Counterfeit Logo Recognition Using Machine learning and Deep Learning Algorithms

K. Pavani [1], Ms. P.Syamala[2]

#1 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.

.#2 Student in the Department of MCA, SRK Institute of Technology, Vijayawada.

Abstract: The identification of counterfeit logos is a crucial application in computer vision, focusing on ascertaining the authenticity of a logo. The fast expansion of digital platforms, e-commerce, and online branding has rendered the improper use of business logos a significant issue affecting company identification, trust, and security. The manual verification of logos is laborious and susceptible to inaccuracies, underscoring the need for an automatic and sophisticated detection solution. This research presents a system that utilizes Machine Learning (ML) and Deep Learning (DL) techniques to classify logos as either authentic or counterfeit.

The system employs image pre-processing methods like scaling, normalization, and data augmentation to improve model effectiveness. Conventional machine learning methods, such as Support Vector Machines (SVM) and Random Forest, are utilized in conjunction with deep learning architectures like Convolutional Neural Networks (CNN) for feature extraction and categorization. The CNN model independently discerns essential aspects from logo images, eliminating the necessity for manual feature engineering.

The trained model's performance is evaluated using measures including accuracy, precision, recall, and F1-score to ensure reliable outcomes. This system is designed to provide rapid and accurate predictions, making it suitable for real-time applications like brand protection and fraud detection. The suggested technique greatly enhances the ability to effectively detect counterfeit logos and is crucial for maintaining brand authenticity in digital environments..

Index terms - *— Fake Logo Detection, Machine Learning, Deep Learning, Convolutional Neural Network (CNN), Support Vector Machine (SVM), Random Forest, Computer Vision, Image Classification, Feature Extraction, Brand Protection, Fraud Detection, Data Augmentation, Logo Authentication, Artificial Intelligence, Digital Security.*

1. INTRODUCTION

In today's digital era, logos play a crucial role in representing the identity of a brand. They are widely used in websites, advertisements, products, and social media platforms. However, with the rapid growth of online businesses and digital content, the problem of fake logo creation and misuse has increased significantly. Fake logos can mislead customers, damage brand reputation, and cause financial losses to companies. Traditional methods of logo verification rely on manual inspection, which is time-consuming, inefficient, and prone to human error. As the number of digital images is increasing rapidly, manual verification is not sufficient to handle large-scale logo authentication tasks. Therefore, there is a need for an automated and intelligent system that can accurately distinguish between original and fake logos.

Machine Learning (ML) and Deep Learning (DL) techniques provide powerful solutions for image classification problems. ML algorithms such as Support Vector Machine (SVM) and Random Forest use handcrafted features for classification, while DL models like Convolution Neural Networks (CNN) automatically learn features from images without manual intervention. These methods improve accuracy and efficiency in detecting counterfeit logos. This project focuses on developing a fake logo detection system using ML and DL techniques. The system processes logo images, extracts meaningful features, and classifies them as genuine or fake. The main goal is to enhance brand protection, reduce fraudulent activities, and provide a reliable automated solution for logo verification in real-world applications.

2. LITERATURE SURVEY

The rapid growth of digital platforms and online branding has increased the need for automated systems to identify counterfeit logos effectively. Various researchers have contributed significantly to the fields of computer vision, machine learning, and deep learning, which form the foundation for fake logo detection systems.

Krizhevsky, Sutskever, and Hinton introduced the groundbreaking paper “ImageNet Classification with Deep Convolutional Neural Networks” in 2012. Their work demonstrated the effectiveness of Convolutional Neural Networks (CNNs) in large-scale image classification tasks. The proposed AlexNet architecture significantly improved image recognition accuracy and inspired the adoption of deep learning techniques in object detection and logo classification applications.

LeCun, Bengio, and Hinton published the influential paper “Deep Learning” in Nature Journal in 2015. The study highlighted the capabilities of deep neural networks in automatically learning complex features from large datasets. Their research emphasized the importance of deep learning in computer vision, speech recognition, and image analysis, providing a strong theoretical foundation for modern logo detection systems.

Dalal and Triggs presented “Histograms of Oriented Gradients for Human Detection” in 2005. Their work introduced the Histogram of Oriented Gradients (HOG) feature descriptor, which became a widely used technique for object detection and feature extraction. HOG features are useful in identifying shape and edge patterns in images and have influenced several machine learning-based image recognition systems.

Lowe introduced the “Distinctive Image Features from Scale-Invariant Keypoints (SIFT)” algorithm in 2004. The SIFT method extracts robust and scale-invariant features from images, enabling reliable object recognition under varying image conditions. This technique contributed significantly to image matching and logo recognition tasks in computer vision applications.

Cortes and Vapnik proposed the “Support-Vector Networks” algorithm in 1995, introducing the Support Vector Machine (SVM) classifier. SVM became one of the most effective supervised machine learning algorithms for classification tasks due to its high accuracy and ability to handle high-dimensional data. It is widely used in image classification and counterfeit logo detection systems.

These studies collectively provide the theoretical and technical background for developing an efficient fake logo detection system using machine learning and deep learning techniques. Their contributions in feature extraction, image classification, and object recognition have greatly influenced modern computer vision applications..

3. METHODOLOGY

i) Proposed Work:

The proposed system seeks to create an automated and intelligent solution for detecting fake logos through the application of Machine Learning (ML) and Deep Learning (DL) techniques. In contrast to the current systems, this method minimizes manual intervention by automatically categorizing logos as either fake or genuine using trained models. The design of the system aims to enhance accuracy, speed, and reliability in the process of logo authentication. In this system, logo images are initially gathered and subjected to preprocessing steps, including resizing, normalization, and data augmentation, to improve image quality and consistency.

Following preprocessing, feature extraction is conducted utilizing both traditional ML techniques such as HOG and SIFT, as well as advanced Deep Learning models like Convolutional Neural Networks (CNN), which autonomously identify significant patterns within the images. The features that have been extracted are subsequently employed to train classification models, including Support Vector Machine (SVM), Random Forest, and CNN. These models undergo evaluation based on performance metrics such as accuracy, precision, recall, and F1-score to guarantee dependable results.

ii) System Architecture:

The identification of counterfeit logos is a crucial application in computer vision, focusing on ascertaining the authenticity of a logo. The fast expansion of digital platforms, e-commerce, and online branding has rendered the improper use of business logos a significant issue affecting company identification, trust, and security. The manual verification of logos is laborious and susceptible to inaccuracies, underscoring the need for an automatic and sophisticated detection solution. This research presents a system that utilizes Machine Learning (ML) and Deep Learning (DL) techniques to classify logos as either authentic or counterfeit.

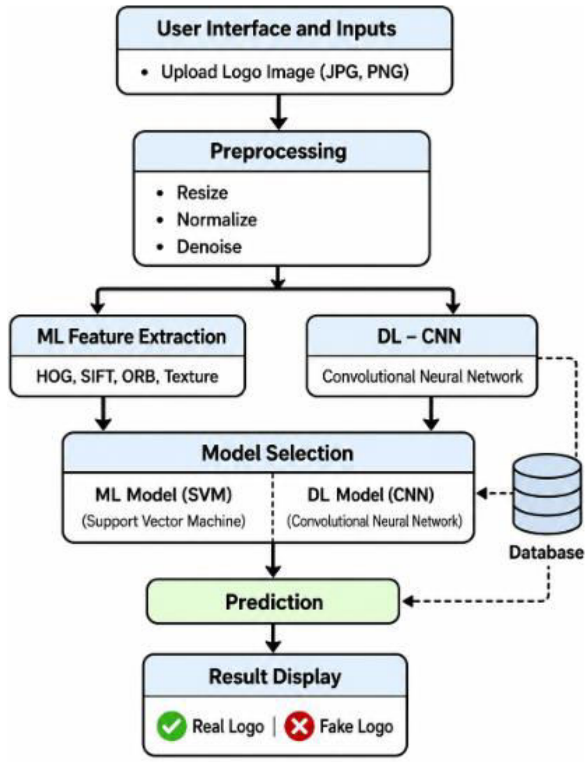
The system employs image pre-processing methods like scaling, normalization, and data augmentation to improve model effectiveness. Conventional machine learning methods, such as Support Vector Machines (SVM) and Random Forest, are utilized in conjunction with deep learning architectures like Convolutional Neural Networks (CNN) for feature extraction and categorization. The CNN model independently discerns essential aspects from logo images, eliminating the need for manual feature engineering.

The trained model's performance is evaluated using measures including accuracy, precision, recall, and F1-score to ensure reliable outcomes. This system is designed to provide rapid and accurate predictions, making it suitable for real-time applications like brand protection and fraud detection. The suggested technique greatly enhances the ability to effectively detect counterfeit logos and is crucial for maintaining brand authenticity in digital environments.The architecture of the Fake Logo Detection System is engineered to ascertain the authenticity of a logo, distinguishing between genuine and counterfeit, through the application of Machine Learning (ML) and Deep Learning (DL) methodologies. The procedure commences with the User Interface and Inputs, wherein users upload logo pictures in formats including JPG or PNG. The uploaded photos are subsequently forwarded to the Preprocessing stage, where image improvement techniques such as scaling, normalization, and denoising are employed to augment image quality and optimize model performance.

Subsequent to preprocessing, the system executes feature extraction utilizing two distinct methodologies. The ML Feature Extraction module employs conventional computer vision techniques, including HOG (Histogram of Oriented Gradients), SIFT (Scale-Invariant Feature Transform), ORB (Oriented FAST and Rotated BRIEF), and texture analysis, to derive significant image features. The Deep Learning - CNN module concurrently acquires significant visual characteristics from logo images through Convolutional Neural Networks.

The collected features are transmitted to the Model Selection phase, where Machine Learning models, including Support Vector Machine (SVM), and Deep Learning models, such as Convolutional Neural Networks (CNN), are employed for classification. The trained models engage with the database for the storage and retrieval of picture data and acquired parameters.

The Prediction module subsequently examines the input logo and produces classification outcomes. The Result Display module presents the output as either a "Real Logo" or "Fake Logo. This design facilitates precise, efficient, and instantaneous logo verification, appropriate for brand protection and fraud detection applications.



. Fig1 proposed architecture

iii) Modules:

The process of data collection is a crucial phase in the project aimed at detecting fake logos through Machine Learning (ML) and Deep Learning (DL). The effectiveness and precision of the system are primarily influenced by the quality and quantity of the dataset gathered. In this initiative, logo images are sourced from a variety of platforms, including Kaggle datasets, official company websites, online image repositories, and through web scraping methods. The dataset comprises both authentic and counterfeit logo images to facilitate effective model training. The assembled dataset features various logo alterations, encompassing modifications in size, color, orientation, brightness, and background conditions.

Counterfeit logos may consist of manipulated, edited, distorted, or imitation versions of original brand logos. To enhance the diversity of the dataset and augment the number of training samples, data augmentation techniques such as rotation, flipping, scaling, and cropping are employed. Once the data is collected, it is systematically organized and categorized into labels such as “Fake” and “Genuine.” Accurate labeling is essential, as supervised Machine Learning and Deep Learning models necessitate correctly classified data during the training phase. Subsequently, the collected data is partitioned into training and testing sets for the purposes of model development and evaluation. In summary, the data collection procedure guarantees that the system is equipped with adequate and high-quality data to train dependable models for precise fake logo detection.

Data collection represents a crucial phase in the creation of a fake logo detection system utilizing Machine Learning (ML) and Deep Learning (DL). The effectiveness of this system is significantly influenced by the presence of a large, varied, and accurately labeled dataset. In this initiative, logo images are gathered from several sources, including public datasets on Kaggle, official company websites, online image repositories, research datasets, and web scraping techniques. Both authentic and counterfeit logo images are incorporated to guarantee balanced training of the model.

The dataset comprises logos from various brands under different image conditions, including varying resolutions, sizes, colors, lighting conditions, orientations, and backgrounds. Counterfeit logos may consist of edited, duplicated, distorted, low-quality, or altered versions of original logos. Collecting such variations enables the model to understand the real-world distinctions between genuine and fake logos. To enhance the size and quality of the dataset, data augmentation techniques are employed. These techniques include image rotation, flipping, scaling, cropping, brightness adjustment, and zooming. Data augmentation increases the number of training samples and enhances the model’s capacity to generalize on previously unseen data. Following data collection, all images are systematically organized and labeled into categories such as “Fake” and “Genuine.” Accurate labeling is vital since supervised Machine Learning and Deep Learning models necessitate correctly classified data during the training phase.

The dataset is subsequently partitioned into training, validation, and testing sets. The training dataset is utilized to train the models, the validation dataset is employed for fine-tuning model parameters, and the testing dataset is used to assess the final performance of the model. Preprocessing tasks such as resizing, normalization, and noise removal are also conducted post-data collection to ensure consistency in image input. High-quality and well-balanced datasets contribute to improved model accuracy, reduced bias, and enhanced reliability of predictions.

iv) Algorithms:

1) **Logistic Regression**

Logistic Regression is a supervised machine learning algorithm used for binary classification tasks. It predicts the probability of an input image belonging to a particular class, such as real or fake logo, using a logistic function. The algorithm is simple, efficient, and suitable for baseline classification performance.

II. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful supervised learning algorithm used for image classification. It identifies the optimal hyperplane that separates different classes with maximum margin, enabling accurate classification of authentic and counterfeit logos even in high-dimensional feature spaces.

III. Random Forest Algorithm

Random Forest is an ensemble machine learning algorithm that combines multiple decision trees to improve classification accuracy and reduce overfitting. Each tree independently predicts the class label, and the final prediction is determined through majority voting, making the model robust and reliable for logo detection tasks.

IV. Histogram of Oriented Gradients (HOG)

HOG is a feature extraction technique used in computer vision for capturing edge and shape information from images. It computes gradient orientations in localized portions of an image, helping machine learning models identify structural patterns present in logos.

V. Scale-Invariant Feature Transform (SIFT)

SIFT is a feature extraction algorithm that detects and describes important key points in images. It is invariant to image scaling, rotation, and illumination changes, making it highly effective for recognizing logos under different viewing conditions.

VI. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a deep learning architecture specifically designed for image processing and classification. CNN automatically extracts hierarchical features from images through convolution, pooling, and fully connected layers, enabling accurate detection and classification of fake and genuine logos.

EXPERIMENTAL RESULTS

The experimental evaluation of the proposed system was carried out using IoT network traffic datasets containing both normal and malicious traffic records. The system was implemented using Python, Flask framework, and machine learning libraries such as Scikit-Learn, NumPy, and Pandas. Machine learning algorithms including Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) were trained using preprocessed IoT traffic data to perform anomaly detection and attack classification. The experimental results demonstrate that the proposed system successfully identifies abnormal traffic behavior and classifies various cyber-attacks such as DoS, DDoS, malware attacks, reconnaissance attacks, botnet activities, and data theft attacks with improved accuracy and reduced false positive rates.

The developed web-based application provides real-time anomaly detection, live input analysis, traffic monitoring, attack confidence evaluation, and risk reference analysis through an interactive user interface. Experimental outputs show that the system effectively analyzes network traffic samples, detects suspicious activities, and generates attack alerts along with confidence scores and traffic feature analysis. The system also supports manual live input analysis and reference-based risk prediction for identifying high-risk and low-risk network traffic conditions. The obtained results confirm that the proposed machine learning framework provides efficient, scalable, and intelligent security protection for IoT environments against modern cyber threats.

Accuracy: A test's accuracy is its capacity to distinguish healthy from ill cases. Find the percentage of instances with genuine positives and negatives to assess test accuracy.

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

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Classification accuracy or positive cases constitute precision. The formula for accuracy is:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: A model's recall measures its ability to recognize all appropriate machine learning class instances. The ratio of accurately predicted positive observations to total positives indicates a model's class instance detection skill.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: Mean Average Precision ranks quality. It considers the number and order of relevant ideas. Calculating MAP at K uses the arithmetic mean of each user or query's Average Precision (AP).

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

F1-Score: A high F1 score suggests an accurate machine learning model. Integrating recall and precision improves model correctness. Accuracy measures how often a model predicts a dataset correctly.

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

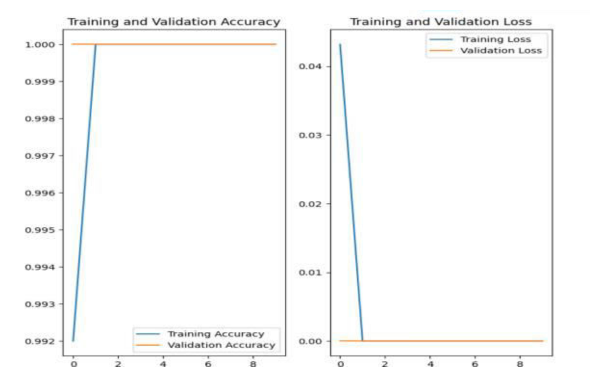


Fig2 The training and validation graphs illustrate that the Deep Learning model has attained exceptionally high accuracy with minimal prediction error in the fake logo detection project. This signifies robust model performance and efficient learning ability for logo classification tasks

Fig 3 The image illustrates the effective operation of the counterfeit logo detection system utilizing Deep Learning. The CNN model proficiently examines the uploaded logo image and determines if it is authentic or fraudulent. In this instance, the system accurately recognizes the uploaded Google logo image as a counterfeit, showcasing rapid prediction speed and precise classification ability

Fig 4 The illustration showcases the effective functioning of the counterfeit logo detection system that employs Deep Learning methodologies. The CNN model proficiently identifies the submitted Lay's logo as an authentic image, exhibiting rapid prediction capabilities and dependable performance.

4. CONCLUSION

The detection of counterfeit logos through Machine Learning (ML) and Deep Learning (DL) offers a sophisticated and intelligent approach for distinguishing between fake and authentic logos. This project effectively illustrates how image processing, feature extraction, and classification algorithms can be employed to automate the logo verification process. Conventional Machine Learning methods, including Logistic Regression, Support Vector Machine (SVM), and Random Forest, assist in the classification of logos based on the features extracted. In contrast, Deep Learning models such as Convolutional Neural Networks (CNN) enhance accuracy by facilitating automatic feature learning.

This system diminishes the need for manual intervention, reduces the likelihood of human error, and delivers quicker and more dependable results when compared to traditional verification techniques. By utilizing preprocessing methods, feature extraction, model training, and performance assessment, the system successfully classifies logos under various image conditions, including noise, scaling, and background changes. In summary, the proposed fake logo detection system bolsters brand protection, aids in fraud prevention, and offers a scalable solution for practical applications in areas such as e-commerce, trademark verification, and digital security.

This project underscores the significance of ML and DL technologies in effectively addressing contemporary challenges in image classification and counterfeit detection. The project further emphasizes the significance of high-quality datasets, appropriate preprocessing, and effective model training in obtaining reliable outcomes. Despite the existence of challenges such as dataset imbalance, computational expenses, and the availability of counterfeit logo samples, the system continues to deliver robust performance and practical applicability.

5. FUTURE SCOPE

The prospective future of the fake logo detection initiative utilizing Machine Learning (ML) and Deep Learning (DL) appears highly encouraging, given the growing demand for brand safeguarding and digital security. In the coming years, the system could be improved by implementing more sophisticated Deep

Learning frameworks like ResNet, EfficientNet, and Vision Transformers, which would enhance both detection precision and processing efficiency. Additionally, the incorporation of larger and more varied datasets would assist the model in effectively managing intricate real-world logo variations and newly generated counterfeit logos. The project has the potential to evolve into real-time applications by linking the system with mobile applications, e-commerce platforms, and cloud-based services for immediate logo verification. Prospective enhancements may also encompass multilingual brand identification, the detection of partially altered logos, and AI-driven counterfeit analysis.

Furthermore, the integration of blockchain technology could significantly bolster authenticity verification and secure brand management. Moreover, the system could be integrated with object detection and image segmentation methodologies to identify fake logos directly from videos, advertisements, and product packaging. With ongoing advancements in artificial intelligence and computer vision, the fake logo detection system is poised to become increasingly accurate, scalable, and applicable across various sectors, including cybersecurity, digital marketing, trademark protection, and online commerce.

The system can also be integrated with blockchain technology to ensure secure brand authentication and facilitate anti-counterfeit tracking systems. Artificial Intelligence-based monitoring tools may automatically scan social media, websites, and e-commerce platforms to detect unauthorized logo usage and trademark infringements. In the future, explainable AI (XAI) techniques could be incorporated to offer detailed explanations for model predictions, thereby enhancing user trust and transparency.

REFERENCES

1. Krizhevsky, A., Sutskever, I., & Hinton, G. E. authored the paper titled "ImageNet Classification with Deep Convolutional Neural Networks," which was published in the Advances in Neural Information Processing Systems (NIPS) in 2012.
2. LeCun, Y., Bengio, Y., & Hinton, G. contributed to the field with their work "Deep Learning," appearing in the Nature Journal, Volume 521, in 2015.
3. Dalal, N., & Triggs, B. presented their research "Histograms of Oriented Gradients for Human Detection" at the IEEE Conference on Computer Vision and Pattern Recognition in 2005.
4. Lowe, D. G. introduced "Distinctive Image Features from Scale-Invariant Keypoints" in the International Journal of Computer Vision in 2004.
5. Cortes, C., & Vapnik, V. discussed "Support-Vector Networks" in the Machine Learning Journal in 1995.
6. Breiman, L. published "Random Forests" in the Machine Learning Journal in 2001.
7. Goodfellow, I., Bengio, Y., & Courville, A. released the book "Deep Learning" through MIT Press in 2016.
8. Simonyan, K., & Zisserman, A. explored "Very Deep Convolutional Networks for Large-Scale Image Recognition" at the International Conference on Learning Representations (ICLR) in 2015.
9. He, K., Zhang, X., Ren, S., & Sun, J. authored "Deep Residual Learning for Image Recognition," which was presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) in 2016.

10. Szegedy, C., Liu, W., Jia, Y., et al. wrote "Going Deeper with Convolutions," which was featured at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) in 2015.

11. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. introduced "You Only Look Once: Unified Real-Time Object Detection" at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) in 2016.

12. Ren, S., He, K., Girshick, R., & Sun, J. presented "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" in the IEEE Transactions on Pattern Analysis and Machine Intelligence in 2017

Author Profiles

Mrs. K. Pavani is working as an Assistant and Head of Department of MCA, in SRK Institute of technology in Vijayawada. She completed her MCA and M.Tech in Computer Science. She has 10 years of teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. Her areas of interest include AI and ML, etc.

Ms. P.Syamala is an MCA Student in the Department of Computer Applications at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. She has Completed Degree in B.Sc. (Computer Science) from Little Flower Degree College, Chinthalapudi. Her area of interests are DBMS and Machine Learning with Python.