Real-Time Comparative Evaluation of ML and DL Models for Object Classification on a Unified Dataset

Shreyas Pankaj Hadawale[1], Sujal Dattatray Sargar[2], Aryan Shankar Salunkhe[3]
Prof. Pradeep Shirke[4], Prof. Avantika Jadhav[5]
Students, Department of Computer Engineering[1-3]
Vidyalankar Polytechnic, Mumbai, India
shreyas.hadawale@vpt.edu.in, sujal.sargar@vpt.edu.in, aryan.salunkhe@vpt.edu.in
pradeep.shirke@vpt.edu.in, avantika.jadhav@vpt.edu.in

Abstract: *Our project presents a comparative study of multiple machine learning and deep learning models for object detection and classification using a common dataset of classes with same number of attributes defined to each model. The models include YOLO, (SVM) Support Vector Machine, Decision Tree, and Random Forest , KNN (K-Nearest Neighbours) , CNN (Convolutional Neural Networks) , ResNet18 , MobieNet. And the reason behind choosing these kind of models is the ability of the algorithm to handle complex type of images and learning those images in particular way . Hence through this comparision it helps in understanding the strengthsand limitations of each model approach .Images we have added to the dataset are belonging to categories such as animals like cats , dogs , watches , birds , human faces and car's for the training and testing.We developed a Web-based interface or frontend developed using Flask allows the users to upload images or capture them through a webcam for real time testing.Also user are allowed to drag and drop the images from their device individually . Hence users are enable to compare and difference betweent the models easily and efficiently . Each model generates predictions and corresponding accuracy metrics,which ensures a fair and real comparison of performance under several identical conditions. The results or output we get in last let us know the comparative and efficient analysis between deep learning models as well as machine learning models.*

Keywords: machine learning

I. INTRODUCTION

Object detection and image classification are the core concepts and techniques which we used in this project to recognize and identify the image based on the trained dataset of images to each model.The concept deep learning and machine learning helps the computers to learn different elements from the images based on the attributes like edges ,patterns, background colour, images contrast,etc. as well as to classify objects automatically without the much involvement of the human .These techniques and methods are mostly applied in real-world applications such as security systems ,automated systems , medical devices ,Defence Power System , etc

Traditional machine learning algorithms like SVM , KNN , Random Forest , Decision Tree are straightforward to work with the well structured data and are easy to understand and interpret from them .While the Deep learning models such as YOLO , CNN and ResNet18 are actually more powerful and comparatively smarter than the machine learning algorithms for studying complex images as they have the ability to learn advanced and detailed features form the data and have higher efficiency and accuracy than the ML algorithms

Our project aims to compare different-different object detection and classification models based on the same dataset of images for training. And for that we have created a Web-based interface or frontend for the user to test the models by giving input through webcam and also through drag and drop functionality that is by uploading images. After users

input the system focuses on comparing accuracy and performance to identify which model is best suited for predicting and detecting the object among all the models

II. EASE OF USE

A. System Accessibility and User Interaction

Our system is designed and developedin such a way that it should be easily accessible and understandable for users in fact to anyone.The Web Interface made using Flask allows the user to either use the upload form device (drag and drop) functionality or by using a webcam for real time detection. Hence the user does not need to be worried and to be know how the models actuals work and detects the objects,Therefore, after submitting the images thew system automatically shows the prediction for all the models in a clear and systematic way

B. Maintaining System Reliability and Consistency

We have trained and tested all the models using the same dataset and preprocessing steps .Before training the Images are resized and normalized to maintain consistency among all the images in the dataset hence it leads to receiving input in the same format for every models. Fixed model parameters are used during evaluation to ensurereliable and repeatable results. We didn't retrained or adjusted the models parameters while testing the models . This ensures the results are reliable and consistent in the system and is fully functional.

III. SYSTEM PREPARATION AND DATA ORGANIZATION

Proper system preparation and dataset organization are must and important for accurate and précisedmodel functioning. Image datasets, model scripts, training and testing functions, web interface files, are maintained and organized separately in the mainproject structure

A. Abbreviations and Terminology

TheNeural-Network is a futuristic and advanced technique of the deep learning model that automatically extracts important features from the images based on the parameters like edges, background colour,patterns,etc. Hence it requires less human interaction.For ex, YOLOwhich is a real time object detection algorithm which is used in real world applications like Defence System ,Automation Systems,Medical Device, etc.

B. Dataset and Measurement Standards

Before training the images from the dataset they are processed and are maintained by same colour contrast ,resizedimages , etc or we can say they are normalization .Hence by following uniform measurements the comparisons between the models remains fair and clear and the system becomes stable and precise .

C. Data Processing Workflow

We pre-processed the images before they sent to model for training , it contains activities like resizing so that all the images are setted to a similar size and normalization so that training becomes more faster and efficient hence improving the efficiency of the models performance.

D. Challenges and Solutions

We face several issues while selecting the dataset , as we get to see too much difference in each images like difference in background , focus on main object , lighting conditions , etc and also the imbalance in the dataset , as few classes have higher number of dataset than other , which might cause imbalance in the models accuracy later.

IV. SYSTEM ARCHITECTURE

We divided the system architecture into separate modules such as the user interface , model evaluation , image pre-processing and result defining . We designed and developed the model or components architecture in such a way like whenever in further if there's a need to add the new models or dataset then it can be easily added and the system can

easily gets updates without affecting the system performance and degrading the system .Our frontend communicates with a flask-based backend , which will be mainly used for receiving the images form the user .After this the data is pre-processed and normalize by the models and defined accuracy and its final prediction for the image .This activity is done by all the models and in final these results are sent to the user's interface so that the user can differentiate between the models based on the accuracy and prediction done by the models .In this way the system architecture is designed very carefully and functionally responding to user's actions .

V. MODELS USED

A. YOLO
Yolo is the abbreviation for (You Look Only Once) it's a advanced or you can say a deep learningbased object detection model which is designed and developed for the real-time applications .Yolo easily detects objects and predicts their bounding boxes .We used Yolo model in our project because it simply detects and classifies the objects in a single step hence maintaining the balance and efficiency in the predictions .

B. Convolutional Neural Networks (CNN)
The Convolutional Neural Network (CNN) is a deep learning model which isespeciallydesigned and developed to work on the images type of data. The working of the Convolutional neural network is pretty simple as it applies filters for the images so that it learns important visual features like edges, background, patterns, color, etc. from the image automatically.

C. Support Vector Machine (SVM)
The Support Vector Machine is a type of classical machine- learning algorithm model which is commonly used forthe classificationtasks .The main idea of the SVM is to create a plane or a decision line that differs the distance between data points and the different classes .

D. ResNet18
We also used ResNet-18which is a deep convolutional neural- network which isbuilt and developed to address the problem of gradient vanishing .It allows us to use it important feature like residual connections which helps us to skip the certain layers to learn form the images. This feature helps in learning deep and effectively and hence making the model to improves its efficiency .

E. Random Forest
Random Forest can be defined as using multiple or we can say that combining multiple Decision Trees to improves its accuracy , efficiency and prediction by degrading overlifting .In this system, we used the Random Forest model to evaluate how ensemble methods perform on the same dataset.

F. Decision Tree
We used the decision tree model which is a type of the machine-learning algorithm that can make easy decisions by splitting the data into the smaller number of groups depending upon the important values .In this project we used Decision Tree as a basic traditional model to analyze classification performance.

G. K-Nearest Neighbours (KNN)
MobileNet-18 is a lightweight deep learning model which is mainly known for its fantastic efficiency and low resource cost .In our project , the reason behind the use of the MobileNetto compare how the other models performs to such heavier architectures . Its Lightweight component make us to use it. K-Nearest Neighbors (KNN) is a simple machine learning algorithm model that can assigns a class to a data by just simply based on the majority class among its nearest neighbors .Since KNN does not need a complex training phase hence quick evaluation can be done .

H. MobileNet

MobileNet-18 is a lightweight deep learning model which is mainly known for its fantastic efficiency and low resource cost. In our project , the reason behind the use of the MobileNetto compare how the other models performs to such heavier architectures . Its Lightweight component make us to use it.

V. RESULTS AND DISCUSSION

All models were tested with the same dataset and the same set of testing images, whether taken by direct image upload or with a live webcam. This ensured the same fairness in contrasting the models with one another. For the performance of the models, parameters like accuracy, predictive confidence, and the time it took to obtain the output were compared. While testing the models, it was observed that models like YOLO and neural networks performed better with complex images, especially in the presence of multiple objects or with varied background images. However, models like Random Forest and SVM performed equally well with simpler object classes and ideally distinguishable features. Also, the Decision Tree model was easier to understand and produced quicker output, but it failed to perform with complex image scenarios. All these phenomena specifically define the trade-off among accuracy, understanding, and efficient processing of different models.

VI. CONCLUSION

The proposed project presents a comprehensive comparison of object detection and classification models using a common dataset and testing setup. Supplying the same input images to YOLO, neural network models, Support Vector Machine, Decision Tree, and Random Forest made a non-biased and coherent analysis of their performances possible. This web-based platform allows real-time testing with image uploads or live webcam inputs, which makes the evaluation process practical and user-friendly. We can easily perceive from the results that deep learning models perform better when dealing with complex images and multiple objects, while classical machine learning models provide faster execution with better interpretability on simpler tasks. Overall, the proposed framework provides useful insights that can help in selecting an appropriate model based on the requirements of accuracy, efficiency, and application complexity.

ACKNOWLEDGMENT

REFERENCES

[1] OpenAI, *ChatGPT: Large Language Model for Technical Assistance*, https://chat.openai.com/, accessed Jan. 2026.

[[2] Wikipedia contributors, "Object detection," *Wikipedia, The Free Encyclopedia*, https://en.wikipedia.org/wiki/Object_detection, accessed Jan. 2026.

[3] Wikipedia contributors, "Convolutional neural network," *Wikipedia, The Free Encyclopedia*, https://en.wikipedia.org/wiki/Convolutionsal_neural_network, accessed Jan. 2026.