MIDOCEAN UNIVERSITY

FACULTY OF INFORMATICS

## Real-time Object detection of Personal Protective Equipment (PPE) in video Streaming Using YOLO Deep Learning Technique

BY

**Younis Arrabi  
Mohammad Al Johani**

A Thesis Submitted to

Department of Informatics in Partial Fulfillment of the Requirements for the Master Degree of Master of Artificial Intelligence

Under the Supervision of

### Prof. Dr. Yasser Abdelhamid

### Associate Professor, College of Informatics

### Midocean University

## Acknowledgements

Special thanks to Dr. Prof. Dr. Yasser Abdelhamid for his invaluable guidance, encouragement, and insightful feedback that significantly contributed to the completion of this research. his support has been a constant source of inspiration.

Would also like to thank the faculty and staff of informatics college in Midocean university for providing an excellent academic environment and the necessary resources to conduct this research. Special thanks to my colleagues and friends, whose support and stimulating discussions have enriched my research experience.

Deeply indebted to my family for their unwavering support and encouragement. Their patience and understanding have been a great source of strength during this challenging journey, thank you all for your contributions and support.

## Abstract

Magnificent progress in Artificial Intelligence computer vision techniques has been achieved during the last decades. Object detection should benefit from these improvements to discover about how workers are safe in different workplace environments. Because of the continuous increase in the number of researchers who are affiliated with computer vision and apply object detection techniques and deep learning methods, a relatively new discipline called Personal protective equipment auto detection has emerged. It aims to use methods, tools and techniques of deep learning in object detection to create and learn a model that can detect personal protective equipment in real-time and in automatic way.

Ensuring the proper use of Personal Protective Equipment (PPE) is critical in all workplace with accident high hazard settings; to reduce the risk, this project develops a real-time system for detecting Personal Protective Equipment in video streams using You Only Look Once (YOLO) deep learning techniques. The system identifies various Personal Protective Equipment requirements, such as helmets and vests, with high accuracy and minimal latency. By leveraging the efficiency of latest YOLO version, the solution achieves effective real-time performance, making it suitable for integration into existing surveillance infrastructures. This approach offers a scalable method for enhancing workplace safety monitoring and compliance.

Accuracy of detecting Personal Protective Equipment in most studies do not achieve the state-of-the-art accuracy in contrary with deep learning performances on datasets. The objective of this project is to develop a DL-based object detection model for the real-time detection of safety tools, including helmets, vests, gloves, safety shoes and safety goggles using different datasets.it will be designed to detect tools using object detection deep learning.

This can be decomposed into the following sub-objectives:

• Selecting and implementing proper algorithms from the recent literature for detecting objects.

• Evaluating the selected algorithms based on metrics such as accuracy, precision, and recall.

The research work introduced two machine learning models, namely Convolutional Neural Networks (CNNs), and You Only Look Once (YOLO), to detect PPE. The most used deep learning techniques in object detection are You Only Look Once (YOLO) and Convolutional Neural Networks so they must be included in our research. And their performance and advantage and limitation taken into considerations in selecting final Technique.

Our results showed that the YOLO models outperformed CNN Model. The accuracy of YOLO 94.68% and that of CNN was 90.43

The results show that accuracy of YOLO in Object detection reaches the state-of-the-art accuracies on datasets and can be efficiently used for detecting PPE of the workers and in real-time systems to reduce risk of accidents.

## Table of Contents

[List of Abbreviations xiv](#_Toc208473)

[List of Figures xvi](#_Toc208474)

[List of Tables xviii](#_Toc208475)

**Chapter 1: Introduction** ................................................................................ 1

1.1 Problem Definition .................................................................................... 2

1.2 Thesis Objectives ....................................................................................... 3

1.3 Thesis Contribution .................................................................................... 4

1.4 Thesis Organization.................................................................................... 5

### Chapter 2: Overview on Personal protective equipment Detection........... 6

2.1 Introduction ................................................................................................ 6

2.2 Personal protective equipment.................................................................... 6

2.3 Applications of Personal protective equipment.......................................... 7

2.3.1 Detection............................................................................................ 8

2.3.2 Clustering ......................................................................................... 9

2.3.3 Relationship Mining ........................................................................ 10

2.3.4 Discovery with Model ..................................................................... 11

2.3.5 Distillation of Data for Human Judgment ........................................ 12

2.4 Goals of Personal protective equipment Detection................................... 12

2.5 Detection of Personal protective equipment ............................................. 13

2.5.1 Students Previous Grades and Class Performance ........................... 14

2.5.2 Students e-Learning Activity ........................................................... 14

2.5.3 Students Demographics ................................................................... 15

2.5.4 Students Social Information ............................................................ 16

2.5.5 Students Environment ..................................................................... 16

2.5.6 Instructor Attributes......................................................................... 17

2.5.7 Course Attributes ............................................................................. 18

2.5.8 Course Evaluations .......................................................................... 18

2.5.9 Students Experience Information .................................................... 18

2.6 Dataset of Educational Data Mining ......................................................... 20

2.6.1 Student Performance Dataset .......................................................... 20

2.6.2 Educational Process Mining (EPM): A Learning Analytics Dataset 22

2.6.3 Students' Academic Performance Dataset ....................................... 23

2.6.4 Student Performance: Standardized Testing Results ....................... 24

2.6.5 The Open University Learning Analytics Dataset (OULAD) ......... 25

2.7 Performance Measures in Educational Data Mining ................................ 26

2.7.1 Accuracy ......................................................................................... 26

2.7.2 Kappa Statistic ................................................................................. 26

2.7.3 Receiver Operating Characteristic ................................................... 27

2.7.4 Precision and Recall ........................................................................ 28

2.7.5 Mean Absolute Error ....................................................................... 28

2.7.6 Root Mean Square Error .................................................................. 29

2.8 Summary .................................................................................................. 29

**Chapter 3: Related Work** ………………………………………………... 31

3.1 Introduction ..............................................................................................31

3.2 History ..................................................................................................... 31

3.3 Predicting Student Performance Using Statistical Analysis .................... 33

3.4 Predicting Student Performance Using Machine Learning ..................... 35

3.4.1 Predicting Student Performance Using Artificial Neural Networks 36

3.4.2 Predicting Student Performance Using Decision Trees .................. 38

3.4.3 Predicting Student Performance Using Support Vector Machines. 39

3.4.4 Predicting Student Performance Using Logistic Regression ........... 40

3.4.5 Predicting Student Performance Using Naïve Bayes ....................... 40

3.4.6 Predicting Student Performance Using k-Nearest Neighbor ........... 41

3.5 Predicting Student Performance Using Both Statistical Analysis and

Machine Learning ........................................................................................ 42

3.6 Predicting Student Performance Using Deep Learning ........................... 44

3.7 The-State-of-Art ...................................................................................... 45

3.8 Summary ................................................................................................. 47

**Chapter 4: Background on Machine Learning** .......................................... 48

4.1 Introduction .............................................................................................. 48

4.2 Conventional Programming vs. Machine Learning .................................. 48

4.3 Supervised Learning ................................................................................ 51

4.3.1 Bias-variance tradeoff ..................................................................... 53

4.3.2 Bias and Variance Test .................................................................... 54

4.3.3 Hypothesis Space............................................................................. 55

4.4 Unsupervised Learning ............................................................................ 56

4.5 Semi-supervised Learning ........................................................................ 57

4.6 Active Learning ........................................................................................ 59

4.7 Reinforcement Learning .......................................................................... 59

4.7.1 Difference between Reinforcement Learning and Supervised

Learning ................................................................................................... 60

4.7.2 Relation between Reinforcement Learning and Dynamic

Programming ........................................................................................... 61

4.7.3 Component of Reinforcement Learning .......................................... 61

4.7.4 Deep Reinforcement Learning ........................................................ 62

4.8 Artificial Neural Network ........................................................................ 63

4.8.1 Common Structure of Artificial Neural Network ............................ 65

4.8.2 Backpropagation .............................................................................. 67

4.8.3 Types of Artificial Neural Networks ............................................... 67

4.9 Decision Trees and Random Forests ........................................................ 69

4.10 k-Nearest Neighbor ................................................................................ 71

4.11 Logistic Regression ................................................................................ 72

4.12 Support Vector Machine ........................................................................ 73

4.13 Naïve Bayes ........................................................................................... 74

4.14 Deep Learning ....................................................................................... 75

4.15 Summary ............................................................................................... 76

**Chapter 5: Proposed Approach** ................................................................. 78

5.1 Introduction .............................................................................................. 78

5.2 Approach Stages ...................................................................................... 78

5.2.1 Data Acquisition ............................................................................ 79

5.2.2 Data Cleaning ................................................................................. 80

5.2.3 Feature Extraction and Selection ................................................... 81

5.2.3.1 Filter Methods ....................................................................... 82

5.2.3.2 Wrapper Methods .................................................................. 83

5.2.3.3 Hybrid Methods .................................................................... 83

5.2.3.4 Embedded Methods ............................................................... 83

5.2.3.5 The Chosen Feature Selection Methods ................................ 84

5.2.4 Data Encoding and Standardization ................................................ 84

5.2.5 Training and Evaluation of Machine Learning Based Methods ...... 85

5.3 Summary .................................................................................................. 85

### Chapter 6: Experimental Results and Discussion .................................... 86

6.1 Introduction .............................................................................................. 86

6.2 Dataset Description .................................................................................. 87

6.2.1 Demographic Information .............................................................. 89

6.2.2 Activities Information .................................................................... 89

6.2.3 Modules and Presentations Information ......................................... 90

6.3 Preparing Combinations of Data .............................................................. 90

6.4 Dealing with Imbalanced Data ................................................................ 91

6.5 Parameters of ML models ........................................................................ 92

6.6 Performance Measures ............................................................................ 94

6.6.1 Confusion Matrix ............................................................................ 94

6.6.2 Accuracy ....................................................................................... 105

6.6.3 Precision ....................................................................................... 107

6.6.4 Recall or Sensitivity (True Positive Rate) ..................................... 107

6.6.5 F1 Score ......................................................................................... 109

6.6.6 Area Under the Curve and Receiver Operating Characteristic .... 110

6.7 Most Important Features for Predicting Student Performance ............... 120

6.7.1 Feature Selection Using Embedded Methods ............................... 120

6.7.2 Feature Selection Using Filter Methods ....................................... 125

6.8 Summary ................................................................................................ 125

**Chapter 7: Conclusion and Future Work** ................................................ 127

7.1 Important Features ................................................................................. 127

7.2 Most Efficient Models in Educational Data Mining ............................... 129

7.3 Future Work ........................................................................................... 129

References ................................................................................................... 130

# List of Abbreviations

* **Acc.:** Accuracy
* **ANN:** Artificial Neural Network
* **ANOVA:** Analysis of Variance
* **API:** Application Programming Interface
* **CM:** Confusion Matrix
* **CSV:** Comma Separated Values
* **DL:** Deep Learning
* **DM:** Data Mining
* **DT:** Decision Tree
* **EDM:** Educational Data Mining
* **FCBF:** Fast Correlation Based Filter
* **FSFS:** Feature Selection Using Feature Similarity
* **FN:** False Negative
* **FP:** False Positive
* **FPR:** False Positive Rate
* **ITS:** Intelligent Tutoring System
* **KNN:** K-Nearest Neighbor
* **LA:** Learning Analytics
* **LMS:** Learning Management System
* **LR:** Logistic Regression
* **MAD:** Mean Accuracy Decrease / Mean Absolute Difference
* **MAE:** Mean Absolute Error
* **MDI:** Mean Decrease in Impurity
* **MGSACO:** Microarray Gene Selection based on Ant Colony Optimization
* **ML:** Machine Learning
* **MPMR:** Maximum Projection and Minimum Redundancy
* **NB:** Naïve Bayes
* **OULAD:** Open University Learning Analytics Dataset
* **PSP:** Predicting Student Performance
* **RF:** Random Forest
* **RMSE:** Root Mean Square Error
* **ROC:** Receiver Operating Characteristic
* **RRFS:** Relevance Redundancy Feature Selection
* **SMOTE**: Synthetic Minority Oversampling Technique
* **SU**: Symmetrical Uncertainty
* **SVM:** Support Vector Machine
* **TN:** True Negative
* **TP:** True Positive
* **TPR:** True Positive Rate
* **UCI:** University of California, Irvine
* **VLE:** Virtual Learning Environment

# List of Figures

|  |  |
| --- | --- |
| **2.1** Educational Data Mining Process ……………………………………… | 7 |
| **2.2** Applications of Educational Data Mining ……………………………… | 8 |
| **2.3** Distribution of Papers that used different attributes for PSP problem …. | 19 |
| **2.4** Student Performance Data Set …………………………………………... | 21 |
| **2.5** Educational Process Mining (EPM): A Learning Analytics Data Set … | 23 |
|  |  |
| **3.1** Most Commonly Used ML Algorithms in EDM ……………………… | 35 |
|  |  |
| **4.1** Machine Learning Subfields …………………………………………… | 50 |
| **4.2** Supervised Learning ……………………………………………………. | 51 |
| **4.3** Unsupervised Learning …………………………………………………. | 56 |
| **4.4** Reinforcement Learning ………………………………………………... | 59 |
| **4.5** Simple ANN ……………………………………………………………. | 66 |
| **4.6** The relation between performances of deep learning algorithm compared to other traditional machine learning algorithms and the amount of data … | 76 |
|  |  |
| **5.1** The Proposed Approach …………………………….…………………... | 78 |
|  |  |
| **6.1** OULAD Database Schema ……………………………………………... | 88 |
| **6.2** A Bar Chart Shows Imbalanced Data …………………………………… | 91 |
| **6.3** Comparing Precisions …………………………………………………...107  **6.4** Comparing Recalls ………………………………………………………108  **6.5** Comparing F1-Scores ……………………………………………………109  **6.6** ROC curve of the ANN model on demographics without using SMOTE110  **6.7** ROC curve of the ANN model on demographics using SMOTE ………111  **6.8** ROC curve of the ANN model on demographics and scores …………..111  **6.9** ROC curve of the ANN model on demographics, scores, and VLE ……112  **6.10** ROC curve of the RF model on demographics ………………………..112  **6.11** ROC curve of the RF model on demographics and scores ……………113 **6.12** ROC curve of the RF model on demographics, scores, and VLE ………113 **6.13** ROC curve of the k-NN model on demographics ………………………114  **6.14** ROC curve of the k-NN model on demographics and scores …………114 | |

**6.15** ROC curve of the k-NN model on demographics, scores, and VLE ……115

**6.16** ROC curve of the LR model on demographics ………………………… 115

**6.17** ROC curve of the LR model on demographics and scores ……………116

**6.18** ROC curve of the LR model on demographics, scores, and VLE ………116

**6.19** ROC curve of the NB model on demographics …………………………117

**6.20** ROC curve of the NB model on demographics and scores ……………117

**6.21** ROC curve of the NB model on demographics, scores, and VLE ………118 **6.22** ROC curve of the SVM model on demographics ……………………….118

**6.23** ROC curve of the SVM model on demographics and scores …………119

**6.24** ROC curve of the SVM model on demographics, scores, and VLE …….119

**6.25** Comparing AUCs …………………………………………………..…….120 **6.26** MDI of Demographics ……………………………………………...……121

**6.27** MAD of Demographics ………………………………………………….121

**6.28** MDI of Demographics and Scores ……………………………………….122

**6.29** MAD of Demographics and Scores ………………………………………123

**6.30** MDI of Demographics, Scores, and VLE …………………………..……124

**6.31** MAD of Demographics, Scores, and VLE ………………………………...124

# List of Tables

**2.1** Student Performance Data Set Description …………………………21

**2.2** Students' Academic Performance Dataset Description ……………… 23

**2.3** School data file fields Description ………………………………… 25

**3.1** Summary of researches that used Machine Learning Based models … 43

**6.1** Positions of True Positives, False Positives, True Negatives, and False

Negatives in the Confusion Matrix …………………………………… 94

**6.2** CM of the ANN Model on Demographics without Using SMOTE … 96

**6.3** CM of the ANN Model on Demographics Using SMOTE …………. 96 **6.4** CM of the ANN on Demographics and Scores ……………………… 97

**6.5** CM of the ANN on Demographics, Scores. And VLE ……………… 97 **6.6** CM of the RF on Demographics only ……………………………….. 98

**6.7** CM of the RF on Demographics and Scores ………………………… 98

**6.8** CM of the RF on Demographics, Scores and VLE ………………….. 99 **6.9** CM of the k-NN on Demographics ………………………………….. 99

**6.10** CM of the k-NN on Demographics and Scores ……………………… 100

**6.11** CM of the k-NN on Demographics, Scores, and VLE ……………… 100

**6.12** CM of the LR on Demographics …………………………………….. 101 **6.13** CM of the LR on Demographics and Scores ………………………… 101 **6.14** CM of the LR on Demographics, Scores, and VLE …………………. 102

**6.15** CM of the NB on Demographics ……………………………………. 102

**6.16** CM of the NB on Demographics and Scores ………………………... 103

**6.17** CM of the NB on Demographics, Scores, and VLE ………………… 103 **6.18** CM of the SVM on Demographics ………………………………….. 104

**6.19** CM of the SVM on Demographics and Scores ……………………… 104 **6.20** CM of the SVM on Demographics Scores, and VLE ……………….. 105

**6.21** Comparing Accuracies of Models on Different Data Combinations ... 106

**6.22** Important Features Using Filter Methods …………………………… 125

## Chapter 1 Introduction

Safety of the worker during working hours become critical requirement over all kinds of projects, scaling from Small to mega size, companies and governments seeking to have safe workplace environment for the workers, and this have benefit on both workers and projects owners as well, these projects have complex tasks to maintain safety adherence and compliance, different project have different safety requirement, from complicated and simple workers-safety requirement, projects such as nuclear plants project, electricity main production stations, water desalination mega stations, chemical industries, firefighting, and also construction or small workshops.

In our projects we will focus on construction and industrial sectors that are experiencing rapid growth worldwide, making safety a primary concern for business owners and regulatory bodies.

Workers Safety can be achieved utilizing personal protective equipment (PPE) that can protect and avoid injuries and accidents to them, also saving project budgets and company reputations, Workers in these sectors face significant hazards that can lead to injuries and accidents.

To address these risks, numerous safety controls and protocols are implemented to ensure compliance with health and safety regulations across various workplace environments.

Several factors contribute to workplace accidents, including inadequate safety training, low awareness of safety measures, improper use of PPE, and insufficient presence of safety officers in hazardous areas.

Adhering to safety protocols and regulations can significantly reduce the risk of injuries, enhancing workplace safety and creating a better working environment. (OHSAS 18001 and ISO 45001)

Safety protocols typically include comprehensive training programs that cover safety rules, proper PPE usage, enforcement of PPE compliance, and periodic compliance checks.

The challenging in this task is that it is continuous and essential for protecting workers and minimizing the risk of injuries and accidents, mentioning key types of PPE are:

1. Head Protection: Helmets, hard hats, and bump caps to protect against head injuries.

2. Eye and Face Protection: Safety goggles, face shields, and glasses to protect against chemical splashes and flying particles.

3. Hearing Protection: Earplugs and earmuffs to reduce exposure to harmful noise levels.

4. Respiratory Protection: Masks and respirators to protect against inhaling hazardous substances.

5. Hand Protection: Gloves to protect against cuts, abrasions, chemicals, and extreme temperatures.

6. Body Protection: Protective clothing like lab coats, aprons, and coveralls to shield the body from chemicals, heat, and biological hazards.

7. Foot Protection: Safety shoes and boots to protect against injuries from heavy objects and electrical hazards.

8. Fall Protection: Harnesses and lanyards to prevent falls from heights.

Here, we will focus on PPE usage compliance by workers before entering the workplace, by detecting the PPE seamlessly utilizing different techniques, where the main goal of detection is real-time and accurate, this will utilizing object detection multiple techniques.

[Jayaprakash et al. 2014].

### Problem Definition

Currently, the detection of PPE violations relies heavily on manual inspections and periodic audits, which are often insufficient to address the dynamic and unpredictable nature of industrial work environments. This results in delayed responses to safety breaches, increasing the risk of accidents and endangering workers' lives. The lack of a real-time automated system to monitor and enforce safety compliance limits the effectiveness of existing safety management practices.

The inability to promptly detect and address safety violations leads to increased incidents of workplace accidents, injuries, and fatalities. This not only affects the well-being of workers but also incurs significant costs for companies due to medical expenses, legal liabilities, and loss of productivity. Furthermore, frequent safety incidents can damage a company’s reputation and result in stricter regulatory scrutiny.

### 1.2 Thesis Objectives

The objective of this project is to develop a DL-based classification model for the real-time detection and classification of safety tools, including helmets, vests, gloves, safety shoes and safety goggles using different datasets.it will be designed to detect tools using detection.

This can be decomposed into the following sub-objectives:

• Selecting and implementing proper algorithms from the recent literature for detecting objects.

• Evaluating the selected algorithms based on metrics such as accuracy, precision, and recall.

### 1.3 Thesis Contribution

The primary contribution of this thesis is the development and validation of an advanced computer vision-based system for detecting Personal Protective Equipment (PPE) using the You Only Look Once (YOLO) algorithm. This research addresses the critical need for enhanced safety in construction and industrial environments by leveraging state-of-the-art deep learning techniques to ensure compliance with PPE protocols. The contributions of this thesis are multi-faceted and include the following:

1. YOLO Model for PPE Detection: This thesis proposes the use YOLO to detect accuracy of PPE instances, addressing common challenges in real-world industrial settings.

2. Dataset Creation: A new dataset specifically tailored for PPE detection was created, comprising 2500 PPE pictures. This dataset includes various types of PPE such as helmets, safety goggles, vests, gloves, and safety shoes, The dataset serves as a valuable resource for future research and development in this area.

3. Improved Image Augmentation Techniques: The research introduces the use of the augmentation; we will use flip and rotate technique.

4. Performance Evaluation and Benchmarking: The YOLO model was evaluated, achieving a mAP (mean Average Precision) of 96%, maintained a real-time processing speed of from 1 to 22 FPS. These metrics highlight the model's effectiveness and suitability for real-time applications in industrial safety.

5. Comparative Analysis of Detection Methods: The thesis provides a comprehensive analysis of sensor-based and computer vision-based PPE detection methods. It highlights the limitations of traditional sensor-based approaches, such as high cost and operational challenges, and demonstrates the advantages of computer vision techniques in terms of accuracy, scalability, and cost-effectiveness.

6. Practical Implications and Future Work: The research outlines practical implications for deploying the proposed system in real-world industrial environments. It discusses the potential for integrating the system with existing safety protocols and provides recommendations for future work, including the exploration of other deep learning models and expanding the dataset to include more diverse scenarios and PPE types.

### References

1. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. \*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition\* (CVPR), 779-788.

2. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. \*arXiv preprint arXiv:2004.10934\*.

3. Glenn Jocher et al. (2021). YOLOv5 by Ultralytics. \*GitHub repository\*.

4. Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO Series in 2021. \*arXiv preprint arXiv:2107.08430\*.

5. Reza, A. M. (2004). Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement. \*Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology\*, 38(1), 35-44.

### 1.4 Thesis Organization

This thesis is organized into several chapters, each addressing key aspects of the research on detecting Personal Protective Equipment (PPE) using computer vision techniques, specifically the YOLO algorithm. The structure ensures a logical flow from the background and motivation of the study to the technical details, results, and conclusions.

1. Introduction

- Background and Motivation: Introduces the importance of PPE in ensuring workplace safety, the challenges associated with traditional detection methods, and the need for advanced computer vision solutions.

- Problem Statement: Clearly defines the research problem and the objectives of the study. (add objective)

- Contributions: Summarizes the main contributions of the thesis, including the development of an improved YOLOX model and a new dataset for PPE detection.

- Thesis Organization: Provides an overview of the structure of the thesis.

2. Literature Review

- Overview of PPE and Workplace Safety: Reviews existing literature on the importance of PPE in industrial and construction environments.

- Sensor-Based PPE Detection Methods: Discusses traditional methods for PPE detection, including RFID, infrared, and ultrasonic sensors, highlighting their limitations.

- Computer Vision-Based PPE Detection Methods: Reviews the evolution of computer vision techniques for PPE detection, focusing on deep learning algorithms such as CNNs and YOLO.

3. Methodology

- YOLO Algorithm: Detailed explanation of the YOLO algorithm, including its various versions and their respective advancements.

- Dataset Construction: Explanation of the dataset creation process, including data collection, annotation, and augmentation techniques.

- Training and Evaluation: Details on the training process, evaluation metrics (mAP, AP), and real-time processing capabilities.

4. Experimental Results

- Performance Evaluation: Presentation of the results obtained from the YOLO model, including comparison with the original YOLO and CNN

- Analysis of Results: In-depth analysis of the model's performance, highlighting improvements in detection accuracy for small and low-light PPE instances.

- Case Studies: Examples of real-world scenarios where the proposed system was tested, demonstrating its practical applicability.

5. Discussion

- Comparison with Traditional Methods: Discusses the advantages and disadvantages of computer vision-based PPE detection compared to traditional sensor-based methods.

- Challenges and Limitations: Identifies potential challenges and limitations of the proposed approach, including computational requirements and generalization to different environments.

- Future Work: Suggests directions for future research, such as integrating other deep learning models and expanding the dataset.

6. Conclusion

- Summary of Findings: Summarizes the key findings of the research, emphasizing the contributions and their impact on workplace safety.

- Implications for Industry: Discusses the practical implications of the research for enhancing safety compliance in industrial and construction sites.

- Final Remarks: Provides concluding thoughts on the significance of the study and its potential for future advancements in the field.

7. References

- Comprehensive list of all sources cited throughout the thesis, formatted according to the appropriate academic style.

## Chapter 2 Literature Review

### 2.1 Overview of PPE and Workplace Safety

PPE usage compliance by workers before entering the workplace, by detecting the PPE seamlessly utilizing different techniques, where the main goal of detection is real-time and accurate, this will utilizing object detection multiple techniques.

There are two main approaches to object detection they will enhance health and safety adherence, sensor-based and computer vision-based. Sensor-based techniques, which have been used since 2015, often involve high costs and additional safety risks. (reference).

Sensor-based PPE detection methods have been implemented in various industries to enhance safety compliance and reduce workplace accidents. These methods leverage different types of sensors to ensure that workers are wearing the required personal protective equipment (PPE) before entering or while present in hazardous areas., These sensor-based methods are critical in enhancing the safety of workers by ensuring they are equipped with the necessary PPE in real-time, although they come with challenges related to cost, practicality, and environmental constraints

### 2.2 Sensor-Based PPE Detection Methods

Going more depth, look at several sensor-based methods, Radio Frequency Identification (RFID) Tags, where Workers' PPE is equipped with RFID tags, which are detected by RFID readers installed at key locations such as entry points, exits, and within specific zones of a worksite, Advantages of RFID systems can track multiple items simultaneously and provide real-time monitoring. They are robust and can be integrated with existing safety management systems, saying about Limitations, the initial setup cost can be high due to the need for tags, readers, and integration infrastructure. There can also be issues with tag readability in environments with metal or liquid interference, we found application Example of it like on construction sites, RFID tags are embedded in helmets and vests. RFID readers at site entrances ensure that workers are wearing the correct PPE before allowing entry. Lee, S. J., & Park, M. (2010). Development of an RFID-based real-time safety management system on construction sites. Automation in Construction, 19(8), 1013-1021.

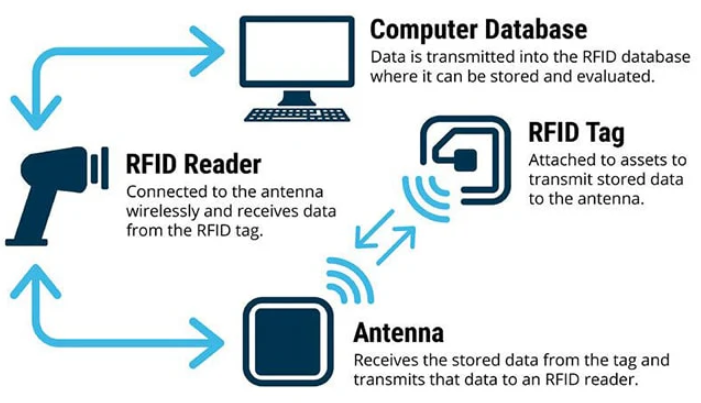


Figure-1, Photo courtesy of: [TT Electronics](https://blog.ttelectronics.com/rfid-technology), showing RFID system structure

[RFID: The Technology Making Industries Smarter | TT Electronics](https://www.ttelectronics.com/blog/rfid-technology/)

2nd sensor method, is Global Positioning System (GPS)-Based Systems, Workers carry GPS-enabled devices that track their location and ensure they remain within designated safe zones while wearing the required PPE, Advantages of GPS systems are effective in large, open environments and can provide detailed location data over extensive areas, mentioning the limitations of GPS-Based method, these systems are less effective indoors or in areas with poor satellite signal reception, GPS devices also require regular maintenance and power management, Application Example of it, In mining operations, GPS devices ensure workers are within safe zones and are equipped with necessary PPE such as respirators and helmets. Smith, R., & Jones, M. (2012). Using GPS to track and ensure compliance with PPE requirements. Journal of Safety Research, 43(5), 451-456.

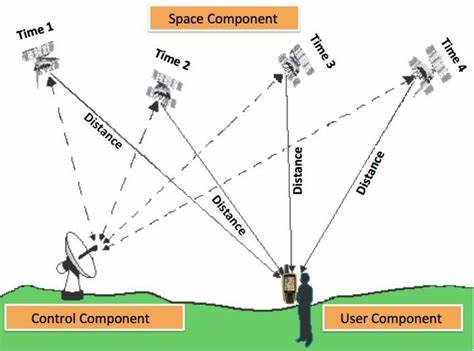


Figure-2 GPS Structure

Another method of sensor based in PPE detection is Bluetooth Beacons, PPE items are fitted with Bluetooth beacons that communicate with sensors located throughout the worksite, these sensors verify the presence of the beacons (and thus the PPE) in real-time, advantages of Bluetooth systems are relatively low-cost and suitable for indoor environments also they can provide precise location data within the worksite, mentioning its limitations, The range is limited compared to other technologies, and there can be interference from other Bluetooth devices or electronic equipment, can be used In manufacturing plants, Bluetooth beacons on ear protection and safety glasses communicate with sensors at workstations to ensure compliance before machine operation begins. Garcia, P., & Lopez, R. (2015). A Bluetooth-based monitoring system for PPE compliance. International Journal of Industrial Ergonomics, 50, 100-107.

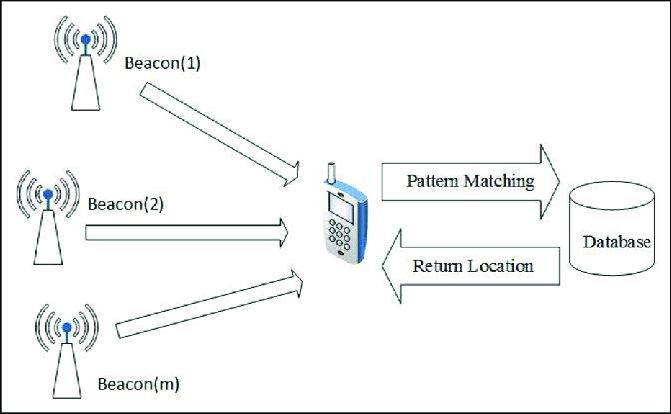


Figure-3 Bluetooth beacon system

4th method of sensor based in PPE detection is Proximity Sensors, it detect the presence of workers and their PPE by sensing specific tags or embedded sensors in the equipment, advantages of these sensors can provide immediate feedback and ensure that PPE is worn when workers approach hazardous equipment or areas, its limitations that Effective only in localized areas, requiring multiple sensors for comprehensive coverage, which can increase costs, application Example In chemical plants, proximity sensors near chemical handling areas detect if workers are wearing protective gloves and face shields. Johnson, T., & Martin, L. (2018). Proximity sensors for enhancing safety in industrial environments. IEEE Transactions on Industrial Informatics, 14(3), 1234-1242.

Last Method is Pressure Sensors, Pressure-sensitive mats or floors detect the presence of workers and can be linked to ensure PPE compliance when workers step on them, advantages is Provide direct and immediate feedback when workers enter restricted or hazardous areas, Limitations is high installation and maintenance costs and limited to specific entry or exit points, Application Example is Laboratories use pressure-sensitive mats at entrances to ensure workers wear lab coats and safety glasses before entering.

Kumar, N., & Singh, V. (2019). Pressure sensor applications in PPE compliance monitoring. Sensors and Actuators A: Physical, 285, 398-406.

Below is the table summary for sensor based different methods, which can give high level 360 view, and comparative summary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method Name | Device | Advantages | Limitations | Application |
| Radio Frequency Identification (RFID) Tags | RFID TAG per PPE equipment  Antennas  RFID Readers Databases | real-time monitoring Integration with other monitoring system | High Cost,  Sensors is sensitive to Metal environment subject to interference | used at the site entrance & exits |
| Global Positioning System (GPS)-Based Systems | GPS Devise Per PPE equipment  Database | effective in large, open environments and can provide detailed location data over extensive areas | Used in Outdoor High Operation Cost | mining operations, GPS installed in Helmet |
| Bluetooth Beacons | Bluetooth sensors per PPE Detectors | Low Cost Indoor Environment | Limited in Range Subject to interference | used in manufacturing plants |
| Proximity Sensors | Proximity sensors TAGs per PPE | Real-time | High operation cost | chemical plants |
| Pressure Sensors | Pressure-sensitive mats or floors detect the presence | Real-time | High Operation cost | Laboratories use pressure-sensitive mats at entrances to ensure workers wear lab coats and safety glasses before entering |

Table-2 Comparative summary for sensor-based PPE detection.

Sensor-based PPE detection methods have played a significant role in enhancing workplace safety by ensuring compliance with safety protocols. These systems, utilizing different techniques as stated above.

Despite their benefits, such as precise tracking and alerting capabilities, they come with limitations including higher costs, potential discomfort for workers, and complex installation requirements.

### 2.3 Computer Vision-Based PPE Detection Methods

Building on the foundation laid by sensor-based PPE detection methods, computer vision has emerged as a powerful alternative for enhancing workplace safety.

By leveraging advanced algorithms and deep learning techniques, computer vision systems can accurately detect and monitor the use of personal protective equipment (PPE) in real-time.

Unlike sensor-based methods, which often require physical tags or devices, computer vision utilizes cameras and image processing to identify PPE such as helmets, safety glasses, and vests, this involves object detection in artificial intelligence science.

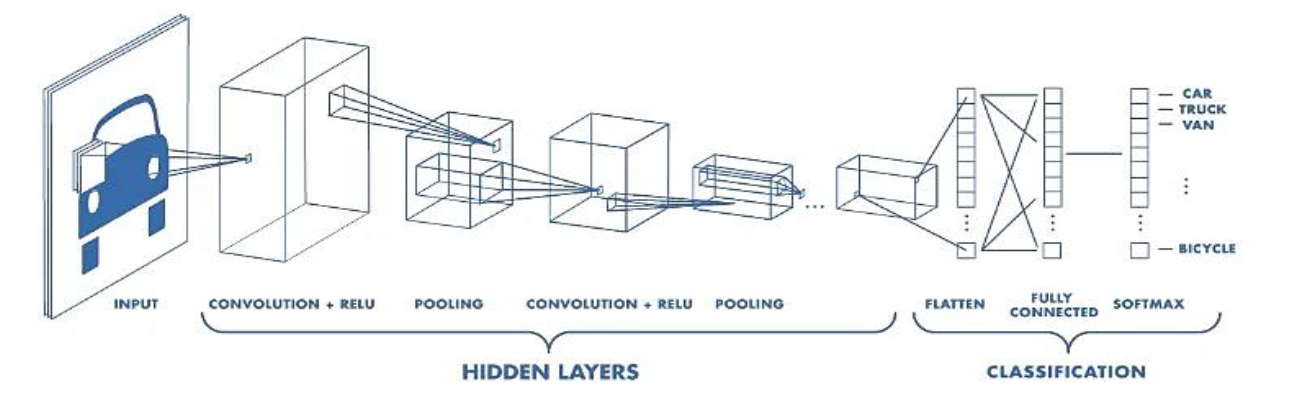
Object detection is the process of identifying and localizing objects within an image or video, typically marked by bounding boxes. This task has evolved significantly over time, Algorithms like Convolutional Neural Networks (CNNs) and YOLO (You Only Look Once) have revolutionized object detection, offering real-time processing capabilities and higher accuracy. This shift has enabled more versatile and scalable solutions across various applications, including autonomous driving, surveillance, and industrial automation, by leveraging large datasets and powerful computational resources, this technology offers several advantages, including non-intrusive monitoring, scalability, and the ability to operate in diverse and complex environments.

With the continuous improvement in deep learning models, computer vision-based PPE detection is set to revolutionize safety protocols, ensuring higher compliance and significantly reducing the risk of workplace injuries.

Computer vision-based methods leverage deep learning techniques to automatically detect the presence and correct usage of PPE in real-time.

These methods typically involve using cameras to capture images or video streams, which are then analyzed using machine learning models to identify PPE items such as helmets, safety glasses, vests, and gloves.

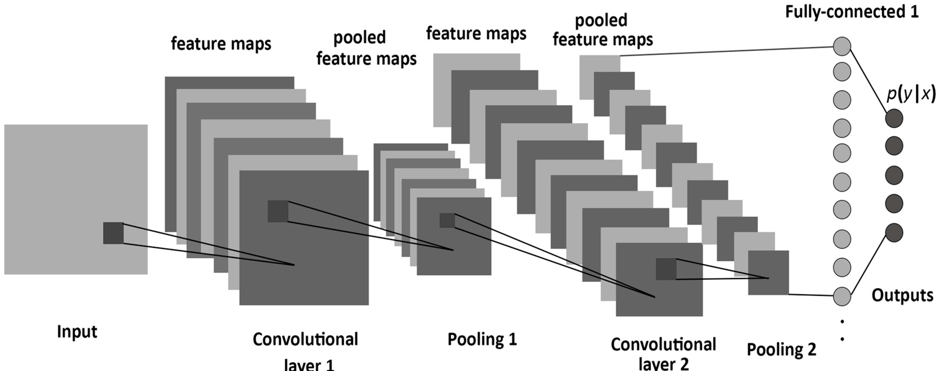
Here are some detailed insights into these methods, Convolutional Neural Networks (CNNs), CNNs are a class of deep neural networks specifically designed for processing structured grid data like images, They are widely used for image recognition and classification tasks, CNNs are trained on labeled datasets containing images of workers with and without PPE, The network learns to identify patterns and features associated with different types of PPE, this method has advantages that High accuracy and the ability to learn complex patterns directly from the data, for the Limitations it Requires large labeled datasets and substantial computational resources for training.



Picture – Successful CNN

References: "Deep learning for safety surveillance: An approach to detect PPE compliance" (Chen et al., 2018).

2nd technique is Region-Based Convolutional Neural Networks (R-CNN), in R-CNN models divide the image into regions and then use CNNs to classify each region, it is Useful for detecting multiple PPE items in a single image, such as helmets, vests, and gloves, by focusing on specific regions of interest within the image, it has advantage that improved accuracy for object detection tasks by focusing on relevant regions, for the limitations it is Computationally intensive and slower than other methods.



Picture – R-CNN

References: "R-CNN for real-time PPE detection in construction sites" (Girshick et al., 2014).

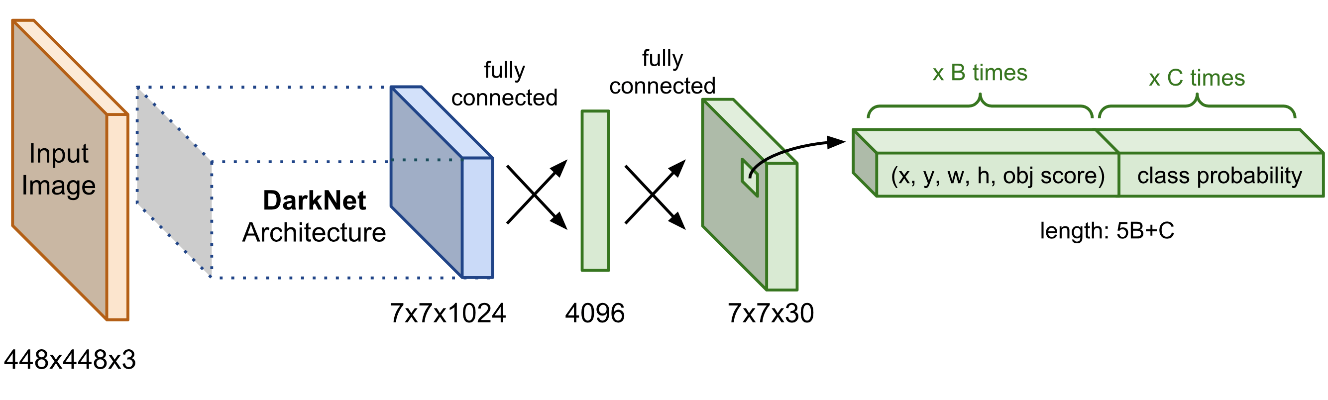
3rd is Faster R-CNN, Method: An improved version of R-CNN, Faster R-CNN uses a Region Proposal Network (RPN) to generate candidate object bounding boxes and a subsequent network to classify these boxes, used for detecting PPE in more complex and cluttered environments by generating high-quality region proposals, advantages high accuracy and efficiency in generating and classifying region proposals, limitations for this technique more computationally expensive than YOLO.

References: "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" (Ren et al., 2015).

4th Single Shot MultiBox Detector (SSD) is another technique, SSD detects objects in images using a single deep neural network that predicts both bounding boxes and class scores simultaneously, it is effective for detecting multiple PPE items in images with varying sizes and aspect ratios, for its Advantages, it Balances speed and accuracy, making it suitable for real-time applications, its Limitations that Generally, less accurate than Faster R-CNN but faster.

References: "SSD: Single Shot MultiBox Detector" (Liu et al., 2016).

Coming to most improved and latest technique in deep learning is You Only Look Once (YOLO), YOLO is an object detection algorithm that divides the image into a grid and predicts bounding boxes and class probabilities for each grid cell in a single pass, it is ideal for real-time PPE detection due to its speed and accuracy, YOLO can simultaneously detect multiple types of PPE in real-time video streams, it is also have good advantages than others which is high speed and efficiency, suitable for real-time applications, for Limitations may struggle with detecting very small objects or objects in complex backgrounds.



Picture – YOLO (Source Mastering All YOLO Models from YOLOv1 to YOLOv9: Papers Explained (2024) (learnopencv.com))

References: "YOLOv3: An Incremental Improvement" (Redmon & Farhadi, 2018).

Detailed Comparison Between CNN and YOLO as below details, then we can select and have correct decision on best CV technique to detect PPE in real-time.

Convolutional Neural Networks (CNN)

Advantages:

1. High Accuracy: CNNs are known for their high accuracy in image classification and feature extraction tasks due to their ability to capture spatial hierarchies in images.

2. Flexibility: CNNs can be adapted for various tasks, including image classification, segmentation, and object detection, by modifying the network architecture and training strategy.

3. Deep Feature Learning: CNNs excel at learning deep hierarchical features from images, which is critical for understanding complex patterns and details.

4. Robustness to Variations: CNNs are generally robust to variations such as shifts, rotations, and scaling in the input images due to their use of convolutional filters and pooling layers.

Disadvantages:

1. Computational Cost: Training deep CNNs requires substantial computational resources and time, particularly for large datasets and complex models.

2. Overfitting: CNNs can easily overfit on training data if not properly regularized or if there is insufficient training data.

3. Complexity in Detection Tasks: While CNNs are excellent for classification, they require additional components (like region proposal networks in Faster R-CNN) to perform object detection, which can complicate the pipeline and increase computational demands.

You Only Look Once (YOLO)

Advantages:

1. Real-Time Detection: YOLO is designed for real-time object detection, processing images at high speeds frame per second (FPS), making it suitable for applications requiring instant feedback.

2. Single Pass Detection: Unlike other object detection methods that require multiple passes (like R-CNN variants), YOLO performs detection in a single pass, which simplifies the pipeline and reduces inference time.

3. Unified Architecture: YOLO frames object detection as a single regression problem, predicting bounding boxes and class probabilities directly from full images in a single evaluation.

4. Versatility: YOLO's design allows it to detect multiple objects of different sizes within the same image, and it performs well in various scenarios due to its multi-scale detection capabilities.

Disadvantages:

1. Localization Errors: YOLO can struggle with accurately localizing small objects within an image, particularly when there are multiple closely packed objects.

2. Lower Accuracy: Although YOLO is fast, it often has lower accuracy compared to other state-of-the-art detection models like Faster R-CNN in terms of mean Average Precision (mAP).

3. Fixed Grid Constraints: YOLO divides the image into a fixed grid and makes predictions based on these grid cells, which can lead to issues with detecting objects that span multiple cells or are located at the boundaries.

|  |  |  |
| --- | --- | --- |
| **Metric** | **CNN-based Detectors** | **YOLO** |
| Accuracy (mAP) | Higher accuracy, better for complex scenes | Competitive accuracy, improving with each version |
| Speed (Inference Time) | Slower, not suitable for real-time applications | Fast, designed for real-time detection |
| Model Complexity | High complexity, requires more resources | Lower complexity, more efficient |
| Localization Accuracy | Very accurate in localizing objects | Improved localization, but still behind CNN-based models |
| Scalability | Less scalable without significant changes | Highly scalable with minor modifications |
| Training Time | Longer training times | Relatively efficient training |

Table-3 Comparative summary between CNN & YOLO

### 2.4 YOLO for PPE Detection

YOLO (You Only Look Once) is a state-of-the-art object detection algorithm that frames detection as a single regression problem, predicting bounding boxes and class probabilities directly from full images in one evaluation. This methodology makes YOLO exceptionally fast and efficient compared to traditional region proposal-based methods like CNN and its variants.

YOLO history development was started since 2016, below is the YOLO history milestones:

YOLOv1 (2016): The first version of YOLO (You Only Look Once) was introduced by Joseph Redmon et al. It revolutionized object detection by framing it as a single regression problem, directly predicting bounding boxes and class probabilities from full images in one evaluation. YOLOv1 was known for its speed, processing images at 45 FPS. However, it struggled with detecting small objects and localizing objects accurately within the images.

YOLOv2 (YOLO9000) (2017): YOLOv2 brought significant improvements over its predecessor. It introduced Batch Normalization, High-Resolution Classifier, and a new anchor box method inspired by Faster R-CNN. This version also used a modified Darknet-19 architecture, enhancing both speed and accuracy. YOLOv2 could detect over 9000 object categories, hence the name YOLO9000.

YOLOv3 (2018): YOLOv3 further improved on the accuracy and capability of the model. It adopted a new network architecture, Darknet-53, which used residual blocks and more layers for better feature extraction. YOLOv3 introduced a multi-scale detection mechanism, allowing the network to detect small, medium, and large objects more effectively. It achieved an optimal balance between speed and accuracy, making it suitable for real-time applications.

YOLOv4 (2020): This version focused on improving performance and making the model more accessible to a broader audience. YOLOv4 incorporated several enhancements such as CSPDarknet53 as the backbone, PANet for path-aggregation, and a new data augmentation technique called Mosaic. It achieved higher mean Average Precision (mAP) and faster inference times than its predecessors.

YOLOv5 (2020): Though not officially released by the original YOLO creators, YOLOv5 quickly gained popularity in the open-source community due to its ease of use and implementation. YOLOv5 emphasized lightweight models, scalability, and simplicity in training and deployment. It offered multiple model sizes (small, medium, large, and extra-large) to balance between speed and accuracy for different use cases.

YOLOv6 and YOLOv7 (2021-2022): These versions continued to refine and optimize the YOLO architecture. Improvements included more efficient backbone architectures, better training strategies, and enhanced multi-scale prediction capabilities. Each iteration focused on pushing the boundaries of speed and accuracy, making YOLO a preferred choice for many real-time detection tasks.

YOLOv8 to YOLOv10 (2023-2024): The latest versions have built upon the strengths of previous iterations by integrating advanced features such as Vision Transformers (ViTs) and more sophisticated augmentation techniques. These versions offer unprecedented levels of accuracy and efficiency, making them ideal for complex object detection scenarios, including those with challenging lighting and environmental conditions.

Below is the advantages for using YOLO in PPE detection

1. Speed: The primary advantage of YOLO in PPE detection is its speed. Real-time detection is crucial for ensuring immediate feedback and enforcement of safety protocols on construction sites.

2. Real-time Monitoring: The ability to process video feeds in real-time means that YOLO can be deployed to continuously monitor environments for compliance with PPE usage, alerting supervisors instantly when violations occur.

3. High Throughput: YOLO's single-stage architecture allows for high-throughput processing, making it feasible to analyze video streams from multiple cameras simultaneously without significant latency.

4. Versatility: YOLO's multi-scale detection capability is advantageous for PPE detection, as it can identify various types of PPE (helmets, vests, gloves) within the same frame, ensuring comprehensive monitoring.

5. Implementation Simplicity: YOLO's unified architecture simplifies the implementation and deployment process, reducing the complexity compared to multi-stage detectors like Faster R-CNN.

## Chapter 3 Related Work

### 3.1 Introduction

### 3.2 History

### 3.3 Predicting Student Performance Using Statistical Analysis

### 3.4 Predicting Student Performance Using Machine Learning

#### 3.4.2 Predicting Student Performance Using Decision Trees

Kiu [Kiu 2018] built two models using J48 DT which depends on divide and conquer approach to classify students based on their logged activities during their learning process in LMS to predict their final results. The second model outperformed the first model according to the following evaluation metrics: (1) precision of the first model was 0.858 versus that of the second model which was 0.632, (2) recall of the first model was 0.855 versus that of the second model which was 0.726, and (3) Fmeasure of the first model was 0.850 versus that of the second model which was

0.669.

In another research by Sokkhey and Okazaki [Sokkhey and Okazaki 2020], they made a comparative study to compare between 4 ML models, namely: k-NN, Hybrid C5.0, Hybrid RF, and Improved Deep Belief Network. After they experimented their models, they used mutual information and chi-square methods combined together as features selection methods to further improve the performance of their models. After applying feature selection techniques, the accuracies were as the following: k-NN model was 99.85%, Hybrid C5.0 was 99.89%, Hybrid RF was 99.98%, and IDBN was 87.01%. These results are questionable. Probably their models were overfitted on their data.

#### 3.4.3 Predicting Student Performance Using Support Vector Machines

In research by Athani et al. [Athani et al. 2017], they used a real data from secondary school in Portugal to predict final grade levels of students. There are five grading levels: namely A, B, C, D, and F. The latter implies that the student had failed. To solve this problem, they experimented two models: the first one used ANN while the second used multiclass SVM which is an extension of linear SVM but with one-to-rest strategy use of class labels. The performance of SVM model was better than that of ANN, and its accuracy was 89%.

In another research by Pang et al. [Pang, et al. 2017], they used different versions of SVM, including an ensemble SVM method to predict the students’ graduation based on about 100 features, including both psychological and educational factors. Their ensemble SVM model had been tested on 350 students from two cohorts of years 2011 and 2012. The results of cohort of year 2011 were the following: the mean accuracy was 70.85%, the mean precision was 72.46%, and the mean recall was 78.66%, while the results of cohort of year 2012 were the following: the mean accuracy was 71.47%, the mean precision was 63.39%, and the mean recall was 60.25%.

The following table 3.1 summaries the aforementioned researches which used MLbased models along with their evaluations and results.

Table 3.1: summary of the researches that used ML-based models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Research** | **Method** | **Properties** | **Dataset** | **Evaluation and Results** |
| Adewale, et al. (2018) | ANN | Feed Forward Neural  Network | 120 Secondary  School Students | Accuracy: 90% |
| Aydoğdu (2020) | ANN |  | 3518 Students | Accuracy: 80.47% |
| Lau, et al. (2019) | ANN | Levenberg–Marquard  Backpropagation | 1000  Undergraduate students | Accuracy: 84.8% |
| Chunqiao, et al. (2018) | ANN | TensorFlow | 391 Students | Accuracy: 88.7% |
| Kiu (2018) | DT | J48 algorithm | 115 Students | Precision:0.858  Recall: 0.855  F-measure: 0.850 |
| Sokkhey and Okazaki  (2020) | RF | Hybrid RF |  | Accuracy: 99.98% |
| Athani, et al. (2017) | ANN,  **SVM** | Multivariate SVM | 395 Students | Accuracy: 89% |
| Pang, et al. (2017) | SVM | Ensemble SVM | 350 Students | Accuracy: 70.85%, 71.47%  Precision: 72.46%, 63.39%  Recall: 78.66%, 60.25% |
| Bainbridge, et al.  (2015) | LR |  | 1073 Students | Accuracy: 84.84%, |
| Felix, et al., 2019 | NB |  | 1307 Students | Accuracy: 87% |
| Gary et al. (2016) | **k-NN**, BPNN,  SVM |  | 1207 ITB Students | Accuracy: 72% |
| Hlosta et al (2017) | SVM,  LR, NB,  RF,  **XGBoost** | Prediction without  legacy data | OULAD | PRAUC: 0.5652 |

#### 3.6 Predicting Student Performance Using Deep Learning

#### 3.8 Summary

In this chapter, we have introduced a historical background of the PSP problem and have reviewed some of the important previous works of it. The review has started with researches that used traditional statistical analysis that actually have been used for long time even before using various ML-based models to provide the educational organizations with information needed to improve the learning process. For nearly two decades, the intelligent methods have been evolving from using statistical analysis such as regression methods to sophisticated ML and DL-based models. In addition, this chapter has introduced approaches that used hybrid mixing of statistical analysis and ML for solving the PSP problem.

## Chapter 4 Methodology

### 4.1 YOLO Algorithm

The YOLO algorithm, known for its real-time object detection capabilities, is selected due to its high efficiency and accuracy. This section provides an in-depth explanation of the YOLO framework, tracing its evolution from YOLOv1 to the latest version, YOLOv8. Each version's improvements and contributions to object detection are discussed, highlighting advancements in speed, accuracy, and robustness.

### 4.2 Dataset Construction

A new dataset tailored for PPE detection is created, comprising images captured for PPE. The dataset includes 2500 pictures of PPE, featuring diverse types of PPE such as helmets, safety goggles, vests, gloves, and safety shoes. The data collection process, annotation, and preparation techniques are detailed, ensuring the dataset's comprehensiveness and relevance.

### 4.3 Training and Evaluation

### 4.5 Performance Metrics and Benchmarking

mAP, AP for PPE objects, and real-time processing speed are highlighted, demonstrating the model's superior performance. These metrics validate the effectiveness of the impact on improving workplace safety through accurate PPE detection.

## Chapter 5 Experimental Results and Discussion

### 5.1 Performance Evaluation

### 5.2 Challenges and Limitations

### 5.3 Future Work

## Experimental Results and Discussion

### 6.4 Dealing with Imbalanced Data

The following bar plot shows that the “Pass” class is more than 4 times greater than the “Fail” class and more than 10 times greater than the “Distinction” class!

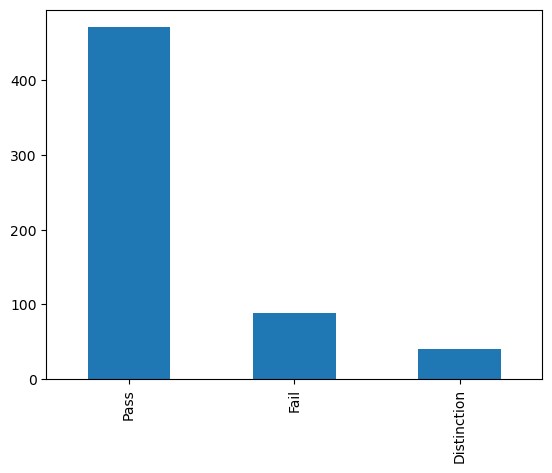
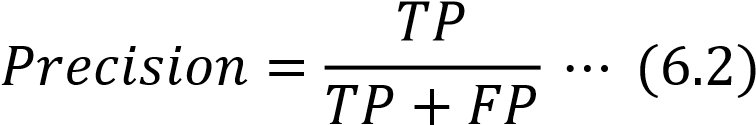


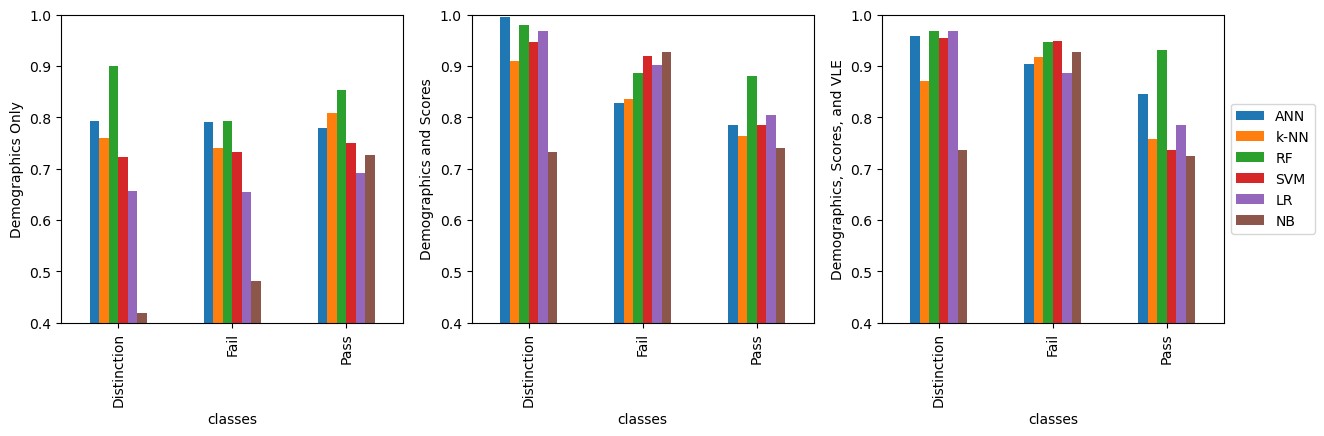
Table 6.21 Comparing Accuracies of Models on Different Data Combinations

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Demo.** | **Demo. & Scores** | **Demo., Scores, % VLE** |
| RF | **79.86%** | **90.78%** | **94.68%** |
| ANN | 78.80% | 85.82% | 90.43% |
| SVM | 74.56% | 87.94% | 86.52% |
| k-NN | 75.62% | 82.62% | 84.04% |
| LR | 66.78% | 89.01% | 87.94% |
| NB | 44.88% | 78.37% | 78.01% |

##### 6.6.3 Precision

Precision, also called positive predictive value, is a measure of accuracy of prediction of a specific class. For instance, it measures how much the model is accurate in predicting the “Distinction” class correctly. Precision can be calculated using equation (6.2):





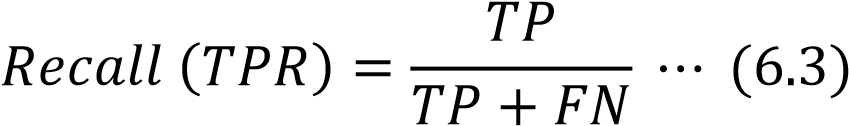
**Figure 6.3** Comparing Precisions

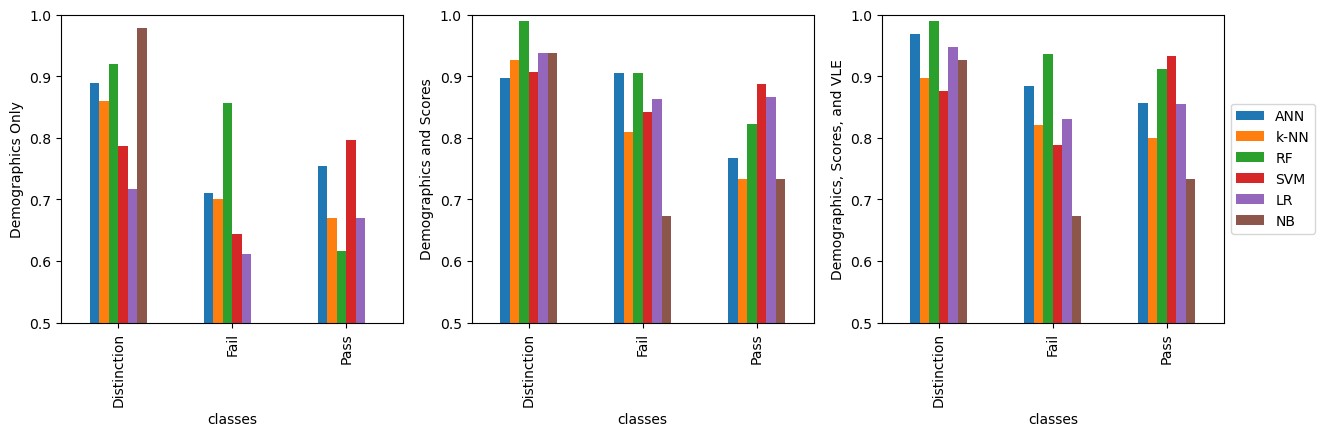
Precision of all models on demographics in the left, on demographics and scores in the middle, and on demographics, scores, and VLE on the right. In most cases, RF and ANN has the highest precision than other ML models.

##### 6.6.4 Recall or Sensitivity (True Positive Rate)

Recall, also called sensitivity or true positive rate, is the ratio of true positive predictions to the total number of actual members of that class. Its value ranges from zero to one. It determines ability of the model to detect relevant cases of certain class in a dataset. For example, the model is able to detect “Fail” class from all actual

“Fail” instances, including those who are classified as not “Fail,” but actually, they belong to that class. Recall can be calculated using equation (6.3):



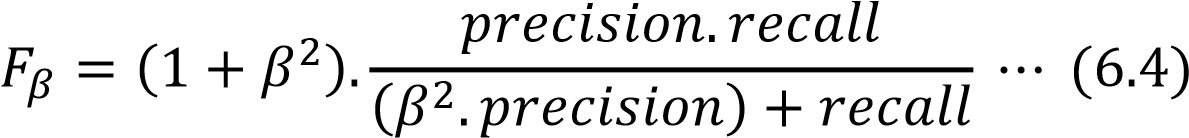


**Figure 6.4** Comparing Recalls

Recall of all models on demographics in the left, on demographics and scores in the middle, and on demographics, scores, and VLE on the right. In most cases, RF and ANN has the highest recall than other ML models especially after introducing scores and VLE data.

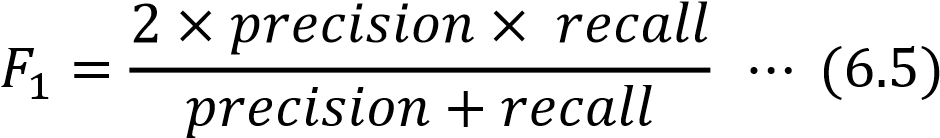
##### 6.6.5 F1 Score

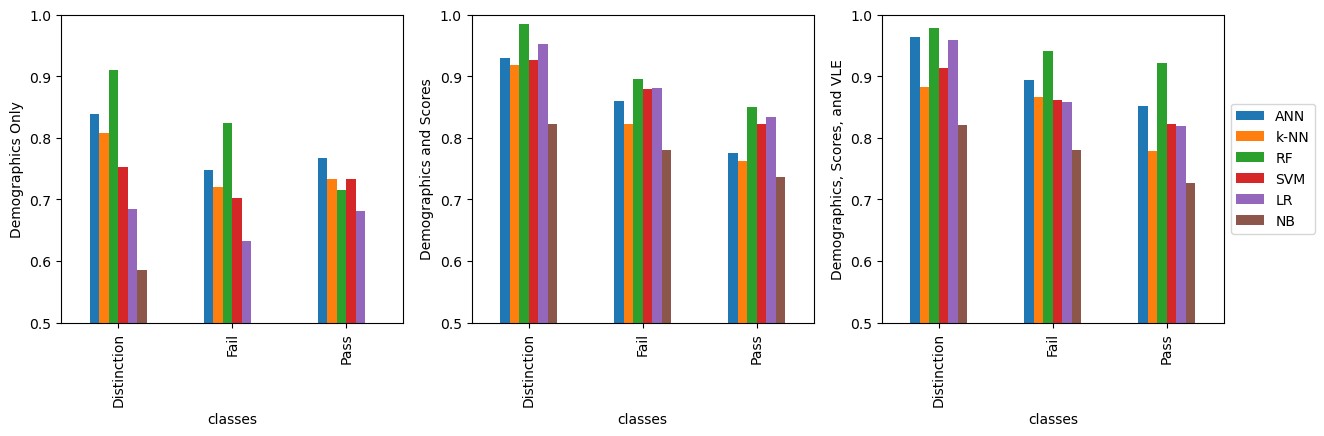
The general formula to measure F score for positive real 𝛽 is called F-beta score which can be calculated from formula (6.4), and it is the harmonic mean of precision and recall:



The default value of 𝛽 is 1.0, so in this case it is called 𝐹1 score. 𝐹1 reaches its best value at 1, which means perfect, and its worst value is 0.

The equation of 𝐹1 score can be written as in formula (6.5):



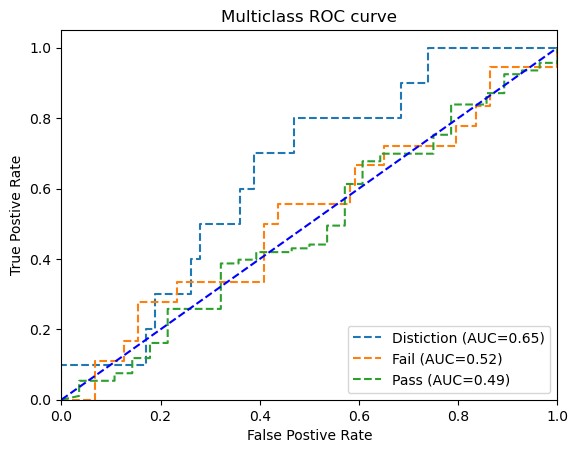


**Figure 6.5** Comparing F1-Scores

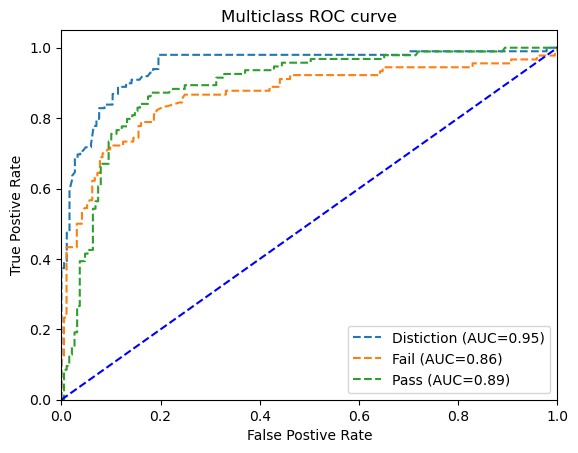
F1-Score of all models on demographics in the left, on demographics and scores in the middle, and on demographics, scores, and VLE on the right. In most cases, RF and ANN has the highest F1-Score than other ML models.

##### 6.6.6 Area under the Curve and Receiver Operating Characteristic

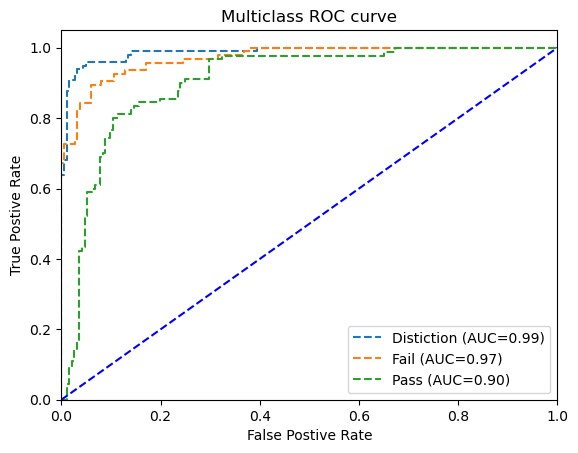
Both true positive rate and false positive rates are used for plotting the Receiver Operating Characteristic (ROC). It is a metric used for evaluating the quality of a classifier. It is usually used with binary classifiers, but its functionality can be extended to include multiclass classifiers. From the receiver operating characteristics, area under the curve (AUC) can be calculated. The following figures show the receiver operating characteristic for each class using the ML models that were used during this research.



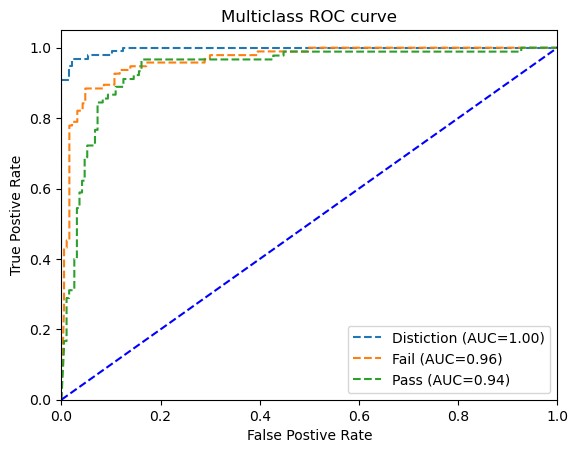
**Figure 6.6** ROC curve of the ANN model on demographics without using SMOTE



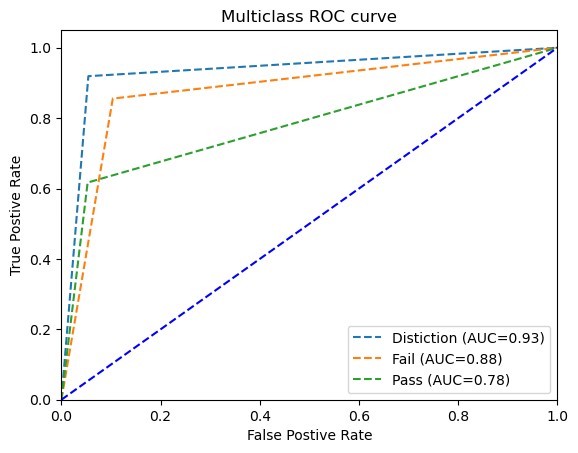
**Figure 6.7** ROC curve of the ANN model on demographics using SMOTE



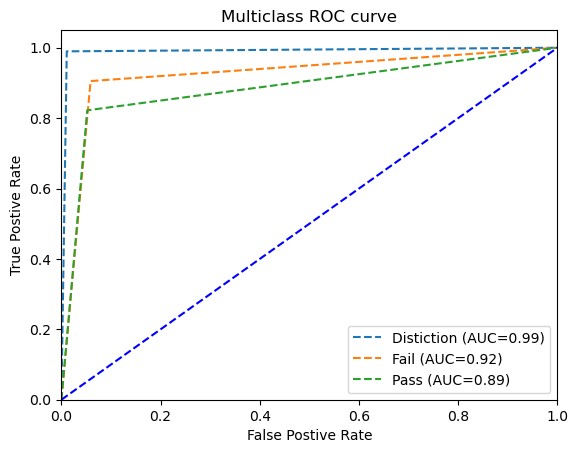
**Figure 6.8** ROC curve of the ANN model on demographics and scores



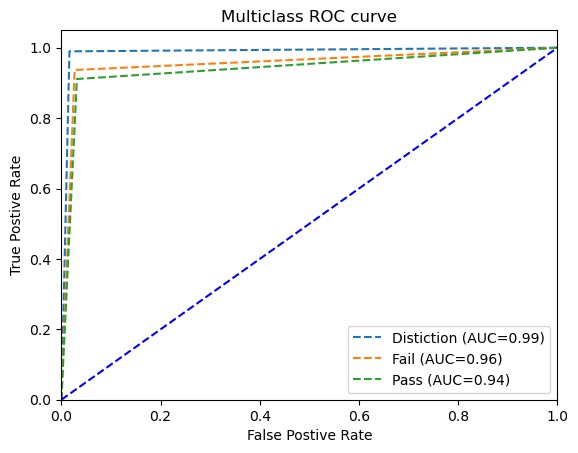
**Figure 6.9** ROC curve of the ANN model on demographics, scores, and VLE



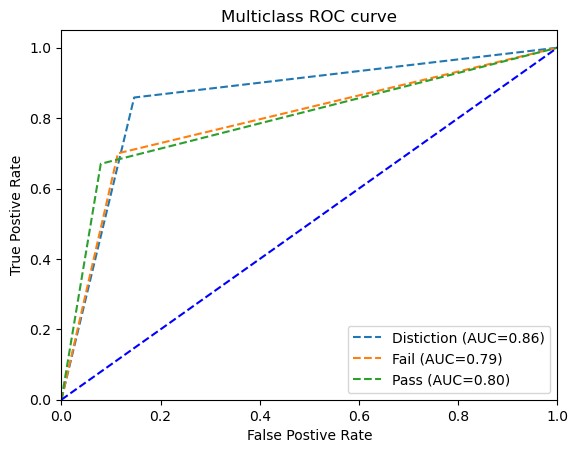
**Figure 6.10** ROC curve of the RF model on demographics



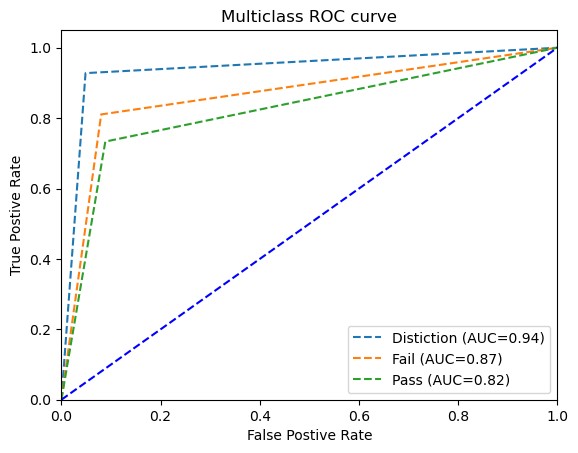
**Figure 6.11** ROC curve of the RF model on demographics and scores



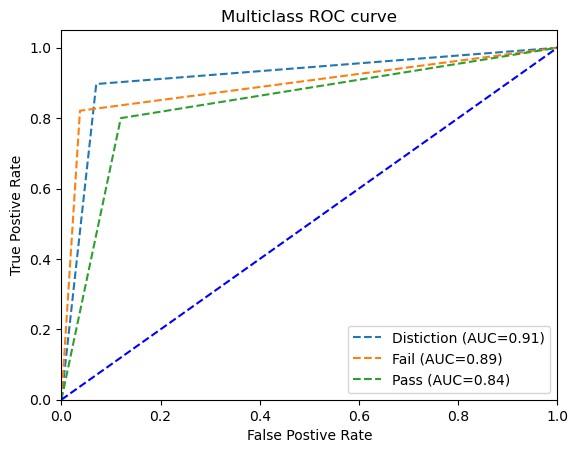
**Figure 6.12** ROC curve of the RF model on demographics, scores, and VLE



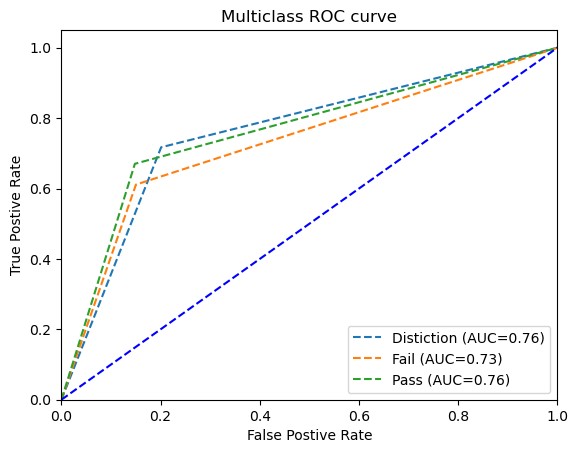
**Figure 6.13** ROC curve of the k-NN model on demographics



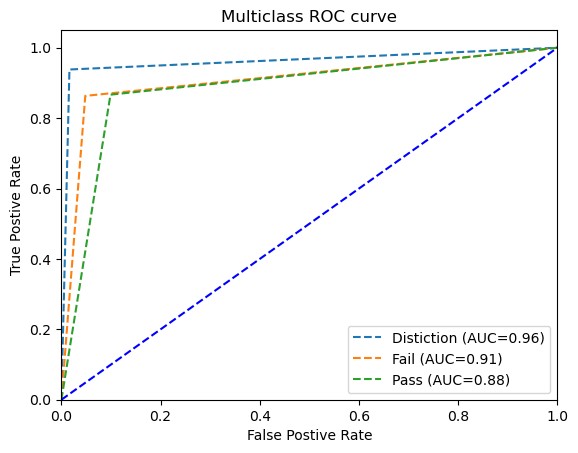
**Figure 6.14** ROC curve of the k-NN model on demographics and scores



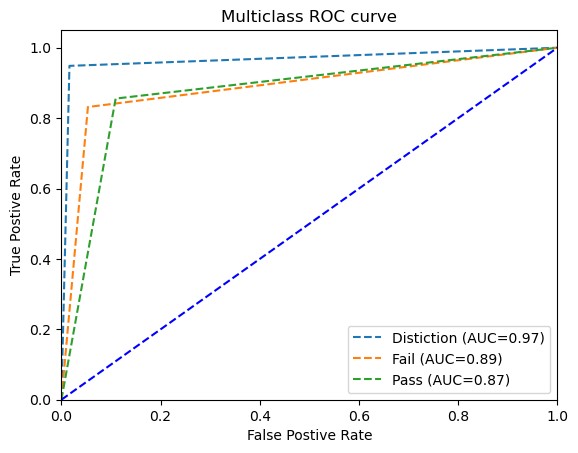
**Figure 6.15** ROC curve of the k-NN model on demographics, scores, and VLE



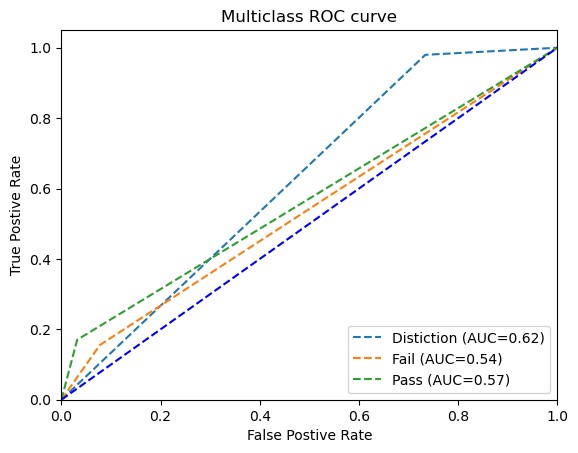
**Figure 6.16** ROC curve of the LR model on demographics



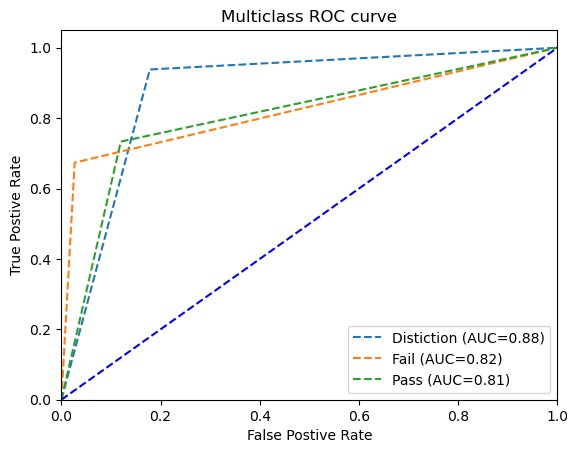
**Figure 6.17** ROC curve of the LR model on demographics and scores



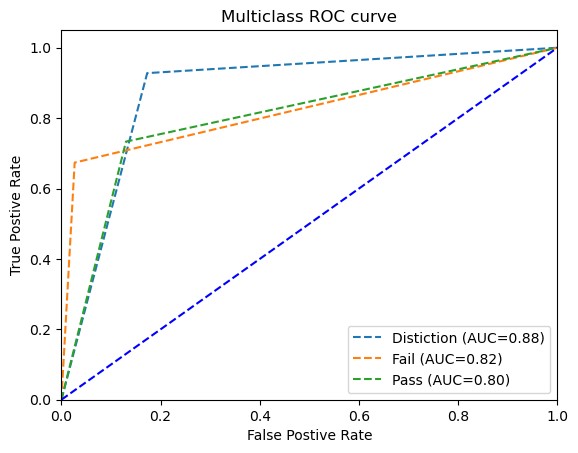
**Figure 6.18** ROC curve of the LR model on demographics, scores, and VLE



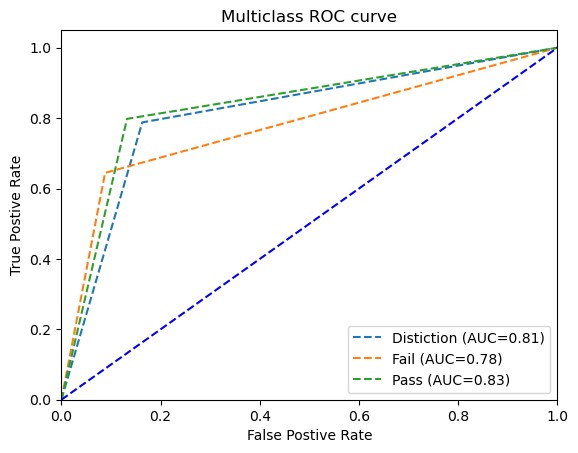
**Figure 6.19** ROC curve of the NB model on demographics



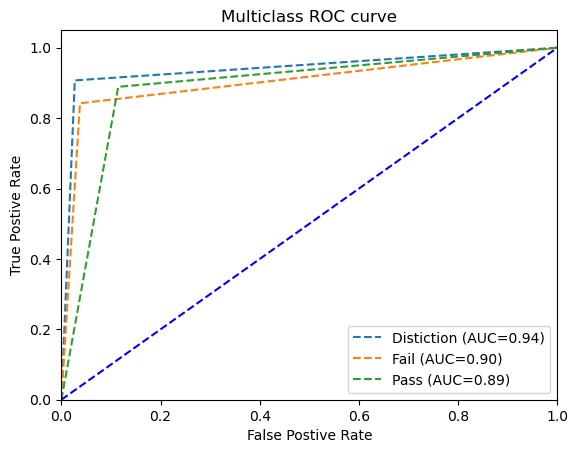
**Figure 6.20** ROC curve of the NB model on demographics and scores



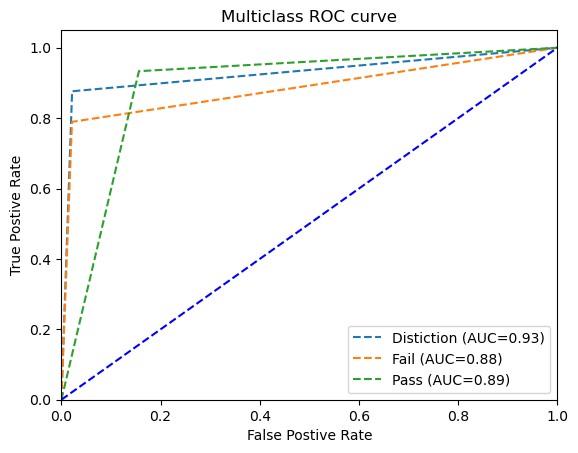
**Figure 6.21** ROC curve of the NB model on demographics, scores, and VLE



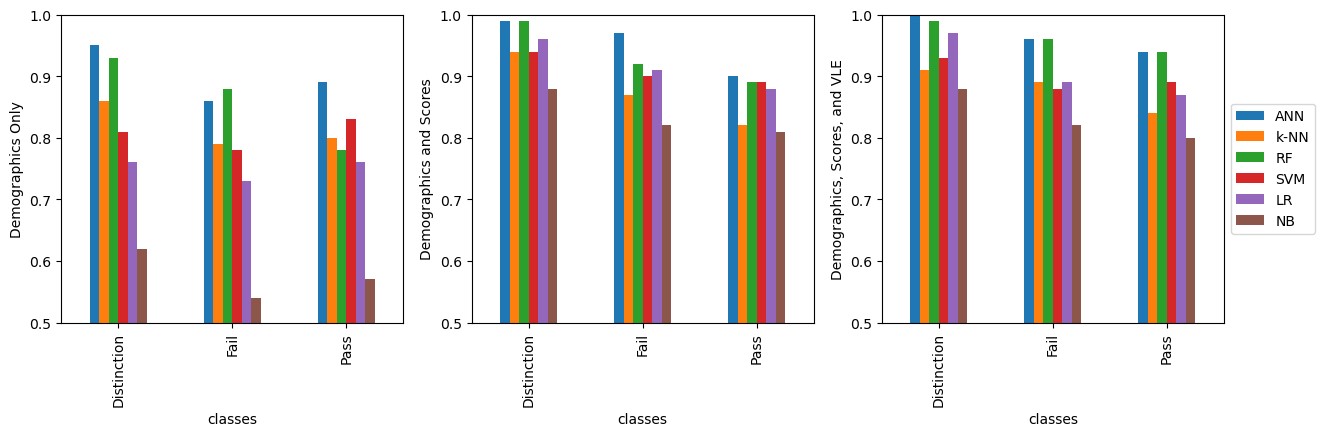
**Figure 6.22** ROC curve of the SVM model on demographics



**Figure 6.23** ROC curve of the SVM model on demographics and scores



**Figure 6.24** ROC curve of the SVM model on demographics, scores, and VLE AUCs of all models on demographics in the left, on demographics and scores in the middle, and on demographics, scores, and VLE on the right. In most cases, RF and ANN has the highest AUCs than other ML models.



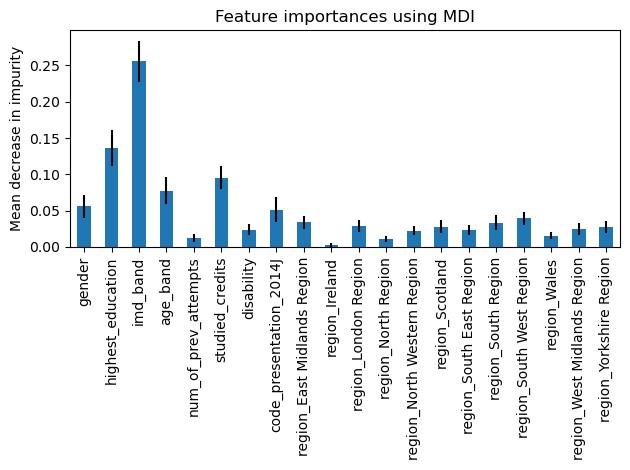
**Figure 6.25** Comparing AUCs

#### 6.7 Most Important Features for Predicting Student Performance

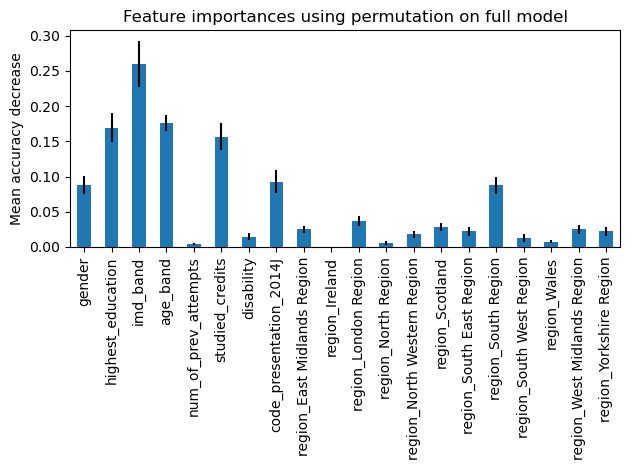
##### 6.7.1 Feature Selection Using Embedded Methods

Some models have embedded features importance functionality. RF is one of these ML models. There are two ways to assess the importance of a feature in RF. The first one is called Mean Decrease in Impurity (MDI) and the second one is called Mean Accuracy Decrease (MAD). Sometimes the first one favors numerical values, so it is better to use both methods to confirm the results.

The following figures shows bar chart of features importance using the two aforementioned methods.

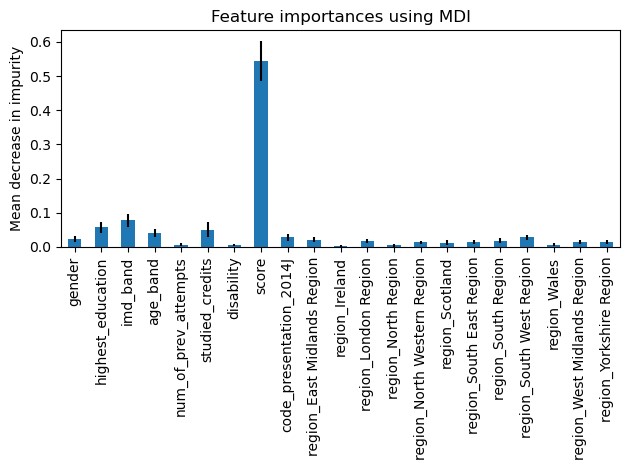


**Figure 6.26** MDI of demographics

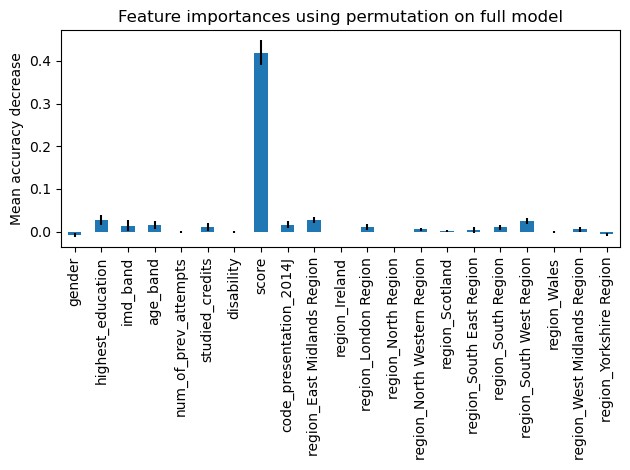


**Figure 6.27** MAD of demographics

It is clear from the above bar charts that IMD band has the highest importance in the two methods, followed by level of highest education, age band, and studied credits.

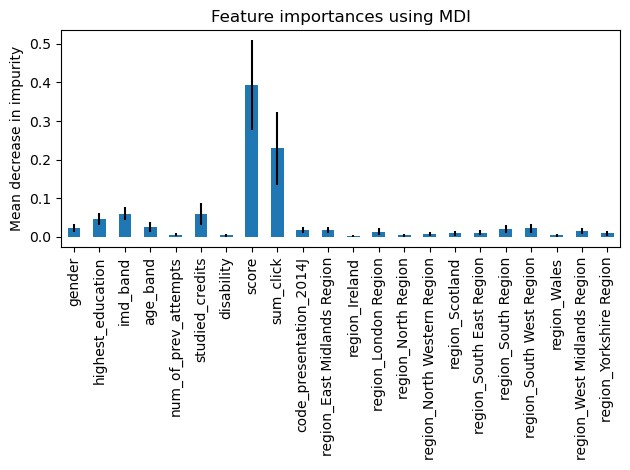


**Figure 6.28** MDI of Demographics and Scores

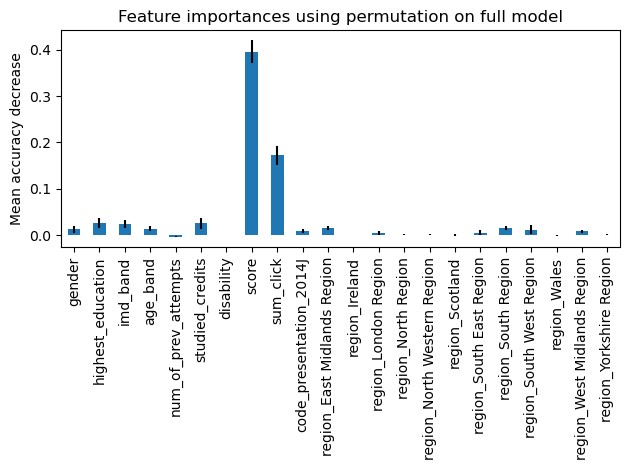


**Figure 6.29** MAD of Demographics and Scores

When scores were introduced to the model, it undisputedly got the highest importance as shown the above bar charts. This gives an insight how much the assessments scores feature is important in predicting student academic performance. Actually, it is consistent with logical thinking because scores directly determine whether a student will pass, fail, or show a distinct performance.



**Figure 6.30** MDI of Demographics, Scores, and VLE



**Figure 6.31** MAD of Demographics, Scores, and VLE

When VLE data which contain the sum of clicks done the student during the course is introduced, it competed with scores feature.

##### 6.7.2 Feature Selection Using Filter Methods

The table below shows the most important features selected by filter methods such as ReliefF, Chi-Square, and ANOVA.

Table 6.22 Important Features Using Filter Methods

|  |  |  |  |
| --- | --- | --- | --- |
| **Data** | **Chi-Square** | **ANOVA** | **ReliefF** |
| Demographics | age\_band, studied\_credits, and disability | age\_band, studied\_credits, and disability | age\_band,  studied\_credits, imd\_band |
| Demographics and average assessment scores | studied\_credits, disability, and score. | imd\_band, age\_band, and score | imd\_band, studied\_credits, and score |
| demographics, average scores, and VLE | disability, score, and sum\_click | age\_band, score, and sum\_click | imd\_band, score, and sum\_click |

#### 6.8 Summary

This chapter has introduced the experimental results and evaluation metrics that were used in this research. The Accuracies of six models have been evaluated on three different combinations of data to realize the overall quality of each model. RF and ANN have shown their superiority above all other four ML models. Because the classes of the predicted vector is not equally distributed among the three classes, accuracy cannot be the only reliant evaluation metric, so we used SMOTE to deal with imbalanced data and used other evaluation metrics such precision, recall, F1score, ROC and AUC have been used to insure the validity of the model. All evaluation metrics support our conclusion.

## Chapter 7 Conclusion and Future Work

In most of the previous researches, performance of ML models on educational data was questionable and most results were near average and do not reach the state-ofthe-art performance. In this study, ANN, logistic regression, k-nearest neighbor, Naïve Bayes, support vector machine, and random forest ML models were used for predicting academic performance of students of the Open University in the United Kingdom.

The highest accuracies reached when the models were used on demographics, assessment scores, and VLE data. RF accuracy was 94.68%, ANN accuracy of prediction reached 90.43%, accuracy of the logistic regression model was 87.94%, accuracy of support vector machine was 86.52%, k-nearest neighbor was 84.04%, and accuracy of the accuracy of Naïve Bayes was 78.01%. The RF and ANN models outperformed all other models’ accuracies and also other evaluation metrics such as precision, recall, and F1 score.

### 7.1 Summary of Findings

Machine Learning models can reach the-state-of-the-art performance in predicting student academic performance using their demographic, assessment data, and their online learning environment activity, but using demographic data only will give – in most cases – a near average performance.

The following conditions must be met to get the-state-of-the-art prediction accuracy using ML-based models on educational data:

1. The input features must be relevant to student academic performance such as assessments scores, their activity on the VLE, highest level of education, their age, and deprivation level of regions where they live.
2. Relatively large dataset with a large number of observations must be used to allow the model to “learn” from various training examples, and hence improving its performance on test set.
3. Data should be balanced to avoid because imbalanced data cause the model to favor the most represented class in the dataset Important features using different data combinations:
   * Demographics only: o Age band o Studied credits o Disability o IMD band
   * Demographics and assessments scores: the previous features plus the scores
   * Demographics, assessments scores, and VLE: the previous features plus the sum of clicks

### 7.2 Implications for Industry

The results show that RF and ANN can perform very well on educational data as they do on data that come from other field because they can scale with complex datasets. It is important though to give relevant features such as scores and VLE activity. This study supports other studies that suggest using RF and ANN on educational data to predict student academic performance.

### 7.3 Implications for Industry

### 7.3 Future Work

Further investigations can be made to predict student performance using their VLE data using same ML or DL models. Conducting more researches that study other factors affecting student performance such as course attributes and instructors attributes. Also, statistical analysis can be made to find correlations between factors such age, gender, presence of disability, living in deprived areas, and so forth, and their influences on the performance of students. Moreover, recurrent neural networks

(RNNs) can be used for detecting students who are about to withdraw or dropout. Finally, building a dataset that contain more features that are related to VLE and assessments to investigate their role in predicting student performance.

## References

### [Abiodun, Oludare Isaac, et al., 2018]

Abiodun, Oludare Isaac, Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Nachaat AbdElatif Mohamed, and Humaira Arshad. “State-of-the-art in artificial neural network applications: A survey”. Heliyon 4. 2018.

### [Abu Saa, Amjed, et al., 2019]

Abu Saa, Amjed, Mostafa Al-Emran, and Khaled Shaalan. “Factors Affecting Students’ Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques”. Tech Know Learn 24, 567–598., 2019.

### [Adewale, Amoo M., et al., 2018]

Adewale, Amoo M., Alaba O. Bamidele, and Usman O. Lateef. “Predictive modelling and analysis of academic performance of secondary school students: Artificial Neural Network approach”. International Journal of Science and Technology Educational Research 9(1), 1-8, 2018.

### [Ali S, Rasheed Mansoor, and S Perumal, 2022]

Ali S, Rasheed Mansoor, and S. Perumal. “Multi-class LDA classifier and CNN feature extraction for student performance analysis during Covid-19 pandemic”. International Journal of Nonlinear Analysis and Applications, 13(1), pp. 13291339., 2022.

### [Alpaydin, Ethem, 2010]

Alpaydin, Ethem (2010). Introduction to Machine Learning. London: The MIT Press. ISBN 978-0-262-01243-0

### [Amershi, Saleema, and Cristina Conati., 2006]

Amershi, Saleema, and Cristina Conati. “Automatic Recognition of Learner Groups in Exploratory Learning Environments”. Proceedings of ITS 2006, 8th International Conference on Intelligent Tutoring Systems, 463-472., 2006.

### [Amrieh, Elaf Abu, et al., 2016]

Amrieh, Elaf Abu, Thair Hamtini, and Ibrahim Aljarah. “Mining educational data to predict student’s academic performance using ensemble methods”. International Journal of Database Theory and Application 9(8), 119-136., 2016.

### [Anjum, Nadia, and Srinivasu Badugu., 2020]

Anjum, Nadia, and Srinivasu Badugu. “A study of different techniques in educational data mining”. In Advances in decision sciences, image processing, security and computer vision. Springer. pp. 562– 571., 2020.

### [Asiah, Mat, et al., 2019]

Asiah, Mat, Nik Zulkarnaen Khidzir, Safaai Deris, Nik Nurul Hafzan Mat Yaacob, Mohd Saberi Mohamad, and Siti Syuhaida Safaa. “A review on predictive modeling technique for student academic performance monitoring”. In proceedings of materials science, engineering and chemistry (MATEC) web of conferences, 255. 2019.

### [Asif, Raheela, et al., 2017]

Asif, Raheela, Agathe Merceron, Syed Abbas Ali, and Najmi Ghani Haider. “Analyzing undergraduate students’ performance using educational data mining”.

Computers and Education., 2017

### [Athani, Suhas, et al., 2017]

Athani, Suhas, Sharath A. Kodli and Mayur N. Banavasi, and P.G. Sunitha Hiremath. “Student performance predictor using multiclass support vector classification algorithm”. 341-346., 2017.

### [Aydoğdu, Şeyhmus, 2020]

Aydoğdu, Şeyhmus. “Predicting student final performance using artificial neural networks in online learning environments”. Education and Information Technologies 25, 1913–1927., 2020.

### [Bainbridge, Jay, et al., 2015]

Bainbridge, Jay, James Melitski, Anne Zahradnik, Eitel J. M. Lauría, Sandeep Jayaprakash, and Josh Baron. “Using learning analytics to predict at-risk students in online graduate public affairs and administration education”. Journal of Public Affairs Education, 21(2), 247-262., 2015.

### [Baker, Ryan, and Kalina Yacef, 2009]

Baker, Ryan, and Kalina Yacef. “The State of Educational Data Mining In 2009: A Review and Future Visions”. JEDM-Journal of Educational Data Mining, 1 (1), 317., 2009.

### [Baker, Ryan, and Paul Salvador Inventado, 2014]

Baker, Ryan, and Paul Salvador Inventado. (2014). “Educational Data Mining and Learning Analytics”, In: Larusson J., White B. (eds) Learning Analytics. Springer, New York, NY. pp. 61-75., 2014.

### [Bakhshinategh, Behdad, et al., 2018]

Bakhshinategh, Behdad, Osmar R. Zaiane, Samira ElAtia, and Donald Ipperciel. “Educational data mining applications and tasks: A survey of the last 10 years”. Education and Information Technologies 23, 537–553., 2018.

### [Beal, Carole, et al., 2006]

Beal, Carole, Lei Qu, and Hyokyeong Lee. “Classifying learner engagement through integration of multiple data sources”. Proceedings of the 21st National Conference on Artificial Intelligence., 2006.

### [Bellman, Richard, 1957]

Bellman, Richard. “A Markovian Decision Process”. Journal of Mathematics and Mechanics 6. 679-684., 1957.

### [Berland, Matthew et al., 2014]

Berland, Matthew, Ryan S. Baker, and Paulo Blikstein. “Educational data mining and learning analytics: Applications to constructionist research”. Technology, Knowledge and Learning 19, 205–220., 2014.

### [Bishop, Christopher, 2006]

Bishop, Christopher. “Pattern Recognition and Machine Learning”, Springer., 2006. **[Brahim, Ghassen Ben, 2022]**

Brahim, Ghassen Ben. “Predicting Student Performance from Online Engagement Activities Using Novel Statistical Features”. Arabian Journal for Science and Engineering, 2022.

### [Breiman, Leo, et al., 1984]

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. “Classification and regression trees”. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software., 1984.

**[Breiman, Leo, 1996]**

Breiman, Leo. “Bagging predictors”. Machine learning 24(2), 123-140., 1996.

### [Breiman, Leo, 2001]

Breiman, Leo. “Random Forests”. Machine Learning 45(1), 5–32., 2001. **[Broomhead, David S., and David Lowe., 1988]**

Broomhead, David S., and David Lowe. “Radial basis functions, multi-variable functional interpolation and adaptive networks”. Royal Signals and Radar Establishment Malvern (United Kingdom)., 1988.

### [Burgos, Concepción, et al., 2018]

Burgos, Concepción, María L. Campanario, David de la Peña, Juan A. Lara, David

Lizcano, and María A. Martínez.. “Data mining for modeling students’ performance: A tutoring action plan to prevent academic dropout”. Computers and Electrical Engineering., 2018.

### [Chapelle, Olivier, et al., 2006]

Chapelle, Olivier, Bernhard Scholkopf, and Alexander Zien. "Semi-supervised learning (chapelle, o. et al., eds.; 2006) [book reviews]." IEEE Transactions on Neural Networks 20.3., 2009.

### [Chunqiao et al., 2018]

Chunqiao, Mi, Peng Xiaoning, Deng Qingyou. “An Artificial Neural Network Approach to Student Study Failure Risk Early Warning Prediction Based on TensorFlow”. Advanced Hybrid Information Processing (pp.326-333) 2018.

### [Cortez, Paulo and Alice Silva, 2008]

Cortez, Paulo and Alice Silva. “Using Data Mining to Predict Secondary School Student Performance”. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUture

BUsiness TEChnology Conference (FUBUTEC 2008), pp. 5-12, Porto, Portugal, April, 2008,

### [Dada, Emmanuel Gbenga, et al., 2019]

Dada, Emmanuel Gbenga, Joseph Stephen Bassi, Haruna Chiroma, Adebayo Olusola Adetunmbi, and Opeyemi Emmanuel Ajibuwa. "Machine learning for email spam filtering: review, approaches and open research problems." Heliyon 5(6), 2019.

### [Dahl, George E., et al., 2013]

Dahl, George E., Tara N. Sainath, and Geoffrey E. Hinton. "Improving deep neural networks for LVCSR using rectified linear units and dropout." In 2013 IEEE international conference on acoustics, speech and signal processing, pp. 8609-8613. IEEE, 2013.

### [Deng, Li, and Dong Yu, 2014]

Deng, Li, and Dong Yu. "Deep learning: methods and applications." Foundations and trends in signal processing 7(3–4), 197-387., 2014.

### [Baggiyam, N. Dhana, and R. Pankajam, 2017]

Baggiyam, N. Dhana, and R. Pankajam. "Social intelligence in relation to academic achievement." International Journal of Research-GRANTHAALAYAH 5(3), 1822., 2017.

### [Dokas, Paul, et al., 2002]

Dokas, Paul, Levent Ertoz, Vipin Kumar, Aleksandar Lazarevic, Jaideep Srivastava, and Pang-Ning Tan. "Data mining for network intrusion detection." In Procceding NSF Workshop on Next Generation Data Mining, pp. 21-30., 2002.

### [Felix, Igor, et al., 2019]

Felix, I.; Ana Paula Ambrosio, Jacques Duilio, Eduardo Simões. “Predicting student outcome in moodle”. In Proceedings of the Conference: Academic Success in Higher Education, Porto, Portugal, pp. 1–2, 14–15 February, 2019.

### [Gary, Geraldine, et al., 2016]

Gray, Geraldine, Colm McGuinness, Philip Owende, and Markus Hofmann. “Learning Factor Models of Students at Risk of Failing in the Early Stage of Tertiary Education”. Journal of Learning Analytics, 3(2), 330-372., 2016.

### [Geman, Stuart, et al., 1992]

Geman, Stuart, Élie Bienenstock, and René Doursat. "Neural networks and the bias/variance dilemma". Neural Computation 4(1), 1–58., 1992.

### [Goodfellow, Ian, et al., 2016]

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. “6.5 Back-Propagation and Other Differentiation Algorithms”. Deep learning. MIT press, pp. 200–220., 2016.

### [Heise, Natascha, et al., 2020]

Heise, Natascha, Carolyn A Meyer, Brendan A Garbe, Heather A Hall, and Tod R Clapp. “Table Quizzes as an Assessment Tool in the Gross Anatomy Laboratory”. Journal of medical education and curricular development vol. 7. 29 Jul. 2020.

### [Hinton, Geoffrey E., and Terrence Joseph Sejnowski, 1999]

Hinton, Geoffrey E., and Terrence Joseph Sejnowski, eds. Unsupervised learning: foundations of neural computation. MIT press, 1999.

### [Hinton, Geoffrey E., et al., 2006]

Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets". Neural computation 18(7), 1527-1554., 2006.

### [Hinton, Geoffrey E., 2009]

Hinton, Geoffrey E. "Deep belief networks". Scholarpedia 4(5) 5947, 2009. **[Hinton, Geoffrey E., 2007]**

Hinton, Geoffrey E. "Learning multiple layers of representation". Trends in cognitive sciences 11(10), 428-434., 2007.

### [Hlosta, Martin, et al., 2017]

Hlosta, Martin, Zdenek Zdrahal, and Jaroslav Zendulka. "Ouroboros: early identification of at-risk students without models based on legacy data". In Proceedings of the seventh international learning analytics & knowledge conference, pp. 6-15., 2017.

### [Ho, Tin Kam, 1995]

Ho, Tin Kam. "Random decision forests". In Proceedings of 3rd international conference on document analysis and recognition, vol. 1, pp. 278-282. IEEE, 1995.

### [Hu, Bin, 2017]

Hu, Bin. "Teaching quality evaluation research based on neural network for university physical education". In 2017 International Conference on Smart Grid and Electrical Automation (ICSGEA), pp. 290-293. IEEE, 2017.

### [Hussain, Sadiq, et al., 2019]

Hussain, Sadiq, Zahraa Fadhil Muhsion, Yass Khudheir Salal, Paraskevi Theodorou, Fikriye Kurtoglu, and G. C. Hazarika. "Prediction Model on Student Performance based on Internal Assessment using Deep Learning". iJET 14(8), 4-22., 2019.

### [James, Gareth M., 2003]

James, Gareth M. "Variance and bias for general loss functions." Machine learning 51(2), 115-135., 2003.

### [James, Gareth, et al., 2013]

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An introduction to statistical learning. Vol. 112, pp. 3-7. New York: springer, 2013. **[Jayaprakash, Sandeep M., et al., 2014]**

Jayaprakash, Sandeep M., Erik W. Moody, Eitel JM Lauría, James R. Regan, and Joshua D. Baron. "Early alert of academically at-risk students: An open source analytics initiative". Journal of Learning Analytics 1(1), 6-47., 2014.

### [Kaelbling, Leslie Pack, et al., 1996]

Kaelbling, Leslie Pack, Michael L. Littman, and Andrew W. Moore. "Reinforcement Learning: A Survey". Journal of Artificial Intelligence Research 4, 237–285., 1996.

### [Kingma, Diederik P., and Max Welling, 2013]

Kingma, Diederik P., and Max Welling. “Auto-encoding variational bayes”. arXiv preprint arXiv:1312.6114., 2013.

### [Kingma, Diederik P., and Jimmy Ba, 2014]

Kingma, Diederik P., and Jimmy Ba. “Adam: A method for stochastic optimization”. arXiv preprint arXiv:1412.6980., 2014.

### [Kiu, Ching Chieh, 2018]

Kiu Ching Chieh. “Supervised Educational Data Mining to Discover Students’

Learning Process to Improve Students’ Performance”. In: Tang S., Cheah S. (eds)

Redesigning Learning for Greater Social Impact. Springer, Singapore. pp 249-258., 2018.

### [Koutroumbas, Konstantinos and Theodoridis, Sergios, 2008]

Koutroumbas, Konstantinos and Theodoridis, Sergios. Pattern Recognition (4th ed.). Burlington. ISBN 978-1-59749-272-0., 2008.

### [Kumar, Suhas, 2015]

Kumar, Suhas. “Fundamental Limits to Moore’s Law”. arXiv preprint arXiv:1511.05956., 2015.

### [Kuzilek J., Martin Hlosta, Zdenek Zdrahal 2017]

Kuzilek J., Hlosta M., Zdrahal Z. Open University Learning Analytics dataset Sci. Data 4:170171 doi: 10.1038/sdata.2017.171 (2017).

### [Lau, E., et al., 2019]

Lau, E., L. Sun, and Qingping Yang. “Modelling, prediction and classification of student academic performance using artificial neural networks”. SN Applied Sciences. 2019.

### [Li, Shuping, and Taotang Liu, 2021]

Li, Shuping, and Taotang Liu. “Performance Prediction for Higher Education Students Using Deep Learning”. 2021.

### [Von Luxburg, Ulrike, and Bernhard Schölkopf., 2011]

Von Luxburg, Ulrike, and Bernhard Schölkopf. “Statistical learning theory: Models, concepts, and results”. Handbook of the History of Logic. 10: Section 2.4, 2011.

### [McCulloch, Warren S., and Walter Pitts., 1943]

McCulloch, Warren S., and Walter Pitts. “A logical calculus of the ideas immanent in nervous activity”. Bulletin of Mathematical Biophysics 5, 115–133. 1943.

### [Mehryar Mohri, et al., 2018]

Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar. “Foundations of Machine Learning”. The MIT Press, 2018.

### [Merceron, Agathe, and Kalina Yacef., 2008]

Merceron, Agathe, and Kalina Yacef. “Interestingness Measures for Association Rules in Educational Data”. Proceedings of the First International Conference on Educational Data Mining, 57-66., 2008.

### [Miljanovic, Milos., 2012]

Miljanovic, Milos. “Comparative analysis of recurrent and finite impulse response neural networks in time series prediction”. Indian Journal of Computer Science and Engineering 3(1) 180-191., 2012.

**[Mitchell, Tom M., 1997]**

Mitchell, Tom M. “Machine Learning”. McGraw Hill. pp. 2., 1997.

### [Mnih, Volodymyr, et al., 2013]

Mnih, Volodymyr, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. "Playing atari with deep reinforcement learning." arXiv preprint arXiv:1312.5602., 2013.

### [Mohebali, Behshad et al., 2020]

Mohebali, Behshad, Amirhessam Tahmassebi, Anke Meyer-Baese, and Amir H. Gandomi. "Probabilistic neural networks: a brief overview of theory, implementation, and application." Handbook of Probabilistic Models (2020): 347367.

### [Mood, Alexander M., et al., 1973]

Mood, Alexander M., Duane C. Boes, and Franklin A. Graybill. Introduction to the Theory of Statistics (3rd Edition). McGraw-Hill. Section 2.2. 1973.

### [Moore, Gordon E., 1998]

Moore, Gordon E. “Cramming more components onto integrated circuits.” Proceedings of the IEEE 86(1), 82-85., 1998.

### [Moreno-Marcos, Pedro, et al., 2020]

Moreno-Marcos, Pedro, Ting-Chuen Pong, Pedro Merino, and Carlos DelgadoKloos. “Analysis of the Factors Influencing Learners’ Performance Prediction with Learning Analytics”. IEEE Access 8, pp. 5264-5282., 2020.

### [Noprianto, Eko, 2017]

Noprianto, Eko. “Extrovert versus Introvert Students: What EFL Learning Strategy do They Use”. Asian TEFL, 2(2), 119-135., 2017.

### [O’Mahony, Niall, et al., 2019]

O’Mahony, Niall, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, and Joseph Walsh. “Deep learning vs. traditional computer vision.” In Science and Information Conference, pp. 128-144. Springer, Cham, 2019.

### [Pang, Yulei, et al., 2017]

Pang, Yulei, Nicolas Judd, Joseph O'Brien, and Michael Ben-Avie. “Predicting students' graduation outcomes through support vector machines”. 1-8. 2017.

### [Phauk, Sokkhey and Takeo Okazaki, 2020]

Phauk, Sokkhey and Takeo Okazaki. “Developing Web-based Support Systems for Predicting Poor-performing Students using Educational Data Mining Techniques”. International Journal of Advanced Computer Science and Applications. 2020.

**[Quinlan, J. R., 1986]**

Quinlan, J. R. Induction of decision trees. Machine learning, 1(1), 81-106., 1986.

**[Quinlan, J. Ross, 2014]**

Quinlan, J. Ross. C4.5: programs for machine learning. Elsevier., 2014.

### [Rivas, Alberto, et al., 2019]

Rivas, Alberto, Jesús M. Fraile, Pablo Chamoso, Alfonso González-Briones, Sara Rodríguez, and Juan M. Corchado. “Students performance analysis based on machine learning techniques”. In International Workshop on Learning Technology for Education in Cloud, pp. 428-438., Springer, Cham, 2019.

### [Rokach, Lior, and Oded Z. Maimon, 2007]

Rokach, Lior, and Oded Z. Maimon. “Data mining with decision trees: theory and applications”. World scientific 69, 2007.

### [Romero, Cristóbal, and Sebastián Ventura., 2010]

Romero, Cristóbal, and Sebastián Ventura. “Educational data mining: a review of the state of the art”. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618., 2010.

### [Salloum, Said A., et al., 2019]

Salloum, Said A., Mostafa Al-Emran, Khaled Shaalan, and Ali Tarhini. “Factors affecting the E-learning acceptance: A case study from UAE”. Education and Information Technologies, 24(1), 509–530., 2019.

### [Sammut, Claude, and Geoffrey I. Webb, 2011]

Sammut, Claude, and Geoffrey I. Webb, eds. “Encyclopedia of Machine Learning”. Springer Science and Business Media. pp. 100–101., 2011.

### [Samuel, Arthur L., 1959]

Samuel, Arthur L “Some Studies in Machine Learning Using the Game of Checkers”. IBM Journal of research and development, 3(3), pp.210-229., 1959.

### [Scheines, Richard, et al., 1994]

Scheines, Richard, Peter Spirtes, Clark Glymour, and Christopher Meek. TETRAD II: Tools for Discovery. Lawrence Erlbaum Associates, Hillsdale, NJ., 1994.

### [Schmidhuber, Jürgen., 2015]

Schmidhuber, Jürgen. “Deep learning in neural networks: An overview”. Neural Networks. 61, 85–117., 2015.

### [Schwenker, Friedhelm, et al., 2001]

Schwenker, Friedhelm, Hans A. Kestler, and Günther Palm. “Three learning phases for radial-basis-function networks”. Neural Networks. 14 (4–5): 439–458., 2001.

### [Settles, Burr, 2009]

Settles, Burr. "Active Learning Literature Survey". Computer Sciences Technical Report 1648. University of Wisconsin–Madison., 2009.

### [Shahiri, Amirah Mohamed, and Wahidah Husain, 2015]

Shahiri, Amirah Mohamed, and Wahidah Husain. “A review on predicting student's performance using data mining techniques”. Procedia Computer Science 72, 414422., 2015.

### [Sherimon, Vinu, and Sherimon Puliprathu Cherian, 2017]

Sherimon, Vinu, and Sherimon Puliprathu Cherian. “Building a multiple linear regression model to predict students’ marks in a blended learning environment”. In Interactive Mobile Communication, Technologies and Learning, pp. 903-911. Springer, Cham, 2017.

### [Shi, Shaohuai, et al., 2016]

Shi, Shaohuai, Qiang Wang, Pengfei Xu, and Xiaowen Chu. “Benchmarking stateof-the-art deep learning software tools”. In 2016 7th International Conference on Cloud Computing and Big Data (CCBD), pp. 99-104. IEEE, 2016.

### [Siemens, George, 2013]

Siemens, George. “Learning analytics: The emergence of a discipline”. American Behavioral Scientist 57(10), 1380-1400, 2013.

### [Silver, David, et al., 2016]

Silver, David, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser et al. “Mastering the game of Go with deep neural networks and tree search”. nature 529(7587), 484-489., 2016.

### [Specht, Donald F., 1991]

Specht, Donald F. “A general regression neural network." IEEE transactions on neural networks 2(6), 568-576., 1991.

### [Stehman, Stephen V., 1997]

Stehman, Stephen V. “Selecting and interpreting measures of thematic classification accuracy”. Remote sensing of Environment 62(1), 77-89., 1997.

### [Tokic, Michel, and Günther Palm., 2011]

Tokic, Michel, and Günther Palm. “Value-difference based exploration: adaptive control between epsilon-greedy and softmax”. In Annual conference on artificial intelligence, pp. 335-346. Springer, Berlin, Heidelberg, 2011.

### [Troussas, Christos, et al., 2020]

Troussas, Christos, Akrivi Krouska, and Maria Virvou. "Using a multi module model for learning analytics to predict learners’ cognitive states and provide tailored learning pathways and assessment." In Machine Learning Paradigms, pp. 9-22. Springer, Cham, 2020.

### [Hamza, Turabieh, et al., 2021]

Turabieh, Hamza, Sana Al zwari, Mahmoud Rokaya, Wael Alosaimi, Abdullah Alharbi, Wajdi Alhakami, and Mrim Alnfiai. “Enhanced Harris Hawks optimization as a feature selection for the prediction of student performance”.

Computing 103, 1417–1438. 2021.

### [Umair, Sajid, and Muhammad Majid Sharif, 2018]

Umair, Sajid, and Muhammad Majid Sharif. “Predicting students grades using artificial neural networks and support vector machine”. In Encyclopedia of Information Science and Technology, Fourth Edition, pp. 5169-5182. IGI Global, 2018.

### [Van Otterlo, Martijn, and Marco Wiering, 2012]

Van Otterlo, Martijn, and Marco Wiering. "Reinforcement learning and markov decision processes." In Reinforcement learning, pp. 3-42. Springer, Berlin, Heidelberg, 2012.

### [Vapnik, Vladimir, 2013]

Vapnik, Vladimir. The nature of statistical learning theory. Springer science & business media, 2013.

### [Weng, John J., 1993]

Weng, John J., Narendra Ahuja, and Thomas S. Huang. “Learning recognition and segmentation of 3-D objects from 2-D images”. In 1993 (4th) International Conference on Computer Vision, pp. 121-128. IEEE, 1993.

### [Wolpert, David H., 1996]

Wolpert, David H. “The lack of a priori distinctions between learning algorithms”. Neural computation 8(7), 1341-1390., 1996.

### [Wook, Muslihah, et al., 2017]

Wook, Muslihah, Zawiyah M. Yusof, Mohd Zakree, and Ahmad Nazri. “Educational data mining acceptance among undergraduate students”. Education and Information Technologies 22(3), 2017.

### [Zhang, Wei, et al., 1990]

Zhang, Wei, Kazuyoshi Itoh, Jun Tanida, and Yoshiki Ichioka. “Parallel distributed processing model with local space-invariant interconnections and its optical architecture." Applied optics 29(32), 4790-4797., 1990.

### [Zhu, Xiaojin Jerry, 2005]

Zhu, Xiaojin Jerry. "Semi-supervised learning literature survey”. University of Wisconsin-Madison Department of Computer Sciences, 2005.