ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG721

Improving Online Learning Outcomes: A Novel Approach to Detecting Drowsiness and Sustaining Engagement

Vishvash C[1], Vivek Ganga NarayanRao[2], Monal Digeshwar Bhiwgade[3], Ritendu Bhattacharyya[4], Bharani Kumar Depuru[5]
[1]Research Associate, AISPRY Pvt Ltd, Hyderabad, India.
[2]Research Associate, AISPRY Pvt Ltd, Hyderabad, India.
[3]Research Associate, AISPRY Pvt Ltd, Hyderabad, India.
[4]Mentor, Research and Development, AISPRY Pvt Ltd, Hyderabad, India.
[5]Director, AISPRY Pvt Ltd, Hyderabad, India

Abstract:- In today's digital learning era, a common yet significant challenge emerges: participants often fall asleep during lecture videos, leading to missed concepts and overwhelming frustration as they struggle to find where they left off. This research delves into a solution designed to enhance comprehension and ease the mental stress associated with these interruptions.

Our journey started with data collection from the Internet focusing on Indian and Malaysian contexts we categorized the data into two states active and drowsy, with this dataset in hand we proceeded to a meticulous preprocessing phase, and data augmentation was employed to enhance the diversity of our dataset while image normalization standardized the inputs data balancing was meticulously implemented to provide an unbiased representation of both classes to our model

We then set the stage for intense competition among three advanced models: YOLO-V8, DenseNet201, and ResNet152. Each model underwent a preliminary evaluation over 10 epochs, where YOLO-V8 emerged as the frontrunner with a compelling accuracy of 92% on test data. This promising result spurred us to push further, training YOLO-V8 over two extended phases.

This paper chronicles the path from identifying a widespread issue to developing a solution that enhances educational comprehension and reduces participant stress. By incorporating the YOLO-V8 model into educational platforms, we introduce an innovative method to detect drowsiness and sustain engagement, ensuring every participant remains engaged in the digital classroom.

Keywords:- *Drowsiness Detection, Online Learning Engagement, YOLOv8, MobileNetV2, EfficientNetB0, AI in Digital Classrooms, Student Attention Monitoring.*

I. INTRODUCTION

In the ever-changing world of online education, an unexpected but widespread concern has emerged: pupils falling asleep during lecture videos. Consider this scenario. A determined student eager to learn the nuances of a complicated topic sits down to hear an online lecture. As the minutes pass, their eyelids become heavier, and despite their best attempts, they fall asleep. When they wake up, they are bewildered, having missed important parts of the presentation, leaving them feeling confused and dissatisfied. This situation is more than just an annoyance; it is a serious impediment to effective learning and comprehension.

The project methodology followed here is the open source CRISP-ML(Q) methodology from 360DigiTMG(ak.1) [Fig.1] where CRISP-ML(Q) [1] stands for CRoss Industry Standard Practice for Machine Learning with Quality assurance. CRISP-ML(Q) can broadly be defined as a methodology designed to deal with a Machine Learning solution's project lifecycle [13].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG721

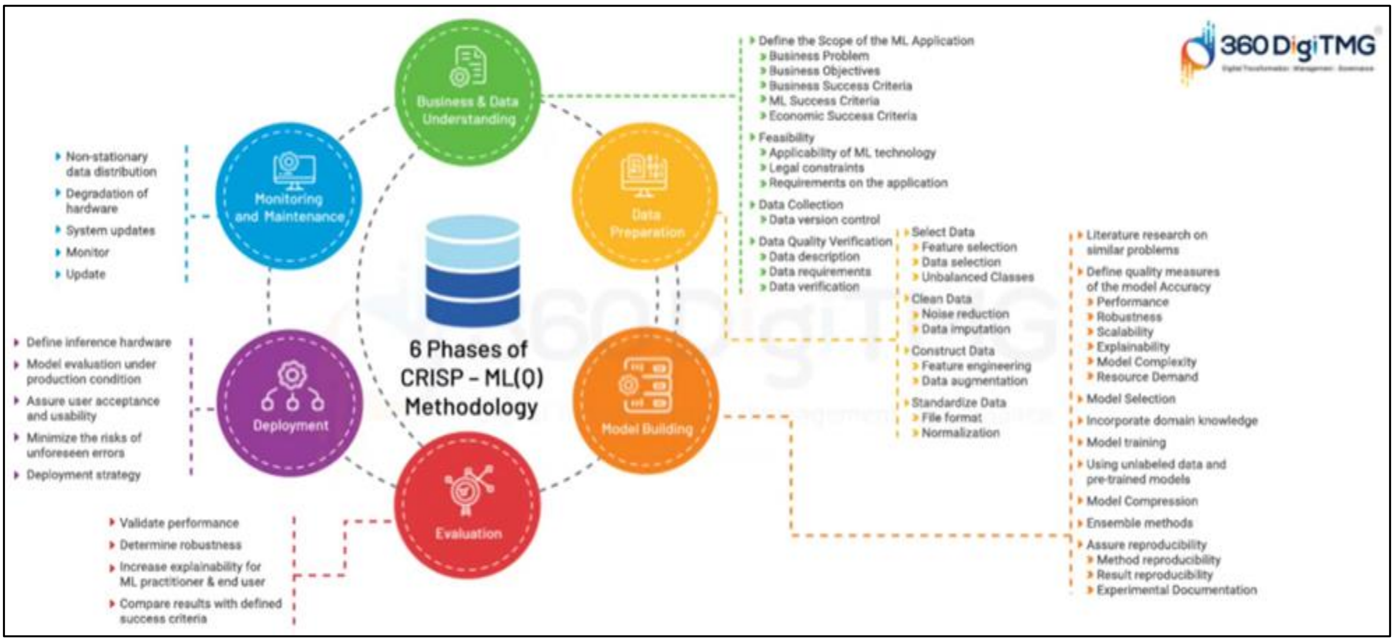

Fig 1 CRISP-ML (Q) Methodological Framework, outlining its key components and steps visually. (Source:-Mind Map - 360DigiTMG)

The conventional classroom provides an atmosphere rich in engagement, interaction with others, and the in-person presence of a teacher, every one of which assists pupils remain alert. However, whereas the digital classroom provides exceptional ease and flexibility, it frequently lacks these features, resulting in distractions and, all too often, unintended sleep. The repercussions are clear: lost content, fragmented knowledge, and stress as students try to catch up on what they've missed.

Recognizing this prevalent issue, we set out on a quest to discover a solution that would allow online learners to keep their concentration and efficiently understand the information. Our journey began with an in-depth look at the unique behaviors and issues that students encounter, with a special emphasis on students in India and Malaysia. These locations, with their various educational landscapes, provided a rich framework for our research, allowing us to collect insights and data that would serve as the cornerstone of our study.

Our approach utilizes the power of machine learning, following a structured methodology akin to CRISP-ML(Q), ensuring systematic and effective problem-solving from various perspectives. This methodical framework encompasses critical stages: understanding business requirements, interpreting data, preparing data, modeling, evaluating outcomes, and deploying solutions. Adhering to this structured approach guided us through each phase of our study, from problem recognition to the delivery of a robust solution. Initially, our focus was on data collection. We meticulously sourced a diverse dataset from online platforms, comprising numerous instances of children exhibiting both attentiveness and drowsiness. This dataset authentically mirrored real-world scenarios and served as the foundation of our research.

Simply having raw data isn't enough; it needs thorough processing to ensure our models can learn effectively. This necessity brought us to the preprocessing stage, where we applied various techniques to enhance the quality and usefulness of the data. By augmenting the data, we expanded our dataset, introducing variations that simulate different learning conditions. Image normalization helped standardize inputs, ensuring they were consistent and reliable. We also balanced the dataset to avoid bias, making sure active and drowsy states were equally represented.

We utilized an enhanced dataset to develop a machine learning model aimed at detecting signs of drowsiness in online students this innovative system serves as an early alert mechanism subtly notifying students to maintain attentiveness during their classes our objective was to introduce a technological solution that enhances the educational environment and encourages sustained engagement

In the development of our drowsy detection system for online classes, we conducted a thorough evaluation of several cutting-edge deep learning models, each with distinct characteristics. After rigorous testing of YOLO-V8, DenseNet201, ResNet152, MobileNet V2, and EfficientNetB0, YOLO-V8 emerged as the standout performer due to its exceptional accuracy and reliability. Encouraged by its initial performance, we proceeded to further refine and train the YOLO-V8 model, resulting in significant enhancements in accuracy and validation metrics

This article chronicles our journey from identifying the problem to finding answers to the problems college students need to sleep well. We are committed to improving skills for our research, reducing stress, also making learning more productive, and demonstrating that a generation is capable of constructing collaboration that provides pliability.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG721

II. METHODS AND TECHNOLOGY

➢ *Data Dimension*

Table 1 *Data Dimension*

Number of Images	1,05,500
Number of Videos	28
Number of Images from videos (2fps)	4500
Total number of Images	**1,10,000**

➢ *Model Architecture*

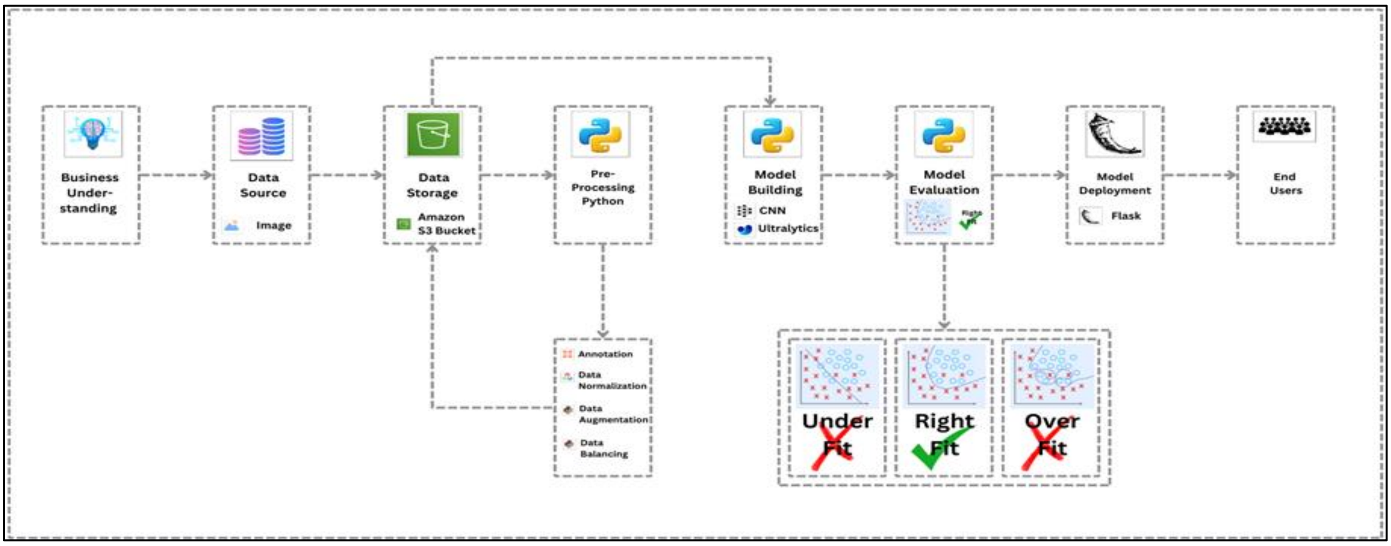

Fig 2 Architecture Diagram: Explain the workflow of the drowsiness detection module of AiTutor (Source: ML Workflow - 360DigiTMG)

The architecture in the form of a flowchart [Fig.2] shows the working of drowsiness detection models covering an entire circulation from problem identification to end-user deployment each stage of the process is designed to ensure success and the use of ML models by providing a clear path from the initial idea to the implementation strategy.

The total basis of the project is based on the acknowledgment of meeting the needs of the business. The goal here is to decide what has to be done and in this process, the group can escalate the success and impact of the result.

Once the problem is well understood the next phase is to gather the required data for this project the primary data source consists of images these images serve as the raw data which will be processed and analyzed throughout the project the standard and relevance of the data collected are critical as they directly influence the effectiveness of the machine learning model developed in later stages.

Once the required data is congregated it is stored securely and efficiently in an Amazon S3 bucket this cloud-based solution [2] provides easy and reliable storage making data simply accessible for further work and analysis can be simply accessed for processing and review.

The initial phase is to prepare the collected data for modeling this is done by using Python to perform various preprocessing tasks including annotation profile normalization profile enhancement and profile evaluation annotation involves tagging images with relevant information while data normalization measures data against a standard quantity data management improves the dataset through various changes and data analysis ensures that the dataset is evenly distributed among different groups this preliminary step is important for the development of a good model and a good study.

In the orienting phase, the dataset is utilized to construct an ML model whereas Python is utilized to build models using methods such as CNN [3] and tools like Ultralytics which are specifically functional for image processing. This phase intends to construct a strong model that could be used to work more precisely and classify material

In developing a drowsiness detection system for online classes testing is important to evaluate the model's effectiveness this step determines if the model is an appropriate fit overfit or underfit. Upon completing this assessment, the model proceeds to the deployment phase where it is implemented using Flask [4]. This web framework simplifies creating web applications allowing

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG721

end users to interact with the model through an intuitive API.

Finally, the end-user step marks the end of the process by which the model is tailored to the user's requirement. This step is mandatory because it ensures the product not only works but also benefits the user by resolving identified business hassle.

- *Methodology*

Our quest to tackle the challenge of scholars falling asleep during online lectures was driven by the crisp-ML(Q) methodology. This structured framework provided a roadmap that ensured every aspect of our research from data collection to model deployment was methodical and purposeful, aiming ultimately to enhance educational comprehension and alleviate learner frustration.

- *Data Collection*

Our adventure commenced with an extensive exploration of online assets focusing specifically on educational contexts in India and Malaysia. This deliberate desire allowed us to seize a huge spectrum of learning environments and pupil behaviors vital for building a comprehensive dataset. As seen in [Fig.3], We collected 65,000 Active images and 45,000 Drowsy images with Indian and Malaysian faces. Our dataset was meticulously curated to encompass two number-one state participants actively engaged in mastering lively and those showing signs of drowsiness. This initial phase laid the groundwork for know-how of the nuanced demanding situations confronted utilizing students in numerous online mastering settings.

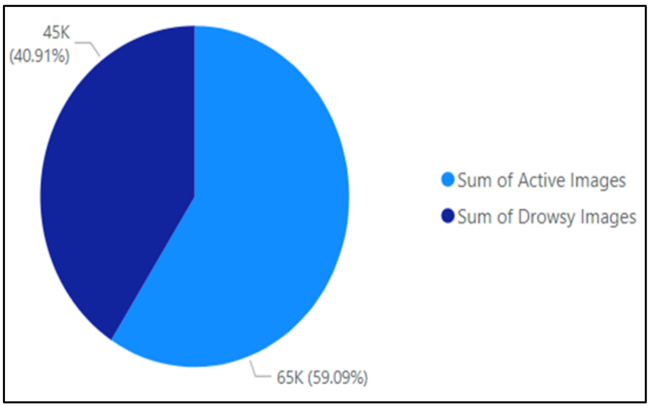

Fig 3 Data Collection for Active and Drowsy Images

- *Data Preparation*

The dataset we collected was extensive but exhibited an imbalance containing 45000 images of active subjects and 65000 images of drowsy subjects balancing is also an important phase of the process without balancing data the model could have become influenced toward the overrepresented class leading to skewed prediction and reduced accuracy in real-world scenarios recognizing the significance of high-quality data for our models we undertook essential data preparation steps one such step involved data augmentation a technique is known to enhance the performance and robustness of machine learning model in computer vision projects specifically we applied methods such as 2-degree clockwise flipping horizontal rotation and 2x scaling this augmentation enriched our dataset making the need for additional data collection unnecessary.

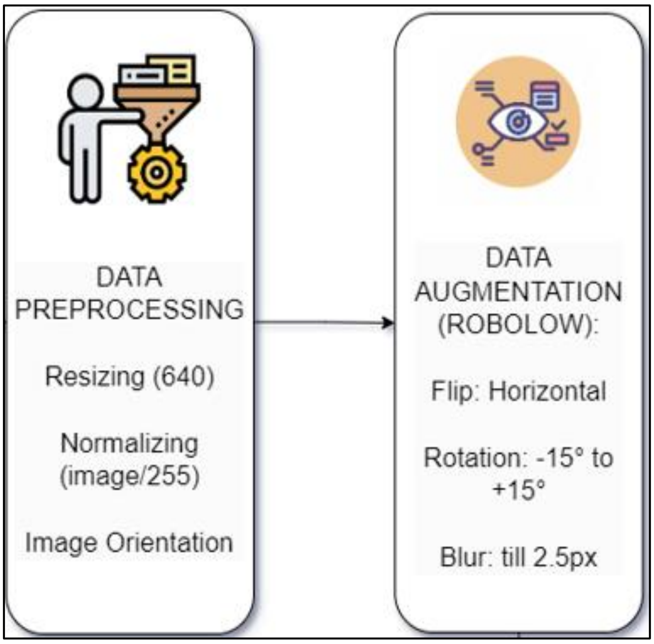

Fig 4 Data Preprocessing and Data Augmentation

Moreover, we applied picture normalization to standardize pixel value throughout the dataset promoting stability and improvement. The learning efficiency of our models to save potential biases we cautiously balanced the dataset by way of manner of ensuring identical illustrations of active and drowsy lessons this planned method bolstered the robustness of our models and also ensured equity in our predictions. We used data augmentation as seen in [Fig.4] and balanced the data with 1,00,000 images in each class as seen in [Fig.5].

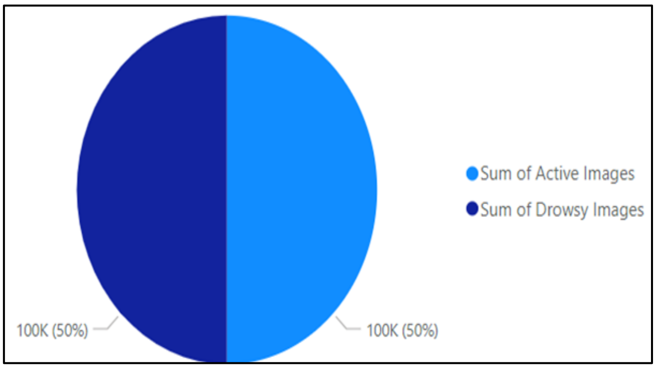

Fig 5 Balanced data of Active and Drowsy Images after Data Augmentation

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG721

➢ *Model Building*

In the model-building phase, we explored five distinct models: YOLO-V8, DenseNet201, ResNet152, MobileNet V2, and EfficientNetB0 selected by their distinctive ability and relevance concerning image classification in our context of identifying drowsiness in virtual class students.

- *DenseNet201:*

DenseNet201 [6] is a deep convolutional neural network known for its densely linked layers. It connects each layer with all other layers in a feed-ahead style, permitting characteristic reuse and efficient learning. DenseNet201 excels in getting fine-grained details from pictures, making it appropriate for obligations requiring particular photo classification and feature extraction. However, its architectural complexity and computational demands had been balanced in opposition to further models apt for real-time processing and on-the-spot feedback.

- *ResNet152:*

ResNet152 represents an advanced extension of the ResNet structure [7], prominent via its residual studying strategies and bypass connections. Those connections help alleviate the diminishing gradient hassle during the training, facilitating efficient deep neural network training. ResNet152 is acclaimed for its precision in picture recognition tasks, in particular in figuring out minute visual information and patterns. However, its significant depth and computational load have been evaluated against the real-time processing needs of our application.

- *MobileNetV2:*

MobileNetV2 is made to cater to cellular smartphone and edge vision packages, emphasizing computative capability, without compromising an immoderate quantity appertaining to accuracy. It makes use of lightweight depth-wise convolutions [8] to lessen the variety of features and operations, making it light and appropriate for deployment on devices with limited computative resources. While it gives an extraordinary overall performance in useful resource-restrained environments, its trade-off between velocity and model accuracy led us to remember greater robust models for our precise real-time detection needs.

- *EfficientNetB0:*

EfficientNetB0 is a cutting-edge neural community architecture (EfficientNets) that grows intensively, with width, and high resolution primarily based mostly on a compounded scale method [9]. It gets fine accuracy with fewer parameters and computations in comparison to standard architectures. Its robust overall performance in numerous picture category tasks, inclusive of face popularity, has been attractive. however, its suitability for our actual-time detection requirement was evaluated against models highly optimized for instant responsiveness and non-forestall tracking.

- *YOLO-V8*

YOLO-V8 is a cutting-edge object detection model [10] known for its real-time processing capabilities and high efficiency. Unlike traditional models that segment an image into grids for detection, YOLO-V8 processes the entire image simultaneously, predicting bounding boxes and class probabilities in one go. This feature makes it ideal for scenarios requiring rapid detection and classification, such as determining if a student in an online class is attentive or drowsy. YOLO-V8's ability to handle real-time video streams and perform robustly in complex environments makes it perfectly suited for monitoring student attentiveness during online learning sessions.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG721

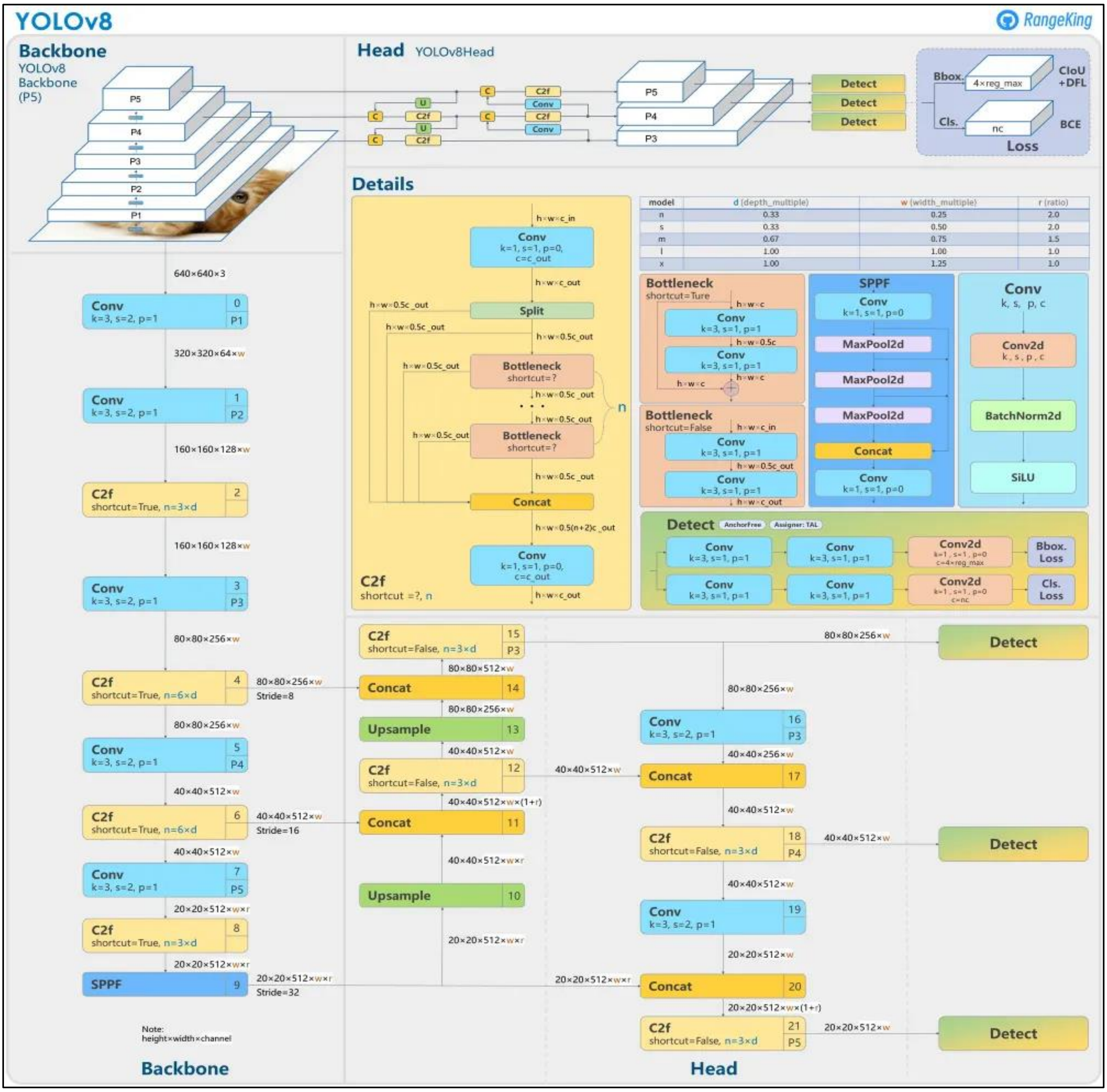

Fig 6 YOLOv8 Architecture, Visualization made by GitHub user RangeKing (Source:-YOLOv8 Architecture)

After comprehensive evaluation and experimentation with those models, YOLO-V8 emerged as the most efficient choice for our drowsy detection device in online instructions. Its ability to process video streams in real-time, correctly come across and classify objects (which include facial states like drowsiness), and manage varying lights and environmental conditions have been pivotal factors. YOLO-V8's streamlined architecture [Fig.6] and efficient inference pipeline ensure timely indicators and interventions to maintain pupil engagement and recognition during online learning classes.

➢ *Model Evaluation*

In this section, we delve into the evaluation of our models, focusing on their performance metrics and how these insights guided us to identify the most effective model for detecting drowsiness in online learning environments.

- *Initial Evaluation*

To assess our models for detecting drowsiness in online studying environments, we focused on both speed and accuracy. This evaluation helped us become aware of the ideal version that excels in these key overall performance areas. via comprehensive testing and analysis, we received valuable insights that guided our selection of the best model for this particular utility.

Models	Epochs	Hyperparameter	Training Accuracy	Validation Accuracy	Testing Accuracy
yolov8	10	lr = 0.01 (default), imgsz=224, cache='disk', batch=256, patience=10	92.40	92.00	91.80
MobileNet V2	30	lr = 0.00001 imgsz=224, cache='disk', batch=256, patience=10	97.66	91.92	91.31
EfficientNetB0	30	lr = 0.0001 imgsz=224, cache='disk', batch=256, patience=10	80.52	80.36	76.78
Resnet 152	10	lr = 0.01 imgsz=224, cache='disk', batch=256, patience=10	87.40	84.30	84.40
Densenet 201	10	lr = 0.01 imgsz=224, cache='disk', batch=256, patience=10	76.90	67.80	72.40

Fig 7 Model Evaluation using Different Pre-Trained Models before the Data Augmentation

From the above results [Fig.7], YOLO-V8 demonstrated a compelling performance with an accuracy of 91.80% on the test data after 10 epochs, indicating its potential for real-time drowsiness detection in online learning environments. However, given the high initial training accuracy of MobileNet V2, further extended training sessions were warranted to explore the model's full potential.

- *Extended Training and Refinement*

Encouraged by way of yolo-v8s preliminary standard overall performance, we prolonged the teaching period for this version throughout tiers of different epochs to further refine its overall accuracy.

Models	Epochs	Hyperparameter	Training Accuracy	Validation Accuracy	Testing Accuracy
yolov8	30	lr = 0.01 (default), imgsz=224, cache='disk', batch=256, patience=10	97.40	96.10	97.20
yolov8	50	lr = 0.01 (default), imgsz=224, cache='disk', batch=256, patience=10	99.10	98.90	98.90

Fig 8 Hyperparameter Tuning for the best Model before the Data Augmentation

Upon extending the training to 30 and 50 epochs as seen in [Fig.8], YOLO-V8 exhibited huge improvements. At 30 epochs, it carried out a testing accuracy of 97.20%, and at 50 epochs[11], it further progressed to an excellent 98.90%. This steady development with extended epochs demonstrated the version's capability to examine and generalize effectively from the data.

Models	Epochs	Hyperparameter	Training Accuracy	Validation Accuracy	Testing Accuracy
yolov8	30	lr = 0.01 (default), imgsz=224, cache='disk', batch=256, patience=10	99.10	99.00	99.00

Fig 9 Hyperparameter Tuning for the best Model after Data Augmentation

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG721

To further validate the robustness of yolo-v8, we conducted additional experiments with augmented data. This extended dataset coupled with the same hyperparameters yielded super outcomes yolo-v8 performed an outstanding accuracy of 99.1% at the training images, 99.0% on the validation data, and 99.0% on the test set all inside just 30 epochs as seen in [Fig.9]. The version demonstrated consistent overall performance with an early stop triggered on the tenth epoch indicating efficient learning and exceptional generalization abilities. These results reinforce the effectiveness of yolo-v8 as the most desirable choice for our drowsiness detection gadget in online studying environments.

- *Best Model Selection*

With the complete assessment, yolo-v8 emerged as the exceptional model for our drowsiness detection gadget. Its excessive accuracy across training, validation, and testing phases underscores its robustness and suitability for real-time software in online studying platforms; the version's potential to address complex atmospheres and provide timely alerts ensures that students can hold attentiveness during their online classes.

Finally, after a thorough evaluation, iterative training of yolo-v8 has ready us with a dependable device to address the typical issue of sleepiness in online studying leading towards progressed academic results and decreased learner frustration.

III. RESULTS AND DISCUSSIONS:

This AI system intends to notify the scholars about their attention throughout their online classes. This system will classify the state of the user and notify them whether they are active or in a drowsy condition. This will ameliorate the awareness of students throughout their lecture and also their engagement with AiTutor.

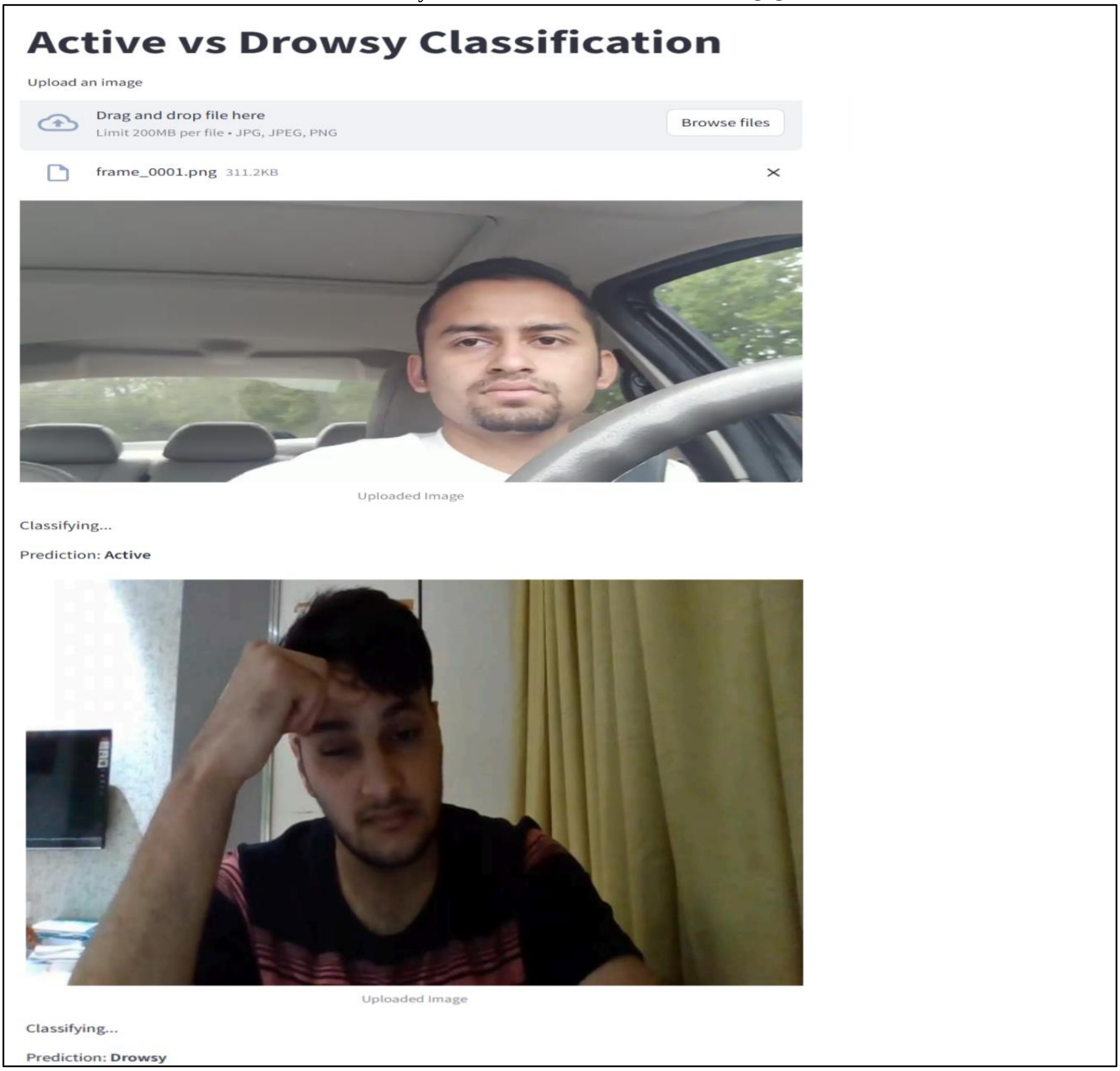

Fig 10 Deployment of Drowsy Detection System

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24AUG721

The model enables users to upload images directly or capture them using a webcam. Once an image is provided, the model processes it to determine whether the depicted state is active or drowsy [Fig.10]. This functionality allows users to easily evaluate the state by simply supplying an image.

For webcam systems, the model continuously monitors the user's state by analyzing their eye movements and facial expressions in real-time. As the webcam captures the live feed, the model processes each frame to detect signs of activity or fatigue. Based on motion detection and analysis, the model provides regular notifications to keep the user informed about their current state, helping them stay alert and take necessary actions if they feel sleepy. This real-time monitoring is particularly useful for applications where constant awareness is crucial, such as driving or operating technology.

IV. CONCLUSION

Inside the circle of online schooling, ensuring students stay alert and engaged during lectures is important. Our studies address this issue by developing a device to come across drowsiness and hold students focused by means of collecting data from students in India and Malaysia. We recognized patterns that imply when students are attentive and when they are turning into drowsy. This progressive solution helps you save the learners from lacking crucial parts of their classes and reduces the pressure of seeking to catch up later.

By integrating this device into online scholastic structures, we aim to create extra productive and interesting learning surroundings. Our research highlights the capability of the usage of smart technology to intensify online education by retaining learners engaged and alert. We can improve their mastering experience and assist them in obtaining better effects. This technique not only addresses common troubles but also paves the way for further progress in online mastering.

Incorporating a drowsiness detection AI model into online job interviews can boost engagement by tracking candidate alertness and concentration. This approach helps ensure that candidates stay focused, which can result in a more accurate assessment of their responses and overall demeanor. The system could also suggest breaks if it detects signs of tiredness, potentially enhancing performance. Nonetheless, it is essential to address privacy issues carefully to ensure that the technology is implemented in a transparent and respectful manner.

FUTURE SCOPE

Future enhancements for this task consist of integrating the system with CCTV in physical school rooms and online learning structures for real-time monitoring. This integration will enable the detection of any malpractice during examinations, which includes candidates using devices, failing to maintain right eye contact, or behaviors like looking down or sideways that might imply trying to find answers on a mobile device.

Advancements in model architectures, such as CNNs and RNNs, will improve accuracy. Extending support to mobile and tablet devices will enhance accessibility. Additional features like interaction rate and blink rate detection will provide a more comprehensive analysis. Implementing real-time alerts, personalized models, multimodal data fusion, and privacy-preserving techniques will further enhance functionality. Emphasis on scalability, user feedback, and continuous improvement will ensure the system's effectiveness across various sectors.

ACKNOWLEDGMENTS

We affirm our use of the CRISP-ML(Q) and ML Workflow which are openly available on the official 360digitmg website with explicit consent from 360digitmg.

REFERENCES

[1]. Inna Kolyshkina and Simeon, Interpretability of Machine Learning Solutions in Public Healthcare: The CRISP-ML Approach, 2021, Volume 4. https://doi.org/10.3389/fdata.2021.660206

[2]. B., Mallikarjuna and D., Arunkumar Reddy, Cloud Storage for Data Sharing: Infrastructure as Service (IaaS) in Cloud Environment (February 7, 2018). 2018 IADS International Conference on Computing, Communications & Data Engineering (CCODE) 7-8 February, Available at SSRN: http://dx.doi.org/10.2139/ssrn.3169039

[3]. Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, Jun Zhou, A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects, Published in IEEE Transactions on Neural Networks and Learning Systems (Volume: 33, Issue: 12, December 2022), Publisher: IEEE, DOI: https://doi.org/10.1109/TNNLS.2021.3084827

[4]. Relan, K. (2019). Deploying Flask Applications. In: Building REST APIs with Flask. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-5022-8_6

[5]. Mingle Xu, Sook Yoon, Alvaro Fuentes, Dong Sun Park, A Comprehensive Survey of Image Augmentation Techniques for Deep Learning, Pattern Recognition, Volume 137, 2023, 109347, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2023.109347

[6]. Adhinata, F., Rakhmadani, D., Wibowo, M., & Jayadi, A. (2021). A Deep Learning Using DenseNet201 to Detect Masked or Non-masked Face. JUITA: Jurnal Informatika, 9(1), 115 - 121. doi: http://dx.doi.org/10.30595/juita.v9i1.9624
[7]. Prabhakaran, A.K., Nair, J.J., Sarath, S. (2021). Thermal Facial Expression Recognition Using Modified ResNet152. In: Thampi, S.M., Gelenbe, E., Atiquzzaman, M., Chaudhary, V., Li, KC. (eds) Advances in Computing and Network Communications. Lecture Notes in Electrical Engineering, vol 736. Springer, Singapore. https://doi.org/10.1007/978-981-33-6987-0_32
[8]. K. Dong, C. Zhou, Y. Ruan, and Y. Li, "MobileNetV2 Model for Image Classification," 2020 2nd International Conference on Information Technology and Computer Application (ITCA), Guangzhou, China, 2020, pp. 476-480, doi: https://doi.org/10.1109/ITCA52113.2020.00106
[9]. Md. Alamin Talukder, Md. Abu Layek, Mohsin Kazi, Md. Ashraf Uddin, Sunil Aryal, Empowering COVID-19 detection: Optimizing performance through fine-tuned EfficientNet deep learning architecture, Volume 168, 2024, 107789, ISSN 0010-4825, doi: https://doi.org/10.1016/j.compbiomed. 2023.107789
[10]. Sohan, M., Sai Ram, T., Rami Reddy, C.V. (2024). A Review on YOLOv8 and Its Advancements. In: Jacob, I.J., Piramuthu, S., Falkowski-Gilski, P. (eds) Data Intelligence and Cognitive Informatics. ICDICI 2023. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-99-7962-2_39
[11]. Shivam Sinha, T.N.Singh, V.K.Singh, A.K.Verma. Epoch Determination for Neural Network by self-organized map (SOM), *Comput Geosci* 14, 199-206 (2010). https://doi.org/10.1007/s10596-009-9143-0