Fake Product Detection

Malampati Jai Durga Bhavani[1], Pulipati Sneha[2], Anandula Mahitha[3], Mandadi Akshitha[4]
[1,2,3,4]*Department of Computer Science Engineering (Artificial Intelligence and Machine Learning), Sreyas Institute of Engineering and Technology, Hyderabad, India*

Abstract -In recent years, There are many more counterfeit products. It became easier in the production of fake products. The main purpose is to check whether the items are real or fake. The aim is to check the product images, price, product description and barcodes if they match, can be conclude it as original items. This helps the people for not buying such counterfeit products and by reducing such things the people can trust and can buy anything. Not only the general products this play a major role in medical, it is very dangerous. This paper will provide a better solution for counterfeit products using Artificial Intelligence. These fake products are mainly in the online shopping system. This project has the potential to greatly reduce the impact of counterfeit goods on online platforms. It offers a scalable, automated solution that can work across different product categories and languages. To test this System, we collect the data from the different online platforms and declare it as real or fake. This gives the trust to people who shop on online and doesn't encourage the fake products.

Keywords: Artificial Intelligence, Counterfeit, Medical, Barcodes, Real and fake products.

1.INTRODUCTION

Now a days, it is easier to buy anything online rather than offline. But with that benefit comes a big problem: fake products. Counterfeit items such as clothes, electronics, medicines, beauty products etc. might look like real products but they always fall short in quality and safety. Counterfeit electronics can overheat or catch fire, and fake beauty products might cause skin damage or allergic reactions. Generally, to identify fake products people do things like scanning barcodes, but this method is slow, not trustworthy and don't work well when dealing with millions of products.

That's where Artificial Intelligence comes in. AI can help us find and stop fake products faster and smarter. It can look at tons of data, spot tiny details that humans might miss, and learn to tell the difference between real and fake over time.

Image Checking: AI looks at the photo of the product that we took and compare it to the real product and helps in picking whether the product is real or fake like by noticing when a logo is a little different or the colours may be quite different from the original ones.

Text Scanning: AI reads the product descriptions, reviews, seller information to catch fake products like difference in wording, fonts, patterns that often shows up in scams.

Tracking Origins: When combined with technologies like blockchain or IOT, AI can help trace where a product actually came from, helping verify if it's genuine.

1.1 Problem Statement

With the growing popularity of online shopping, the problem of fake products being sold on e-commerce platforms is becoming worse. These fakes can put people in danger and result in large financial losses for both customers and real sellers, in addition to hurting well-known companies. Using product images, descriptions, user ratings, barcodes, and seller profiles, this study aims to create an advanced system that can recognize fake products on its own. The goal is to help consumers make safer transactions and increase their confidence in online shopping by differentiating authentic products from counterfeit ones.

1.2 Objective

The objective of this project is to develop a smart AI system that identifies fake products by looking closely at things like product images, logos, packaging and descriptions. Using machine learning and computer vision, the system will be trained to tell the difference between real and counterfeit items with a high level of accuracy.

So, the idea is to create a tool that checks the product's validity in real time whether it is being sold online or in the store, so that people can buy the products trustfully.

Literature

a. Image-Based Detection

Convolutional Neural Networks (CNNs) are widely used to detect differences in product appearance (logos, packaging, design).

Applications: clothing, branded accessories, electronics, cosmetics.

Transfer learning with pre-trained models like ResNet, VGG, and EfficientNet is common.

Example paper:

"Image-based counterfeit detection using deep learning" – IEEE, 2021.

b. Text and NLP-Based Detection

Natural Language Processing (NLP) techniques are used to detect fake product listings or reviews on e-commerce platforms.

Models: LSTM, BERT, RoBERTa for analyzing product descriptions, seller behavior, or customer reviews.

Example paper:

"A deep learning approach for fake review detection" – Elsevier, 2020.

c. Multimodal Learning

Combines visual data (images) and textual data (descriptions, metadata).

Enhances detection accuracy by cross-validating multiple sources of information.

Example approach:

Using joint CNN + RNN or multimodal transformers.

d. Anomaly Detection

Unsupervised learning methods like autoencoders, isolation forest, and clustering to detect unusual patterns in sales or product features.

Useful when labeled data is scarce.

Applications by Sector

E-commerce: Fake listings, fake reviews, and seller fraud detection.

Pharmaceuticals: Detecting counterfeit medicines via image analysis or QR traceability.

Luxury goods: Visual verification of logos, stitching, material, etc.

Electronics: Verifying packaging and part authenticity.

Data Challenges in the Literature

Limited access to high-quality, labeled datasets of fake products.

Need for domain-specific training (e.g., fake luxury bags vs fake drugs).

Counterfeiters evolve, requiring continuous learning and model adaptation.

Complementary Technologies

Blockchain for supply chain integrity.

IoT and RFID to track product movement and verify origins.

Computer vision + OCR to analyse labels and serial numbers.

Limitations and Gaps in Current Research

Limited generalizability across product categories.

Adversarial examples can trick AI models.

Dependence on high-quality images and structured data.

Few real-time detection systems in production.

Representative Studies and Tools

"Detecting Counterfeit Product Listings on Online Marketplaces" – SIGIR 2020

"Fake Product Review Monitoring Using Deep Learning Techniques" – Springer 2021

Amazon's Project Zero, Alibaba's AI-based anti-counterfeiting tools.

2. RELATED WORK

The detecting fake product has been the smarter now a days and thanks to technology. This product has many cool ways to be done. This product has the most powerful methods to throughout images to instance this there is one AI tool called Deep logo, this can recognize or match the brand logos which are there in product photos. If the product is made in an exact imitation of something valuable which has the slightly of or else it poorly copied logo, then the AI will spot it either it can have small difference (just small difference) between the people it probably can't notice it.

The Another method or technique called Siamese network (it has nothing related to cats!) Anyways, it goes through images. Text also plays a main role in this product. It is also used to read the product reviews (AI) and one picks the fake spots throughout the odd patterns. For example, sometimes reviews sound like overall perfect and upon repeats of strange phrases.

It somewhat like giving or instructing the computer to come to a sense when something fells "off". Another AI tools like BERT have the feature to scan the product descriptions and to seller the profiles. All

these tools may raise some of the concerns, for example, if there is a fine jewellery or best spoke furniture or luxury cars etc.. All these will be with an offered price, may look better to fact or seller has a little knowledge about it.

Another side Blockchain technology is used to add extra layer to this product. The blockchain method can track the product of the journey from hand-hand or from manufacture to persons hand and makes hard find the fake product. Parallel, smart contracts can securely and eternally save data of the product and carrying the product manufactured data and location of the product.

This specifically used to cost the product like healthcare and food industry where geniuses are required. Close but no cigar, some elegant systems are connected to each of these projects. to better find the fake ones, all in one system can combine the sell a background, review proof, phots analysis, and so on. These kinds of tools are had been in the major projects like amazon and Alibaba to directly identify and number of questionable tasks in the announcements.

3. METHODOLOGY

When identifying if a product is real or counterfeit, the Counterfeit Monitoring System improves testing efficiency. The majority of systems that predict product failures include machine learning, statistical, and algorithm techniques.

We need historical data that includes examples of fake data in order to determine which items are most likely to be counterfeit. Using this data, an algorithm is shown to detect patterns and generate exact predictions about a product's quality.

AI-based counterfeit surveillance systems may be able to detect over 70% of all issues, per a recent study on defect prediction. As a result, only 35% to 60% of standard manual tests were found. This illustrates how artificial intelligence (AI) may significantly increase the precision of detecting defective or fake products.

3.1 NLP (NATURAL LANGUAGE PROCESSING)

Using a text-based AI model, such as natural language processing (NLP), to review the text on product packaging and logos is a different successful approach. An NLP model has been learned to identify suspicious or fake text patterns with a large data set which includes both real and fake logos each other with their packaging.

In between identifying and decoding human language, this means creating models and algorithms that can create it. These systems have a wide range of applications, including machine translation, sentiment analysis, classification of text, and exchange system design.

By this method, the system uses a two-step pipeline. In the first phase, which is called feature extraction, important features, or hints, are chosen from the input text. After that, the text is cleaned up to divide it into tokens, they're almost easy, clear words that the model can understand.

When analysing a piece of text, the feature extractor checks whether certain important words or phrases appear. If it finds one, it marks that feature as present.

3.2 Hog Based Logo Detection

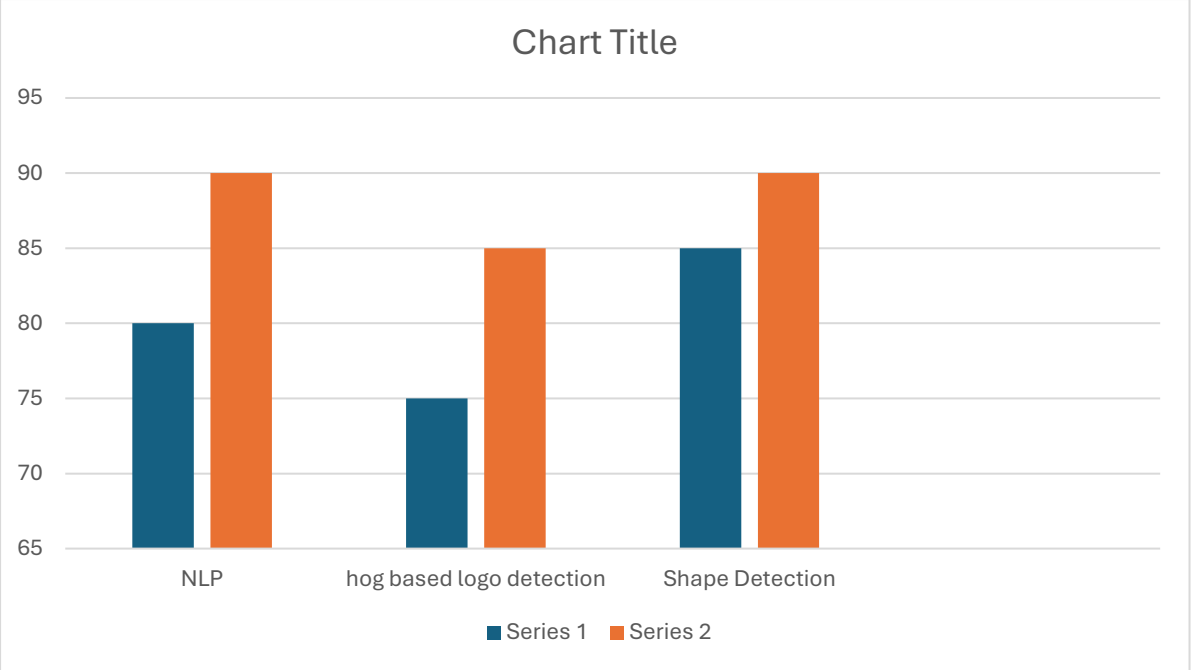
The History of Oriented Gradients (HOG) approach includes both training and testing phases. During training, an SVM-based (Support Vector Machine) logo classifier is trained separately for each logo class. Then in the testing phase, the system scans scene images using a sliding to detect and identify logos that belong to a specific class.

HOG is a feature descriptor used in image processing and computer vision to detect objects in a picture. It measures the frequency with which particular gradient directions basically, changes in edges or intensity directions occurs in particular, limited areas of a picture. These patterns help the computer system to understand the form and texture of object in the image. And the fields of computer vision and pattern recognition, HOG is widely used to characterize the local researchers Dalal and Triggs developed a term at the 2005 CVPR conference, and it immediately became well-known**.**

3.3 Shape Detection

Shape recognition plays a key role in image processing. It focuses on identifying and understanding shapes within different parts of an image especially areas that stand out because of difference in brightness, colour, or texture. This approach is especially useful for recognizing real logos using shape-based techniques. Additionally, shape features generally show out more than other components like texture when using local feature approaches like Scale-Invariant Feature Transform

(SIFT). Some kinds of objects are better differentiated by their shape than by their texture, which helps to explain this given possibility of noisy or irregular local variables, form recognition often returns higher quality results. Finding and identifying things in a photograph by their form is the primary objective of shape detection. This activity is often performed by computer vision, a branch of artificial intelligence that is focused on assisting computers in understanding and decoding visual data from their Surrounding. According to the specific task at hand and the features of the objects being studied, there are several Algorithms and methods for create verification that may be used.



- Series1 is actual percentage of NLP and series2 is the also of NLP percentage in between.
- Series1 is actual percentage of hog-based logo detection and series2 is the also of percentage of hog-based logo detection in between.
- Series1 is actual percentage of Shape Detection and series2 is the also of percentage of Shape Detection in between.

4. PROPOSED SYSTEM

4.1 Image-Based Product

This system helps identify fraudulent items simply with the aid the pictures. The image can either be captured using a mobile phone to ensure the analysis conducted will yield credible outcomes.

In addition to packaging and design, it also scans barcodes, OQ codes, logos and text on the label. Smart AI technology now can authenticate a product label not just by reading it but also by recognizing the image which goes along with it analyses the packaging, logos, barcodes, and compares it with the databases of authentic product and decides if everything is in order.

It possesses the intelligence to identify warning signs, including grammatical error, incorrect labelling, counterfeit logos, and faulty OQ codes. When discrepancies arise, the product is marked as potentially counterfeit. To assist user in making educated choices, the system can also assign and "authenticity score" that indicates the likelihood of the product being genuine. It is an uncomplicated and unobtrusive solution that can be implemented on a large scale, making it ideal for applications such as online retail, customs inspections, or aiding everyday shoppers in steering clear of fakes.

Algorithm

CNN algorithm that has been trained to differentiate between fake and real photos.

Step 1: Gathering Information

Gather product photos and evaluate them as authentic or fake.

Step 2: Preparing the data

All photos should be resized to a specific size, such as 128 by 128 pixels.

Pixel values should be levelled (scaled between 0 to 1).

Divide the dataset into sets for training and validation.

Step 3: Design of the CNN Model

1. Layer of Input: (128, 128, 3)
2. Conv2D: 3x3, ReLU, 32 filters
3. (2x2) MaxPooling2D
4. Conv2D: ReLU, 3x3, 64 filters
5. (2x2) MaxPooling2D
6. Conv2D: ReLU, 3x3, and 128 filters
7. (2x2) MaxPooling2D
8. Make it flat
9. Dropout rate: 0.5
10. Dense: ReLU, 128 units
11. Output: 1 unit sigmoid

Step 4: Assemble the Model

Adam is the optimizer.

Binary The loss is called cross entropy.

Metrics: Accurate

Step 5: Model Training Train the model over ten to twenty epochs using training data.

Use the validation data to confirm.

Step 6: Examine and stock

Evaluate the model's accuracy and save it for further use.

Step 7: Project a Novel Image

Reduce and normalize a new image as part of the organizing step.

Make predictions using the trained model.

- Output is fake if it exceeds 0.5;
  else, it is real.

4.2 Text Description

The proposed approach for based on text identify of product labels, packaging or online listings, by simply uploading/ scanning images of a product's label or description and having the text extracted using optical character recognition (OCR). Following extraction of text, the system displays key characteristics like product name, brand information, ingredients and other specific variables, which can be used for product identification by using high-level Natural Language Processing (NLP) techniques.

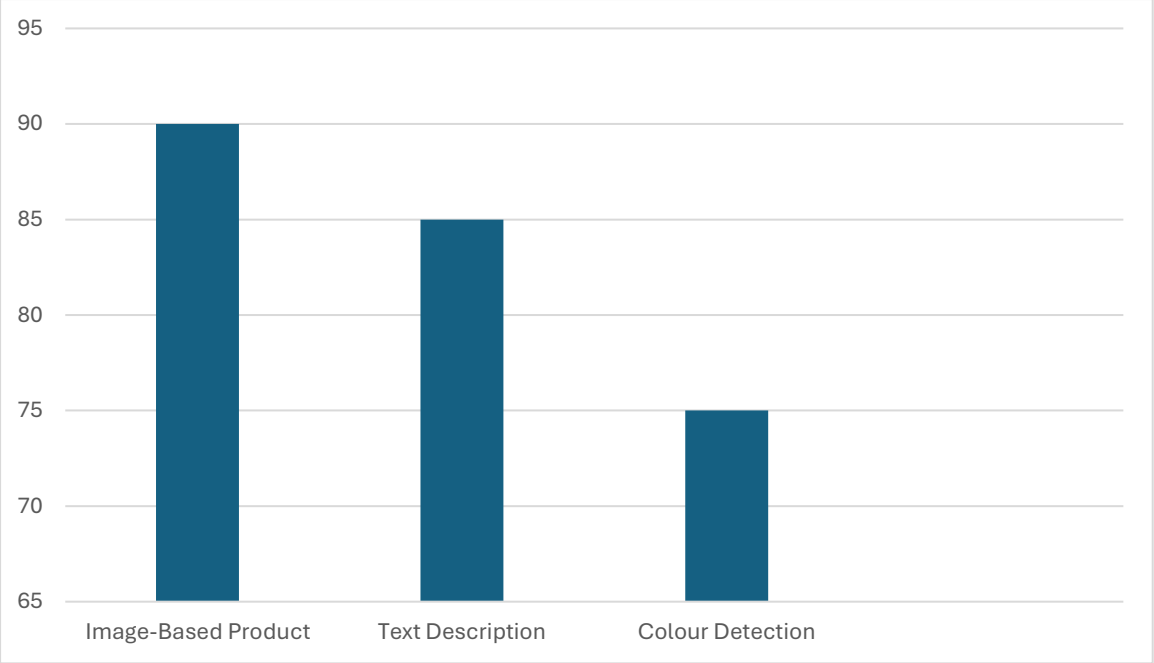
With this data, the system compares the language extracted to a database of real, verified product descriptions generated from producers or verified other sources. It checks for any errors that commonly indicate the occurrence of fake goods, spelling of brand name, inaccuracy in product information, missing or changed ingredients or invalid product standards. In addition to these tests, an algorithm also checks the language of the description and its phrasing to find out if the text is written in a strange way or for any unusual patterns that indicate the product is fake. If the system finds any significant differences in the extracted text from the actual source records, it flags the product as being fake. Based on the defects it detects; the system provides an authenticity score that users can use to determine how likely it is that the product in question is authentic. This tool makes it very easy, efficient and scalable to detect fake goods in online shopping where product descriptions are most likely the first thing a buyer will encounter with the price. By using text analysis on an incredibly diverse range of objects such as words, phrases, paragraphs and illustration, it helps to ensure that only real products come to market, thus protecting both consumers and businesses from the harm caused by fake products.

4.3 Colour Detection

This system is capable to assist in identifying fake products the checking the colour on a label, the packing, and the product itself. It uses sophisticated image processing technology to check the primary colour of vital areas such as logos and branding, logo and branding colours are meticulously checked as the user are provided with an image photo which serves as the base colour image of trusted sources.

Well-known brands focus on uniform so many brands mark sticks, and difference in colour; value of light, level differences, or strength differences are all red flags. Pay can balances every business requirement to make sure these colours are used appropriately and aligned with standard operations branding colours.

If the colours are not what is deemed original, the system marks low authenticity score. In addition, it would flag it as possibly fake to the user. It is a novel non-invasive approach to easily flagged counterfeit product, especially useful for e-commerce.



5. APPLICATIONS

An available way of solving the problem and to keep an end to it, stopping the customers from buying the fake or parallel good products is a fake typical program. It gives customers and businesses persons in asserting that this is a good product and not selling the twin. This product looks more trustability using different intelligent technologies. The one of the most used(often) method used to scan the barcodes or QR codes on the codes and examining to find whether it is matching to original one or to manufacture database, this app match up the product vapidness. There the program users use there is another feature called Visual Testing, there the program users use their phone cameras, layout, draft colour, typesetting or point size which one has been stored the system. In this period, it is determining the small differences also.

A few of these applications are Blockchain technology, which stores permanent and secure product records and supplies a digital certificate that tests the product is availability, other features making use the phone's camera or special lights like UV to search for holograms or hidden security markings on the package. While these security signals can be hard to replicate, they greatly contribute in the identification of criminal activity. The software also retains exhaustive data on actual items, including how they should seem and how typical fake look, making compared less difficult.

6. CONCLUSION

This study presents a fresh and practical way to apparatus the growing issue of fake products by searching Artificial intelligence, image recognition, and text analysis. The aim is to create a vacuum, user-friendly app that lets people to take a photo of a product logo and quickly find out if it's true or not. While the theory is solid, there's still room to grow. To make the system even fashionable and more reliable, we need to gather more distinct data and improve the way that data is designated and to organized it. We also program to boost the speed and precision of our algorithms and develop our algorithms and develop our ability to pull useful data from online platforms, especially commerciality's.

In the lengthy run, this kind of technology could really make a diversity. By giving everyday users, a feather in their caps (powerful tool in their hands), we can help to create a secure, more than a transparent market. It's not just about spotting dummy goods it's about empowering people to make informed against and building trust in the products they buy.

REFERENCES

[1] Varghese, Richu Jacob, Shilpa Shaji Nellikkakunnel, and Sruthy Sunilkumar. "An AI-Based Fake Products Identification System." (2022).

[2] Rajalakshmi, B., Anushka Dharan, Jainam Jain, and Shivswami Chavan. "Fake Product Identification System." In *2024 5th International Conference for Emerging Technology (INCET)*, pp. 1-6. IEEE, 2024.

[3] Shreekumar, T., Puneet Mittal, Sukhwinder Sharma, Rajesh N. Kamath, Sreeja Rajesh, and B. Nruthya Ganapathy. "Fake product detection using blockchain technology." *Journal of Algebraic Statistics* 13, no. 3 (2022): 2815-2821.