



HAL
open science

Accident Detection System Using Deep Learning

J. Amala Ruby Florence, G. Kirubasri

► **To cite this version:**

J. Amala Ruby Florence, G. Kirubasri. Accident Detection System Using Deep Learning. 5th International Conference on Computational Intelligence in Data Science (ICCIDS), Mar 2022, Virtual, India. pp.301-310, 10.1007/978-3-031-16364-7_23 . hal-04381284

HAL Id: hal-04381284

<https://inria.hal.science/hal-04381284v1>

Submitted on 9 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

ACCIDENT DETECTION SYSTEM USING DEEP LEARNING

¹Amala Ruby Florence J, ²Kirubasri G

¹PG Scholar, Department of Computer Science and Engineering,

²Assistant Professor, Department of Computer Science and Engineering,

^{1,2}Sona College of Technology, Salem, Tamil Nadu, India.

¹amalarubyflorence.20dsm@sonatech.ac.in, ²kirubasri.cse@sonatech.ac.in

Abstract- India has around one and a half lakh persons die due to road accidents per year. One of the main reasons for this number is no timely availability of help. Automatic accident detection can shorten the response time of rescue agencies and vehicles around accidents to improve rescue efficiency and traffic safety level. The ability to detect and track vehicle can be used in applications like monitoring road accidents. The proposed system uses YOLOv5 deep learning algorithm to detect the vehicles from the real time CCTV surveillance video. The primary focus of the system is to build a model that detects various class of vehicle by using custom dataset. The dataset consists of 1000 images with various condition such as rainfall, low visibility, luminosity, and weather conditions. The proposed framework uses YOLOv5 to detect vehicle with improved efficiency in real time object detection system. This model can further be extended to analyse and classify accidents using 3D Convolutional Neural Network based on the severity of the accident and to alert the nearest hospital.

Keywords—*Object identification, YOLOv5, Accident Detection, Convolutional Neural Network, 3D-CNN Algorithm, webapp.*

I. INTRODUCTION

India loses 1.5 lakh lives due to road accidents every year according to recent survey. According to recent statistics provided by Ministry of Road Transport and Highways, India tops the world in road crash deaths and injuries. It has 1 per cent of the world's vehicles but accounts for 11 per cent of all road crash deaths, witnessing 53 road crashes every hour killing 1 person every 4 minutes. The country accounts for about 4.5 lakh road crashes per annum, in which 1.5 lakh people die.

Many proposed approaches for detecting these incidents include traffic flow modelling, vehicle activity analysis, and vehicle interaction modelling. Over-speeding is caught using surveillance cameras and radars. Despite all these precautions, the main cause of

this proportion is a lack of timely assistance. Traffic flow modelling techniques are used in traditional detection approaches. The performance of traffic flow depends on data quality. In addition, these techniques are based on a single type of feature that is unable to match the accuracy and real-time performance requirements of accident detection. Image recognition technology provides high efficiency, flexibility in installation, and minimal maintenance costs for traffic situation awareness.

Automatic accident detection can increase rescue efficiency and traffic safety by reducing the time it takes for rescue agencies and vehicles to respond to an accident. Existing accident detection system suffer from tradeoff between computational overhead and detection accuracy. Some of the early detection techniques, such as YOLO CA, Gaussian mixture model (GMM), Single Shot Multi-Box Detector (SSD), and Convolutional Neural Network (CNN) have similar accuracy and precision compared to other models. Vision Based Detection requires a large storage capacity, and several parameters such as weather conditions and image resolution must be considered to attain accuracy. Vision-based crash detection systems have expanded significantly in recent years as technology progresses.

This system intends to detect accidents using a vehicle collision system in the initial phase. This methodology was proposed to develop an effective model for detecting different classes of vehicles such as cars, motorcycles, buses, and other vehicles. This research focuses on using CCTV surveillance cameras to detect vehicles in real time. The data was gathered from a variety of sources such as YouTube videos and search engines. The model is trained with a custom dataset using the YOLOv5 object detection algorithm that is based on computer vision. As a result, the proposed model is effective at detecting objects in real time under a variety of situations including rainfall, poor visibility, daytime, nighttime.

The motivation of this paper is to shorten the response time of rescue agencies and vehicles around accidents to improve rescue efficiency and traffic safety level. The major objective of this project is to create a model to detect road accidents, analyse and classify the severity of the accident and to build an

alerting system. The model is evaluated using several measures such as precision, recall, mAP, F1 score etc.

II. LITERATURE ANALYSIS

A real-time automatic detection accident detection system. Boxy Vehicle dataset was used for training the model. The images from the dataset were annotated with 2D Bounding box. Custom annotation is made on images for training ResNet152 in YOLOv3. Mini YOLO is used in detection as its efficient and has high recall. Bounding Box and probabilities are obtained for each frame. Soft labeling was done to choose highest probable class. The processed images are tracked using SORT for the status of damage throughout the entire frames. The Velocity and position of each objects are accurately obtained. For each detected frame it is classified in a model as damaged or undamaged. The system is trained on SVM with radial basis kernel to identify the damage status. This model considered only the static motion of accidents. Sequence model was not addressed in this model and field testing was performed [1].

An unsupervised method for detecting traffic accidents in first-person footage. Main innovation is that they used three separate ways to discover anomalies by forecasting the future locations of traffic location and then evaluating the predictive accuracy, performance, and consistency metrics. One of the system's biggest flaws is that it assumes fixed cameras and unchanging backgrounds in videos, which is fine for surveillance but not for vehicle-mounted cameras. Second, it presents the issue as a one-class classification problem, which relies on time-consuming human annotation and only identifies anomaly categories that have been expressly trained. They provide a new dataset of different traffic incidents, An Accident Detection (A3D), as well as another publicly available dataset was used to evaluate this approach [2].

Identifying images of different vehicle crashes, the model uses deep learning methods such as Convolutional Neural Networks. Once training the model with a large amount of visual data is done the model can more accurately classify the severity of the crash by minimizing the classification error. The optimal strategy in generating the model is determined by comparing the accuracy of the model utilizing various Activation functions with the optimizers in CNN. By comparing activation functions with each optimizer, this work focuses on the comparison of various activation functions and optimizers and their effect on the prediction accuracy in classifying [10].

Recently, Zhenbo Lu proposed a framework using residual neural network (ResNet) combined with attention modules was proposed to extract crash-related appearance features from urban traffic videos which were further fed to a spatiotemporal feature fusion model, Conv-LSTM. The model achieved an accuracy of 87.78% on the validation dataset and an acceptable detection speed [12].

Sergio Robles-Serrano proposed DL-based method capable of detecting traffic accidents on video. The model considers that visual components occur in a temporal order to represent traffic collision events. As a result, the model architecture is composed of a visual extraction of features phase followed by a transient pattern identification. Convolution and recurrent layers are used in the training phase to learn visual and temporal features utilizing built-from-scratch and public accessible datasets. In public traffic accident datasets, an accuracy of 98 percent was attained in the detection of accidents, demonstrating a strong capacity in detection irrespective of the road structure [13].

Kyu Beom Lee proposed an Object Detection and Tracking System (ODTS) used in conjunction with the Faster Regional Convolution Neural Network (Faster R-CNN) in this paper for Object Detection and the Conventional Object Tracking algorithm was introduced and implemented for automatic monitoring and detection of unexpected events on CCTVs in tunnels. With a sample of event photos in tunnels, this model has been trained to achieve Average Precision value of 0.85, 0.72, and 0.91 for target items like cars, person, and fire. The ODTS-based Tunnel CCTV Accident Detection System was then tested using four accident recordings, one for each accident, based on a trained deep learning model. As an outcome, the technique can detect all incidents in under ten seconds [7].

In another deep learning-based detection real-time system with the use of a Computer Vision algorithm. The system uses YOLOv4 to perform quicker object detection in real time, and it was tested in a variety of situations, including rainfall, poor visibility, daytime, nighttime and snow using the proposed dataset. In the first phase, the dataset's pre-processing processes are completed. Different types of vehicle photos are recorded after the first phase. The Precision, Mean Average Precision (map) and Average Intersection Over Union (IoU) of the model with which it can recognize the various types of Vehicles were found in the final phase, based on the properties computed or computed in the earlier phases. When we compare the standard results to our results, we can see that the average precision of class Bus is far higher than any other model, that of class Motorcycle is significant compared to almost all other models, that of class Car is greater than YOLOv2's performance, and only the average precision of class Truck is lower than the other models due to various types of trucks labelled in the

dataset [14].

Shilpa Jahagirdar proposed an automatic road accident detection algorithm that use real-time surveillance videos to detect accidents. None of these systems are robust enough to address all accident instances since they do not consider variable lighting circumstances, changing weather conditions, and different traffic patterns. Many of these strategies have been detailed and compared in this study. For Motion detection, Background subtraction, optical flows were utilized. For feature extraction the parameters that were considered were velocity orientation, acceleration, vehicle area, Histogram, Traffic trajectory. For Accident Recognition Machine learning tools, statistical calculation and comparison to thresholds were made. The major challenge in this paper was to differentiate collision and occlusion factor that needs to be addressed particularly and it will help to reduce false accident detection rate [15].

III. PROPOSED APPROACH

A Convolution neural Network (CNN) is a type of artificial neural network used in scene classification, object detection, segmentation, and recognition and processing. CNN is the basic building block of almost all the object detection models. Most of the models like Mask-R CNN, SSD or YOLO are built using convolution neural networks. It is also called as ConvNet. These neural networks are best suited for image, video analysis and also in document recognition, speech synthesis and speech recognition.classification and location algorithms can be divided into two kinds namely Two stage model and One stage model.

A. Two stage model

R-CNN, Fast R-CNN, Faster R-CNN, and Faster R-CNN with FPN are examples of two-stage models. These techniques use selective research and the Region Proposal Network (RPN) to choose approximately 2000 proposal regions in the image, and then use CNN to detect objects based on the attributes retrieved from these regions. Although these region-based models effectively find objects, extracting solutions takes a long time. In R-CNN, they used a selective search to suggest regions for each item in the input image which was extensively time consuming. Fast R-CNN significantly enhanced R-CNN's efficiency but it still utilizes selective search as regional proposal network. Fast R-CNN share the convolutional layers for region proposal.

Faster RCNN employed a fully convolutional neural network for region proposal.

B. One stage model

YOLO (You Only Look Once) and SSD (Single Shot Multi Box Detector) are the example One stage model. These algorithms implement location and classification by one CNN, which can provide end to end detection service. Because of eliminating the process of selecting the proposal regions, these algorithms are very fast and still have guaranteeing accuracy.

YOLOv3 is built on the Darknet Architecture and features 53 layers that have been trained on the ImageNet dataset. Upsampling and residual connections are used in YOLOv3. At three distinct scales, the detection is carried out. Although it is more effective at recognizing smaller things, it takes longer to process than earlier generations.

The CSPDarknet53 backbone, spatial pyramid pooling extra module, PANet path-aggregation neck, and YOLOv3 head make up the YOLOv4 architecture. The backbone of the YOLOv5 model is a Focus structure with a CSP network. The SPP block and PANet make up the neck. It uses GIoU-loss and has a YOLOv3 head.

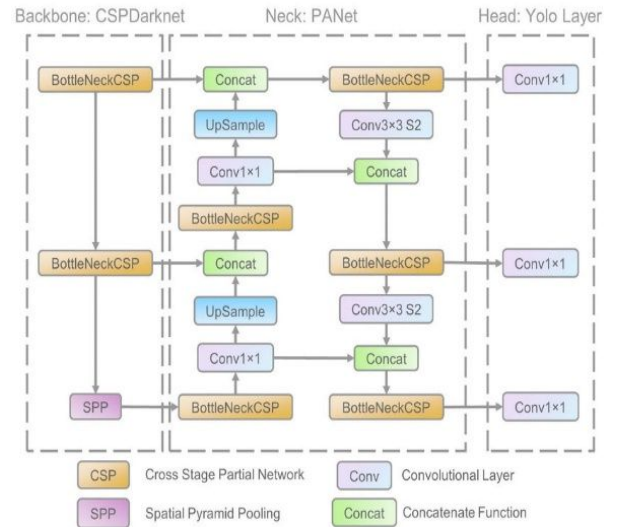


Figure 1. Architecture of YOLOv5 Framework

C. Architecture of YOLOv5

The YOLOv5 model consists of Focus structure and CSP network as the backbone. The neck is composed of SPP block and PANet. It has a YOLOv3 head using GIoU-loss. YOLOv5 is a convolutional neural network (CNN) that detects objects in real-time with great accuracy. This approach uses a single neural network to process the entire picture, then separates it into parts and predicts bounding

boxes and probabilities for each component. These bounding boxes are weighted by the expected probability. It then delivers detected items after non-max suppression.

IV. EXPERIMENTAL DESIGN

All the experiments were conducted on Google Colab which used virtual GPU with 12.72GB Main memory and 68.40 GB Disk space. Video was processed using cv2 and program was written in python 3.8.

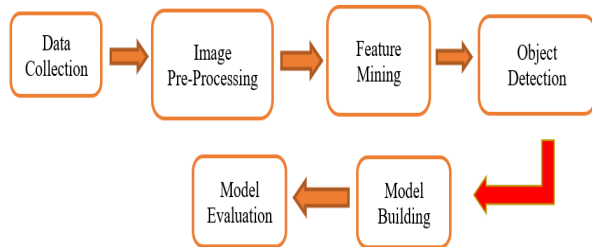


Figure 2. Flow Diagram of proposed model

A. Data Collection

The dataset for this project is a collection of 100 mp4 films containing CCTV footage of various road incidents from around the world, which was gathered from YouTube videos and the CADP dataset. These videos have various types of vehicles such as Car, Bus, Motorcycle, Truck etc. The object was classified into four classes namely Car, Motorcycle, Bus, and others. These videos were shot in various weather and sunlight conditions, such as broad daylight, night, hail, snow, rain etc.

B. Image Pre-Processing

In this Phase the images are annotated based on categories manually using online tool makesense. From the annotated image the labels of bounding box were collected and stored. Once label and respective images were collected the image and label dataset is split in the ratio of 80:20 for training and validation set. A dataset yaml was prepared to contain all the classification types and the path of training and validation set. We then download YOLOv5 darknet framework and the standard weight files for training and testing. Now we create a custom cfg file based on the needs of our training model, which includes four classes: car, bus, motorcycle, and others. So, in our standard YOLOv5 cfg file, we set batch=64, subdivisions=64, width=416, height = 416, max batches = 8000 (number of

classes*2000), steps = 7200 (90 percent of max batches), and set filters = 27(number of classes + 5) *3 in every convolutional layer above yolo layer and set classes = 4 in every yolo layer below convolutional layer. We are now making changes to the YOLOv5 darknet framework's standard Makefile. Set GPU to 1, CUDNN to 1, OPENCV to 1, and CUDNN HALF to 0. Then we first break these videos of our dataset into frames. We do this with the help of cv2. Now for every second of each video we have 5 frames, now we have to annotate these frames and label the various kinds of vehicles traversing on the roads in various weather conditions.

The makesense tool then generates a.txt file for each image including the label of the class along with the coordinates of the x-centre, coordinates of the y-centre, width of the rectangle box, and height of the rectangle box for each of the vehicles indicated in the frame. Now, we divide these.txt and.jpg files in 80:20 ratios for training, testing and save them in the file path images and name them individually.

C. Feature Mining

By using the custom YOLOv5 darknet framework, we train our model for 8000 iterations with 30 epochs. At the end when our model training is completed, we get the 'best and last weights' file which can be trained and tested on our dataset and can be used to detect vehicle in videos.

D. Object Detection

After training our model we use the best weights file that we got from training our model and use our custom cfg file to test the model with the predefined classes of vehicle such as car, motorcycle, bus, others with the predicted accuracy percentage.

E. Model Building

Once the model is trained and tested, for better accuracy we alter the filters and custom cfg file to improve the performance.

F. Model Evaluation:

The accuracy of object detection model is evaluated using performance measures such as Precision, Recall, F1 Score and mAP.

V. PROPOSED ARCHITETURE

The videos are collected from the CCTV surveillance camera and split into frames with dimension 1280 x 720. YOLO v5 expects annotations for each image in form of .txt file where each line of the text file describes a bounding box to define the location of the target object.

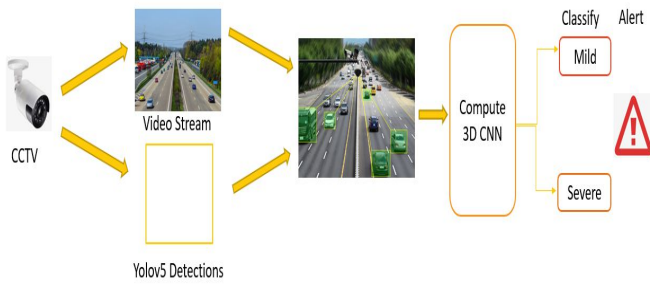


Figure 3. Architectural Design of proposed model

The images are annotated and labeled in the format of ‘class x_center y_center width height’. The model is trained with the custom YOLOv5 darknet framework weights. we get the best and last weights file for our custom dataset that was to used to train the classes of vehicle. The model is tested with best weights file and evaluation of the model is done to validate and to improve the efficiency of vehicle identification.

This will be extended to identify the accident images and classify the accidents as mild and severe using 3D CNN Algorithm. Based on the classification, with the help of webapp we build a mail alerting system.

VI. RESULTS AND DISCUSSION

Table 1. Analysis of Various CNN Models

| S.No. | Method | Average Precision | Average IoU |
|-------|--------------|-------------------|-------------|
| 1 | Fast R-CNN | 80.3 | 0.63 |
| 2 | Faster R-CNN | 86.79 | 0.69 |
| 3 | SSD | 89.65 | 0.71 |
| 4 | YOLOv3 | 91.17 | 0.73 |
| 5 | YOLO-CA | 91.51 | 0.74 |
| 6 | YOLOv5 | 91 | 0.915 |

YOLO performs well when compared to Fast R-CNN. In comparison to v4 and v5, the time spent training in YOLOv3 is considerably lesser. When all factors are considered, v4 has some positive performance characteristics and requires less training time than v5.

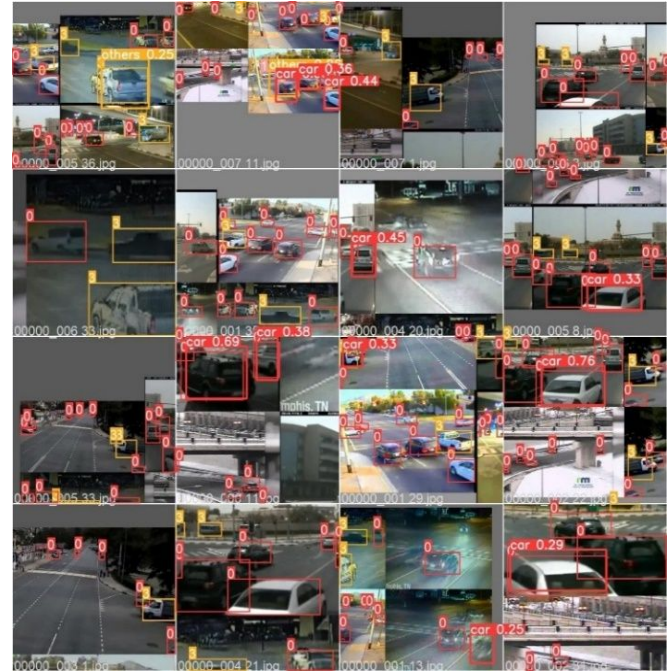


Figure 4. Results of Detected Image Frames

VII. ACCIDENT DETECTION MODULE

The model detects the object and classifies according to its class. By detecting the classes this model can be further extended to detect and classify the accidents based on the severity by using 3D-CNN. Using the 3D volumetric representation of the image, 3D-CNN extracts spatial information from the inputs in three dimensions. It takes advantage of interslice context leading to improved performance. Once the model is trained to classify the accidents based on its severity it can further be employed in real time monitoring and alerting by building a webapp Api to alert nearest hospital with the location details from the mounted surveillance camera along with the severity information to quicken the response time and to improve the traffic safety level

VIII. CONCLUSION

In this we have built a object detection system to identify the various types of vehicles using YOLOv5 from the real time CCTV surveillance video. After training the model we were able to classify the vehicle. We trained the model with 1000 images in future we can increase the dataset to improve the efficiency of the system. By Comparing the results, we can see the positive trend in the precision and accuracy level. YOLOv5 provides great accuracy in detection. As YOLOv5 was tested with larger batch size, it has higher inference speed than most of the detectors. In further step we can build end to end system to identify the accident images and to analyze the severity of it using 3D-CNN Algorithm.

REFERENCES

- [1] Manu s.pillai, Manju khari, “Real-time Image Enhancement for an Automatic Automobile Accident Detection through CCTV Using Deep Learning”, Springer 2021.
- [2] Nancy.P, Dilli Rao.D, Babuaravind.G, Bhanushree.S “Highway Accident Detection And Notification Using Machine Learning”IJCSMC 2020
- [3] Yuanlong Yu, Miaoxing Xu, Jason Gu, “Vision-based traffic accident detection using sparse spatio-temporal features and weighted extreme learning machine”.IEEE 2019.
- [4] Daxin Tian, Chuang Zhang Xuting, Duan, XixianWang, “An Automatic Car Accident Detection Method Based on Cooperative Vehicle Infrastructure Systems”,IEEE 2019.
- [5] Akshit Diwan Vandity, Gupta Chaitanya Chadha“Accident Detection Using Mask R-CNN”. IJMTST 2021.
- [6] Durges Kumar Yadav Renu Ankita Iftisham Anjum,”Accident Detection Using Deep Learning”, IJECCE 2020.
- [7] Kyu Beom Lee, HyuSoungShin , “An application of a deep learning algorithm for automatic detection of unexpected accidents under bad CCTV monitoring conditions tunnels ”. IEEE 2019.
- [8] Singh, Chalavadi, Krishna Mohan , “Deep spatio-temporal representation for detection of road accidents ” IEEE 2019
- [9] Y. Yao, M. Xu, Y. Wang, D. J. Crandall, and E. M. Atkins“Unsupervised traffic accident detection in first-person videos” IEEE 2020
- [10] Saiteja Mothe, Adusumalli Surya Teja, Bh avitha Kakumanu, Ravi.”A Model for Assessing the Nature of Car Crashes using Convolutional Neural Networks”, IEEE 2020.
- [11] Gokul Rajesh, AmithaRossy,Benny“A Deep Learning based Accident Detection System”, ICCSP 2020.
- [12] Zhenbo Lu, Wei Zhou ,Shixiang Zhang, Chen Wang “A New Video-Based Crash Detection Method: Balancing Speed and Accuracy Using a Feature Fusion Deep Learning Framework” IJAT 2020.
- [13] Sergio Robles-Serrano, German Sanchez-Torres, John Branch-Bedoya “Automatic Detection of Traffic Accidents from Video Using Deep Learning Techniques” MDPI 2021.
- [14] Vibhanshu Singh Sindhu “Vehicle Identification from Traffic Video Surveillance Using YOLOv4” IEEE 2021.
- [15] Shilpa Jahagirdar “Automatic Accident Detection Techniques using CCTV Surveillance Videos: Methods, Data sets and Learning Strategies” IJEAT 2020.
- [16] Singh, Chalavadi, Krishna Mohan , “Deep spatio-temporal representation for detection of road accidents ” IEEE 2019.
- [17] Y. Yao, M. Xu, Y. Wang, D. J. Crandall, and E. M. Atkins“Unsupervised traffic accident detection in first-person videos” IEEE 2020.
- [18] Saiteja Mothe, Adusumalli Surya Teja, Bh avitha Kakumanu, Ravi.”A Model for Assessing the Nature of Car Crashes using Convolutional Neural Networks”,IEEE 2020.
- [19] Gokul Rajesh, AmithaRossy,Benny“A Deep Learning based Accident Detection System”, ICCSP 2020.
- [20] Zhenbo Lu, Wei Zhou , Shixiang Zhang, Chen Wang “A New Video-Based Crash Detection Method: Balancing Speed and Accuracy Using a Feature Fusion Deep Learning Framework” IJAT 2020.