idea development, and streamlining communication processes [4]. Through the application of NLP and machine learning algorithms, these systems are capable of understanding contextual language, producing human-like text, and processing large-scale consumer information to uncover valuable trends and behavioral patterns. As a result, businesses can create more personalized, context-aware, and engaging brand experiences that enhance customer satisfaction and strengthen consumer relationships. Beyond content generation, AI-based frameworks also facilitate essential branding activities such as automated brand name generation, semantic similarity assessment, sentiment evaluation, and predictive performance analysis. These capabilities allow organizations to make informed, data-driven branding decisions, optimize marketing strategies, and respond effectively to evolving market demands. Previous research indicates that incorporating AI into branding practices not only improves operational efficiency and fosters innovation but also establishes a systematic and customer-focused approach for developing distinctive and competitive brand identities in today's dynamic digital environment [5,6].

2. Literature Survey

Mendes, et al. [7] investigated the capability of multilingual transformer-based models to predict affective properties from textual data, focusing on emotional dimensions such as valence and arousal. Their work evaluated different transformer architectures, including DistilBERT and XLM-RoBERTa, to analyze the impact of model complexity on prediction performance. Experimental findings revealed that larger transformer models captured semantic and emotional information more effectively, although they required greater computational resources. The study highlighted the balance between computational efficiency and predictive accuracy for language understanding applications. Wang, et al. [8] introduced a semantic-oriented framework for identifying duplicate and highly similar textual content using advanced natural language processing techniques. Their approach combined Word2vec embeddings with Latent Dirichlet Allocation (LDA) to capture both semantic similarity and topic-level relationships. This integrated methodology significantly improved the detection of paraphrased and contextually related text compared with conventional keyword-matching approaches. Their results demonstrated the advantages of combining distributed word representations with probabilistic topic modeling for semantic comparison tasks. Harry Ph. Sophocleous, et al. [9] explored the growing role of artificial intelligence in entrepreneurship by examining its influence on business decision-making, operational efficiency, and customer intelligence. Their research emphasized the value of AI-driven analytics in supporting strategic planning while also discussing important challenges such as ethical concerns, algorithmic bias, implementation complexity, sustainability, and data privacy. Through comprehensive analysis, they concluded that successful AI adoption requires alignment with organizational goals and responsible technological integration.

Sadhuram, et al. [10] presented an AI-based question answering system designed to improve information retrieval through a structured NLP framework. Their

Received: 30-05-2026 | Accepted: 01-07-2026 | Published: 08-07-2026

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

methodology incorporated several stages, including query analysis, passage retrieval, sentence ranking, answer extraction, and summarization. The proposed system was evaluated using more than 422 SQUAD articles and approximately 87,599 cross-domain questions, demonstrating reliable performance across diverse information retrieval scenarios. Their work illustrated the effectiveness of integrating NLP and AI techniques for scalable question answering applications. Michael Gerlich, et al. [11] analyzed the influence of artificial intelligence on consumer behavior by comparing AI-generated recommendations with those provided by human social media influencers. Using a mixed-methods research design involving surveys from 478 participants and semi-structured interviews, they investigated trust, user perception, and recommendation effectiveness across multiple domains. Their findings suggested that AI systems perform well in analytical decision-support tasks, whereas human influencers remain more effective in establishing emotional engagement and authenticity. Jiang, et al. [12] proposed a hybrid deep learning framework that integrated Word2vec-based feature extraction with Long Short-Term Memory (LSTM) networks for enhanced text classification and prediction. Word embeddings generated contextual semantic representations, while the LSTM model captured sequential dependencies within textual data. The authors also compared the proposed architecture against benchmark models including RF, LR, SVM, k-Nearest Neighbors (KNN), and Hash Trick (HT). Experimental evaluation confirmed that the hybrid Word2vec-LSTM model consistently achieved superior predictive accuracy and robustness.

Talha Ahmed Khan, et al. [13] examined practical implementations of artificial intelligence across industrial and commercial sectors through detailed real-world case studies. Their research highlighted improvements in operational efficiency, decision-making quality, and resource optimization resulting from AI adoption. In addition, they discussed implementation challenges such as system integration, ethical considerations, and data security concerns. Their study provided comprehensive insights into the opportunities and practical limitations associated with deploying AI technologies in organizational environments. Nurmambetov, et al. [14] developed a sentiment analysis framework based on Logistic Regression (LR) for binary text classification. Their methodology employed a systematic preprocessing pipeline consisting of language detection, tokenization using the Natural Language Toolkit (NLTK), stop-word elimination, and Stanford-based lemmatization. These preprocessing techniques substantially improved text representation and classification performance. The study demonstrated that conventional machine learning models can achieve strong sentiment prediction accuracy when supported by effective feature engineering. Liu, et al. [15] proposed an enhanced paraphrase generation model that extended the traditional sequence-to-sequence (S2S) architecture by integrating an attention mechanism with topic-aware guidance. Topic information extracted using LDA served as prior knowledge to preserve semantic consistency during sentence rewriting. The combination of attention mechanisms and topic modeling produced paraphrases with greater fluency, contextual relevance, and diversity. Their findings demonstrated that incorporating semantic guidance significantly improves the quality and usefulness of automated text generation systems.

3. Proposed Methodology

This research provides an end-to-end intelligent framework for automated brand name generation, semantic evaluation, and

Received: 30-05-2026 | Accepted: 01-07-2026 | Published: 08-07-2026

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

category prediction using NLP and ML. Initially, the textual dataset is collected and subjected to multiple preprocessing operations to improve data quality and prepare meaningful input for feature extraction. The processed text is then explored through various analytical techniques, including target distribution analysis, word cloud visualization, N-gram analysis, part-of-speech tagging, and document length analysis, to understand the linguistic characteristics of the dataset. As illustrated in Figure. 1, transformer-based DistilRoBERTa generates contextual semantic embeddings, while categorical encoding, numerical feature extraction, and SMOTE collectively enhance feature representation and address class imbalance. Multiple classification algorithms, namely LR, RFC, and SVM, are trained and evaluated to identify the most effective prediction model. Finally, the selected model is deployed through an interactive web application that supports real-time inference, predicts brand categories, computes relevance scores, and stores prediction results for efficient and scalable brand analysis.

Data Collection and Text Preprocessing

The system begins by loading the textual dataset containing brand-related information in CSV format. The raw text undergoes several preprocessing operations, including cleaning, lowercasing, tokenization, stop-word removal, and lemmatization using NLTK and WordNet. These steps eliminate noise, normalize the textual content, and produce a high-quality corpus suitable for subsequent feature extraction

Exploratory Analysis and Feature Engineering

After preprocessing, the cleaned data is analyzed to understand its linguistic structure and distribution. Visualization techniques such as target distribution plots, word clouds, N-gram analysis, part-of-speech tagging, and document sequence length analysis are performed to identify meaningful textual patterns. The extracted insights support effective feature engineering for machine learning.

Semantic Feature Extraction and Model Training

The processed text is transformed into contextual semantic representations using the DistilRoBERTa transformer model. Additional categorical and numerical features are generated, while SMOTE is applied to balance minority classes and improve model robustness. The enriched feature set is then used to train LR, RFC, and SVM, and their performance is evaluated using multiple classification metrics.

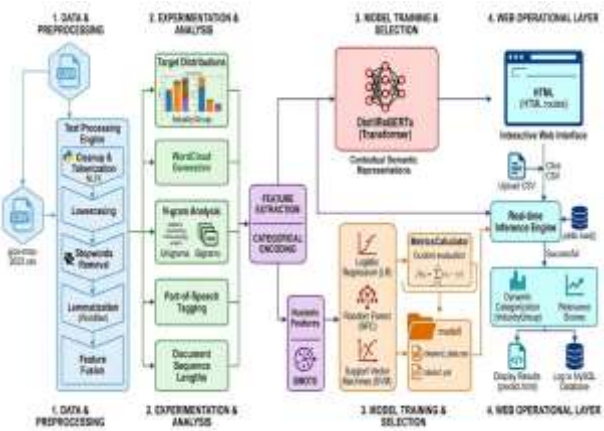


Figure. 1: Proposed System Architecture

Model Evaluation and Selection

The trained classifiers are compared using evaluation measures such as accuracy, precision, recall, F1-score, and relevance metrics. Performance analysis identifies the model that provides the best generalization and prediction capability. The selected model, along with the preprocessing artifacts, is saved for deployment and future inference.

Web-Based Deployment and Real-Time Prediction

The optimized prediction model is integrated into an interactive HTML-based web

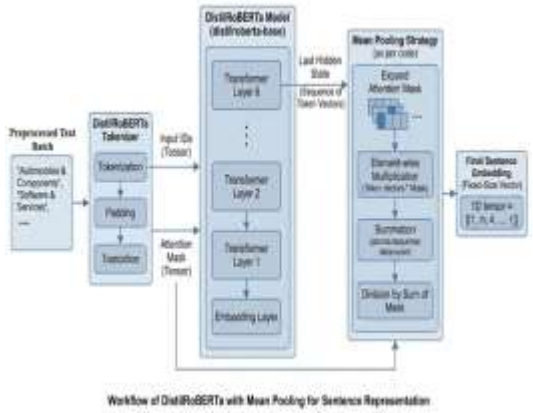
Received: 30-05-2026 | Accepted: 01-07-2026 | Published: 08-07-2026

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

application that enables both single and batch brand name evaluation. Users can upload CSV files or enter individual brand names to receive real-time category predictions and relevance scores. The prediction results are displayed dynamically and recorded in the MySQL database, providing an efficient, scalable, and user-friendly platform for automated brand name analysis.

3.1 DistilRoBERTa

DistilRoBERTa, is a lightweight and highly efficient version of RoBERTa created through knowledge distillation. It retains nearly 95% of RoBERTa language understanding capability while being significantly faster and smaller, making it ideal for real-time applications. Built on transformer self-attention, it captures deep semantic meaning and contextual relationships across text. Its compact architecture ensures quick inference without compromising on accuracy. Because of this balance between speed and performance, DistilRoBERTa is widely used for text encoding, classification, semantic analysis, and large-scale NLP deployments.



Workflow of DistilRoBERTa with Mean Pooling for Sentence Representation

Figure. 2: Internal working flow of DistilROBERTa features extraction.

In this research, DistilRoBERTa is used as a feature extractor to convert each cleaned text entry into a dense semantic embedding vector. The code loads the pretrained "distil Roberta-base" model and tokenizer from Hugging Face, tokenizes each text, applies padding/truncation, and passes it through the transformer, as shown in figure 2. Instead of using the CLS token, the pipeline applies mean pooling across token embeddings to obtain a context-rich sentence representation. These embeddings are then saved and fed into machine learning classifiers (LR, RFC, SVC) for predicting the Industry Group labels. By using DistilRoBERTa this way, the system gains powerful contextual understanding while keeping processing fast and efficient for a live Flask application.

Pre-processed Text Input: The cleaned text generated from the NLP preprocessing stage is collected into a unified list of processed sentences. This ensures all noise, stop words, and unnecessary symbols are removed before transformer encoding.

Tokenization + Attention Mask Creation (DistilRoBERTa Tokenizer): Each processed sentence is converted into token IDs while generating an attention mask that marks real tokens vs. padding. Padding and truncation ensure all sequences follow a consistent length for batch processing.

Embedding Layer Conversion: The token IDs are passed into the embedding layer of DistilRoBERTa, where each token is mapped to a high-dimensional vector. These embeddings serve as the initial semantic representation for the model.

Transformer Encoder Processing (Layers 1–6): The embeddings flow through six transformer layers where self-attention and feed-forward networks refine token-level representations. Each layer adds deeper contextual understanding to every token vector.

Extraction of Last Hidden States: DistilRoBERTa outputs a sequence of hidden states representing each token after the final transformer layer. These vectors contain the

Received: 30-05-2026 | Accepted: 01-07-2026 | Published: 08-07-2026

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

rich semantic information needed for classification.

Mean Pooling for Sentence Representation: The hidden states are multiplied by the attention mask, summed across the sequence, and divided by the number of valid tokens. This produces a single fixed-size embedding that represents the entire sentence.

Feature Matrix Formation for Balancing: All sentence embeddings are stacked to form the complete feature matrix. These feature vectors are now ready for the next stage of the pipeline SMOTE balancing.

SMOTE Balancing on Feature Embeddings: SMOTE oversamples minority classes using the feature vectors, ensuring all Industry Group categories have equal representation. This balanced dataset is then used for ML model training.

4. Results and Description

Figure 3 shows the confusion matrix for the Distil RoBERTa-WE-SVC on Industry Group classification. The matrix predominantly follows a perfect diagonal structure, with a few sparse misclassifications concentrated in classes with overlapping business terminology. These errors reflect the strict margin-based decision boundaries characteristic of SVCs, particularly when handling closely related categories. Even with these isolated deviations, overall classification remains highly accurate due to the dense and meaningful representation learned by the embeddings. The visualization confirms SVC's strong performance in high-dimensional semantic spaces.

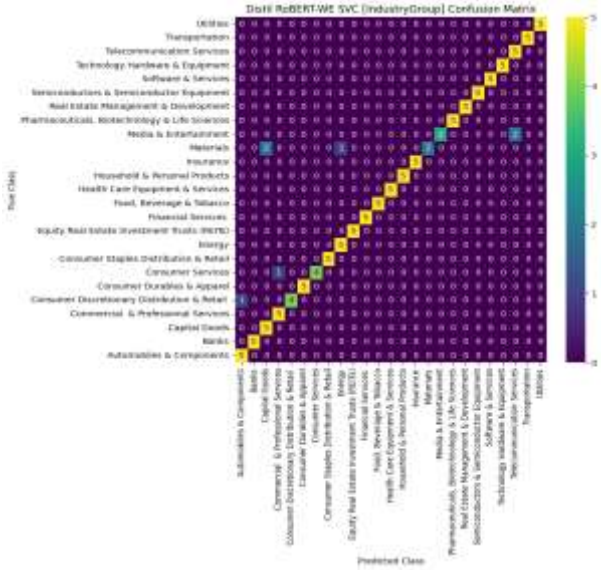


Figure. 3: Distil ROBERT-WE-SVC of Confusion Matrix for Industry Group

Figure. 4: Prediction Results Page

Figure 4 shows the results interface where predicted Industry Group labels are displayed alongside their corresponding input attributes. The table-format presentation organizes sector, industry codes, descriptions, and model-generated outputs into a structured view. This arrangement enables users to verify predictions against actual descriptions and interpret model behavior across multiple entries. The interface supports seamless scanning of numerous records, making large-scale validation efficient and transparent. It completes the prediction pipeline by translating backend model outputs into an accessible, user-facing representation.

Table. 1: Overall Performance Comparison of Classification models

Received: 30-05-2026 | Accepted: 01-07-2026 | Published: 08-07-2026

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
LR	100.00	100.00	100.00	100.00
RFC	99.20	99.33	99.20	99.19
SVC	94.40	95.71	94.40	93.97

Table 1 presents the overall performance comparison of three classification models trained on Distil RoBERTa word embeddings for Industry Group prediction. The LR model achieves a perfect score of 100% across accuracy, precision, recall, and F1-score, demonstrating that the Distil RoBERTa embedding space is highly linearly separable for this task. The RFC follows closely with strong performance achieving 99.20% accuracy and similarly high precision, recall, and F1-scores indicating excellent non-linear decision boundary learning. In contrast, the SVC records comparatively lower performance, with 94.40% accuracy and an F1-score of 93.97%, revealing slight difficulty in distinguishing semantically similar industry descriptions. These results, computed on the balanced test set after SMOTE oversampling, confirm LR as the most effective classifier for Industry Group categorization when paired with transformer-based embeddings.

5. Conclusion

This research successfully demonstrates the effectiveness of combining transformer-based language representations with conventional machine learning algorithms to achieve accurate text classification. By employing DistilRoBERTa embeddings, the framework extracts meaningful contextual and semantic information from textual descriptions, resulting in improved prediction accuracy. Comprehensive preprocessing operations, including text normalization, tokenization, and lemmatization, enhance the quality and consistency of the input data before feature extraction. The use of SMOTE effectively mitigates class imbalance, enabling balanced learning across different industry categories. Machine learning models such as LR, RFC, and SVC deliver strong classification performance when trained on these contextual semantic features. In addition, the framework incorporates extensive EDA, visualization methods, and caching mechanisms to streamline data analysis and improve computational efficiency. A Flask-based web application provides an intuitive interface that supports real-time prediction and seamless user interaction. Overall, the proposed framework offers a robust, scalable, and practical solution for deploying intelligent text classification systems in real-world applications.

References

[1] Arora, S.; Kalro, A.D.; Sharma, D. A comprehensive framework of brand name classification. J. Brand Manag. 2015, 22, 79–116.
[2] Moro Visconti, R. Domain Name Valuation: Internet Traffic Monetization and IT Portfolio Bundling. 2017. Available
[3] Eskiev, M. Naming as one of the most important elements of brand management. SHS Web Conf. 2021, 128, 01028.
[4] Patel, S.B.; Lam, K.; Liebrenz, M. ChatGPT: Friend or foe. Lancet Digit. Health 2023, 5, e102.
[5] Doshi, R.H.; Bajaj, S.S.; Krumholz, H.M. ChatGPT: Temptations of progress. Am. J. Bioeth. 2023, 23, 6–8.

Received: 30-05-2026 | Accepted: 01-07-2026 | Published: 08-07-2026

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

[6] George, A.S.; George, A.H. A review of ChatGPT AI's impact on several business sectors. Partn. Univers. Int. Innov. J. 2023, 1, 9–23.

[7] Mendes, G.A.; Martins, B. Quantifying valence and arousal in text with multilingual pre-trained transformers. In Proceedings of the European Conference on Information Retrieval, Dublin, Ireland, 2–6 April 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 84–100. Jiang, H.; Hu, C.; Jiang, F. Text Sentiment Analysis of Movie Reviews Based on Word2Vec-LSTM. In Proceedings of the 14th International Conference on Advanced Computational Intelligence (ICACI), Wuhan, China, 22 July 2022; pp. 129–134.

[8] Wang, X.; Dong, X.; Chen, S. Text duplicated-checking algorithm implementation based on natural language semantic analysis. In Proceedings of the IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 5 June 2020; pp. 732–735.

[9] Sophocleous, H.P. Harnessing Big Data and Artificial Intelligence for Entrepreneurial Innovation: Opportunities, Challenges, and Strategic Implications. Encyclopedia 2025, 5, 122. https://doi.org/10.3390/encyclopedia5030122

[10] Sadhuram, M.V.; Soni, A. Natural language processing based new approach to design factoid question answering system. In Proceedings of the 2nd International Conference on Inventive Research in Computing Applications (ICIRCA), Virtual, 15–17 July 2020; pp. 276–281.

[11] Gerlich, M. The Shifting Influence: Comparing AI Tools and Human Influencers in Consumer Decision-Making. AI 2025, 6, 11. https://doi.org/10.3390/ai6010011

[12] Jiang, H.; Hu, C.; Jiang, F. Text Sentiment Analysis of Movie Reviews Based on Word2Vec-LSTM. In Proceedings of the 14th International Conference on Advanced Computational Intelligence (ICACI), Wuhan, China, 22 July 2022; pp. 129–134.

[13] Khan, T.A.; Ali, S.M.; Ali, K.M.; Aziz, A.; Ahmad, S.; Anwar, A.; Khan, S.A. Harnessing Artificial Intelligence for Optimum Performance in Industrial Automation. Eng. Proc. 2024, 76, 105. https://doi.org/10.3390/engproc2024076105

[14] Nurmambetov, D.; Dauylov, S.; Bogdanchikov, A. Kazakh Names Generator Using Deep Learning. Her. Kazakh-Br. Tech. Univ. 2021, 17, 171–177.

[15] Liu, Y.; Lin, Z.; Liu, F.; Dai, Q.; Wang, W. Generating paraphrase with topic as prior knowledge. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 2381–2384.

[16] Kumar, Anuj. "Optimization of Machining Parameters in Turning Operations for Surface Roughness and Tool Wear." International Journal of Engineering Research and Science & Technology, Vol. 12, No. 4, 2016, pp. 47-64

Received: 30-05-2026 | Accepted: 01-07-2026 | Published: 08-07-2026