Human Abnormal Behavior detection using Deep Learning

Likith S R
School of CSE&IS (BCA)
Presidency University
Bangalore, India
anandlikith111@gmail.com

Nabila Nausheen
School of CSE&IS (BCA)
Presidency University
Bangalore, India
nabilazehra27@gmail.com

Rajesh Prasad
School of CSE&IS (BCA)
Presidency University
Bangalore, India
rajesh.p3807@gmail.com

Shashank Kumar
School of CSE&IS (BCA)
Presidency University
Bangalore, India
sahu.sk30@gmail.com

Rohan Sharma
School of CSE&IS (BCA)
Presidency University
Bangalore, India
rohan.sharma6004@gmail.com

***Abstract*— The rapid growth of video surveillance systems has amplified the need for automated techniques to detect abnormal human behaviors in real-time for public safety applications. The objective of this study is to develop an accurate and efficient deep learning framework to identify anomalous human activities from video data. We investigate state-of-the-art deep convolutional neural network architectures, capable of learning discriminative features directly from raw video frames without relying on hand-crafted features. To overcome the shortage of training samples, transfer learning from large-scale action recognition datasets is utilized. The deep models are extensively evaluated on one of the benchmark datasets, UCF-Crime, totaling 79,959 video frames spanning 7 abnormal behavior classes. Performance is assessed using standard metrics including frame-level accuracy, area under the ROC curve, and loss values. Among the models investigated, DenseNet201 achieves superior results, yielding 91.89% accuracy and 0.0034 loss, with the ROC-AUC curve and confusion matrix analysis further validating its effectiveness for robust abnormal behavior detection in real-world video surveillance scenarios.**

Keywords*— *Human Abnormal behavior detection, CCTV, DenseNet201, ResNet50, MobileNetV2, Anomaly Detection, Behavior Recognition

I. INTRODUCTION

A. Background

In the constantly changing field of public safety and law enforcement, using state-of-the-art technologies has become essential. Deep learning, a kind of artificial intelligence that imitates the human brain's functions to comprehend intricate data patterns, is one such technological frontier. Detecting illegal activity and abnormal human behaviour is a crucial application in this field. Because of the sophisticated methods that criminals are employing to elude traditional law enforcement measures, such technology is necessary. Deep learning algorithms enable rapid and precise analysis of vast data sets, including social media activity and security footage.

Neural networks—hierarchical structures modelled after the human brain—are the fundamental building blocks of deep learning, allowing it to derive meaningful insights from unprocessed data. Large datasets are used to train these networks so they can identify patterns that point to criminal activity or departures from social standards. By means of ongoing improvement and modification, these algorithms are capable of detecting minute indicators and deviations that would escape human scrutiny, thereby augmenting preventive and preventative efforts against criminal activity.

In addition, the incorporation of deep learning into current platforms for surveillance and predictive analytics enables law enforcement to foresee and stop illegal activity before it happens. Real-time detection of suspicious patterns or behaviours enables authorities to quickly take action and reduce possible risks to public safety. Deep learning's use to the detection of illegal activities, however, brings up ethical and privacy issues. It is crucial to find a balance between the demands of personal freedom and security. To ensure responsible and ethical use, strong protections and supervision systems must go hand in hand with the creation and implementation of deep learning solutions.

Deep learning projects use AI to detect illegal conduct and anomalous behaviour. Deep learning algorithms, modelled after the human brain, excel at figuring out complicated patterns from huge datasets. By training neural networks on data sources such as social media, surveillance footage, and crime records, the goal is to construct a system that detects minute indicators of potential criminal activities or societal aberrations. Developing preventive measures will equip law enforcement to prepare for and navigate hazards, thereby improving public safety. But putting such technology into practice raises ethical concerns about discrimination, false positives, and breaches of privacy. To ensure responsible and moral use while protecting civil liberties, effective governance, transparency, and accountability measures are required.

Section II Provides the literature review, wherein relevant published research papers have been summarized, informing the project's implementation. Section III Describes the overall high-level design of the model including the architectural and data flow diagrams. Section IV Focuses on describing the various models that have been used along with their architectures. Section V illustrates the findings of the models used, comparing and contrasting their respective metrics. Section VI Concludes the report and discusses future scope.

II. LITERATURE SURVEY

A. Related Work

In this paper [1], the authors present a framework in which they introduce a real time approach to identifying abnormal behaviour in social networking sites making use of user behaviours and their curated profiles with the use of Convolutional Neural Networks (CNN). The proposed framework takes the data through a pipeline of stages: profile verification, one-time password generation and verifying the user using their cookie information. A CNN model learns the

data at a non-linear level by making use semantics, syntax and analysing the data at a token level using Natural Language Processing (NLP). It takes in this data and returns a group of feature vectors. The users are pooled into these vectors that are either labelled as normal or abnormal based on the behaviour. The final output is the percentage of normal and abnormal users in the SW. For training, a variety of URLs have been sourced from the internet, like Facebook,

Twitter, and malicious URL sites like VirusTotal and PhishTank. The final metrics make use of the total number of True positives, True negatives, False positives, and False negatives to determine that their model has a staggering 99.9% accuracy rate, outperforming other models like Logistic Regression and Linear SVM.

This study [2] focuses on utilizing the video data from surveillance cameras to monitor the environment in crowded and uncrowded areas to label any behaviour as normal or abnormal. The study delves into some specific forms of abnormal behaviour starting with loitering where there is a set specific threshold of time that dictates whether a person is aimless loitering around. Other abnormal behaviours investigated were falling; using a depth sensor-based model to detect the change in the stature of a person, abnormal patient behaviour, violence, panic, snatching, and sexual abuse. The study observed that the CNN model performed better than other machine learning methodologies in recognizing specific features from the images. Since they used 3D CNN, it performed better than its 2D counterpart due to having extra motion detection capabilities for a 3D environment. The authors wish to further improve their approach by integrating the use of LSTM to incur real time identification and the ability to alert neighbouring people in case of any victim needing immediate help.

Khosro Rezaee et al [3] present a strategy for real-time monitoring for abnormal behaviour in crowded spaces with the assistance of transfer learning and drones surveilling the area. The model utilizes deep transfer learning and a modified ResNet architecture which process the frames that are captured by the drone technology and stored in the network. The authors make use of two modified ResNet-18 architectures where one of them consumes the video frames and the other monitors the background frames notifying the congestion of the crowd. They use the UMN and the UCSD datasets with 11 and 48 video samples respectively showcasing various crowded scenarios that might occur in real life. The model performed well under its testing achieving an accuracy above 90 percent. The authors plan on improving the pre-processing techniques such as overcoming overcrowding, pixel occlusion and object overlap to better the accuracy and processing times of the videos.

The study [4] focuses on the use of weakly supervised deep neural networks to detect unusual human movement from surveillance videos. The proposed method leverages deep neural networks (DNN) to automatically extract abnormal features from time-series data. By utilizing weakly supervised training, the DNN model is optimized through a devised loss function, enabling the detection and quantification of outliers in time-series data, particularly unnatural human motion. The study proposes a DNN training technique with an approach similar to multiple-instance learning for detecting unnatural human motion. Furthermore, the study details the network structure and training process, illustrating the use of a one-dimensional convolutional neural network (1DCNN) for analysing time-series data. The neural network consists of multiple phases of repeating 1DCNN, ReLU activation and batch normalization layers which outputs higher anomaly scores for data points containing abnormalities.

The paper [5] introduces a real-time image-based recognition system for human activities in indoor environments, aiming to assist disabled individuals, enhance surveillance, track human behaviour, facilitate human-computer interaction, and optimize resource utilization. The proposed system is an Image based Human Recognition (IHAR) system that utilizes images captured by CCTV, filters the images down to enhance the data quality, feature extraction using Principal Component Analysis (PCA) and utilizes various machine learning algorithms and compares their accuracies. The dataset collected comprises 10 different activities, including walking, sitting down, standing up, pick and throw, pushing, pulling, waving hands, clapping hands, and carrying, consisting of 35,530 images. The dataset is divided into different training and testing sets, with experimental results demonstrating high accuracy rates for the recognition algorithms. The study demonstrates that the Random Forest classifier with Canny Edge detector filtering technique achieved the highest accuracy of 97%.

The paper [6] discusses the application of the Long Term Short Memory (LSTM) method combined with Particle Swarm Optimization (PSO) for Human Activity Recognition (HAR) in videos. It highlights the significance of HAR in various fields such as surveillance, gaming, and healthcare and the challenges faced by conventional HAR methods, including slower recognition and low precision rates. The use of wearable devices and smartphones equipped with sensors for monitoring human activities is also mentioned. The document emphasizes the need for a more efficient framework to replace high-configured computational vision and operate with minimal data, particularly for real-time recognition in videos. Furthermore, it delves into the methodology, including the use of the UCF-50 dataset, the PSO-LSTM approach for training and testing, and the results obtained, including loss observed during training and accuracy gained during testing.

The research study [7] focuses on patient monitoring through abnormal human activity recognition using a CNN architecture. The dataset consists of 192 patient videos, with 23,040 frames, acquired from eight volunteers performing eight different activities. The dataset is partitioned into two ratios: 60% for training and 40% for testing, and 70% for training and 30% for testing. The labelling of the dataset was done using the VoTT annotation tool, adding bounding boxes around the patient and class IDs of the activity. The study employed the YOLO network as the backbone CNN model and achieved an accuracy of 96.8% in abnormal action recognition, outperforming a state-of-the-art method. The performance metrics, including recall, accuracy, precision, and F1-Score, were compared for different dataset partition ratios, showing substantial improvement with a 70:30 partition. The study also discussed the limitations of the YOLO model and proposed future work to handle small objects and multiple individuals in patient monitoring.

The study [8] aimed to develop an AI-based weapons detection system using the YOLOv4 Darknet framework for CCTV surveillance in Malaysia. The Open Images V6 dataset, comprising over 3000 images, was utilized for training the model. The training involved two sessions: single class and

multiple class object detection. The results indicated that the single class object detection achieved an average accuracy of 66.67% to 77.78%, while the multiple class object detection achieved up to 100% accuracy on most input images. The study also highlighted the challenges of recognizing different types of weapons due to variations in shape and size. The study proposed potential improvements for the multiple class object detection model, such as the ability to classify more types of weapons and implementation on microcontrollers for broader use.

The study [9] discusses the development of a real-time video surveillance system for detecting abnormal human behaviour. Various methodologies and models are presented, including the use of CNNs, OpenCV, motion influence maps, and YOLOv3 for detecting suspicious activities such as abandoned luggage, gun detection, and unusual human behaviour. Kamthe et al. proposed a semantic approach for defining and detecting suspicious activities, testing the model using CAVIAR (PETS 2004) and PETS 2006 datasets, achieving accuracies of 57% in object detection, 90% in object tracking, 93% in detecting loitering at ATM, and 96% in detecting abandoned bags. These datasets were crucial in training and evaluating the models for detecting abnormal human behaviour in surveillance videos. The document also outlines future enhancements to further improve the system's effectiveness and broaden its scope, focusing on enhanced security systems and real-time video analysis for more precise detection.

The research [10] presents a study on predicting personality traits using textual data in English and Brazilian Portuguese. It aims to address the scarcity of lexical resources for Brazilian Portuguese and evaluate the feasibility of training models with English textual data to predict personality traits in Brazilian Portuguese. The methodology involved data collection from the myPersonality dataset and Twitter, followed by text pre-processing and analysis using Word Embeddings techniques. Machine Learning models were trained and tested for FFM personality traits recognition. The results showed that the Stochastic Gradient Descent (SGD) model performed satisfactorily for predicting personality traits in Brazilian Portuguese textual data. The study also compared the performance of the proposed method with related works and found that the results were close to those obtained by other approaches.

The paper [11] presents SigSegment, a novel algorithm designed to address distracted driving by accurately detecting and segmenting abnormal driving events in real-world videos. Leveraging deep learning techniques like LSTM and CNN, SigSegment effectively identifies distractions such as eating or phone use while driving. Tested on a dataset of 210 video clips from 35 drivers engaged in various tasks, SigSegment achieved notable success in spotting distracted driving behaviours, as evidenced by its performance in the AI City Challenge 2023. Subsequent developments are planned to improve the algorithm's capacity to identify and evaluate anomalies in driving, which could support advanced driver assistance systems in improving road safety. With more study concentrating on extending its capabilities to detect other forms of driver attention or combining it with current driver assistance systems, SigSegment shows potential in the fight against distracted driving and offers insightful information for future traffic safety measures.

The Multilevel Guidance-Exploration Network and Behavior-Scene Matching Method is a unique framework [12] that the authors offer for the detection of anomalies in human behaviour in video data. The method utilizes normalizing flow to guide motion and appearance anomaly detection by merging RGB and skeletal features. To identify abnormalities, a behaviour-scene matching module investigates connections between typical behaviours and scenes. With AUC ratings of 86.9% and 73.5%, respectively, the system achieves state-of-the-art performance on the ShanghaiTech and UBnormal datasets. The method's superiority is confirmed by comprehensive implementation and assessment measures, such as ablation studies and comparisons with sophisticated approaches. Its robustness in scene-related anomaly identification is emphasized by comparison with other state-of-the-art approaches and visualization assessments that show its ability to find abnormalities across different circumstances.

The paper [13] notes the growing interest in behaviour modeling for surveillance and human-computer interactions and investigates the use of real-time video surveillance to identify anomalous human behaviour. It assesses current approaches, focusing on techniques for feature extraction, segmentation, and classification. The efficiency of Convolutional Neural Networks (CNNs) in video analysis is emphasized. There is discussion of several models and methods, such as object detection and suspicious activity identification. To strengthen security in a variety of contexts, including private properties, banks, offices, and airports, the conclusion highlights the effectiveness of the system and outlines recommendations for future improvements. In summary, the document presents a thorough analysis of aberrant behaviour detection, highlighting its potential uses and importance in improving security and safety in various contexts.

The high theft rates and inadequate security measures in Indonesia motivate this study's [14] investigation of the use of CCTV camera video and the YOLO V3 algorithm to identify criminal activities, especially theft, in school settings. Starting with context awareness, the study highlights the stages of data preparation and model training, going into technical details about things like turning on the darknet and configuring parameters. The results show how well YOLO V3 detects human activity, with high accuracy rates—99.1%, in particular. To further improve safety and discourage criminal activity, the study highlights the potential of YOLO V3 to support security measures in educational settings and suggests possibilities for future research, such as object tracking algorithms and integration with current security infrastructure.

The researchers Tserenpurev Chuluunsaikhan, Jong-Hyeok Choi, and Aziz Nasridinov from Chungbuk National University in South Korea [15] have created a revolutionary method for identifying child abuse using real-time video surveillance, which is presented in this study article. The program uses a CAD model with attention networks and Multiple Instance Learning for anomaly scoring, and pre-trained C3D models for feature extraction, using a mix of the UCF-Crime dataset and a recently assembled dataset on child abuse. Evaluation metrics like AUC and ROC curves show how successful the deep learning strategy is; it achieved an astounding AUC score of 84.5%, which is 10% higher than earlier models. The program improves the capacity to recognize cases of child abuse in kindergartens and schools, even in the event of misdetections. To counteract the growing

number of child abuse incidents, the report highlights the significance of using cutting-edge technologies like deep learning. It also makes recommendations for future improvements, like the incorporation of transformer networks and the extension of datasets, to further improve child safety measures.

This study [16] aims to detect and monitor unusual and suspicious behaviour among individuals in crowded locations. The authors utilize a deep learning methodology to remotely identify and track actions that deviate from the norm, providing accurate information regarding the location, time of occurrence, and potential identification of perpetrators. The proposed system leverages the advanced YOLOv8 model for real-time detection of weapons, with a focus on ensuring swift performance by quantizing the model's weights. Upon completion of the study, the authors achieved a precision rate of 92.6%. To train the model, the study amassed a dataset comprising 2986 images of weapons sourced from various platforms such as Google and YouTube, which were then meticulously annotated using the Roboflow website.

This research paper [17] focuses on autonomously monitoring and analyzing crowd behaviour in real-world settings to facilitate efficient crowd management in various public areas like transportation stations and streets. The researchers introduce a transformer-based crowd management monitoring framework named V3Trans-Crowd, which processes video data to extract meaningful insights for categorizing crowd behaviour. This model integrates spatial and temporal considerations to analyze videos depicting crowd movement, enabling the classification of crowd behaviour based on the derived analysis. Additionally, the study employs an enhanced hierarchical transformer tailored for multi-modal tasks. The model is trained using the Crowd-11 dataset, achieving an accuracy of 63.32%.

This study [18] delves into the rising instances of disruptive and offensive behaviours directed towards individuals, particularly women. It introduces a machine learning algorithm known as Darknet 53. The aim of this algorithm is to identify patterns indicative of such behaviours in real-time, distinguishing irregularities from typical behaviour and filtering out inconsistencies from the norm.

This paper [19] introduces an automated system designed to recognize unusual human behaviour captured by CCTV cameras in public spaces. It employs spatio-temporal 3D convolutional neural networks for this purpose. To ensure accuracy, the dataset underwent thorough annotation to filter out noise, facilitating precise localization of anomalies within the video footage. The human-related dataset, comprised of real crime scenes, was then compared against other cutting-edge techniques like Pseudo 3D and ResNet 3D. Through these comparisons, the system achieved an impressive accuracy rate of 97.39%. The experiments were conducted on the UCF-Crime dataset, a widely recognized benchmark video dataset for anomaly detection. This dataset consists of 1900 surveillance videos, including 950 unedited real-world surveillance recordings featuring clear anomalies, as well as normal videos to prevent class imbalance.

The paper [20] discusses the rapid evolution of society, marked by increasing urbanization and population density, which in turn raises concerns regarding safety such as accidents and potential terrorist threats. To address these challenges, there's a growing reliance on intelligent video systems (IVMS) for surveillance, particularly to detect unusual behaviour, like students engaging in misconduct on escalators. The study proposes the use of an OpenPose deep learning network to streamline the process of identifying and analyzing abnormal behaviour by reducing redundant information from human bone facial features during feature extraction. The extracted human skeleton features are then classified using a graph convolution neural network, aiming to minimize the computational complexity of behaviour identification algorithms. Additionally, a sliding window voting method is employed to enhance the accuracy of behaviour classification in practical scenarios. The experimentation involves the utilization of a self-constructed student trajectory dataset alongside the INRIA dataset, resulting in an outstanding accuracy rate of 99.50 percent in diagnosing and classifying abnormal behaviour among students under video surveillance.

III. DESIGN

A. *Architectural Design*

We have proposed a deep learning based approach involving Neural Networks and Transfer learning in order to extract relevant features from the video footage and detect whether there is any abnormal activity and if there is, to categorize it accordingly. Shown in Fig 1 is the proposed architectural design system. Here we specify the steps taken to create our application. The final output returned to the user is the detected behaviour enclosed with bounding boxes to ensure clear communication and accurate results. The proposed system has the following steps:

- Data collection and pre-processing
- Base CNN Model selection
- Transfer Learning
- Feature Extraction
- Anomaly detection layer
- Model training and evaluation
- Deployment and Testing

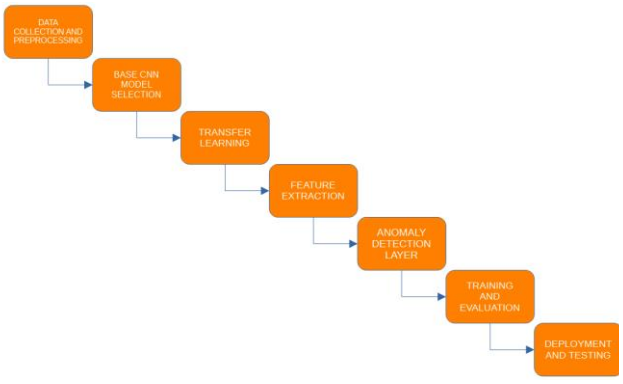

Fig. 1. Architectural Design

B. *Data flow diagram*

Compile a variety of datasets that show human behavior, including both typical and unusual actions. To guarantee representative samples for efficient anomaly identification, preprocess the data by standardizing formats, addressing missing values, and supplementing datasets.

Employ a CNN model that has already been trained and is appropriate for analyzing human behavior while considering the spatial and temporal nuances like ResNet or VGG models.

Start the selected pre-trained CNN model without any classifying layers to apply transfer learning. Utilize the previously acquired information to leverage the model's weights as you fine-tune it based on human behavior data to capture domain-specific elements that are essential for anomaly identification.

Apply the modified CNN layers to extract complex characteristics from data on human behavior. These characteristics must encompass temporal and geographical patterns. On top of the modified CNN architecture, add an anomaly detection layer that is intended to spot departures from typical behavior patterns. This layer allows the model to efficiently identify instances of abnormal behavior. It typically consists of dense neural network components with suitable activation functions.

Utilize labeled datasets of human behavior that have been divided into training, validation, and testing sets to train the model. Throughout the training process, keep an eye on the model's performance and use validation measures to adjust the hyper parameters and avoid overfitting.

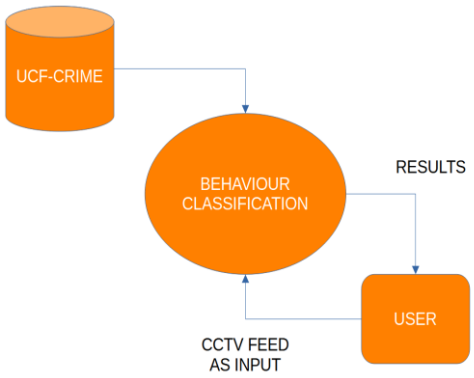

Fig. 2. Level 0 Data flow diagram

IV. Implementation

A. Dataset

The UCF Crime Dataset is an invaluable resource that includes a wide range of criminal activities captured in videos from various regions of the United States. This Kaggle dataset provides information on a variety of crimes, such as abuse, arrest, arson, assault, burglary, explosion, fighting, road accidents, robbery, shooting, shoplifting, stealing, and vandalism. Each class represents a specific type of behaviour or activity depicted in the videos, with important applications in surveillance, law enforcement, and public safety.

In total, there are 1,080,170 Training and 79,959 Testing images used spanning across 7 classes namely, Arson, Assault, Explosion, Normal, Road Accidents, Robbery, Shoplifting and Stealing. Overall, every 10th frame of the videos present in the UCF-Crime dataset have been extracted to maintain variety while minimizing redundancy.

B. DenseNet201

The main innovation of the architecture is its dense connectivity pattern, in which every layer is feed-forward coupled to every other layer (Fig 3). This makes it easier for features to be reused across the network, which improves gradient flow and results in lower parameters and better feature propagation. DenseNet201 expands on this concept by utilizing a more extensive and profound network architecture, with 201 layers overall. Reversed linear unit (ReLU) activations, convolutional procedures, and batch normalization are examples of composite functions that make up each layer of DenseNet201.

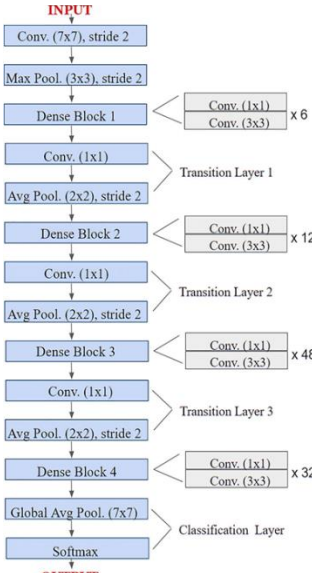

Fig. 3. DenseNet201 Architecture

C. MobileNetV2

The use of linear bottlenecks and inverted residuals, which greatly improve network performance and efficiency, is the primary innovation of the MobileNetV2 architecture. Because each layer's output in this design is directly coupled to its predecessor and to the layer after it (Fig 4), gradient flow and feature reuse are enhanced. The design keeps excellent precision while minimising the amount of parameters. In order to reduce computational costs, MobileNetV2 uses depth wise separable convolutions, which split ordinary convolutions into depth wise and pointwise convolutions. Additionally, it uses batch normalisation and ReLU6 activations to stabilise training and enhance model performance. MobileNetV2, with an emphasis on embedded and mobile vision applications, is made to operate reliably on devices with limited resources.

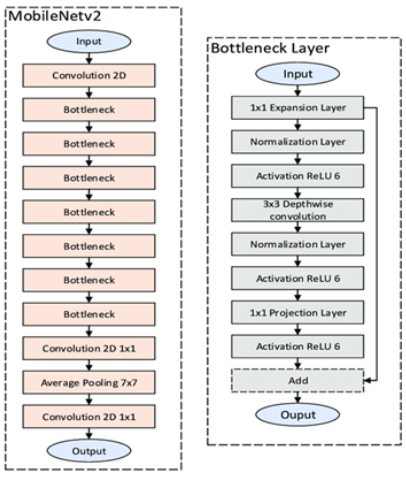

Fig. 4. MobileNetV2 Architecture

D. *ResNet50*

A member of the larger ResNet family, ResNet-50 (Fig 5) introduced the idea of skip connections and residual learning to address the degradation issue that very deep neural networks frequently face. The accuracy of a neural network improves with increasing depth at first, but then starts to rapidly deteriorate, making the training of very deep networks difficult. ResNet-50 uses a unique architectural approach to effectively train very deep networks in order to overcome this problem.

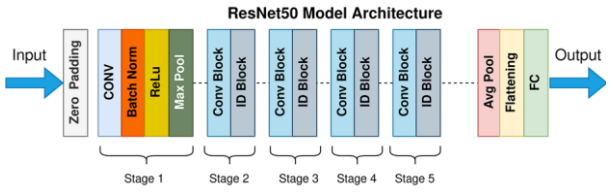

Fig. 5. ResNet50 Architecture

E. *Proposed Model*

The model employs several key components for efficient image classification (Fig 6). ImageDataGenerator from Keras is utilized for real-time data augmentation, expanding the training dataset and enhancing model robustness. The input layer defines the dimensions of the input images. DenseNet201, pre-trained on the ImageNet dataset, serves as the feature extractor, leveraging transfer learning for accurate feature representation. The Global Average Pooling 2D layer reduces the spatial dimensions of feature maps, resulting in a fixed-length vector and lowering the model's parameter count. Dense layers learn complex patterns from the extracted features, with each neuron applying a linear transformation followed by a ReLU activation function. Dropout layers prevent overfitting by randomly deactivating neurons during training, improving generalization. The output layer, a dense layer with a softmax activation function, converts raw outputs into class probabilities. The model is compiled with categorical crossentropy as the loss function, stochastic gradient descent (SGD) as the optimizer, and accuracy and AUC as evaluation metrics.

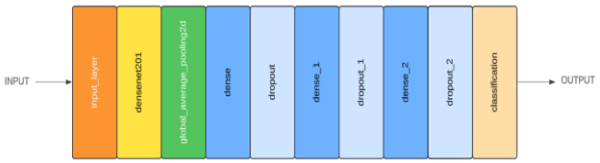

Fig. 6. Proposed Model Architecture

SGD is an iterative version of the classic Gradient Descent algorithm, which minimizes an objective function by repetition. The set of weights and biases that minimize this loss is the objective function, or loss function, in the context of neural networks. Essentially, the SGD optimizer modifies the neural network's weights and biases in a way that is opposite to the gradient of the loss function concerning those parameters. The iterative process of this update aims to progressively minimize the loss function and enhance the network's performance.

Mathematically, the SGD update rule is represented as follows:

$$w_{t+1} = w_t - a\frac{\partial L}{\partial w_t} \quad (1)$$

By adding a portion of the prior update to the present update, the momentum technique adds a "momentum" word to the weight updates. This can lessen oscillations and speed up the optimization process, especially in regions where the gradient is more constant.

$$v_t = \gamma * v_{t-1} + n\nabla w_t$$

$$w_{t+1} = w_t - v_t \quad (2)$$

V. Results and Analysis

A. *DenseNet201*

The DenseNet201 model has achieved 91.89% Validation and Testing accuracy with a loss value of 0.0034 respectively, shown by Fig 7 and Fig 8, using the training history run on 1,080,170 training images and 79,959 testing images belonging to 7 classes. Figures 9 and 10 illustrate some additional metrics that were used to validate the DenseNet201 model, namely ROC-AUC curve and the confusion matrix.

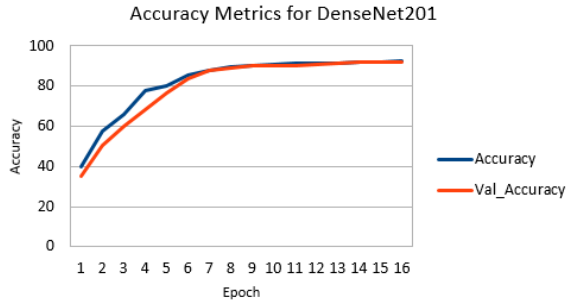

Fig. 7. Accuracy metrics for DenseNet201

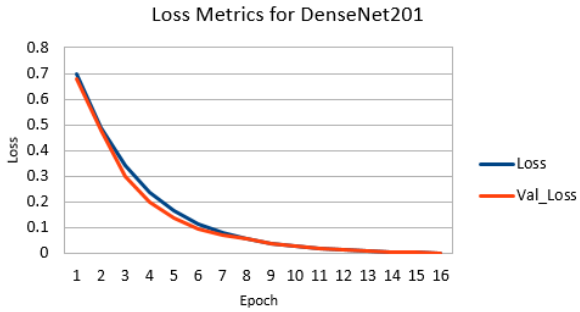

Fig. 8. Loss metrics for DenseNet20

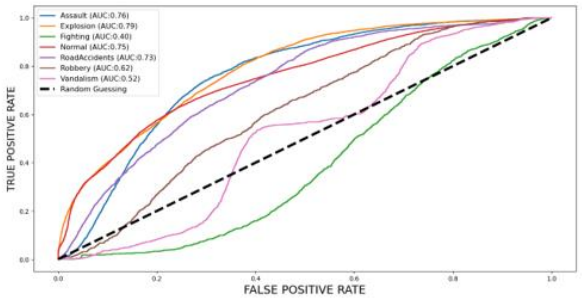

Fig. 9. ROC-AUC for DenseNet20

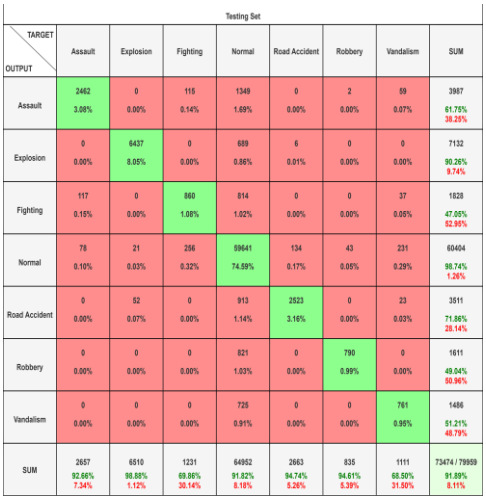

Fig. 10. Confusion matrix for DenseNet20

B. *MobileNetV2*

The MobileNetV2 model has achieved 86% Validation and Testing accuracy with a loss value of 0.085 respectively, shown by Fig 11 and Fig 12, using the training history run on 1,080,170 training images and 79,959 testing images belonging to 7 classes.

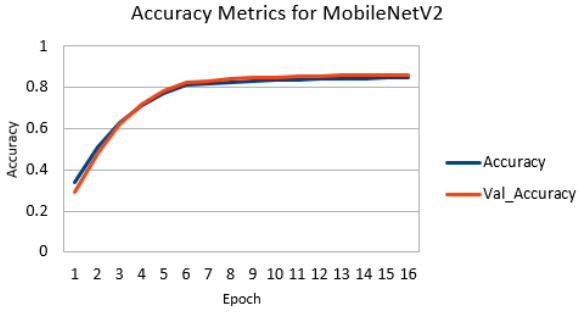

Fig. 11. Accuracy metrics for MobileNetV2

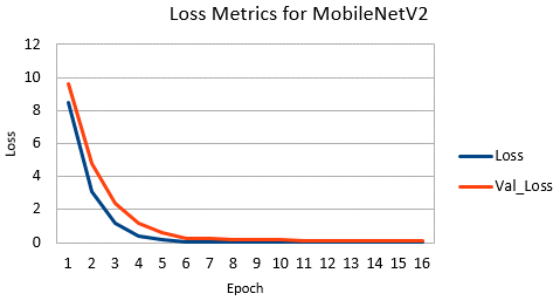

Fig. 12. Loss metrics for MobileNetV2

C. *ResNet50*

The ResNet50 model has achieved 80% Validation and Testing accuracy with a loss value of 0.029 respectively, shown by Fig 13 and Fig 14, using the training history run on 1,080,170 training images and 79,959 testing images belonging to 7 classes.

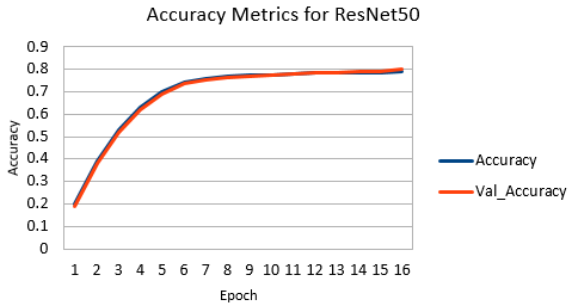

Fig. 13. Accuracy metrics for ResNet50

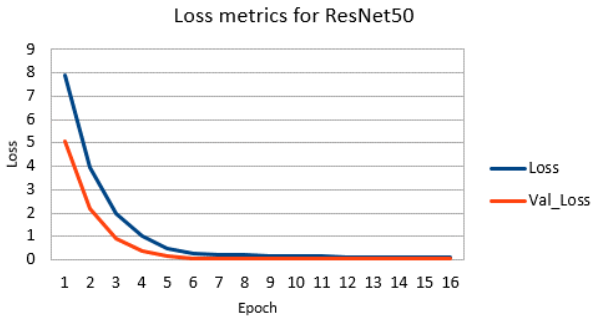

Fig. 14. Loss metrics for ResNet50

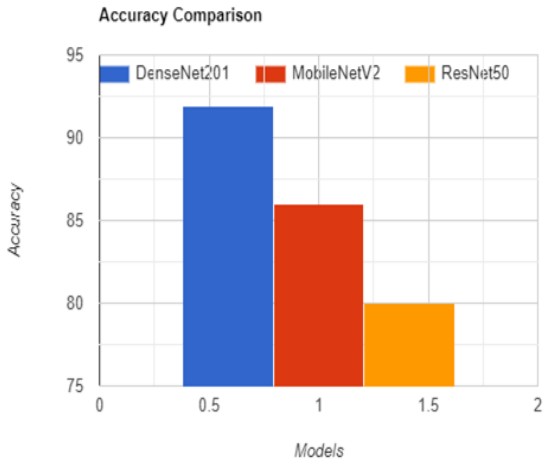

Fig. 15. Accuracy comparison

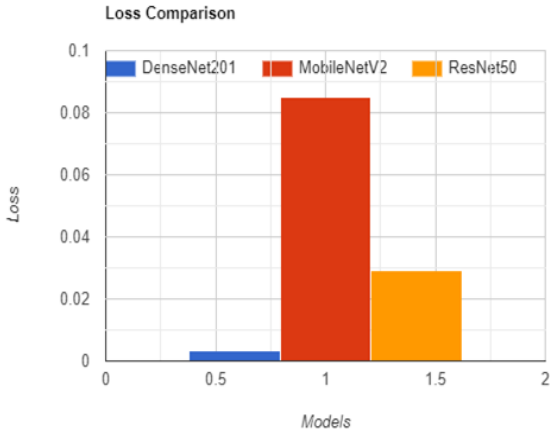

Fig. 16. Loss comparison

VI. Conclusion and Future Scope

The models have been tested using 79,959 frames that have been extracted from videos spanning across 7 classes, under careful monitoring of various metrics such as Accuracy, AUC and Loss values. Further, the best performing model has also yielded the ROC-AUC graph and its confusion matrix to provide further insights into its conclusions. Comparing the three models we have used for transfer learning, DenseNet201 has far surpassed the performance of the other two models, gaining an accuracy of 91.89% (Fig 15) with a loss value of 0.0034 (Fig 16). Thus, after analyzing our results we can conclude that DenseNet201 performs the best under the testing conditions subjected to.

The future scope of this project could involve further enhancements to improve the system's effectiveness and broaden its scope. This could include refining the deep learning models, exploring new datasets, and integrating additional technologies such as spatial and temporal analysis for crowd behaviour, video-based visual transformers for crowd management, and real-time video surveillance for detecting child abuse. Additionally, the project could explore the application of deep learning in monitoring naturalistic driving videos and patient monitoring through abnormal human activity recognition.

References

[1] G. Dinesh Mavaluru, Azath Mubarakali, Bayapa Reddy Narapureddy, et al. Deep convolutional neural network based real-time abnormal behavior detection in social networks Computers & electrical engineering. 2023;111:108987-108987. doi: 10.1016/j.compeleceng.2023.108987.

[2] P. Kuppusamy, V.C. Bharathi. Human abnormal behavior detection using CNNs in crowded and uncrowded surveillance – A survey. Measurement Sensors. 2022;24:100510-100510. doi: 10.1016/j.measen.2022.100510

[3] Khosro Rezaee, Khosravi MR, Maryam Saberi Anari. Deep-Transfer-Learning-Based Abnormal Behavior Recognition Using Internet of Drones for Crowded Scenes. ResearchGate. Published June 2022. Accessed May 17, 2024. doi:10.1109/IOTM.001.2100138

[4] Miki D, Chen S, Kazuyuki Demachi. Unnatural Human Motion Detection using Weakly Supervised Deep Neural Network. ResearchGate. Published September 2020. Accessed May 17, 2024. doi: 10.1109/AI4I49448.2020.00009

[5] Ullah F, Iqbal A, Khan A, Kyung Sup Kwak. An Image-based Human Physical Activities Recognition in an Indoor Environment. ResearchGate. Published October 21, 2020. Accessed May 17, 2024. doi:10.1109/ICTC49870.2020.9289314

[6] M Razmah, Senthil Govindaswamy Ambalavanan, Ramadoss Prabha, A Naveen. LSTM Method for Human Activity Recognition of Video Using PSO Algorithm. ResearchGate. Published December 8, 2022. Accessed May 17, 2024. doi:10.1109/ICPECTS56089.2022.10046783

[7] Nor W, Isa NM. Object Detection: Harmful Weapons Detection using YOLOv4. IEEE Symposium on Wireless Technology and Applications. Published 2021. Accessed May 17, 2024. doi:10.1109/ISWTA52208.2021.9587423

[8] Y. Li, J. Zhang, M. Nie and S. Wang, Hessian-regularized spectral clustering for behavior recognition, 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), Sanya, China, 2020, pp. 156-159, doi: 10.1109/ICHCI51889.2020.00042.

[9] Malik Ali Gul, Muhammad Haroon Yousaf, Nawaz S, Zaka Ur Rehman, Kim H. Patient Monitoring by Abnormal Human Activity Recognition Based on CNN Architecture. Electronics. 2020;9(12):1993-1993. doi: 10.3390/electronics9121993

[10] Santos, Moura F, Lopes L, César M. Words Similarities on Personalities: A Language-Based Generalization Approach for Personality Factors. ResearchGate. Published 2023. Accessed May 17, 2024. doi: 10.1109/ACCESS.2023.3261339

[11] Kwakye K, Seong Y, Aboah A, Yi S. SigSegment: A Signal-Based Segmentation Algorithm for Identifying Anomalous Driving Behaviours in Naturalistic Driving Videos. arXiv.org. Published 2023. Accessed May 17, 2024. https://arxiv.org/abs/2304.09247

[12] Yang G, Luo Z, Gao J, et al. A Multilevel Guidance-Exploration Network and Behavior-Scene Matching Method for Human Behavior Anomaly Detection. arXiv.org. Published 2023. Accessed May 17, 2024. https://arxiv.org/abs/2312.04119

[13] Prof. Shweta Sondawale, Mansi Shinde, Kartiki Nanekar, Shwetali Nalawade, Sanket Pharkute, Video-based Abnormal Human Behavior Detection, International Journal of Research Publication and Reviews, Vol 4, no 5, pp 5711-5717, May 2023

[14] Jumadi, W. B. Zulfikar, F. F. Abdillah, N. A. Dewi, A. R. Atmadja and M. I. Al Amin, Classroom Activities Detection Using You Only Look Once V3, 2023 IEEE 9th International Conference on Computing, Engineering and Design (ICCED), Kuala Lumpur, Malaysia, 2023, pp. 1-6, doi: 10.1109/ICCED60214.2023.10425654

[15] Tserenpurev Chuluunsaikhan, Choi JH, Aziz Nasridinov. Application for Detecting Child Abuse via Real-Time Video Surveillance. ResearchGate. Published September 28, 2022. Accessed May 17, 2024. doi:10.1109/ICISCT55600.2022.10147005

[16] Zuo Y, Aymen Hamrouni, Hakim Ghazzai, Massoud Y. V3Trans-Crowd: A Video-based Visual Transformer for Crowd Management Monitoring. ResearchGate. Published June 2022. Accessed May 17, 2024.

[17] M. Pullakandam, K. Loya, P. Salota, R. M. R. Yanamala and P. K. Javvaji, Weapon Object Detection Using Quantized YOLOv8, 2023 5th International Conference on Energy, Power and Environment: Towards Flexible Green Energy Technologies (ICEPE), Shillong, India, 2023, pp. 1-5, doi: 10.1109/ICEPE57949.2023.10201506.

[18] R. Hasib, A. Jan and G. M. Khan, Real-Time Anomaly Detection for Smart and Safe City Using Spatiotemporal Deep Learning, 2022 2nd International Conference on Artificial Intelligence (ICAI), Islamabad, Pakistan, 2022, pp. 79-83, doi: 10.1109/ICAI55435.2022.9773464.

[19] U. Y. Reddy, M. S. Nikhil, P. S. S. Krishna and S. S, Systematic Harmful Signs Detection for Women'S Safety Using Neural Networks, 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-5, doi: 10.1109/ICCCI56745.2023.10128298.

[20] Ding Y, Bao K, Zhang J. An Intelligent System for Detecting Abnormal Behavior in Students Based on the Human Skeleton and Deep. ResearchGate. Published June 27, 2022. Accessed May 17, 2024. doi:10.1155/2022/3819409