MoveNET Enabled Neural Network for Fast Detection of Physical Bullying in Educational Institutions

Zhadra Kozhamkulova[1], Bibinur Kirgizbayeva[2], Gulbakyt Sembina[3], Ulmeken Smailova[4], Madina Suleimenova[5], Arailym Keneskanova[6], Zhumakul Baizakova[7]

Almaty University of Power Engineering and Telecommunications, Almaty, Kazakhstan[1, 5, 6]
Kazakh National Agrarian Research University, Almaty, Kazakhstan[2]
International Engineering Technological University, Almaty, Kazakhstan[7]
International Information Technology University, Almaty, Kazakhstan[3]
Center of Excellence AEO "Nazarbayev Intellectual Schools", Astana, Kazakhstan[4]

***Abstract*—In this article, we provide a MoveNET-based technique that we think may be used to detect violent actions. This strategy does not need high-computational technology, and it is able to put into action in a very short amount of time. Our method is comprised of two stages: first, the capture of features from photo sequences in order to evaluate body position; next, the application of an artificial neural network to activities classification in order to determine whether or not the picture frames include violent or hostile circumstances. A video aggression database consisting of 400 minutes of one individual's actions and 20 hours of videodata encompassing physical abuse, as well as 13 categories for distinguishing between the behaviors of the attacker and the victim, was created. In the end, the suggested approach was refined and validated by employing the collected dataset during the process. According to the findings, an accuracy rate of 98% was attained while attempting to detect aggressive behavior in video sequences. In addition, the findings indicate that the suggested technique is able to identify aggressive behavior and violent acts in a very short amount of time and is suitable for use in apps that take place in the real world.**

***Keywords*—MoveNET; neural networks; skeleton; bullying; machine learning**

I. INTRODUCTION

The purpose of this project is to scrutinize the problem of aggressive behavior and bullying in schools in order to propose the best possible solution. According to Olweus [1], school bullying is an unwanted aggressive behavior on the part of one or more other students that exposes a victim to negative actions repeatedly and over time. Negative actions can be carried out by physical contact, by words or in other ways, for instance, making faces and obscene gestures, or often ostracizing the victim from the common social community. Generally speaking, bullying may be identified by the three characteristics that are listed below: (1) It is violence-related behavior or purposeful "harm-doing," (2) it is activity that is carried out "repeatedly over time," and (3) it is behavior that occurs in an interpersonal relationship defined by a real or expected imbalance of power [2]. Scientific evidence suggests that bullying affects future mental health functioning of both victims of bullying and those who cause harm/bullying. Apart from physical aggression, bullying also includes psychological pressure, intimidation, rumor spreading, extortion, and mockery.

Aggression may take both direct and indirect forms when it comes to bullying. Direct types of bullying, consisting of an overt demonstration of physical power, can take the form of physical or verbal violence. The term "physical bullying" refers to any kind of physical attack, including in particular striking, shoving, kicking, choking, and any harmful action towards the victim. Bullying victims may be subjected to verbal harassment or intimidation when they are called names, threatened, taunted, teased maliciously, or psychologically intimidated by offensive language. Children may be bullied in a variety of ways, including stealing, vandalizing, making offensive looks or gestures, and making faces [3].

The National report “Factors influencing health and well-being of children and adolescents in Kazakhstan” published by the National Center for Public Health of the Ministry of Health of the Republic of Kazakhstan [4] provides results about health, social conditions and well-being of teenagers aged 11 to 15 years. The study is based on HBSC methodology, a WHO collaborative cross-national survey. The report contains information on social and health indicators that are related to the health and well-being of both children and adolescents. And bullying was defined as one of the risk factors affecting the health and well-being of children in this report.

According to the data published in the National report, 17% of teenagers aged 11 to 15 years were bullied at school one or more times per month. 20% of teenagers from the same age group were involved in bullying others at least once. This rate is higher among 11 and 13 year old boys compared to girls.

The goal of this research is to develop Artificial Intelligence (AI) Solutions in order to utilize them as a basis for designing a prototype of a software-hardware complex that can automatically detect cases of aggressive behavior and potential physical bullying in educational institutions.

The remainder of this paper is structured as follows: The next part discusses cutting-edge physical aggression detection, after which a problem statement is presented and described.

The goals and aims of the research are thoroughly explained in the third part. The human skeleton-based physical aggression detection approach is discussed in the fourth part of this article. The procedure of data collecting and the investigation's outcome are laid out in the fifth part of the report. In the sixth part, findings are discussed, and ongoing issues in the field of violence identification in videos are described. The report is brought to a close with the last part, which discusses the plans for and issues associated with physical bullying detection in video. The creation of an automatic and rapid physical aggression detection system in video security cameras based on human skeleton is the key objective of the work. The developed technique makes it possible to recognize violent events in the video without the need for highly processed hardware.

II. Related Works

The National Center for Educational Statistics (2019) showed that one in five students (20.2%) reported being bullied at school in numerous places, such as a hallway or stairway (43%), in the cafeteria (27%), outside on school grounds (22%), online / text (15%), in the bathroom or locker room (12%), on the school bus (8%) [5].

One of the first systematic studies to collect data on the nature and extent of violence in schools in Kazakhstan was conducted by the United Nations Children's Fund (UNICEF) in 2013, which revealed that 66.2% of schoolchildren were exposed to school violence and discrimination, 63.3% were witnesses, 44.7% were victims, and 24.2% were perpetrators of violence and discrimination against other children in school. [6]

Video analysis is an area of AI and machine learning that has shown good results in recent years and is widely used. Bullying in its various forms poses a serious problem that a vast amount of schoolchildren faces. For various reasons, there are not many scientific investigations in the world which attempt to fix the negative consequences of bullying by means of video analysis. Among the few, slow development in this vector can be mentioned. There are some studies related to cyberbullying and depression detection on online user contents [7-9]. However, there is no evidence about such researches in the Republic of Kazakhstan. This substantiates the novelty of the proposed project.

The current level of development of AI methods for video analysis allows using them to process video footage from school cameras. However, there are not many researchers who study the effectiveness of using AI methods to reduce cases of bullying and its negative effects at school. According to this project, video analysis using AI methods will enable early detection of aggressive behavior. Consequently, the early detection of such cases will facilitate the work of school psychologists in terms of early warning of bullying.

A distinctive feature of this project is its interdisciplinarity: new proposed solutions of AI will push the boundaries of bullying studies to a social phenomenon. The collected data will be used to conduct a psychological study on the effect of bullying on the psychological and emotional health of schoolchildren. The combined use of AI methods and psychology will provide the results that may find application in those areas of life where video analysis is needed.

Using neural network technologies will allow for the intelligent video footage processing in order to assess human behavior and determine aggressive actions.

In the proposed study, software models of artificial intelligence will be trained on the basis of LGD-3D architecture, a two-stream I3D structure. A recent study examined the problem of recognizing aggressive actions based on RGB video data, the I3D architecture showed better results compared to C3D and R3D in all respects.

For video classification, a method based on a neural network with deep convolutional graphs (DCGN) will be used. According to the results of research, this method is superior to alternative ones, such as LTSM and GRU.

The categorization of activities, in addition to the categorization of facial expressions, is an active topic of study that is, nonetheless, fairly difficult. Large variations in action performance brought on by differences in individual's anatomy, as well as temporal and spatial variations (including differences in the pace at which people do actions), are some of the issues that are linked with the categorization of actions [10]. It might be difficult to tell the difference between activities that are just part of the game (such wrestling or hurling things at each other), and those that constitute bullying. Either adjusting the system to disregard activities that are related to typical children's games or integrating numerous algorithms might be the answer to this issue (for example, a combination of the classification of emotions and actions).

In recent years, researchers have raised concerns about physical bullying detection [11-13]. Previous researches [11] used the transfer learning approach on the identification of violent conduct. The authors developed a violence detector that was based on transfer learning and tested it using three different datasets. They had an accuracy rate of 80.90 percent on average when it came to identifying violent content (from a video that was obtained from YouTube). Next study [12] investigated the use of information about irregular mobility to identify violent behavior in surveillance cameras. By using of the Motion Co-occurrence Feature, the authors were able to conduct an analysis on the properties of the motion vectors that were produced in the vicinity of the item (MCF). They utilised the CAVIAR database, but, however, the research did not provide any numerical results about the accuracy on average.

The National Autonomous University of Mexico conducted a study using a systematic observation strategy called SDIS-GSEQ [14-15] to describe the behavioral patterns in children who were identified by the program as "victims" and their changes. The purpose of the study was to investigate the effects of the program on these children.

III. Problem Statement

The purpose of this research is to provide a method for rapid identification of violent incidents captured by video security cameras. The scientific contribution made by this study is the invention of a system for the rapid identification of violent behavior. The objective was accomplished by training a

neural network using a tracking by detection method like MoveNet retrieved points from human skeletons. The following goals need to be completed in order to succeed in this endeavor: a) Development of a Video Dataset with aggressive behaviour scenes; b) Extract human skeleton points using MoveNet; c) Train the neural network applying the extracted points d) Evaluate the trained neural network.

IV. Data

The problem of identifying violent and aggressive conduct may be broken down into a variety of more specific subtasks. Fig. 1 presents the research process as a flowchart for a reference. The flowchart for the study project is divided into its primary components, which are feature extraction, data collection, and classification problem. The section on data characteristics is where the pattern parameters of the perpetrator are defined. The portion responsible for data collection assures the availability of relevant video data, marks up videos according to classifications, stores them in .json format, and trims the marked video sequences that include violent situations in order to produce a dataset. In this section, we present all kinds of data operations from collection to the preparation for neural network training.

A. Data Collection

The first step is to determine the various categories of information that need to be compiled. We came up with different distinct categories of traits to differentiate a victim from an aggressor. These traits may be categorized as either passive or active. In the beginning, we determined their characteristics based on the predetermined classifiers. These characteristics are the ones that should be assessed during the process of data collection. Afterwards, the characteristics of the victim and the aggressive behavior were broken down into 13 different groups.

When searching for video materials that are available for free access on the internet, we applied a variety of search phrases, including "aggression," "physical aggression," "violence," "bullying," "fight," "group fight," and others. After gathering them, the next step was to assign appropriate classes to the spatiotemporal segments included within the videos, and the information about their labeling was saved in the *.json file format. To accomplish this task, we used VGG Image Annotator. Following the completion of the tagging, each of the movies was clipped, and then they were arranged into classes.

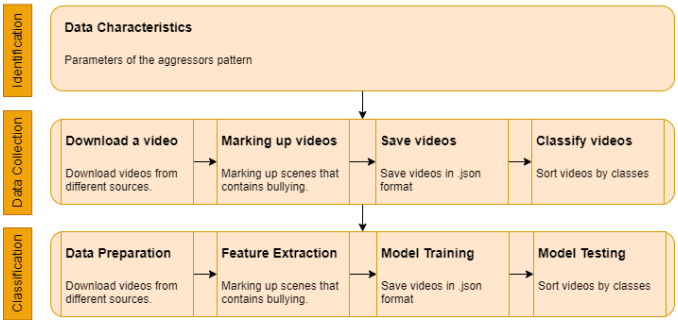


Fig. 1. Flowchart of the research.

B. Dataset

The initial step of our investigation consisted of collecting footage of acts of violence committed by a single individual. There are thirteen categories of violent acts that have been categorized. As a result, there were a total of 80 classifications that were found to belong to either an offender or a victim over the course of the inquiry. In order to put the training model into practice, we needed to determine the activities of a single individual. For such purpose, we broke the project down into 13 courses that each only needs one person to complete. Table I provides an illustration of the thirteen courses that were used in the training of the model. In the course of our research, we developed our very own dataset, which is made up of the thirteen categories that were previously established. The dataset was used for both training and testing of the model that we have suggested. After that, it was put to the test by making use of free datasets of footage of violent acts.

The information shown in Fig. 2 pertains to the videos that were gathered. Videos are gathered in three different formats. Statistical information on the various kinds of video data files that were collected may be seen in Fig. 2(a). The dissemination of video sequences is shown in Fig. 2(b). It was determined that a total of 2,093 short videos illustrating incidents of physical bullying and aggressive behavior should be collected. Approximately 20 hours were spent gathering all of the video data. The following is a list of the file formats that were used to gather the data on the videos:

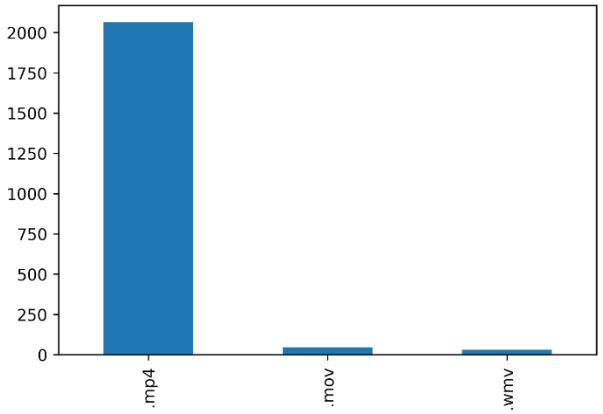
- video in. mp4 format: 2 017 files;
- video in.mov format: 44 files;
- video in.wmv format: 32 files.

TABLE I. Comparison of the Obtained Results

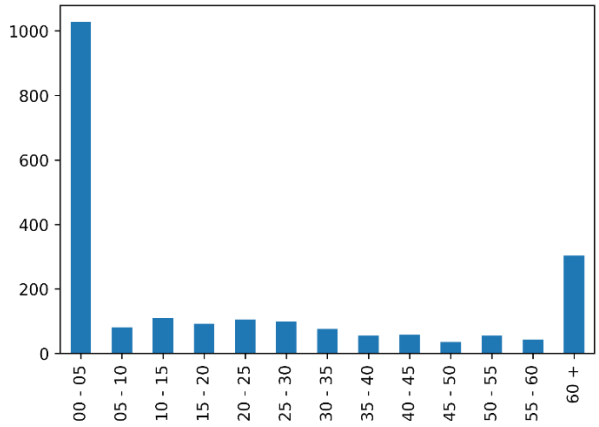
Class id	Class type
0	Large range of hand movements
1	The head is directed towards the victim
2	The body turned to face the victim
3	The shoulders back and the arms back
4	Hands on hips
5	Takes off the outer clothes
6	Kicking
7	Punching
8	Covering the face
9	Legs pointing in different directions
10	A series of bouncing blows
11	Bend over
12	Finger pointing

Throughout the collected data, we also recorded videos of a single person committing physical aggression actions. Additional films are necessary for the first stages of the neural network training. The whole of the brand new video content clocks in at close to four hundred minutes. There are two distinct categories of violent videos, each of which is defined by its intended purpose. The first category of videos depicts acts of violence in crowded places. The second category deals with secluded violence, which can take place between just two people in uncrowded scene and typically involves one

participant acting as a bullier and the other participant acting as a victim.



(a) Video files classified by formats



(b) Collected video files by types

Fig. 2. Collected dataset of videos.

V. Materials and Methods

A. The Proposed Approach

In the next paragraphs, we will discuss our methodology, which is known as the tracking by detection. The suggested system's general design is shown in Fig. 3, which may be seen below. The system may be broken down into three different subproblems. In the first step of this process, we approximate the human stance on each image sequence by applying the MoveNet model to the input image sequence. In the second step, we take each frame and retrieve important points as vectors. MoveNet provides a total of 17 important locations for each frame. As a direct result of this, we are able to generate vectors that include 34 individual components. In the subsequent phase, we combine all of the k vectors into a single vector before passing it on to the step that deals with features extraction and activity identification. In the last step, known as stage three, we train a neural network to solve tasks related to action recognition. There are two different kinds of algorithms for determining the location of a human skeleton based on RGB images: top-down and bottom-up. The first ones will trigger a human detector and examine body joints in boundary boxes that have already been determined. Top-down methods include the ones described in MoveNet [22], HourglassNet [23], and Hornet [24]. There are a few other bottom-up algorithms, such as Open space [25] and PifPaf [26].

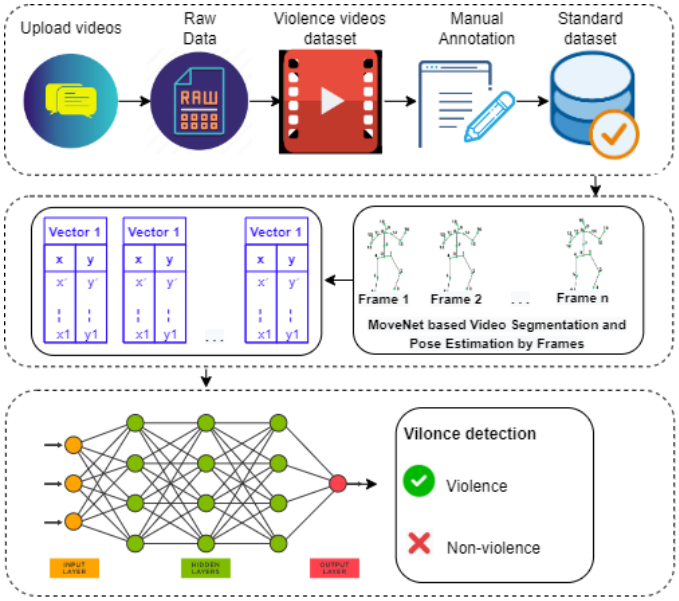


Fig. 3. Flowchart of the study.

We carried out our training using a strategy known as the skeleton approach. The described approach has the potential to reduce the costs associated with processing. A MoveNet based neural network is employed in order to create an accurate appraisal of the figure of either the perpetrator or the victim. Using a MoveNet that has already been pre-trained, a function extraction has the ability to transfer the data obtained in the input space to the target domain. The output of MoveNet represents the human skeleton with 17 primary body points together with their positions and the confidences associated with those sites. There are 17 vital points on the body, including the nose, eyes, ears, shoulders, elbows, wrists, thighs, knees, and ankles. Fig. 4 depicts an instance of 17 key points that MoveNet might obtain and use to train the neural network. These points are applied to the network. The x and y coordinates of the important points are what are used to represent them in the two-dimensional coordinate space.

The following formula illustrates one possible approach to depict the human body:

$$r_b(x_i;\theta), \tag{1}$$

Where θ is neural network parameters, and x_i is training samples. The representation of the human body $rb(x_i; \theta)$ is classified using a layer of fully linked neural networks that have been installed.

It is possible to train the extra neural network by lowering the category cross-entropy loss. This must be done before the network is normalized by the "Softmax" layer. Fig . 5 presents an overview of the architecture of the MoveNet based ANN. In the first step, human activity frames are sent into MoveNet so that crucial points may be extracted. Afterwards, the coordinates of skeleton points are shown and used to represent

them in the feature space. In the final step, the human skeleton's essential points are used to train the network.

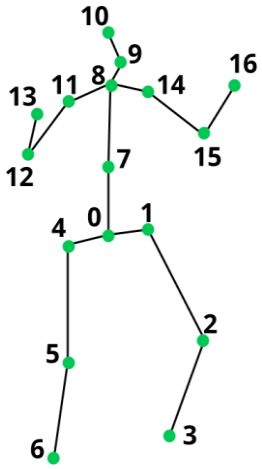


Fig. 4. Extracted points by MoveNet.

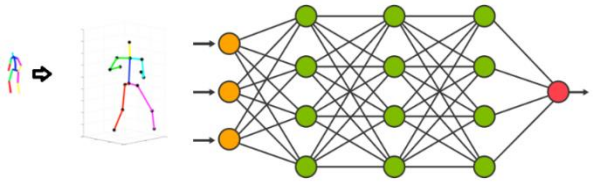


Fig. 5. ANN for MoveNet based physical bullying detection.

As a result, in the first phase of the research, we gather the required data, organized and split it into classes, and afterwards we built a dataset that will be fed into the neural network. The use of MoveNet for the purpose of extracting human skeleton points constitutes the second stage of the study. In order for a neural network to be able to distinguish human activities, we used human skeleton points in the training process. The development of a neural network for the detection of violent actions is the final process of the proposed framework. After that, training and testing the results of the neural network are carried out in order to determine whether or not the proposed approach is suitable for use in the real world.

B. Evaluation

Displaying the outcomes of a prediction model with the help of a confusion matrix is possible. Actual variables are defined by the columns of the confusion matrix, whereas anticipated classes are represented by the rows of the matrix. The matrix displays the number of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) for each class. A number of other efficiency measures, including as accuracy, precision, recall, and F1-score, may be computed with the use of the matrix. Formulae like as precision, recall, F-measure, and accuracy are used in order to assess the outcomes of the suggested methodology, and Eq. (2)-(5) provide an illustration of these respective Eq. [27-29].

$$precision = \frac{TP}{TP + FP}, \tag{2}$$

$$recall = \frac{TP}{TP + FN}, \tag{3}$$

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall}, \tag{4}$$

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP}, \tag{5}$$

We employed a technique called weight-averaging to combine the metrics that were generated for each class into a single variable. This variable weights the values based on the proportion of the class that they represent. In order to validate the prediction models, we resorted to the tried-and-true train/test split. The data set was separated into eighty and twenty percent halves.

VI. Experimental Results

In this part, we provide the research results into the categories of data collecting, feature extraction, and the identification of violent behavior. First subsection depicts human skeleton points' extraction findings; second portion exhibits violent activities detection results. At the conclusion of the second subsection, we compare the achieved findings with the study results that are now considered cutting edge. The research outcome is discussed with the use of evaluation metrics such as confusion matrices, model accuracy, precision, recall, and F1-score.

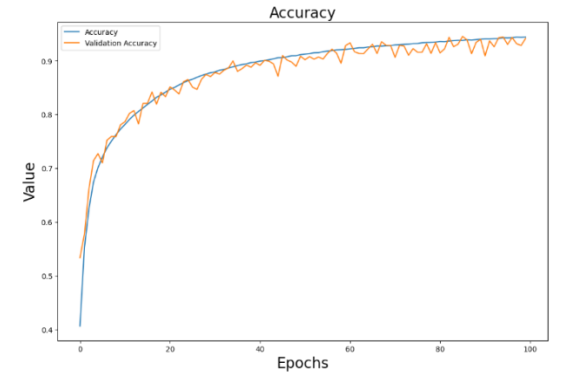
A. MoveNet-based Keypoints Detection

In this part of the article, we retrieved points of the human skeleton from the video sequence. The PoseNET model was used to determine the 17 most important locations. Fig. 6 is a demonstration of human skeletal points that have been retrieved from a live video frame. The extraction of human essential points was performed in a time span of every using a frame as a shot. Due to the rapid nature of the changes that occur in a video feed, the relative positions of the combatants may be immediately adjusted in the event of a conflict. As a direct consequence, there may be many sequenced classes for each participant in the fight. Therefore, the ability to make fast decisions is essential for the identification of violence in videos.

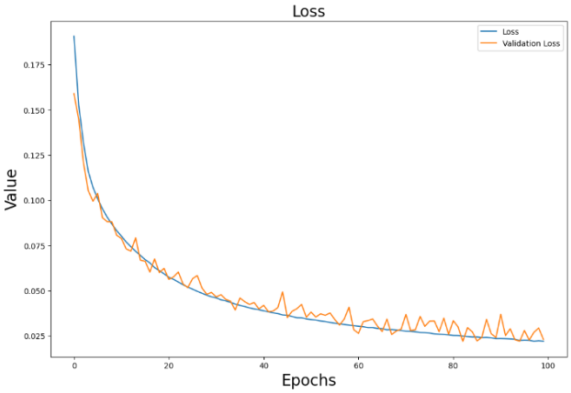
Fig. 6. Testing the proposed framework.

B. Detection of Violent Actions

Throughout the entire period of our experiment, we worked on developing and testing a neural network for violence detection. In order to train the neural network, the MoveNet architecture was applied. Out of the tagged video data, we used it to choose 13 classes for which we then captured further video data. When it came to recognizing aggressive behavior, the constructed action recognition model based on the gathered data performed very well.



(a) Validatoin and test accuracy



(b) Validation and test loss

Fig. 7. Model testing.

The results of the evaluation of the suggested model are shown in Fig. 7. Fig. 7(a) depicts the validation and testing accuracy of the system for the identification of physical bullying throughout the course of eight training epochs. According to the data, the accuracy reaches 98 percent after 8 training epochs have been completed. The values of the neural network loss function are shown in Fig. 7(b) during the course of eight training epochs. According to the data, the amount of validation that is lost is very little even during the beginning stages of the training process.

Fig. 8 depicts the evaluation of the outcomes of classification for a total of 13 different classes. As it can be seen from the graph, every one of the criteria for the evaluation is of an exceptionally high standard. For instance, the accuracy can range anywhere from 0.92 to 0.98, the recall can go anywhere from 0.89 to 1.0, and the F-measure may go anywhere from 0.92 to 0.99. The confusion matrix for the 13 various categories of aggressive behavior are shown in Fig. 9. According to the confusion matrix, the rate of categorization is quite high, and there is a slight misunderstanding between the classifications.

Fig. 8. Confusion matrix of different classes' percentage.

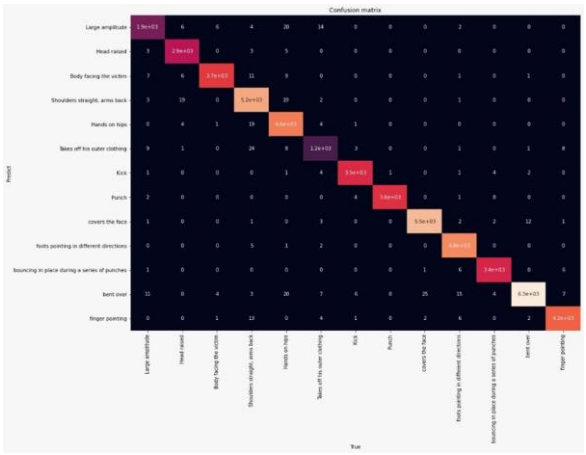


Fig. 9. Confusion matrix for classes.

Fig. 10 illustrates how the proposed framework may be used in a scenario including group fight. In the end, we identify each person's behavior, categorize them, and decide in real time whether they are an aggressor or a victim based on their location, kind of action, and whether they are the bully or the victim. This kind of display of the findings may be helpful for video operators, as it enables them to identify fighting and other forms of physical bullying in real time and to swiftly recognize the attacker and the sufferer in both busy and uncrowded scenes of violence.

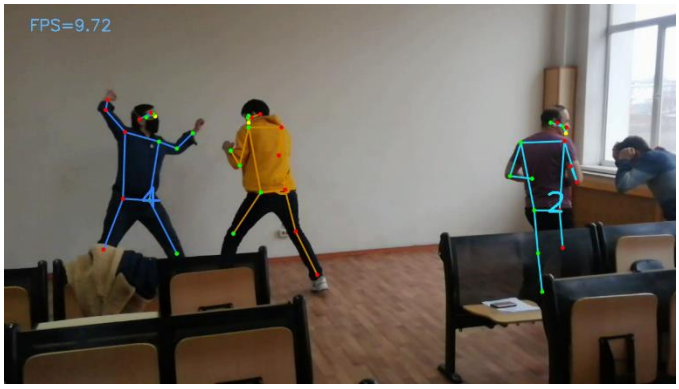


Fig. 10. Model testing.

Table II draws a comparison between the acquired results and the most recent study findings. We analyzed the numerous studies on the detection of physical aggression in the context of three primary assessment criteria: precision, recall, and f-measure. However, the recall and F-score assessment criteria are not used in the majority of the investigations. In situations like this, the Accuracy metric is the most important assessment measure to use when comparing the overall performance of the many recommended methods. In addition, the majority of studies do not include the amount of time spent processing their methods since doing so would be difficult due to disparities in the datasets used and the capabilities of the computer equipment.

TABLE II. COMPARISON OF THE ACHIEVED RESULTS WITH THE OTHER STUDIES

Study	Approach	Precision	Recall	F-score
The proposed approach	**MoveNet based physical bullying detection**	**0.94**	**0.93**	**0.93**
Fenil et. al., 2019, [11]	Bidirectional LSTM	0.94	-	-
Senst et. al., 2017, [12]	Scale-Sensitive Video-Level Representation	0.91-0.94	-	-
Zhang et.al., 2016, [13]	Linear SVM	0.82-0.89	-	-
Sharma & Baghel, 2020 [19]	ResNet-50 and ConvLSTM	0.924	-	-
Cheng et. al., 2020 [30]	Flow Gated Network	0.8725	-	-
Carneiro et. al., 2019 [31]	Multi-Stream CNN	0.8910	-	-
AlDahoul et. al., 2021 [32]	CNN-LSTM based model	0.7335	0.7690	0.7401
Deepak et. al., 2020 [33]	Gradients based violence detection	0.91	0.88	0.88

The findings demonstrate that the suggested method is capable of being used in real-world implementations for the identification of violent behavior by means of security camera footage. The suggested method is more rapid than the model that relies just on pictures due to the use of skeleton points during the training and testing of the neural network. In addition, the developed system will be useful in a variety of settings, including educational institutions and other locations that have video security cameras installed.

VII. CONCLUSION

This study established a physical violence detection based on MoveNet model, which can be employed in real-time and does not need highly processing hardware. The following is the primary benefit offered by the system that has been suggested. To begin with, it is not necessary to provide the system on a regular basis with huge amounts of video footage and photos. Our proposed technology is able to complete tasks more quickly than other systems on the market since it is based on the MoveNet key points of the human skeleton. In this particular instance, this characteristic enables the suggested system to be used for applications that take place in real time and in the actual environment.

The proposed study aims to develop methods for the rapid and accurate identification of violent acts in real time by using video security cameras. In order to accomplish this objective, we would like to provide the following three proposals: Determine the different sorts of violent acts that are shown in a video stream that include both of these categories of violent acts. The first section of the data set consists of the aggressive behavior of a single individual that extends over the course of more than 400 hours. The violent acts committed by a single individual were separated into 13 categories, and the films that were included in the dataset were recorded from a variety of perspectives and acquired using a variety of tools. The second section of the dataset consists of violent acts committed in crowded scenes. The neural network is trained using violent behaviors performed by a single individual, while the suggested system is tested using violent actions performed by a group of bullies.

In order to save time, we employed the MoveNet model to extract skeleton points in order to retrieve an artificial neural network using skeleton key points rather than high-volume video. Using a time interval of one second, skeleton points were retrieved from each frame of the movie. Since human key points are being used, there is no need to load an extremely large number of video frames or photos. The findings of the experiment demonstrate an accuracy of between 95 and 99 percent in the identification of violence based on video; consequently, it is safe to assume that the suggested method is suitable for usage in real-world settings.

REFERENCES

[1] R. Philpot, L. Liebst, K. Møller, M. Lindegaard and M. Levine. “Capturing violence in the night-time economy: A review of established and emerging methodologies,” Aggression and violent behavior, vol. 46, no. 1, pp. 56-65, 2019.

[2] A. Ross, S. Banerjee and A. Chowdhury. “Security in smart cities: A brief review of digital forensic schemes for biometric data,” Pattern Recognition Letters, vol. 138, no. 1, pp. 346-354, 2020.

[3] D. Sultan, A. Toktarova, A. Zhumadillayeva, S. Aldeshov, S. Mussiraliyeva et al., "Cyberbullying-related hate speech detection using shallow-to-deep learning," Computers, Materials & Continua, vol. 74, no.1, pp. 2115–2131, 2023.

[4] G. Sreenu and M. Durai. “Intelligent video surveillance: a review through deep learning techniques for crowd analysis,” Journal of Big Data, vol. 6, no. 1, pp. 1-27, 2019.

[5] P. Vennam, T. Pramod, B. Thippeswamy, Y. Kim and P. Kumar. “Attacks and preventive measures on video surveillance systems: a review,” Applied Sciences, vol. 11, no. 12, pp. 5571, 2021.

[6] R. Nawaratne, D. Alahakoon, D. De Silva and X. Yu. “Spatiotemporal anomaly detection using deep learning for real-time video surveillance,” IEEE Transactions on Industrial Informatics, vol. 16, no. 1, pp. 393-402, 2019.

[7] Sultan, D., Omarov, B., Kozhamkulova, Z., Kazbekova, G., Alimzhanova, L., Dautbayeva, A., ... & Abdrakhmanov, R. (2023). A Review of Machine Learning Techniques in Cyberbullying Detection. CMC-COMPUTERS MATERIALS & CONTINUA, 74(3), 5625-5640.

[8] Narynov, S., Mukhtarkhanuly, D., & Omarov, B. (2020). Dataset of depressive posts in Russian language collected from social media. Data in brief, 29, 105195.

[9] Anand, M., Sahay, K. B., Ahmed, M. A., Sultan, D., Chandan, R. R., & Singh, B. (2022). Deep learning and natural language processing in computation for offensive language detection in online social networks by feature selection and ensemble classification techniques. Theoretical Computer Science.

[10] Z. Shao, J. Cai and Z. Wang. "Smart monitoring cameras driven intelligent processing to big surveillance video data," IEEE Transactions on Big Data, vol. 4, no. 1, pp. 105-116, 2017.

[11] E. Fenil, G. Manogaran, G. Vivekananda, T. Thanjaivadivel, S. Jeeva et al. "Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM," Computer Networks, vol. 151, pp. 191-200, 2019.

[12] T. Senst, V. Eiselein, A. Kuhn and T. Sikora. "Crowd violence detection using global motion-compensated lagrangian features and scale-sensitive video-level representation," IEEE transactions on information forensics and security, vol. 12, no. 12, pp. 2945-2956, 2017.

[13] Omarov, B., Narynov, S., Zhumanov, Z., Gumar, A., & Khassanova, M. (2022). A Skeleton-based Approach for Campus Violence Detection. COMPUTERS MATERIALS & CONTINUA, 72(1), 315-331.

[14] Anand, M., Sahay, K. B., Ahmed, M. A., Sultan, D., Chandan, R. R., & Singh, B. (2022). Deep learning and natural language processing in computation for offensive language detection in online social networks by feature selection and ensemble classification techniques. Theoretical Computer Science.

[15] K. Lloyd, P. Rosin, D. Marshall and S. Moore, "Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures," Machine Vision and Applications, vol. 28, no. 1, pp.361-371, 2017.

[16] P. Bilinski and F. Bremond, "Human violence recognition and detection in surveillance videos," In 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Colorado Springs, CO, pp. 30-36, 2016.

[17] M. Karim, M. Razin, N. Ahmed, M. Shopon and T. Alam. "An Automatic Violence Detection Technique Using 3D Convolutional Neural Network," Sustainable Communication Networks and Application, vol. 55, no. 1, pp. 17-28, 2021.

[18] A. Naik and M. Gopalakrishna. "Deep-violence: individual person violent activity detection in video," Multimedia Tools and Applications, vol. 80, no. 12, pp. 18365-18380, 2021.

[19] M. Sharma and R. Baghel. "Video surveillance for violence detection using deep learning," In Advances in Data Science and Management, ICDSM 2019, pp. 411-420, Singapore, 2020.

[20] M. Asad, J. Yang, J. He, P. Shamsolmoali and X. He. "Multi-frame feature-fusion-based model for violence detection," The Visual Computer, vol. 37, no. 6, pp. 1415-1431, 2021.

[21] Y. Chong and Y. Tay, "Abnormal event detection in videos using spatiotemporal autoencoder," Lecture Notes in Computer Science, vol. 10262, no. 1, pp.189196, 2017.

[22] B. Schmidt and L. Wang, "Automatic work objects calibration via a global–local camera system Robot," Computer-integrated manufacturing, vol. 30, no. 1, pp. 678-683, 2014.

[23] A. Newell, K. Yang and J. Deng. "Stacked hourglass networks for human pose estimation," in European conference on computer vision, ECCV 2016, Amsterdam, The Netherlands, pp. 483-499, 2016.

[24] K. Shrikhande, I., White, D. Wonglumsom, S. Gemelos, M. Rogge et al. "HORNET: A packet-over-WDM multiple access metropolitan area ring network," IEEE Journal on Selected Areas in Communications, vol. 18, no. 10, pp. 2004-2016, 2000.

[25] A. Zanchettin, N. Ceriani, P. Rocco, H. Ding and B. Matthias. "Safety in Human-Robot Collaborative Manufacturing Environments: metrics and Control," IEEE Transactions on Automation Science and Engineering, vol. 13, no. 1, pp. 882-893, 2016.

[26] Kreiss, Sven, Lorenzo Bertoni, and Alexandre Alahi. "Pifpaf: Composite fields for human pose estimation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[27] B. Omarov, N. Saparkhojayev, S. Shekerbekova, O. Akhmetova, M. Sakypbekova et al. "Artificial intelligence in medicine: real time electronic stethoscope for heart diseases detection," CMC-Computers, Materials & Continua, vol. 70, no. 2, pp. 2815-2833, 2022.

[28] Kreuzberger, Dominik, Niklas Kühl, and Sebastian Hirschl. "Machine learning operations (mlops): Overview, definition, and architecture." IEEE Access (2023).

[29] Méndez, Manuel, Mercedes G. Merayo, and Manuel Núñez. "Machine learning algorithms to forecast air quality: a survey." Artificial Intelligence Review (2023): 1-36.

[30] M. Cheng, K. Cai and M. Li, "Rwf-2000: An open large scale video database for violence detection," in 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, pp. 4183-4190, 2021.

[31] Carneiro, S.A., da Silva, G.P., Guimaraes, S.J.F., Pedrini, H.: Fight Detection in video sequences based on multi-stream convolutional neural networks. In: 2019 32nd SIBGRAPI conference on graphics, patterns and images (SIBGRAPI) 2019, pp. 8–15. IEEE

[32] N. AlDahoul, H. Karim, R. Datta, S. Gupta, K. Agrawal et al., "Convolutional Neural Network-Long Short Term Memory based IOT Node for Violence Detection," in 2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, pp. 1-6, 2021.

[33] K. Deepak, L. Vignesh and S. Chandrakala, "Autocorrelation of gradients based violence detection in surveillance videos," ICT Express, vol. 6, no. 3, pp. 155-159, 2020.

[34] C. Duan and X. Li. "Multi-target tracking based on deep sort in traffic scene," Journal of Physics: Conference Series, vol. 1952, no. 2, pp. 022074, 2021.

[35] A. Pramanik, S. Pal, J. Maiti and P. Mitra. "Granulated RCNN and multi-class deep sort for multi-object detection and tracking," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 1, no. 1, pp. 1-11, 2021.