Smart and Explainable Analysis of University Dress Code Compliance using Computer Vision: A Hybrid Approach Combining YOLOv8, EfficientNet-B0, and Grad-CAM

Donatien Kadima Muamba
Université Révérend KIM
Kinshasa/ RDC

Bernice Mambula Zuiya
Université Protestante au Congo
Kinshasa / RDC

Jean-Jacques Kabasubabo Kalombo
ISC / Matadi
MATADI/ RDC

Hénoc Kabengela Umira
Université Protestante au Congo
Kinshasa / RDC

ABSTRACT
The adherence to dress code regulations in universities and higher education institutions is an institutional concern that may require more effective, consistent and transparent monitoring approaches. In this paper, we describe the design and implementation of VestiAI, an explainable artificial intelligence system for automated university dress-code compliance assessment. The proposed hybrid architecture consists of three complementary components : YOLOv8 to detect and localize the clothing items in the images, EfficientNet-B0 to classify the attire in two categories (compliant and non-compliant), and Grad-CAM to explain the visual regions that contribute to the classification decisions. Compared to conventional methods that only offer a final classification result, the presented framework integrates automated assessment with visual explanations, thus improving the interpretability of the decision-making process. The dataset contains 1,337 images which are divided into train, validation and test sets. For clothing detection, YOLOv8 achieved an overall precision of 81.09%, recall of 83.97%, mAP@50 of 86.51%, and mAP@50-95 of 59.27%. For compliance classification, the accuracy, precision, sensitivity and F1-score of EfficientNet-B0 was 81.77%, 79.65%, 86.54% and 82.95% respectively on the 203-image test set. The class-wise detection results also showed the performance variations between the clothing categories, with the best performance on pants with a mAP@50 of 93.99% and the more challenging shoes and hats with the more strict mAP@50-95. Then a web application based on FastAPI was developed to provide a practical interface for image submission and for automatic assessment. The system generates a report detailing the detected garments, the estimated compliance status, and Grad-CAM visualizations that highlight the regions that influenced the classification decision. Results demonstrate the viability of integrating object detection, deep image classification, and explainable artificial intelligence for automatic university dress-code evaluation. The proposed framework aims to support decision making and human validation, not to replace human judgment.

General terms
Artificial Intelligence, Computer Vision, Deep Learning, Image Classification, Object Detection, Explainable Artificial Intelligence

Keywords
Artificial intelligence, computer vision, dress code compliance, YOLOv8, EfficientNet-B0, Grad-CAM, explainability, Deep Learning, academic environment.

1. INTRODUCTION
Artificial intelligence and computer vision [11], [22], [34] are rapidly developing in many fields such as healthcare, security, education, and institutional management. Thanks to advances in deep learning, computers are now able to automatically analyze images and identify objects or visual features with increasing accuracy [5], [31], [33].

Within universities, adherence to dress code rules can be a significant element of discipline and institutional identity. However, it can be difficult to consistently ensure manual monitoring of compliance with dress standards, especially in institutions with large student populations. Moreover, entirely human assessments can be influenced by subjectivity and variations in judgment [10], [11].

The use of computer vision enables an automated approach to analyzing clothing elements present in an image [1], [16], [22]. However, this automatic classification system must go beyond simply delivering a final decision. It is also important to understand the reasons behind this decision in applications based on artificial intelligence. This issue necessitates the integration of explainable artificial intelligence methods, notably Grad-CAM [6], [7], [11], [13].

This work therefore presents the design of an intelligent system, called VestiAI, for the automatic analysis of clothing compliance. The developed approach relies on a pipeline comprising YOLOv8 [4], [16] for detecting the main clothing items, EfficientNet-B0 [5] for binary image classification, and Grad-CAM [6], [13] for providing a visual explanation of the model's decisions.

Therefore, the main objective of this study is to design and implement a system capable of analyzing a clothing image, identifying the garments present, determining whether the outfit is compliant or non-compliant based on the training data, and providing a visualization of the regions that contributed to the classification [4], [5], [6], [11], [23], [24].

2. SCIENTIFIC CONTRIBUTIONS OF THIS STUDY
This work has led to several scientific advances in the field of computer vision and artificial intelligence, with the aim of ensuring compliance with dress code rules in higher education institutions [16], [22].

Our first contribution is the design of an integrated intelligent device combining several complementary processes. These include: identification of clothing using YOLOv8 [4], [16], assessment of their compliance using EfficientNet-B0 [5], and visualization of model decisions using Grad-CAM [6], [13]. The proposed approach does not merely rely on a general classification of the image, but instead employs a multi-step process to increase analysis accuracy and facilitate interpretation of the results.

The second contribution is the addition of explainable artificial intelligence (XAI) to the decision-making process [10], [11], [15]. Grad-CAM enables visualization of the image regions that most contributed to the model's judgment [6], [13]. Explainability of decisions makes the system more transparent, facilitates understanding of its operation, and helps gain user trust—an aspect often overlooked in automated control systems [7], [8], [9].

The third contribution addresses usage and context. To our knowledge, there is little academic work that enables automated compliance checking of clothing [23], [24], [25]. VestiAI precisely presents itself as a response to this need, offering a decision-support tool for academic administrators. The artificial intelligence was trained on data up to October 2023. The system automatically checks whether an outfit complies with the rules in effect, but the competent authority makes the final decision. In fact, artificial intelligence should assist humans, not replace them [10], [11].

The fourth contribution is the development of a functional web application using FastAPI that validates the proposed approach. The application automatically recognizes clothing from the photo provided by the user, checks its compliance, and generates Grad-CAM explanatory maps associated with the predictions [6], [13]. This complete implementation demonstrates that VestiAI can be used in a real-world environment, serving as a proof of concept for future institutional deployments [31], [32], [35].

This study thus opens a new path for the evolution of intelligent clothing analysis systems, combining high detection capability, explanatory procedures, and targeted adaptation to the needs of higher education institutions [11], [15], [23], [24]. It thus contributes to advances in research on explainable computer vision systems in the field of governance and compliance [10], [11], [15].

Table 1: Scientific Positioning of VestiAI

Criterion	*Classical Works*	*VestiAI*
Clothing detection	✓	✓
Compliance classification	✓	✓
Explainable AI (Grad-CAM)	Rarely	✓
Integrated pipeline	Rarely	✓
Operational web interface	Rarely	✓
Application in academic setting	Very little studied	✓

3. MATERIALS AND METHODS

3.1. General System Architecture

The proposed system is based on a multi-step architecture. A clothing image is first provided to the system through a web interface. This image is then analyzed by the YOLOv8 model to identify the various clothing items present, such as hats, shirts, pants, shoes, jackets, or dresses.

The image is then submitted to the EfficientNet-B0 model to perform a binary classification of the outfit. The system then assigns one of the following two classes:

- Compliant
- Non-compliant

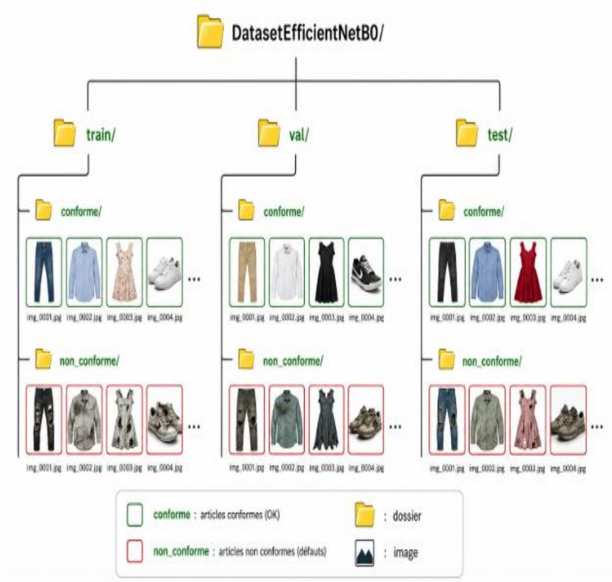


Figure 1. EfficientNetB0 dataset structure.

Finally, the Grad-CAM method is used to generate a heatmap indicating the regions of the image that contributed most to the model's decision.

The system's operation relies on a sequential processing pipeline illustrated in Figure 2. Each module intervenes at a specific stage of the analysis.

The overall system pipeline can be represented as follows :

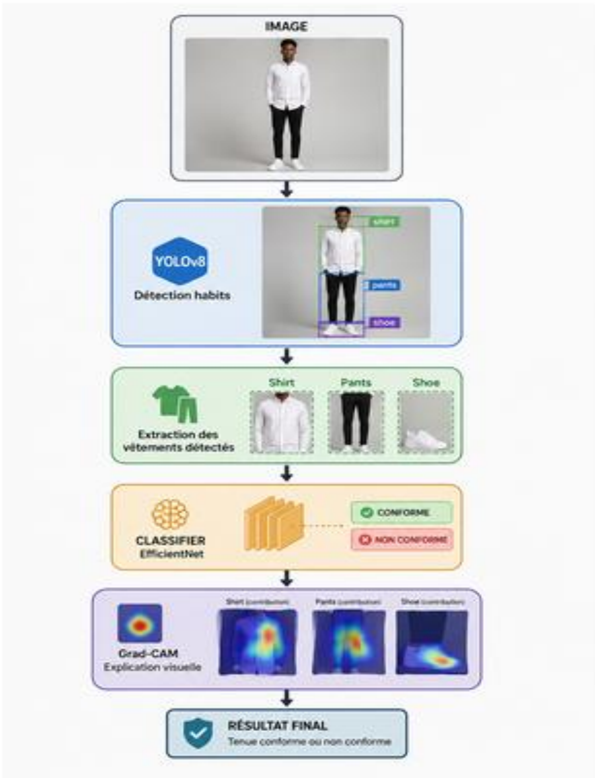


Figure 2. Integrated processing pipeline of the VestiAI system

The proposed architecture thus combines object detection, image classification, and explainability of deep learning models.

4. RESULTS AND DISCUSSION

4.1 Clothing Detection Performance Using YOLOv8

The first stage of the proposed hybrid framework consists of detecting clothing items using YOLOv8. The model was evaluated on the test set using precision, recall, mean Average Precision at an Intersection over Union (IoU) threshold of 0.50 (mAP@50), and mean Average Precision averaged over IoU thresholds from 0.50 to 0.95 (mAP@50–95). The global detection results are presented in Table 2

Table 2. Overall performance of the YOLOv8 clothing detection model

Evaluation Metric	*YOLOv8 Performance*
Precision	81.09%
Recall	83.97%
mAP@50	86.51%
mAP@50–95	59.27%

The model achieved a precision of 81.09% and a recall of 83.97%. These results indicate that YOLOv8 correctly identifies a large proportion of the clothing objects while maintaining a relatively controlled number of false detections. The mAP@50 of 86.51% demonstrates good detection performance when an IoU threshold of 0.50 is used. However, the mAP@50–95 decreases to 59.27%, reflecting the more demanding localization criterion obtained by evaluating the predictions across IoU thresholds ranging from 0.50 to 0.95. This difference indicates that object recognition is generally effective, while the precise alignment of some bounding boxes remains more challenging under strict localization conditions.

To provide a more comprehensive evaluation, the detection performance was also analyzed separately for each clothing category. The class-wise results are presented in Table 3.

Table 3. Class-wise performance of the YOLOv8 clothing detector

Clothing Class	*Precision (%)*	*Recall (%)*	*mAP@50 (%)*	*mAP@50–95 (%)*
Hat	79.83	77.03	84.49	44.98
Jacket	76.41	83.42	87.18	65.45
Shirt	82.23	85.04	87.46	60.53
Pants	91.67	89.15	93.99	72.10
Shorts	74.17	91.21	83.26	48.00
Skirt	82.92	87.03	90.80	70.07
Dress	80.54	84.06	85.91	68.35
Shoe	80.94	74.86	78.95	44.68

The class-wise analysis reveals noticeable differences between clothing categories. The pants category obtained the highest performance, with 91.67% precision, 89.15% recall, 93.99% mAP@50, and 72.10% mAP@50–95. These results indicate that pants were comparatively easier for the detector to identify and localize accurately.

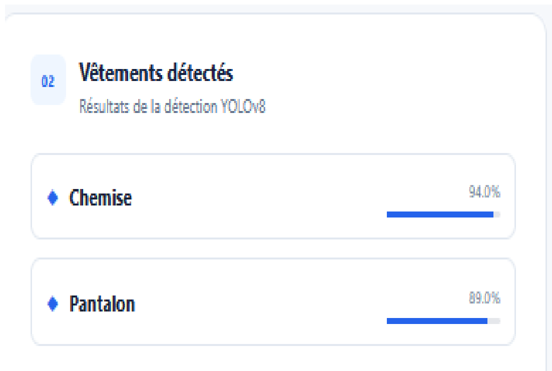
The skirt category also achieved strong performance, particularly with a mAP@50 of 90.80% and a mAP@50–95 of 70.07%. Similarly, the jacket and shirt categories obtained mAP@50 values of 87.18% and 87.46%, respectively. These results demonstrate that YOLOv8 maintains consistent detection capability across several major clothing categories.

The shorts category presents an interesting behavior. Although its recall reached 91.21%, its precision was lower at 74.17%. This indicates that the detector successfully identifies most actual shorts instances but also generates a relatively higher number of false-positive detections for this category. Such behavior suggests that some visually similar clothing patterns may be confused with shorts.

The hat and shoe categories were comparatively more challenging under the strict mAP@50–95 criterion, with scores of 44.98% and 44.68%, respectively. Their recall values were also lower than those of several other categories, reaching 77.03% for hats and 74.86% for shoes. Small object size, partial visibility, occlusion, and variations in appearance may contribute to the difficulty of accurately detecting and localizing these categories. Importantly, their relatively lower mAP@50–95 values do not imply that the objects are systematically missed; rather, they indicate that the predicted bounding boxes are more sensitive to stricter IoU requirements.

Overall, the class-wise results confirm that the performance of the detector is not uniform across all clothing categories. The model performs particularly well for larger and visually distinctive categories such as pants and skirts, whereas smaller or potentially occluded categories such as shoes and hats remain more challenging. This detailed analysis provides a more informative assessment than global metrics alone and identifies specific categories for future improvement.

The YOLOv8 detection results therefore establish a quantitative foundation for the proposed hybrid framework. With an overall mAP@50 of 86.51%, the detector provides reliable clothing localization for the subsequent compliance analysis. The class-wise evaluation additionally highlights the strengths and weaknesses of the detection stage, which is important for interpreting the performance of the EfficientNet-B0 classification stage that follows.



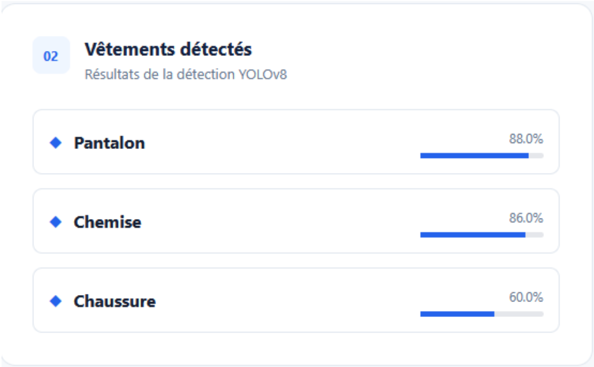


Figure 3. Example of garment element detection by YOLOv8

4.2. Clothing Compliance Classification Using EfficientNet-B0

After the clothing detection stage, EfficientNet-B0 was employed to classify the detected clothing appearance according to the university dress-code requirements. The classifier was evaluated on a test set containing 203 images. The resulting confusion matrix comprised 76 true negatives (TN), 90 true positives (TP), 23 false positives (FP), and 14 false negatives (FN).

Table 4. Confusion matrix of the EfficientNet-B0 classifier

Actual Class	*Predicted Compliant*	*Predicted Non-Compliant*
Compliant	76 (TN)	23 (FP)
Non-Compliant	14 (FN)	90 (TP)
Total	90	113

The classification performance derived from this confusion matrix is presented in Table 5.

Table 5. Performance metrics of the EfficientNet-B0 classifier

Evaluation Metric	*Value*
Accuracy	81.77%
Precision	79.65%
Recall (Sensitivity)	86.54%
F1-score	82.95%

The model correctly classified 166 out of the 203 test images, resulting in an overall accuracy of 81.77%. The precision of 79.65% indicates that most samples predicted as non-compliant were actually non-compliant. The recall reached 86.54%, showing that the classifier successfully identified the majority of non-compliant cases. The F1-score of 82.95% confirms a relatively balanced trade-off between precision and recall.

The confusion matrix provides additional information about the types of classification errors. The model correctly identified 90 non-compliant cases, whereas 14 non-compliant cases were classified as compliant. These 14 false-negative predictions represent cases in which a potential dress-code violation was not detected. In contrast, 23 compliant cases were incorrectly classified as non-compliant, generating false-positive predictions.

From an application perspective, the higher recall than precision is relevant to an automated dress-code compliance system. A high recall reduces the probability of overlooking non-compliant cases. However, the 23 false-positive predictions show that the system may occasionally flag compliant clothing as non-compliant. Such errors are particularly important in practical deployment because an automated system should avoid generating unnecessary alerts or incorrect assessments.

The difference between recall (86.54%) and precision (79.65%) therefore reflects a model that prioritizes the detection of potentially non-compliant cases. This behavior is consistent with a monitoring-oriented application, but it also indicates that further refinement of the classification stage could reduce false-positive decisions.

The F1-score of 82.95% provides a single balanced measure of classification quality and confirms that the EfficientNet-B0 model achieves a satisfactory compromise between detecting non-compliant cases and limiting incorrect positive predictions. Nevertheless, the remaining false positives and false negatives indicate that visually ambiguous clothing patterns remain challenging for the classifier.

These results also demonstrate the complementary roles of the two principal components of the proposed framework. YOLOv8 provides clothing-object detection and localization, while EfficientNet-B0 performs the subsequent compliance-oriented classification. The combination enables the system to move beyond simple object detection toward an application-specific decision regarding dress-code compliance.

Therefore, the EfficientNet-B0 results provide quantitative evidence that the proposed classification stage is capable of distinguishing compliant and non-compliant clothing patterns with an overall accuracy of 81.77% and a sensitivity of 86.54%. The subsequent Grad-CAM analysis further examines whether these classification decisions are based on visually meaningful clothing regions.

The resulting confusion matrix is as follows:

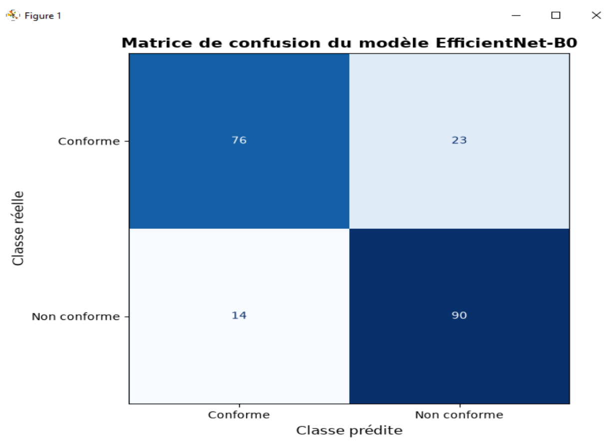


Figure 4. Confusion matrix of the EfficientNet-B0 model

4.3. Explainability Analysis Using Grad-CAM

To improve the interpretability of the classification decisions, Grad-CAM (Gradient-weighted Class Activation Mapping) [4] was applied to the EfficientNet-B0 model. Grad-CAM produces a visual representation of the image regions that contribute most strongly to the predicted class. In the context of VestiAI, this analysis is particularly important because a classification decision regarding dress-code compliance should be supported by visually meaningful clothing characteristics rather than by irrelevant background information.

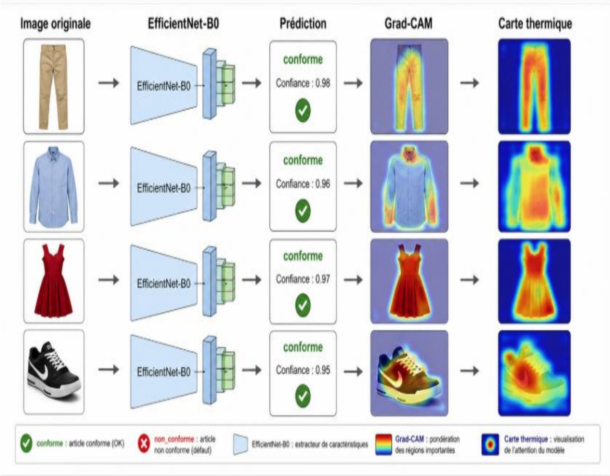


Figure 5. Image analysis process

The generated Grad-CAM visualizations show the regions of the input images that contributed most strongly to the model's predictions. For compliant samples, the activation maps generally highlight relevant clothing regions, such as shirts, trousers, skirts, jackets, or other visible garments.

For non-compliant samples, the activation regions provide an indication of the clothing characteristics that influenced the model toward the non-compliant class.

The qualitative analysis of these visualizations provides an additional level of validation for the classification stage. When the highlighted regions correspond to the relevant garments, the visual explanation is consistent with the intended decision process of the system. Conversely, activation concentrated on irrelevant background regions, faces, or other non-clothing areas would indicate a potential weakness in the learned representation.

Grad-CAM does not modify the classification prediction itself ; rather, it provides an interpretability layer that helps users understand the visual evidence associated with the model's decision. This characteristic is important for a university monitoring application, where automated predictions may require human verification before any administrative action is taken.

The integration of Grad-CAM therefore contributes to the explainability of VestiAI by transforming the output of the EfficientNet-B0 classifier into a visually interpretable representation. Combined with YOLOv8 detection and EfficientNet-B0 classification, this component makes the proposed framework not only capable of detecting and classifying clothing but also capable of providing visual evidence supporting its predictions.

This step improves the interpretability of the system by providing not only a prediction but also a visual indication of the areas that contributed to that prediction.

The confusion matrix shown in Figure 6 allows for a more precise analysis of correct predictions and classification errors of the model.

Figure 6. Grad-CAM visualization of regions that contributed to the model's decision

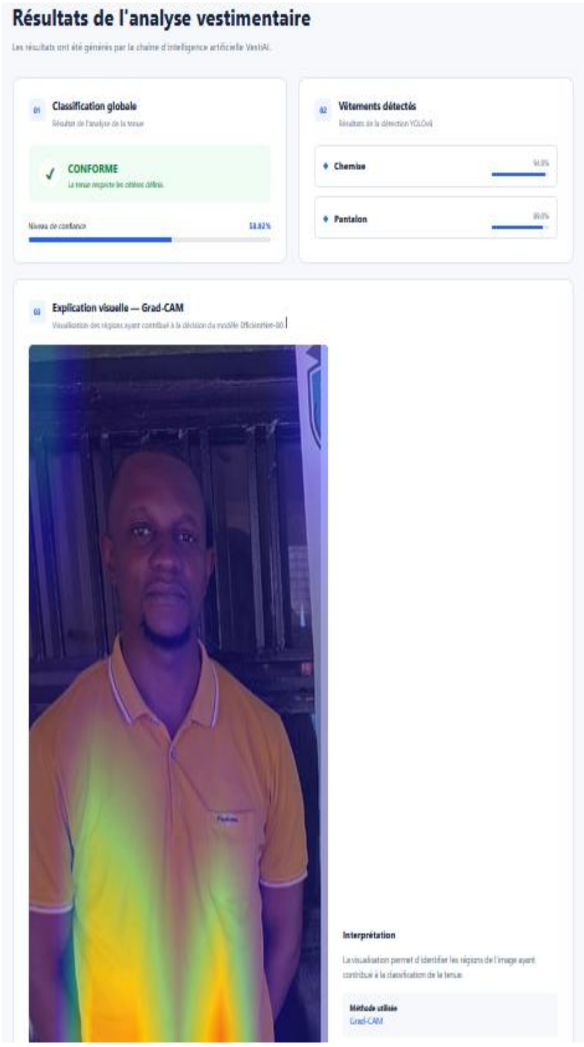
This approach makes it possible to observe whether the model focuses on relevant clothing areas, such as the upper body, pants, skirt, or other visible elements in the image.

Thus, Grad-CAM improves the interpretability of the system and allows the user to better understand the visual elements that influenced the model's decision.

4.4. Integration and Deployment Results

The three principal components of the proposed framework were integrated into a sequential processing pipeline. First, YOLOv8 detects and localizes the clothing items present in an input image. The detected clothing information is then processed by EfficientNet-B0 to determine whether the observed clothing corresponds to the defined compliant or non-compliant categories. Finally, Grad-CAM generates a visual explanation of the classification decision.

The resulting workflow can be summarized as follows:



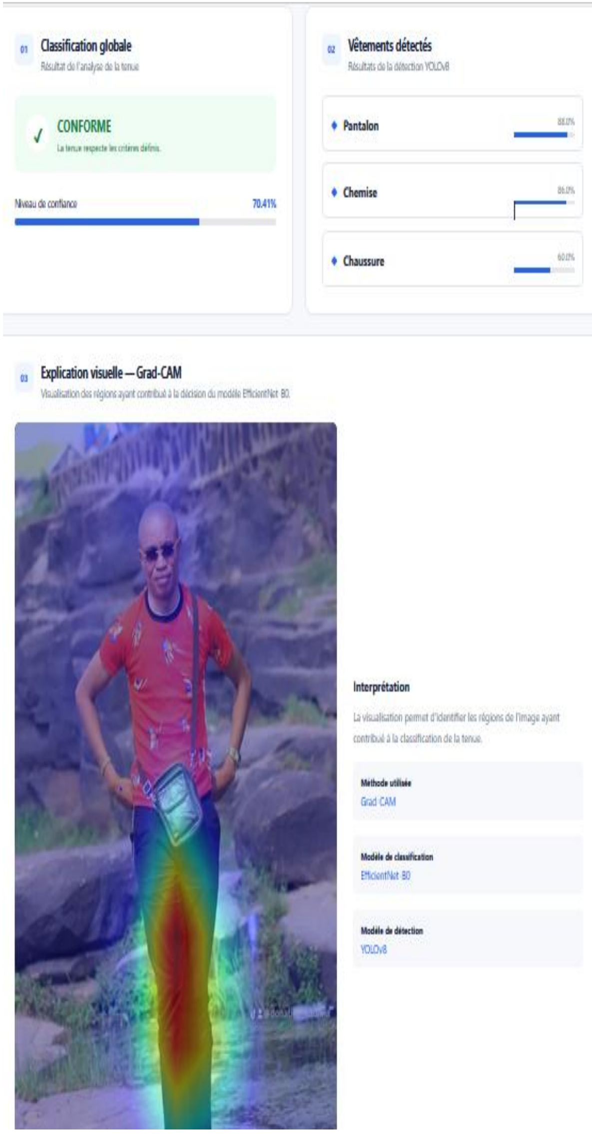


Figure 7. Final web interface of the VestiAI system

This integration enables the system to move from low-level object detection to application-specific decision making. YOLOv8 provides the localization information required to identify relevant clothing regions, while EfficientNet-B0 performs the compliance classification. Grad-CAM complements these predictions by indicating the image regions that contributed to the classification decision.

The experimental results demonstrate that the individual components provide complementary capabilities. YOLOv8 achieved an overall precision of 81.09%, recall of 83.97%, mAP@50 of 86.51%, and mAP@50–95 of 59.27%. EfficientNet-B0 achieved an accuracy of 81.77%, precision of 79.65%, recall of 86.54%, and F1-score of 82.95%. These results indicate that both the detection and classification stages provide a quantitative basis for the final dress-code assessment.

The integration of these components is particularly relevant because clothing compliance cannot be adequately addressed by object detection alone. Detecting a shirt, jacket, skirt, or pair of trousers does not by itself determine whether the complete appearance satisfies the institutional requirements. The subsequent classification stage provides the application-specific decision, while Grad-CAM improves transparency by showing the visual evidence associated with that decision.

The resulting hybrid architecture consequently combines detection, classification, and explainability within a single workflow. This design provides a practical foundation for intelligent dress-code assessment while retaining a human-interpretable component that can support the verification of automated decisions.

Integrating artificial intelligence models into a web application allows the user to submit an image and automatically obtain the analysis results. The final system interface is shown in Figure 7

4.5. Evaluation Scope and Generalization

The experimental evaluation provides quantitative evidence for the performance of the proposed framework on the dataset used in this study. The analysis includes both global and class-wise evaluation of the YOLOv8 detector, as well as confusion-matrix-based evaluation of the EfficientNet-B0 classifier and qualitative analysis using Grad-CAM.

The class-wise YOLOv8 evaluation provides additional insight into the robustness of the detection stage by revealing differences between clothing categories. In particular, pants achieved the highest mAP@50 (93.99%), whereas shoes and hats presented lower mAP@50–95 values (44.68% and 44.98%, respectively). This analysis demonstrates that the evaluation is not limited to a single global performance indicator and allows category-specific weaknesses to be identified.

However, the present study is based on a single primary dataset. Consequently, the reported results should be interpreted within the scope of the evaluated data distribution. Variations in illumination, camera viewpoint, clothing appearance, occlusion, image quality, and the presence of multiple individuals may affect the performance of the proposed system in real-world environments.

Evaluation using independent datasets collected from different university environments and under more diverse acquisition conditions would therefore be valuable for further assessing the generalization capability of the framework. Such external validation constitutes an important direction for future experimental work.

5. COMPARISON WITH EXISTING STUDIES

Table 1. Comparison of existing studies and positioning of VestiAI

Ref.	*Author / Year*	*Application Domain*	*Main Method*	*Explainable AI (XAI)*	*Real-Time*	*Limitations*
[23]	Liu et al., 2016	Clothing Recognition	DeepFashion + CNN	✗	✗	Classification only; no object detection
[24]	Ge et al., 2019	Clothing Detection and Segmentation	DeepFashion2	✗	Partial	High model complexity
[25]	Han et al., 2017	Clothing Compatibility	Bi-LSTM	✗	✗	No garment localization
[26]	Huang et al., 2015	Clothing Image Retrieval	CNN	✗	✗	Unable to detect multiple objects
[28]	Chen et al., 2015	Clothing Attribute Analysis	Deep CNN	✗	✗	Lack of model explainability
[16]	Wang et al., 2023	Object Detection	YOLOv7	✗	✓	No decision interpretability
[17]	Wang et al., 2024	Object Detection	YOLOv9	✗	✓	Optimized for accuracy only
[18]	Wang et al., 2024	Object Detection	YOLOv10	✗	✓	No visual explanation
[19]	Carion et al., 2020	Object Detection	DETR	✗	Partial	Relatively slow inference
[21]	Zhao et al., 2024	Real-Time Object Detection	RT-DETR	✗	✓	Less suitable for resource-constrained devices
Our Approach	**VestiAI**	**University Dress Code Compliance Monitoring**	**YOLO + Attribute Analysis + Grad-CAM**	✓	✓	**First approach integrating object detection, explainability, and university dress code compliance assessment**

Table 1 shows that most previous studies focus on clothing recognition, detection, or recommendation. The latest YOLO models ([16]–[18]) stand out for their remarkable real-time detection performance, while DeepFashion and DeepFashion2 ([23], [24]) are recognized as benchmarks in the field of garment analysis. However, none of these methods combines all of the following:

- Knowledge of clothing;
- Assessment of garment compliance;
- A graphical representation of the model's choices

VestiAI fills this gap by integrating an object detection model coupled with a Grad-CAM-based explainability approach, allowing the system's decisions to be visually justified.

Table 2. Comparison of detection method performances

Ref.	*Model*	*Dataset*	*mAP (%)*	*Real-Time*	*(XAI)*	*Application Domain*
[3]	YOLOv4	COCO	43.5	✓	✗	General Object Detection
[16]	YOLOv7	COCO	56.8	✓	✗	General Object Detection
[17]	YOLOv9	COCO	55.6	✓	✗	General Object Detection
[18]	YOLOv10	COCO	56.4	✓	✗	General Object Detection
[19]	DETR	COCO	42.0	Partial	✗	General Object Detection
[21]	RT-DETR	COCO	54.8	✓	✗	General Object Detection
[23]	DeepFashion	DeepFashion	—	✗	✗	Clothing Classification
[24]	DeepFashion2	DeepFashion2	—	Partial	✗	Clothing Detection
VestiAI	**YOLO + Grad-CAM**	**VestiDataset**	**To be completed after experimental evaluation**	✓	✓	**University Dress Code Compliance Assessment**

Table 2 shows that YOLO architectures are currently the reference models for real-time object detection, due to their excellent precision/speed trade-off. The DeepFashion and DeepFashion2 datasets remain benchmarks for clothing analysis, but they lack any explainability mechanism.

VestiAI stands out from existing approaches by combining fast clothing detection with visual interpretation of decisions using Grad-CAM. This combination improves system transparency and meets the specific requirements of dress code compliance monitoring in the university context.

6. DISCUSSION OF RESULTS

The results show that the proposed approach, combining YOLOv8, EfficientNet-B0, and Grad-CAM, is a relevant solution for implementing an intelligent system for automatic dress code compliance assessment. With an accuracy of 81.77% and an F1 score of 82.95%, the results of the EfficientNet-B0 model demonstrate that it can effectively discriminate between compliant and non-compliant outfits, despite the limited volume of data used for training [5], [31], [34].

Analysis of per-class metrics shows that the "non-compliant" class has a sensitivity (recall) of 86.54%, which is a very good result. This performance indicates that the system successfully identifies most outfits that violate dress code rules. This capability is essential for an application designed to assist universities, as it helps avoid missing cases of non-compliance. However, false positives and false negatives show that the model is not error-free and still has room for improvement [16],[17],[18].

These classification errors can be explained by several factors. The number of observations in the dataset is quite limited, which, on the one hand, prevents the model from generalizing well to diverse situations. Furthermore, the diversity of outfits, variations in people's postures, lighting conditions, background differences, and partial occlusions of clothing all affect the quality of predictions. Additionally, the notion of dress code compliance depends on the specific rules of each institution and may include an element of subjectivity, which poses an additional challenge for any automated system. [23], [24], [25], and [27].

Beyond these limitations, the results obtained clearly demonstrate the feasibility of using deep learning techniques for automatic analysis of dress code compliance in a university context. They show that it is possible to effectively assist academic administrators in their monitoring tasks while reducing the time required for thesis evaluation [1], [5], [16], [31].

Another important contribution of this research is the integration of Grad-CAM. Unlike many approaches that merely provide a prediction, Grad-CAM allows visualizing the areas of the image that most influenced the model's decision, thereby enabling a better understanding of the algorithm's reasoning and verifying that predictions are indeed based on relevant clothing items rather than extraneous information. [13], [6].

In the school context, where adherence to dress code rules can have serious repercussions for students, it is crucial that artificial intelligence algorithms be transparent, traceable, and explainable [10], [11], [15]. By making decisions visually explicit, Grad-CAM increases user trust, facilitates verification of results by university administrators, and promotes more responsible use of artificial intelligence [7], [8], [11].

Finally, the results show that VestiAI is not just an automatic classification system but truly an intelligent decision-support system combining object detection, classification, explainability, and operational deployment in a web application. This multidimensional approach aligns with recent recommendations in the literature on explainable computer vision systems and opens new perspectives for developing reliable artificial intelligence applications in institutional environments [11], [15], [16], [23], [31], [35].

7. LIMITATIONS OF THE STUDY

This study nevertheless presents certain limitations. First, the dataset used includes 1,337 images, which remains relatively limited for training a computer vision system intended to operate under highly varied conditions. Second, the images do not necessarily cover all body types, poses, lighting conditions, and situations that may be encountered in a real university environment.

Furthermore, the classification of dress code compliance depends on the criteria defined for the dataset. The system therefore does not have a general understanding of decency or compliance; rather, it learns the features present in the training examples.

Finally, the performance achieved may vary depending on the quality and context of the images submitted to the system.

8. AREAS FOR IMPROVEMENT

There are a number of opportunities for improvement. If the number of images and the diversity of the dataset increase, the model's generalization ability would improve. It would also be interesting to introduce more variations in morphology, poses, lighting, and contexts.

EfficientNet-B0 can be compared to other deep learning architectures such as EfficientNet-B2, EfficientNet-B3, or other more recent architectures. Annotations with a higher level of precision as well as a more representative dataset can facilitate clothing detection.

Finally, another explainability method besides Grad-CAM could be used jointly to obtain a more complete explanation of the model's decisions. The system can also be deployed on a server to be accessible by various tools at the Islamic University in Congo.

9. CONCLUSION

The objective of this study was to design and develop VestiAI, an intelligent system for analyzing dress code compliance based on computer vision and deep learning. The proposed approach combines automatic clothing detection using YOLOv8, dress code compliance assessment via EfficientNet-B0, and prediction visualization through Grad-CAM [5], [6], [11].

During testing, this technique proved capable of automating the evaluation of clothing outfits in an educational setting. The system can identify the main garments in an image, immediately assess the appropriateness of the outfit, and present the model's choices. With an accuracy of 81.77%, precision of 79.65%, recall of 86.54%, and an F1 score of 82.95%, the results attest to the reliability of the proposed approach as a decision support tool in this domain. The trials conducted as part of this work show that deep learning methods are capable of solving this problem.

This research integrated explainable artificial intelligence (XAI) into the core of the analysis process. Grad-CAM makes it possible to visualize the regions of the image that most influenced the model's decision, thereby increasing transparency, interpretability, and trust in the system [6], [11], [13], unlike methods that merely provide a prediction. This capability is important in an institutional environment where decisions must be clear, controllable, and justifiable.

This research also shows that artificial intelligence can assist decision-making within higher education institutions. This sophisticated system is not intended to replace human judgment, but rather to assist academic leaders in managing attire by automating repetitive tasks, while leaving the final responsibility to the appropriate entity, in line with the principles of ethical and transparent artificial intelligence. [10], [11], [15].

In summary, VestiAI can be improved and extended in many respects. It would be interesting to broaden and diversify the dataset to increase the model's ability to generalize. It would also be interesting to explore more innovative detection and classification structures, integrate a finer analysis of clothing features, and test the system in a real university setting. These advances will strengthen the system and enable new applications in fields that require automatic, precise, and interpretable image processing. [17], [18], [21] and [22].

This is therefore a major advancement for the explainability of computer vision. It proposes a method integrating object detection, classification, explainability, and online deployment to automate the examination of compliance with dress code rules. This makes it possible to design intelligent, transparent, efficient systems tailored to the needs of higher education institutions.

10. BIBLIOGRAPHIE

[1] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. *"You Only Look Once: Unified, Real-Time Object Detection."* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91

[2] Redmon, J., Farhadi, A. *"YOLO9000: Better, Faster, Stronger."* Proceedings of CVPR, 2017, pp. 7263–7271. DOI: 10.1109/CVPR.2017.690

[3] Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M. *"YOLOv4:*

Optimal Speed and Accuracy of Object Detection." arXiv preprint arXiv:2004.10934, 2020.

[4] Jocher, G., Chaurasia, A., Qiu, J. *"YOLO by Ultralytics."* Ultralytics, 2023. Disponible via : https://github.com/ultralytics/ultralytics

[5] Tan, M., Le, Q. *"EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks."* Proceedings of ICML, 2019, pp. 6105–6114.

[6] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. *"Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization."* International Journal of Computer Vision, 2020, 128, pp. 336–359. DOI: 10.1007/s11263-019-01228-7

[7] Ribeiro, M. T., Singh, S., Guestrin, C. *"Why Should I Trust You? Explaining the Predictions of Any Classifier."* Proceedings of ACM SIGKDD, 2016, pp. 1135–1144.

[8] Lundberg, S. M., Lee, S. I. *"A Unified Approach to Interpreting Model Predictions."* Advances in Neural Information Processing Systems (NeurIPS), 2017.

[9] Samek, W., Wiegand, T., Müller, K. R. *"Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models."* IT Information Technology, 2017.

[10] Gunning, D., Aha, D. *"DARPA's Explainable Artificial Intelligence (XAI) Program."* AI Magazine, 2019, 40(2), pp. 44–58.

[11] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., et al. *"Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI."* Information Fusion, 2020, 58, pp. 82–115. DOI: 10.1016/j.inffus.2019.12.012

[12] Doshi-Velez, F., Kim, B. *"Towards A Rigorous Science of Interpretable Machine Learning."* arXiv:1702.08608, 2017.

[13] Chattopadhyay, A., Sarkar, A., Howlader, P., Balasubramanian, V. N. *"Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks."* IEEE Winter Conference on Applications of Computer Vision (WACV), 2018.

[14] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A. *"Learning Deep Features for Discriminative Localization."* CVPR, 2016.

[15] Tjoa, E., Guan, C. *"A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI."* IEEE Transactions on Neural Networks and Learning Systems, 2021.

[16] Wang, C. Y., Bochkovskiy, A., Liao, H. Y. M. *"YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors."* CVPR, 2023.

[17] Wang, C. Y., Yeh, I. H., Liao, H. Y. M. *"YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information."* arXiv:2402.13616, 2024.

[18] Wang, A., Chen, H., Liu, L., et al. *"YOLOv10: Real-Time End-to-End Object Detection."* arXiv:2405.14458, 2024.

[19] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S. *"End-to-End Object Detection with Transformers."* ECCV, 2020.

[20] Zhu, X., Su, W., Lu, L., Li, B., Wang, J. *"Deformable DETR: Deformable Transformers for End-to-End Object Detection."* ICLR, 2021.

[21] Zhao, Y., Lv, W., Xu, S., et al. *"DETRs Beat YOLOs on Real-Time Object Detection."* CVPR, 2024.

[22] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. *"An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale."* ICLR, 2021.

[23] Liu, Z., Luo, P., Qiu, S., Wang, X., Tang, X. *"DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations."* CVPR, 2016.

[24] Ge, Y., Zhang, R., Wang, X., Tang, X., Luo, P. *"DeepFashion2: A Versatile Benchmark for Detection, Pose Estimation, Segmentation and Re-Identification of Clothing Images."* CVPR, 2019.

[25] Han, X., Wu, Z., Jiang, Y. G., Davis, L. S. *"Learning Fashion Compatibility with Bidirectional LSTMs."* ACM Multimedia, 2017.

[26] Huang, J., Feris, R., Chen, Q., Yan, S. *"Cross-Domain Image Retrieval with a Dual Attribute-Aware Ranking Network."* ICCV, 2015.

[27] Liu, S., Feng, J., Song, Z., et al. *"Hi, Magic Closet: Personalized Outfit Recommendation Using Deep Learning."* ACM Multimedia, 2017.

[28] Chen, Q., Huang, J., Feris, R., Brown, L., Dong, J., Yan, S. *"Deep Domain Adaptation for Describing People Based on Fine-Grained Clothing Attributes."* CVPR, 2015.

[29] Yang, W., Luo, P., Lin, L. *"Clothing Co-Parsing by Deep Convolutional Neural Networks."* ICCV, 2014.

[30] Zheng, S., Song, Y., Leung, T., Goodfellow, I. *"Improving the Robustness of Deep Learning for Clothing Recognition."* IEEE Transactions on Image Processing, 2017.

[31] Sze, V., Chen, Y. H., Yang, T. J., Emer, J. *"Efficient Processing of Deep Neural Networks: A Tutorial and Survey."* Proceedings of the IEEE, 2017.

[32] Han, S., Mao, H., Dally, W. J. *"Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding."* ICLR, 2016.

[33] Deng, L., Yu, D. *"Deep Learning: Methods and Applications."* Foundations and Trends in Signal Processing, 2014.

[34] Goodfellow, I., Bengio, Y., Courville, A. *"Deep Learning."* MIT Press, 2016.

[35] Géron, A. *"Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow."* O'Reilly Media, 2022.