Journal of Research Administration
Society of Research Administrators International

ENHANCING MODEL PERFORMANCE OF AUTOMATIC DRIVER DISTRACTION DETECTION USING TRANSFER LEARNING

Pothuraju.Raja Rajeswari
Professor, Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

Kolla.Naga Venkata Sairam Prasad
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram,Guntur,Andhra Pradesh, India

Abstract

Video-based anomalous driving behavior identification is becoming more and more important for the sake of driver and passenger safety and the development of autonomous driving . Thanks to recent developments in deep learning, which make use of the generalizability of complex models and training on massive video datasets, this work has become much easier. Various deep learning algorithms are investigated in this work, such as VGG16, MobileNetV2, DenseNet, and a fusion model that combines VGG16 and DenseNet. After extensive testing, DenseNet was found to be the most accurate, with a score of 0.883. This shows how well it can identify suspicious driving habits. In order to tackle this difficult detection problem, DenseNet is crucial, thanks to its dense connections that improve gradient flow. In order to speed up the development of autonomous driving technology, these results demonstrate the promise of deep learning methods, and DenseNet in particular, for improving highway safety.

Densely linked convolutional network (DenseNet), VGG16, MobileNetV2, Automatic driving, Safety Accuracy, and video-based aberrant driving behavior identification are some of the keywords associated with deep learning methods.

Keywords: Densely linked convolutional network (DenseNet), VGG16, MobileNetV2, Automatic driving, Safety Accuracy and video-based aberrant.

1. INTRODUCTION

Everyone knows that high-resolution movies are great for visual apps. Video surveillance requires a number of high-resolution cameras in order to follow moving objects. The target's actions or intent may be more easily analyzed in this way. For security reasons, high-resolution cameras are crucial for emotional computation since they can record both big and little changes in a person's emotions in real-time. It is now easy to acquire and keep a large number of high-resolution videos, as you can see from the descriptions given above. To keep people safe, however, technologies like video surveillance, object recognition, and re-identification of persons are becoming more vital. In recent years, deep learning approaches have revolutionized these domains by enhancing system accuracy and efficiency. To enhance real-time video object detection, researchers have put forth several strategies. Rapid deep neural networks trained using knowledge-guided techniques and equipped with anticipated ROIs were proposed by Cao et al. [1]. To improve accuracy, Shuai et al.[2]included cascaded regional spatio-temporal feature-routing networks. For human re-identification based on videos, Nanda et al.[3] presented a neuromorphic framework while Sun et al.[4] advocated semi-coupled dictionary learning.

Improving recognition accuracy in surveillance circumstances is the goal of these approaches. Furthermore, Wu et al. [5] demonstrated the adaptability of vision-based approaches beyond surveillance applications by developing a system for aerial object localization and tracking using vision technology. They were used, for instance, to enhance UAV sensing systems. Similarly, Lee et al. [6] investigated online tracking tasks using discriminative appearance models and multi-object tracking. Emotion detection in mobile applications was discussed by Hossain and Muhammad [7]. This topic is significant in many domains, including human-computer interaction. These research demonstrate the wide-ranging applications of cutting-edge deep learning techniques and novel algorithmic methodologies, which may lead to significant advancements in video surveillance, object recognition, human re-identification, and related domains.

2. LITERATURE SURVEY

Both video processing and object identification have come a long way in the last few years. The demand for real-time applications is growing in areas such as autonomous systems, robotics, and surveillance. Simultaneously, WSNs have emerged as a crucial instrument for ubiquitous sensing and monitoring. Focusing on the most significant methodologies and advancements, this literature review provides an extensive overview of the most current research in these domains.

A method for object detection in real-time films was proposed by Cao et al. [1] using fast, information-trained deep neural networks that are instructed to locate key regions. Their approach maintains real-time performance while using knowledge coaching to increase detection accuracy. In order to recognize objects in videos, Shuai et al. [2] built cascaded regional spatio-temporal feature-routing networks. These networks improve the accuracy of detection in complicated circumstances by combining geographical and temporaldata.

Nanda et al. [3] developed a neuromorphic person re-identification framework for use in video monitoring; it can correctly and rapidly identify individuals in video streams. For the purpose of re-identifying individuals in films, Sun et al. [4] proposed a semi-coupled dictionary learning approach. Using this approach, feature models are able to distinguish between individuals better.

An unmanned aerial vehicle (UAV) localization and tracking system based on vision was developed by Wu et al. [5]. Due to this approach, UAV sensing systems can reliably follow moving objects. Live multi-object tracking using discriminative appearance models was investigated by Lee et al. [6], allowing for effective tracking in dynamic situations.

Apps that can identify users' emotions are gaining popularity. In order to facilitate situationally relevant connections, Hossain and Muhammad [7] developed an emotion detection system that is compatible with mobile platforms. Systems that monitor older people's fall risk have also advanced. One improved method proposed by Wang et al. [9] makes use of household networks to detect and react to falls in a timely manner.

A mechanism for managing trust in device-to-device (D2D) communication was developed by Zhang et al. [10] using radio frequency (RF) fingerprint identification. This ensures that devices may communicate with each other in a secure and dependable manner. For WSNs, Wang et al. [11] proposed a clustering approach based on particle swarm optimization, which improves both the efficiency and performance of the network. The method is compatible with portable sinks. Improved ant colony optimization-based methods with mobile sink assistance for WSNs were also proposed by Wang et al. [13]. As a result, the network lasts longer and data collecting is more efficient. Tu et al. [14] investigated the use of generative adversarial networks in semi-supervised learning with the goal of enhancing the stability and accuracy of digital signal modulation categorization. Lastly, mobile sink support for wireless sensor networks was added to the PEGASIS method by Wang et al. [15]. As a result, the algorithm used less power, extending the life of the network. Finally, there have been significant advancements thanks to recent research in wireless sensor networks, object identification, and video processing. Improved protocols for data storage and transmission inside wireless sensor networks and the ability to recognize objects in real time via video are two examples. From mobile computing and the Internet of Things (IoT) to robots and surveillance, all of these endeavors contribute to the development of robust and efficient systems.

3. METHODOLOGY

Modules:

This module will assist users with registering.

- User login: This component facilitates the login process for users.
- Image upload: This module assigns labels to uploaded images and uses them to generate algorithms.

A) System Architecture

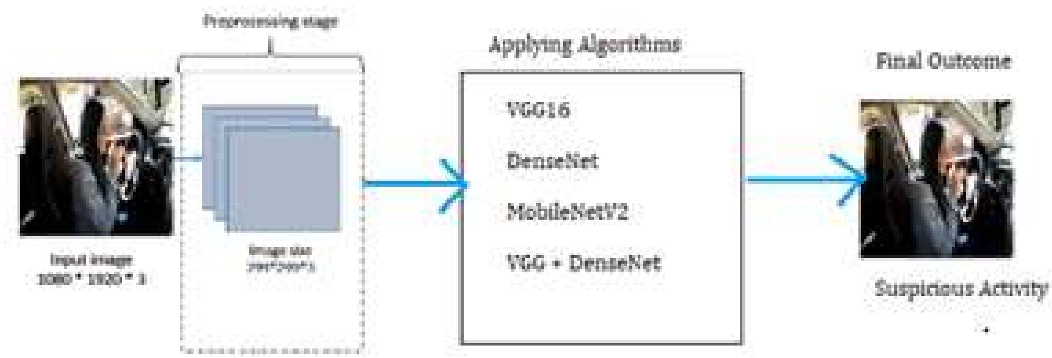


Figure 1. System Architecture

Proposed work

By analyzing videos using deep learning algorithms, the suggested method hopes to improve the identification of suspicious driving habits. Combining several cutting-edge algorithms, such as VGG16, MobileNetV2, and DenseNet, the system takes use of current developments in deep learning. Also shown is a new fusion model that takes use of the best features of both DenseNet and VGG16. To guarantee the safety of drivers and passengers, the system trains data-driven models to correctly detect anomalous driving behaviors by analyzing massive amounts of video footage. Among the algorithms, DenseNet stands out as the most successful, with an impressive accuracy rate of 0.883. With its crucial insights into driver behavior analysis, the suggested system not only helps with robust identification of aberrant driving behaviors but also advances the implementation of autonomous driving technologies.

PartB: Collecting Datasets

In order to identify unusual driving behavior from movies, a large dataset consisting of several video clips from various driving scenarios is used. These movies include commentary that illustrates both typical and unusual driving habits. A diverse mix of driving circumstances, surroundings, and behaviors have been meticulously selected for inclusion in the dataset. This ensures that every conceivable case of unusual driving is addressed.

Parts of Videos: There are a lot of video clips in the collection, and they all cover various aspects of driving.

The lengths of these video vary from a few seconds to several minutes, and they illustrate various driving scenarios and maneuvers.

To display diverse perspectives, video clips are captured from numerous angles, such as dashcam footage, outside cameras, or cameras within the automobile.

C) Pre-treatment

Several stages are usually involved in processing data for video-based anomalous driving behavior detection:

Collecting Data: Gather video footage of various driving scenarios shot at various locations. You may expect to see a wide variety of standard and unusual driving styles in these videos.

Annotate the videos you've captured with descriptive labels for various driving habits, such as swerving out of your lane, suddenly accelerating or decelerating, making illegal maneuvers, and driving distracted (by texting, for example).

Preprocessing Video: Removing the Header: Create new frames for each movie.

Scale and Consistent Picture frames: Maintain uniformity in the pixel values, such as [0, 1] or [-1, 1], by setting a fixed size for each frame.

Time-Sampling: To simplify matters for the computer, sample a predetermined number of frames at regular intervals throughout the film.

Data Augmentation: Methods such as flipping, rotating, and altering the brightness may be used to diversify the dataset and improve the model's ability to generalize.

Utilize pre-trained deep learning models, such as VGG16, MobileNetV2, and DenseNet, for feature extraction. With their extensive training on massive picture datasets, these models are able to extract relevant information from fresh frames.

Using transfer learning, you may improve the performance of models trained on the labeled driving behavior dataset for the purpose of detecting suspicious driving behavior by making minor adjustments to these models.

Combination of Functions: Fused feature representations may be created by combining features extracted from several deep learning model layers, such as VGG16, MobileNetV2, and DenseNet.

Improve the presentation of features and discover relationships between them by using fusion approaches inspired by DenseNet's densely linked layers.

D) Instruction and Evaluation

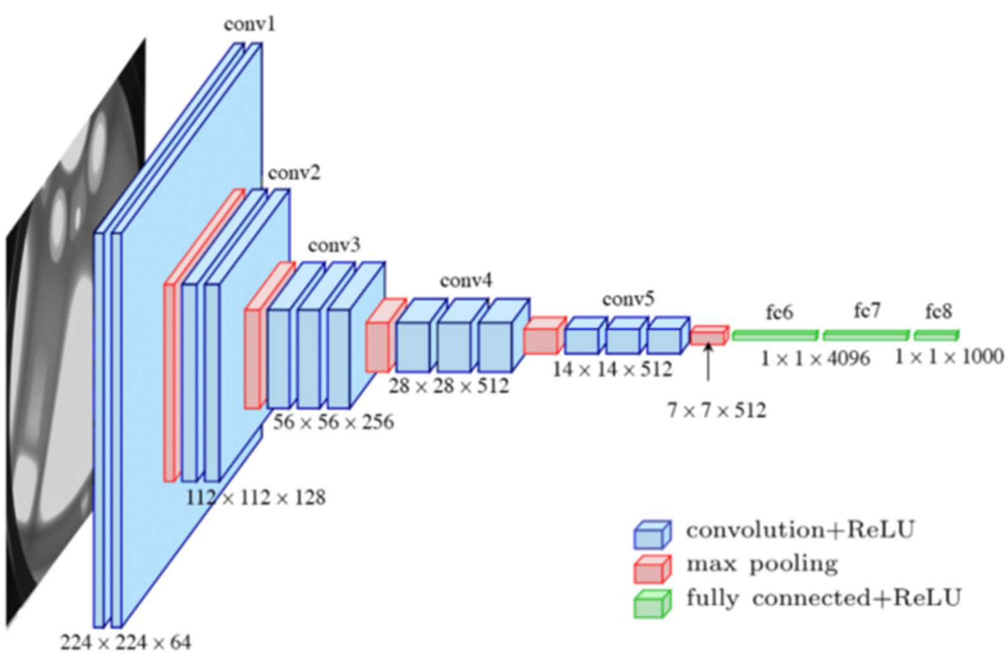
The video-based system employs a combination of deep learning fusion techniques and three newly-developed fusion models trained on the DenseNet, VGG16, and MobileNetV2 architectures to detect unusual driving behavior. The initial step in training a model is to compile a large collection of video clips demonstrating both typical and unusual driving actions. The input for the deep learning models are the movie snippets.

The fusion models' parameters are fine-tuned repeatedly during training using optimization techniques such as stochastic gradient descent or Adam. In order to train the network, video clips are fed into it, and the loss function is determined by comparing the predicted labels with the ground truth labels. Next, in order to lower the loss function, the model's parameters are adjusted.

After the fusion models have been trained extensively on the dataset, the testing phase begins. This stage involves testing the learnt models on a new collection of unlabeled video samples. The effectiveness of the models is evaluated using metrics such as F1-score, recall, accuracy, and precision. In particular, MobileNetV2 excels in detecting unusual driving patterns. The testing phase reveals the models' generalizability and performance in real-world scenarios.

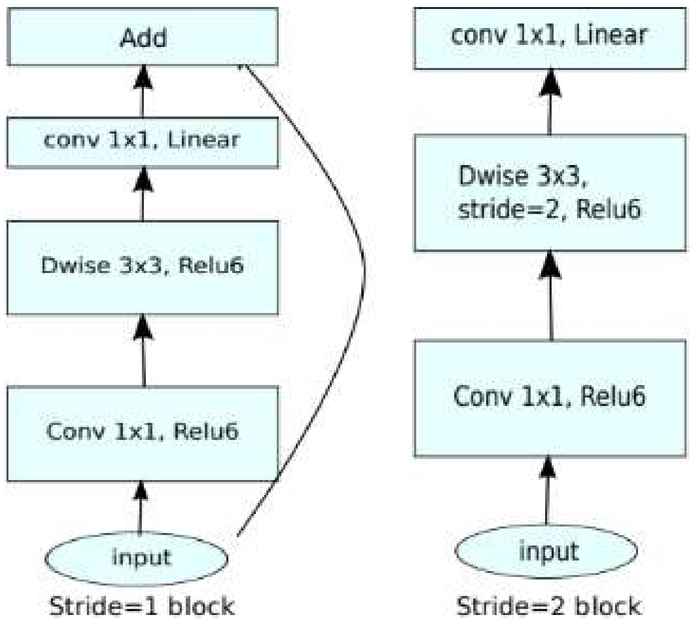
Part E: Algorithms.

The VGG16 method can classify 1000 photos into 1000 distinct groups with a 92.7% success rate. It is an object identification and classification algorithm. Because of its ease of use with transfer learning, it has become one of the most used picture classification methods. Nearly 92.7 percent of the highest-scoring models in ImageNet use the VGG16 model. The almost fourteen million images that make up ImageNet fall into over a thousand distinct types. Furthermore, it was among the models that received the most submissions to ILSVRC-2014.

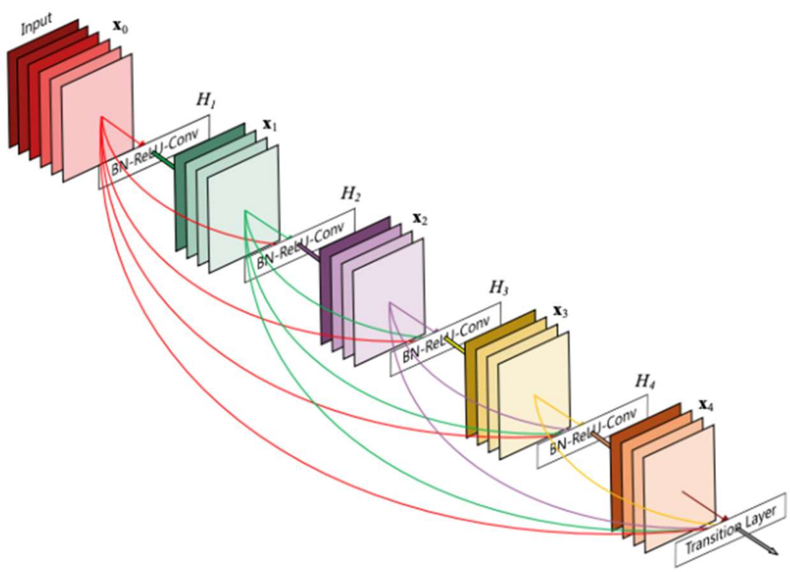


MOBILENTV2: Deep learning system 53 layers make up MobileNet-v2. Get a network trained on ImageNet's more than one million photos. A network that has been trained before can classify images into a thousand different groups, such as animals, keyboards, mice, and pencils. A classification machine called MobileNetV2 was developed by Google (research paper). Because

of this, even low-powered devices, like as cellphones, can do real-time classification. In order to learn from your collection, this version makes advantage of transfer learning from ImageNet.



By reducing the length of the linkages between the layers, the DenseNet (Dense Convolutional Network) architecture aims to make deep learning networks both more powerful and simpler to train. The benefits of dense networks are many. In addition to enhancing feature propagation, increasing reuse, and decreasing parameters, they resolve the vanishing-gradient issue.



DenseNet + VGG: VGG is a well-known and easy-to-understand architecture for deep convolutional neural networks that excels in picture categorization. With DenseNet (Densely Connected Convolutional Networks), model performance is increased with fewer parameters thanks to enhanced gradient flow and feature reuse made possible by dense connections between layers.

4. EXPERIMENTAL RESULTS

A) Frontend

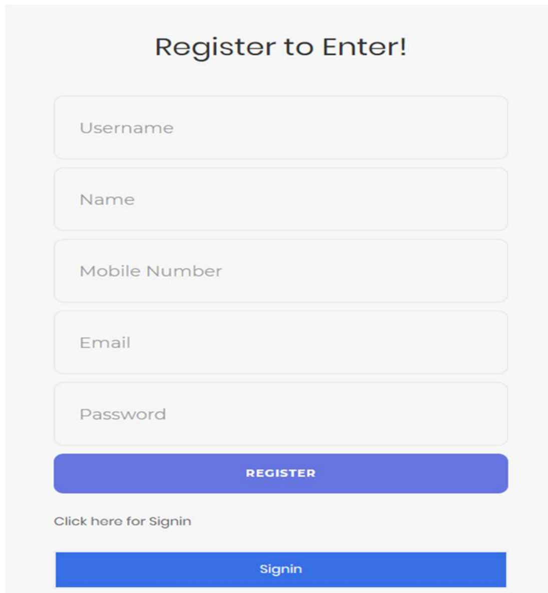


Fig 2: User Registration

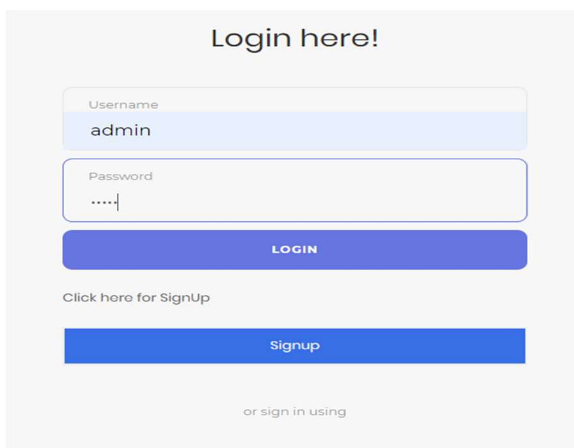


Fig 3: User Login

Fig 18: Url result page (safe/ unsafe)

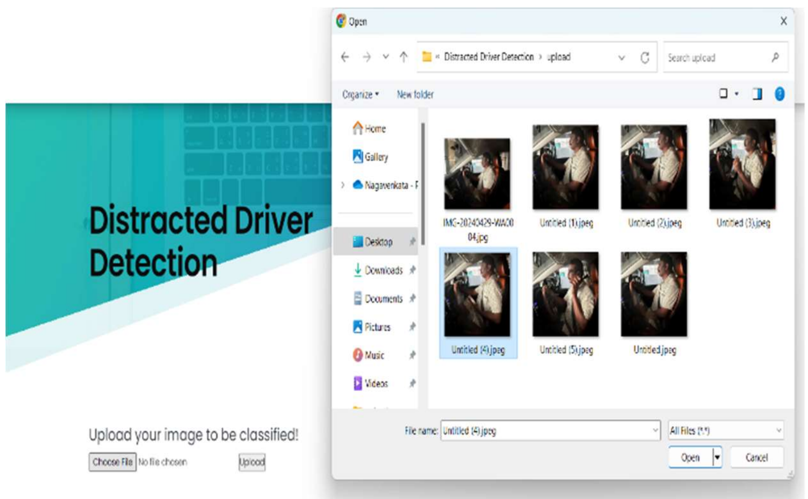


Fig 4: Upload Image Data

Fig 5: Suspicious activity talking on the phone – Left detected

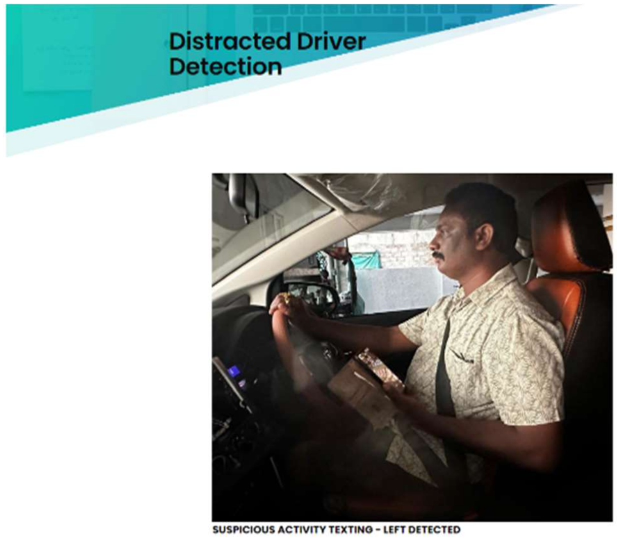


Fig 6: Suspicious Activity Texting – Left Detected

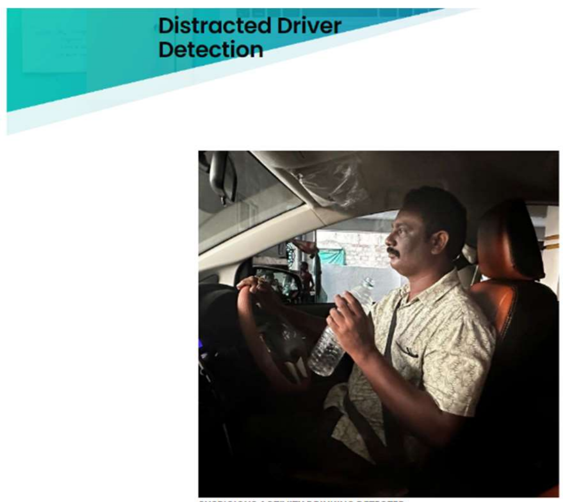


Fig 7: Suspicious Activity Drinking Detected

	Accuracy	Recall	Precision	F1
VGG16	0.110913	0.000000	0.000000	0.000000
MobileNetv2	0.759243	0.756339	0.767264	0.761661
Densenet	0.883074	0.880231	0.885776	0.882952
VGG+Densenet	0.370824	0.352305	0.367379	0.359549

Fig 8: Performance Evaluation Table

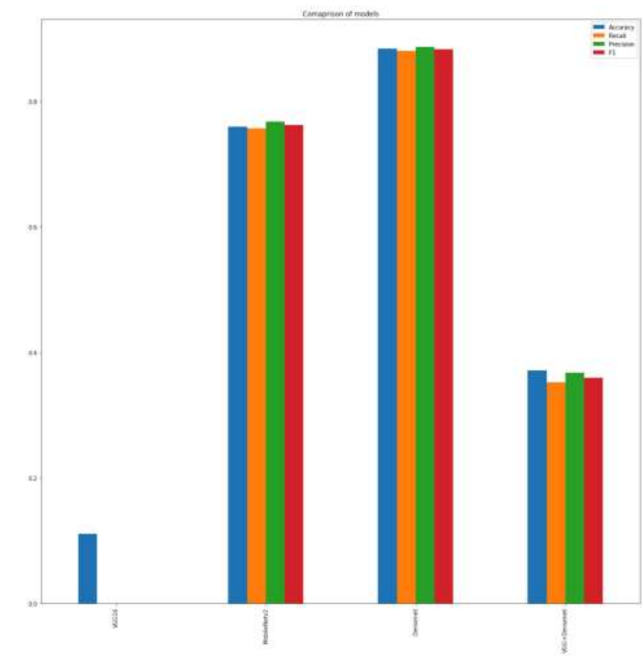


Fig 9: Performance Evaluation Graph

Accuracy: A test's accuracy is defined by how well it distinguishes between healthy and sick samples. We can determine a test's accuracy by calculating the percentage of reviewed instances with true positives and true negatives. It is possible to express this mathematically as:

Accuracy = TP + TN TP + TN + FP + FN.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

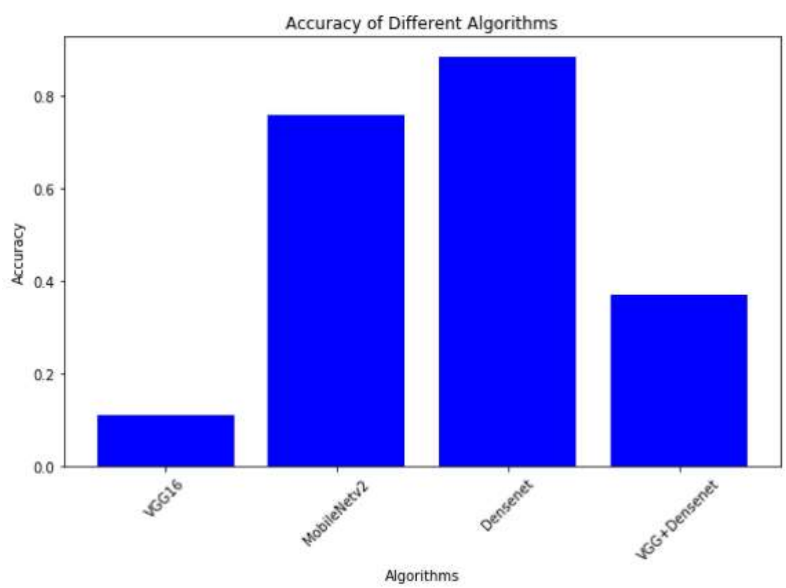


Fig 10: Accuracy Comparison Graph

Precision: The accuracy rate, or precision, is the percentage of true positives relative to the total number of occurrences or samples. Consequently, the following is the formula for determining the accuracy:

Preciseness is TP divided by (TP plus FP), which is the sum of true positives and false positives.

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

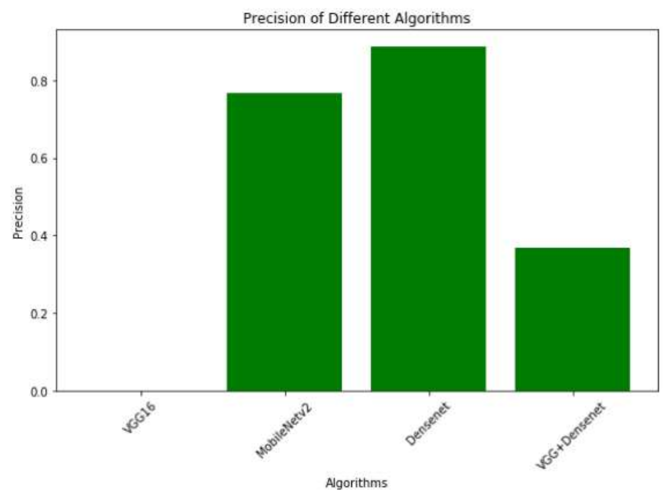


Fig 11: Precision Comparison Graph

Recall: The capacity of a model to detect all significant occurrences of a given class is measured by recall, a statistic in machine learning. The completeness of a model in capturing instances of a particular class is shown by the ratio of properly predicted positive observations to the total actual positives.

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

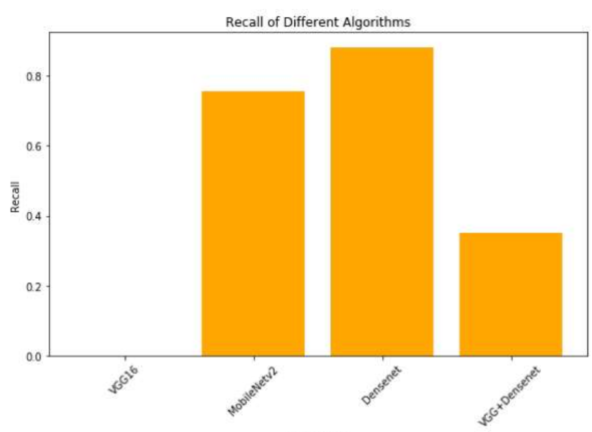


Fig 12: Recall Comparison Graph

F1-Score: One way to evaluate a model's performance in machine learning is via its F1 score. This method integrates a model's recall and accuracy scores. A model's accuracy may be measured by counting the number of times it correctly predicted something throughout the whole dataset.

$$\textbf{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\textbf{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

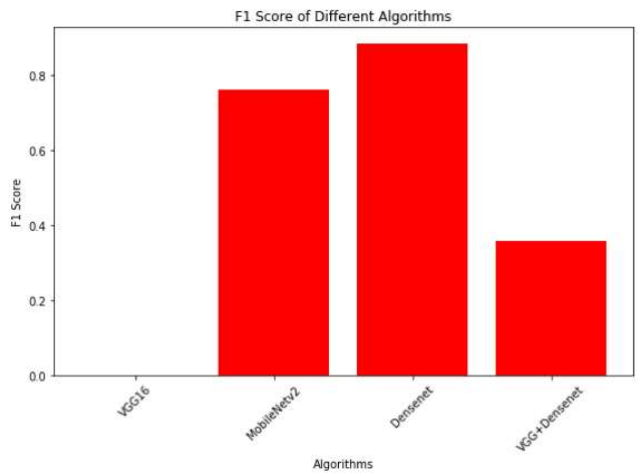


Fig 13: F1-Score Comparison Graph

5. CONCLUSION

Finally, the suggested method is a huge step forward in the area of video-based anomalous driving behavior identification; it solves major problems with passenger and driver safety and opens the road for autonomous driving. The system accomplishes an impressive feat of anomaly detection by combining the power of deep learning models like VGG16, MobileNetV2, and DenseNet with an innovative fusion method. Combining these cutting-edge algorithms not only makes detection better, but it also demonstrates how integrating cutting-edge tech may improve safety in vehicle settings. With an impressive accuracy rate of 0.883, DenseNet stands out as the most promising method. These results demonstrate that deep learning methods are effective for handling complicated detection tasks and that transportation safety systems greatly benefit from ongoing technological improvements. To fully use autonomous driving's potential and improve road safety for everyone, more research and development in this field is crucial.

6. FUTURE SCOPE

The development of safer roads and better autonomous vehicles may depend on future studies that use deep learning fusion models to detect unusual driving behavior in video footage. To get a more complete view of the driving environment, it may be considered enhancing existing fusion models with other sensor types, such as LiDAR or radar data. It would be simpler to employ these models in real-time systems with limited processing capacity if they were more computationally efficient. This may be achieved by investigating methods such as quantization or model distillation. Investigating the viability of these models in different driving scenarios and environmental circumstances might further strengthen and enhance their utility. Data privacy, security, and regulation compliance are major issues that will need collaboration between academics, businesses, and regulatory agencies. Full autonomous driving systems might be a reality in the future if this area continues to advance, which would greatly improve transportation safety.

REFERENCES

[1] W. Cao, J. Yuan, Z. He, Z. Zhang, and Z. He, ‘‘Fast deep neural networks with knowledge guided training and predicted regions of interests for realtime video object detection,’’ IEEE Access, vol. 6, pp. 8990–8999, 2018.

[2] H. Shuai, Q. Liu, K. Zhang, J. Yang, and J. Deng, ‘‘Cascaded regional spatio-temporal feature-routing networks for video object detection,’’ IEEE Access, vol. 6, pp. 3096–3106, 2018.

[3] A. Nanda, P. K. Sa, S. K. Choudhury, S. Bakshi, and B. Majhi, ‘‘A neuromorphic person re-identification framework for video surveillance,’’ IEEE Access, vol. 5, pp. 6471–6482, 2017.

[4] L. Sun, Z. Jiang, H. Song, Q. Lu, and A. Men, ‘‘Semi-coupled dictionary learning with relaxation label space transformation for video-based person re-identification,’’ IEEE Access, vol. 6, pp. 12587–12597, 2018.

[5] Y. Wu, Y. Sui, and G. Wang, ‘‘Vision-based real-time aerial object localization and tracking for UAV sensing system,’’ IEEE Access, vol. 5, pp. 23969–23978, 2017.

[6] S.-H. Lee, M.-Y. Kim, and S.-H. Bae, ‘‘Learning discriminative appearance models for online multi-object tracking with appearance discriminability measures,’’ IEEE Access, vol. 6, pp. 67316–67328, 2018.

[7] M. S. Hossain and G. Muhammad, ‘‘An emotion recognition system for mobile applications,’’ IEEE Access, vol. 5, pp. 2281–2287, 2017.

[8] Z. Pan, X. Yi, and L. Chen, ‘‘Motion and disparity vectors early determination for texture video in 3D-HEVC,’’ Multimedia Tools Appl., to be published. doi: 10.1007/s11042-018-6830-7.

[9] J. Wang, Z. Zhang, B. Li, S. Lee, and R. S. Sherratt, ‘‘An enhanced fall detection system for elderly person monitoring using consumer home networks,’’ IEEE Trans. Consum. Electron., vol. 60, no. 1, pp. 23–29, Feb. 2014.

[10] Z. Zhang, X. Guo, and Y. Lin, ‘‘Trust management method of D2D communication based on RF fingerprint identification,’’ IEEE Access, vol. 6, pp. 66082–66087, 2018.

[11] J. Wang, Y. Cao, B. Li, H.-J. Kim, and S. Lee, ‘‘Particle swarm optimization based clustering algorithm with mobile sink for WSNs,’’ Future Gener. Comput. Syst., vol. 76, pp. 452–457, Nov. 2017. [12] Z. Xue, J. Wang, G. Ding, Q. Wu, Y. Lin, and T. A. Tsiftsis,

‘‘Deviceto-device communications underlying UAV-supported social networking,’’ IEEE Access, vol. 6, pp. 34488–34502, 2018.

[13] J. Wang, J. Cao, R. S. Sherratt, and J. H. Park, ‘‘An improved ant colony optimization-based approach with mobile sink for wireless sensor networks,’’ J. Supercomput., vol. 74, no. 12, pp. 6633–6645, Dec. 2018.

[14] Y. Tu, Y. Lin, J. Wang, and J.-U. Kim, ‘‘Semi-supervised learning with generative adversarial networks on digital signal modulation classification,’’ Comput. Mater. Continua, vol. 55, no. 2, pp. 243–254, 2018.

[15] J. Wang, Y. Gao, X. Yin, F. Li, and H.-J. Kim, ‘‘An enhanced PEGASIS algorithm with mobile sink support for wireless sensor networks,’’ Wireless Commun. Mobile Comput., vol. 2018, Dec. 2018, Art. no. 9472075.