Indicateurs émotionnels et posturaux pour l'intervention pédagogique précoce : modéliser la difficulté comme une déviation de la réussite individuelle de l'étudiant.

Edouard Nadaud[1,2], Lionel Prevost[1,2], Bénédicte Le Grand[2], Bastien Tarot[1,3]

[1] ESIEA, ESIEALAB-LDR (Learning Data Robotics)

[2] Paris 1 Panthéon Sorbonne, CRI (Centre de Recherche en Informatiques)

[3] ESSCA School of Management

[1]prénom.nom@esiea.fr [2]prénom.[Le-]nom@univ-paris1.fr [3]Prénom.NOM@essca.fr

Résumé

Face à la « cécité attentionnelle » inhérente à l'enseignement hybride, nous présentons SEPT, un cadre "privacy-by-design" analysant les trajectoires émotionnelles et physiques captées pendant les sessions de travail des étudiants via une webcam. En modélisant la difficulté comme une déviation de la signature personnelle de réussite plutôt que par des normes de groupe, notre approche améliore la détection de l'échec +17%. Le système permet une alerte précoce robuste dès les premiers tiers de l'activité, sans compromis significatif sur la performance. Respectueux de la vie privée, le cadre SEPT a des performances équivalentes aux méthodes biométriques intrusives pour identifier efficacement les 10-15% d'étudiants en détresse.

Mots-clés

Informatique affective, trajectoires émotionnelles, applications d'apprentissage automatique, IA éthique, analyse de l'apprentissage, mouvements du visage.

Abstract

To address the "attentional blindness" inherent in blended learning, we present SEPT, a privacy-by-design framework that analyzes learners' emotional and physical trajectories from webcam feeds. By modeling difficulty not as a deviation from a group norm, but as a departure from the student's individual "signature of success," our approach improves detection by 17%. The system enables robust early warning within the first 30% of activity. Completely privacy-respecting, SEPT achieves performance equivalent to intrusive biometric methods (AUC 0.67) in effectively identifying students at risk of dropping out.

Keywords

Affective Computing, Emotion Trajectories, Machine-Learning Applications, Ethical AI, Learning analytics, Head movements.

1 Introduction

Dans les environnements d'apprentissage à distance et les environnements d'apprentissage à distance et hybrides, les enseignants sont confrontés à une « cécité attentionnelle » car il leur est impossible de suivre les signaux non verbaux d'une cohorte entière pour identifier les étudiants en détresse. Si l'informatique affective offre des solutions automatisées pour restaurer ce lien pédagogique, le domaine se heurte à deux obstacles majeurs.

Premièrement, de nombreux modèles agrègent les états émotionnels sur une session (ex. « frustration moyenne »), écrasant ainsi la dimension temporelle. Mathématiquement, ces approches échouent à distinguer la « difficulté productive » (une confusion transitoire menant à la compréhension) de la « frustration destructrice » (une difficulté persistante menant au désengagement), car ces deux états peuvent produire des effets et des états mentaux identiques ou proches.

Deuxièmement, les avancées récentes en Apprentissage Profond, telles que les *Transformers* ou les LSTM, modélisent certes ces dynamiques temporelles, mais introduisent une barrière de « soif de données » (Data Hunger). Ces architectures exigent des jeux de données annotés massifs pour éviter le surapprentissage, ce qui les rend inapplicables dans les contextes de classe typiques (« *Small Data* »). De plus, elles reposent souvent sur un traitement vidéo brut, soulevant de graves préoccupations éthiques au regard du RGPD [1] ou l'AI ACT européen [2].

L'objectif de cette étude est de concilier une analyse temporelle granulaire avec des contraintes strictes de confidentialité. Nous proposons un cadre *privacy-by-design* exploitant la dynamique des trajectoires comportementales pour permettre une détection robuste et précoce de la difficulté académique.

L'article est organisé comme suit : la section 2 sera consacrée à l'état de l'art. La section 3 présentera nos principales contributions méthodologiques. Dans la section 4 seront détaillés nos résultats expérimentaux. La section 5 conclura nos travaux et sera suivie d'une discussion et de perspectives en section 6.

2 État de l'art

2.1 Informatique Affective et Engagement

La Théorie Contrôle-Valeur [3] établit que les émotions d'accomplissement modulent l'attention et l'effort cognitif [4]. Ces états sont généralement repérés via le modèle

dimensionnel de Russell [5] (Valence/Arousal). Toutefois, la réalité pédagogique est plus nuancée : des états a priori négatifs comme la confusion peuvent être bénéfiques s'ils viennent à être résolus [4]. Bien que les méta-analyses confirment la corrélation entre succès et émotion positive [3], la réalité" est plus nuancée. La confusion, bien que négative, peut être productive si elle est transitoire et résolue (problème de l'impasse cognitive) ; lorsqu'elle persiste, elle conduit au désengagement.

Dans ce contexte, le concept d'engagement en tant qu'objet multidimensionnel combinant des aspects comportementaux (effort, persistance), émotionnels (intérêt, valeur) et cognitifs (profondeur de traitement) devient central [6]. Les émotions agissent comme les antécédents immédiats de cet engagement : un étudiant peut ressentir de la frustration (émotion) mais maintenir son effort (engagement comportemental) jusqu'à un point de rupture. Cependant, les méthodes traditionnelles pour identifier ces états souffrent de limites significatives : s'appuyer sur l'auto-évaluation par questionnaire a posteriori ne permet pas de passer à l'échelle, tandis que les capteurs de signaux physiologiques comme l'EEG sont intrusifs. Le défi actuel est donc d'inférer ces états internes complexes uniquement à partir de signaux externes observables sans perturber les activités académiques.

2.2 Représentation de l'Engagement en Informatique

Pour relever ce défi, l'analyse de la dynamique temporelle de l'apprentissage nécessite des descripteurs continus (comme la valence et l'arousal) capables de discriminer les états affectifs. Cette approche est ancrée dans la théorie de la Cognition Incarnée (*Embodied Cognition*) [7, 8], qui postule que les processus cognitifs, tels que la charge mentale ou la résolution de problèmes, se manifestent par des ajustements moteurs subtils et involontaires. Dans ce contexte, les mouvements physiques, particulièrement ceux de la tête et du regard, sont critiques.

Bien que le regard soit unanimement reconnu comme le marqueur le plus direct pour le foyer visuel de l'attention [9], l'oculométrie traditionnelle impose des contraintes de calibration et matérielles incompatibles avec un déploiement écologique à grande échelle. Pour répondre à ces limites, l'Estimation de la Pose de la Tête (HPE) a émergé comme une alternative robuste. Des études comparatives récentes démontrent une forte corrélation entre l'orientation de la tête et la direction du regard dans des contextes d'interaction humain-écran, validant l'utilisation de l'HPE comme un estimateur fiable de l'attention en conditions naturelles [10].

Les travaux en analyse de données d'apprentissage ont identifié des rôles sémantiques distincts pour les différents degrés de liberté de la tête [11] :

- Lacet (*Yaw*) : La recherche utilisant des réseaux de neurones convolutifs a établi que la variance de l'angle de lacet discrimine efficacement entre la concentration sur le contenu pédagogique et les distractions environnementales (« vagabondage mental »), offrant une mesure quantitative de la persistance de l'attention [12].
- Roulis (*Roll*) : Des études plus récentes sur la communication non verbale associent l'inclinaison de la tête à un marqueur d'intérêt ou de curiosité, mais aussi, en cas de persistance, à de la confusion ou à une tentative de résoudre une dissonance cognitive [13]. Contrairement aux mouvements de navigation, un roulis signale souvent un changement interne de perspective face à une information ambiguë [14].

Ainsi, la littérature suggère que la combinaison des dynamiques émotionnelles (expressions faciales) et des micromouvements de la tête (HPE) offre une représentation de l'engagement plus riche que l'un ou l'autre canal pris isolément [10].

2.3 Des dynamiques temporelles à la reconnaissance précoce

Bien que l'informatique affective ait fait des progrès significatifs dans la détection d'états émotionnel [15], les modèles pédagogiques prédominants souffrent encore d'un « angle mort temporel ». De nombreuses approches actuelles reposent largement sur l'agrégation statistique, telle que la « valence moyenne » à l'échelle de la session ou la fréquence des émotions basiques d'Ekman pour prédire la performance [16]. Ce lissage statistique est mathématiquement réducteur : une valence moyenne, « neutre », est identique pour les deux profils radicalement opposés de l'étudiant totalement désengagé et de celui oscillant entre une frustration productive et des moments de compréhension [13, 17].

Les limites de l'agrégation sont bien documentées. Une étude majeure analysant les émotions d'élèves de collège n'a trouvé aucune différence significative dans le temps moyen passé dans la confusion ou l'ennui entre les élèves performants et ceux en difficulté [18]. L'absence de corrélation robuste entre les moyennes émotionnelles simples et la performance académique suggère que ce n'est pas la présence d'une émotion qui détermine le résultat, mais sa dynamique. C'est la persistance (durée) et la transition (séquence) des états qui sont prédictives : une brève frustration suivie d'un réengagement est bénigne [19] ; une frustration persistante est délétère. Par conséquent, le défi n'est plus de classifier ce que l'étudiant ressent à l'instant *t*, mais de comprendre comment sa trajectoire émotionnelle évolue dans le temps.

En suivant l'évolution pas à pas des états affectifs, on capture les dynamiques affectives : les émotions surgissent, se transforment ou s'estompent continuellement durant l'apprentissage. Cette vue dynamique s'aligne avec les théories modernes de l'affect et de la cognition, qui postulent une interaction entre émotions, engagement et processus cognitifs dans le temps [20]. Ainsi, connaître la succession des émotions (et pas seulement leur moyenne) donne un aperçu de la manière dont l'apprenant réagit aux défis cognitifs, les régule ou subit leur impact. Les théories de la régulation émotionnelle soutiennent cette vue : la capacité d'un apprenant à gérer la frustration, surmonter la confusion ou maintenir l'engagement se déploie nécessairement dans le

temps à travers une série d'ajustements.
Ces constats motivent la modélisation des signaux affectifs et comportementaux comme des séries temporelles multivariées.

2.4 Modélisation de Séries Temporelles : Vers la Reconnaissance Précoce d'Action

La trajectoire émotionnelle adopte une approche inspirée de la Reconnaissance Précoce d'Action (*Early Action Recognition*) en vision par ordinateur. Dans ce paradigme, l'objectif est de classifier une séquence (par ex., un geste humain) dès ses premières images, sans attendre son achèvement. Transposé à l'éducation, nous postulons qu'une « trajectoire d'échec » (caractérisée, par exemple, par une hésitation prolongée couplée à une désorganisation du regard) possède une signature temporelle identifiable bien avant la validation finale de la réponse par l'étudiant.
De plus, les approches actuelles en Apprentissage Profond (Transformers, LSTM) tente de résoudre cela en modélisant des séquences, mais elles introduisent une nouvelle barrière : la « Voracité en Données ». Ces architectures nécessitent des jeux de données annotés massifs qui les rendant inapplicables dans les contextes de classe typiques de « Small Data » (taille de la cohorte inférieure à 100 individus) où les contraintes RGPD limitent la collecte de données [12]. Contrairement aux actions physiques discrètes, les états cognitifs internes comme la « confusion » manquent de vérité terrain visuelle non ambiguë. Définir un étalon-or (gold standard) nécessite soit une annotation experte coûteuse, qui souffre d'une faible fiabilité inter-juges due à l'interprétation subjective des états, soit des auto-évaluations des étudiants, qui perturbent leurs activités et introduisent un biais.
Nous utilisons ici des mesures de similarité élastique, spécifiquement la DTW (*Dynamic Time Warping*) [21]. Contrairement à la distance Euclidienne, qui exige un alignement temporel rigide, la DTW permet l'alignement non linéaire de deux séries en minimisant les effets de distorsion temporelle. Cette propriété est cruciale pour modéliser le rythme cognitif : deux étudiants peuvent suivre le même chemin intellectuel (lecture → réflexion → résolution) mais à des vitesses différentes. En traitant l'évolution des signaux multimodaux comme une forme dynamique plutôt que comme une séquence d'états indépendants, la DTW permet la quantification de la similarité structurelle entre une tentative en cours et des trajectoires typiques, facilitant ainsi une classification précoce et explicable.
Cette étude poursuit deux objectifs de recherche principaux :

- Proposer une approche robuste et respectueuse de la vie privée pour identifier les signes de difficulté académique en utilisant des données non intrusives et conformes à la confidentialité.
- Valider le potentiel du système pour l'alerte précoce, permettant un soutien opportun avant que l'échec ne survienne.

Nous démontrons que cette approche atteint une utilité d'alerte précoce comparable aux méthodes invasives de l'état de l'art, mais avec une architecture entièrement respectueuse de la vie privée adaptée au déploiement dans le monde réel.

3 Proposition et méthodologie

Pour aller au-delà de l'état de l'art, nous proposons le cadre SEPT (Figure 1) afin de répondre aux questions de recherche ouvertes identifiées ci-dessus. Cela nous permet de modéliser les dynamiques temporelles de la difficulté sans compromettre la vie privée de l'étudiant. Nous postulons que l'étudiant possède une signature cinématique multimodale durant son parcours d'apprentissage, et que la difficulté se manifeste comme une signature de déviation identifiable par rapport à cette dernière.

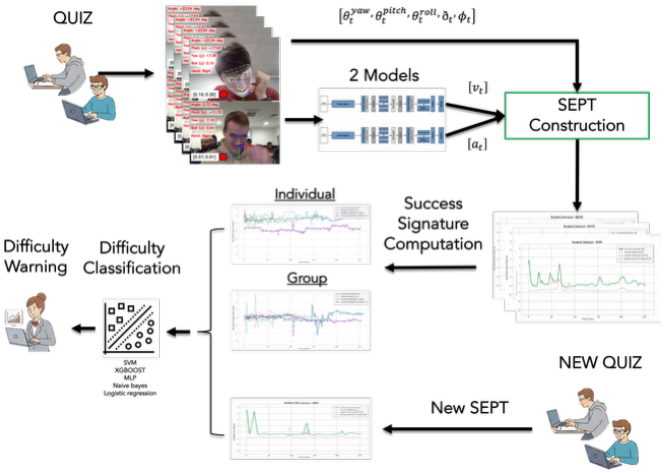


Figure 1 : Prédire la difficulté de l'étudiant pendant l'évaluation en utilisant le cadre SEPT

3.1 Formalisation de l'Espace d'Engage-ment Multimodal

Nous définissons l'engagement multimodal de l'apprenant au temps t comme un vecteur $x_t \in \mathbb{R}^7$ intégrant des états affectifs internes et des marqueurs comportementaux externes dérivés des captures de la webcam. Nous combinons les états internes (l'affect) avec leurs manifestations physiques externes pour capturer l'expérience vécu par l'étudiant :

$$x_t = [v_t, a_t, \theta_t^{yaw}, \theta_t^{pitch}, \theta_t^{roll}, \eth_t, \phi_t]$$

Inférence Affective (Valence v_t/Arousal a_t) : Afin de capturer les dynamiques émotionnelles continues, nous avons développé une architecture de réseau de neurones convolutifs (CNN) spécifiquement optimisée pour l'inférence en temps réel à partir de flux vidéo non contraints (webcam). Face à l'absence de modèles open source, nous avons conçu et entraîné notre propre réseau profond (de bout en bout). Nous nous affranchissons ainsi de la dépendance à des réseaux pré-entraînés génériques souvent sous-optimaux pour cette tâche spécifique. Le modèle repose sur une

architecture efficiente structurée autour de cinq blocs convolutionnels séquentiels.

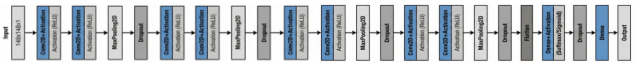


Figure 2 : Architecture du Réseau de Neurones Convolutifs pour la prédiction de la Valence et de l'Arousal

Le modèle a été entraîné sur les jeux de données de référence *AffectNet* [22] et *AffWild* [23]. Il utilise l'erreur quadratique moyenne (MSE) comme fonction de perte et le coefficient de corrélation de concordance (CCC) comme métrique. Comme illustré dans le Tableau 1, nos deux modèles (un pour la valence et un pour l'arousal) atteignent un CCC de 0,714 pour la prédiction de la valence et de 0,644 pour la prédiction de l'arousal. Ces résultats rivalisent avec la performance de l'état de l'art [24], tout en étant spécifiquement calibrés pour la variabilité inhérente aux environnements éducatifs non contrôlés.

Tableau 1. Comparaison de performance des Modèles pour la détection de valence et d'arousal sur *AffectNet* (la meilleure performance est en gras, la seconde soulignée).

	Valence		Arousal	
Modèles	CCC↑	RMSE↓	CCC↑	RMSE↓
CAGE 2024[24]	**0.716**	**0.331**	0.642	**0.305**
VGG-F 2021[22]	0.710	0.356	0.629	0.326
Nos modèles	*0.714*	*0.339*	**0.644**	*0.323*

Caractéristiques de pose de la tête :

- $\theta_t^{yaw}, \theta_t^{pitch}, \theta_t^{roll}$ correspondent aux angles d'Euler de la pose de la tête, indicateurs de l'attention visuelle et de la cognition incarnée (par ex., réflexion, confusion).
- $\eth_t$ est la distance normalisée à l'écran, indiquant l'engagement postural (approche/évitement). Nous n'estimons pas la distance absolue visage-caméra de l'étudiant ; au lieu de cela, nous quantifions les changements de distance à partir de l'échelle de la boîte englobant le visage (détecté en utilisant *OPENCV* [25] et *YoloV5* [26]). Sous le modèle perspectif, la taille apparente de l'image d'un objet varie inversement avec la profondeur, donc la taille de la boîte englobante fournit un proxy pratique pour la variation de distance.
- ϕ_t représente l'estimation de l'angle du regard.

Par conséquent, l'engagement de l'étudiant est modélisé comme SEPT [27] $\mathcal{T}$, défini comme une série temporelle ordonnée de longueur L :

$$\mathcal{T}_i = \{x_1, x_2, \dots, x_L\} \in \mathbb{R}^{7\times L}$$

3.2 Modélisation des Signatures de parcours de réussite : Signatures de Groupe vs. Individuelles

Plutôt que d'entraîner un classificateur binaire sur la classe hétérogène "Échec", nous définissons la difficulté comme une déviation d'une "Signature de Réussite". En adoptant ce point de vue, la difficulté académique est une divergence structurelle par rapport à ces signatures de parcours de réussite établies. Nous définissons une « signature de parcours de réussite » construite à partir des données des étudiants les plus performants. Cette approche offre un avantage décisif pour les contextes éducatifs « *Small Data* » : elle ne nécessite des données denses que pour la classe majoritaire (succès), rendant le modèle robuste à la rareté des échantillons d'échec. Nous formalisons deux topologies de référence distinctes : les signatures de parcours de réussite du groupe (M_{group}) et les signatures de parcours de réussite individuelles (μ_{ind}).

Signature de Parcours de Réussite Individuelle (L'Auto-Signature). Nous postulons que chaque étudiant possède sa propre signature d'engagement de réussite. Par exemple, un étudiant peut présenter une agitation productive (mouvements de tête rythmiques pendant la réflexion), tandis qu'un autre reste immobile. Un modèle global pourrait négliger la spécificité individuelle et interpréter l'agitation du premier comme de la distraction. Pour filtrer ces signatures d'engagement de réussite propres, nous construisons une référence basée sur la propre histoire de succès de l'étudiant. Soit ψ_{ind} l'ensemble des trajectoires historiques pour l'étudiant i associées à une performance réussie (note g ≥ 50% de la note maximale). Puisque la moyenne arithmétique de séries temporelles multivariées temporellement déformées est mal définie, nous définissons la Signature de parcours de réussite individuelle μ_{ind} comme le médoïde de l'ensemble ψ_{ind} des trajectoires de l'étudiant. Formellement :

$$\mu_{ind} = arg\,min_{\mathcal{T}\in\psi_{ind}} \sum_{\mathcal{T}'\in\psi_{ind}} D_{DTW}(T, T')$$

μ_{ind} représente la trajectoire historique spécifique qui est la plus centrale au comportement de réussite de l'étudiant, capturant ses habitudes posturales et émotionnelles en l'absence de difficultés académiques.

Signatures de parcours de réussite de Groupe (La Signature Sociale). Pour adresser le problème de démarrage à froid (étudiants sans historique) et pour capturer des modes universels d'engagement, nous définissons une topologie de groupe. Cependant, nous reconnaissons qu'il n'y a pas une seule « bonne façon » d'apprendre ; le succès est multimodal (par ex., « concentration stoïque » vs. « Exploration dynamique »). Nous agrégeons les trajectoires de réussites de toute la cohorte dans un ensemble $\psi_{all} = \cup_i \psi_{ind}$. Nous appliquons un clustering hiérarchique agglomératif sur la matrice de dissimilarité DTW de ψ_{all} pour partitionner l'espace en K clusters distincts $\{C_1, C_2, \dots, C_k\}$. Pour chaque cluster k, nous identifions son médoïde μ_k .

L'ensemble des Signatures de parcours de réussite de groupe est défini comme la collection de ces trajectoires :

$$M_{group} = \{\mu_1, \mu_2, \dots, \mu_k\}$$

Cet ensemble M_{group} forme une variété représentant les signatures de parcours de réussite de groupe distinctes. Celles-ci ne représentent pas une seule « bonne façon » d'apprendre, mais plutôt les modes distincts d'engagement efficace permis au sein du système éducatif. La note g sert uniquement de filtre transitoire pour peupler ψ_{ind} et n'est jamais stockée. Puisque chaque étudiant quel que soit sa moyenne générale présente des séquences réussies, cette approche évite un profilage de capacité ; elle modélise l'alignement momentané de l'engagement, pas le niveau académique général.

3.3 Quantification de la Difficulté comme Déviation

Nous définissons la difficulté comme une déviation par rapport aux signatures de parcours de réussite établies dans la Section 3.2. Le score de difficulté $\delta(T_{new})$ quantifie la dissimilarité entre une nouvelle trajectoire T_{new} et les signatures de référence en utilisant la déformation temporelle dynamique (DTW).

Déviation de Signature Individuelle (Figure 3). Sous l'hypothèse de l'auto-référence, la difficulté est définie comme la déviation par rapport à la signature historique spécifique de l'étudiant μ_{ind}, filtrant efficacement le comportement personnel :

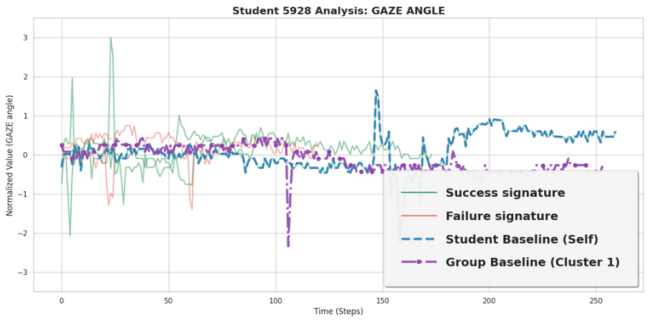
$$\delta_{ind}(T_{new}) = D_{DTW}(T_{new}, \mu_{ind})$$

Cette formulation réduit l'analyse de trajectoire en haute dimension à un scalaire univarié, qui sert d'entrée pour la fonction de décision *f* décrite dans la section suivante.

Déviation de Signature de Groupe (Figure 3). Sous l'hypothèse de la norme sociale, la difficulté est définie comme la distance à la plus proche Signature de Parcours de Réussite valide au sein de la variété de groupe M_{group}. Cette formulation accommode la nature multimodale du succès :

$$\delta_{group}(T_{new}) = min_{\mu_k \epsilon M_{group}}(D_{DTW}(T_{new}, \mu_k))$$

La déviation de signature est calculée pour chacune des sept dimensions de l'engagement.



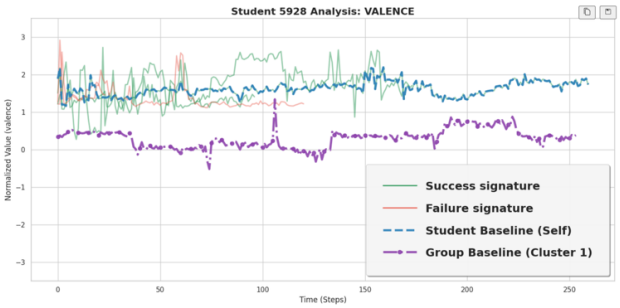


Figure 3 : Déviation de signature de l'Étudiant 5928 (focus sur l'angle du regard et la valence). Les trajectoires d'échec (rouge) montrent une déviation significative de la signature personnelle de l'étudiant (bleu).

Afin d'inférer l'état de difficulté de l'apprenant, le vecteur multivarié des déviations calculé sur les sept dimensions de l'engagement, sert d'entrée à un algorithme de classification supervisée. La valeur cible de cet apprentissage est une variable binaire issue de la vérité terrain objective : la classe positive (y=1 justifiant une alerte) caractérise une trajectoire menant à un échec (score inférieur à 50% de la note maximale), tandis que la classe négative (y=0) désigne une réussite.

Afin d'identifier l'architecture la plus performante pour cette tâche, nous avons évalué et comparé plusieurs familles d'algorithmes, englobant à la fois des modèles linéaires et non linéaires. Plus précisément, nos expérimentations ont porté sur la Régression Logistique et les Machines à Vecteurs de Support (SVM) pour les approches linéaires, ainsi que sur la classification bayésienne naïve, les Forêts Aléatoires, le Perceptron Multicouche (MLP) et l'algorithme XGBoost pour les approches non linéaires.

Dans le cas des modèles linéaire la décision finale du modèle, qui déclenche l'alerte précoce à destination du formateur, repose sur le franchissement d'un seuil de décision définie par un seuil λ. La contribution majeure de notre approche réside dans l'instanciation de ce seuil :

- **Pour l'approche de groupe** : L'algorithme optimise un hyperplan séparateur global et un seuil unique λ_{group} appliqués uniformément à l'ensemble de la cohorte.
- **Pour l'approche individuelle :** L'algorithme intègre une frontière de décision adaptative reposant sur un seuil personnalisé λ_i propre à chaque étudiant *i*. Cette calibration individualisée est cruciale : elle permet de moduler la sensibilité de l'alerte en fonction de la variance comportementale historique inhérente à l'étudiant. Ceci évite les faux positifs pour des apprenants présentant naturellement une forte expressivité physique ou émotionnelle en situation de réussite.

3.4 Stratégie d'Alerte Précoce et Protocole

Pour opérationnaliser la détection de difficultés pour un système de tutorat intelligent (STI) réactif, nous définissons la fonction de prédiction d'échec $f : T \rightarrow \{0,1\}$. Cette

fonction mappe la trajectoire d'un étudiant T_{new} vers un état binaire (0 : Succès/Normal, 1 : Difficulté/Anomalie). Le lien entre la difficulté quantifiée $\delta(T_{new})$ (définie dans la section 3.3 comme la distance aux signatures de référence) et la prédiction f est établi via un seuil de décision λ. Nous postulons qu'un étudiant a des problèmes si sa déviation comportementale excède cette frontière apprise :
$f(T_{new}) = 1[\delta(T_{new}) > \lambda]$

Le choix de cette approche par mesure de similarité élastique (DTW) plutôt que par apprentissage profond (Deep Learning) est motivé par les contraintes spécifiques du contexte pédagogique "*in-the-wild*". Le Tableau 2 positionne SEPT par rapport aux paradigmes dominants de l'état de l'art.

Tableau 2. Positionnement de SEPT face aux études récentes de détection précoce.

Méthode	Caractéristiques techniques	Volume de données requis
Deep Learning (Bi-LSTM, Transformers)[28]	Modélisation de séquences complexes via réseaux profonds.	Très élevé ("Data Hunger").
Ensemble Learning (XGBoost, Random Forest)[29]	Classification basée sur des moyennes statistiques	Moyen.
Notre approche SEPT (OE-DTW)	Alignement élastique et comparaison de signatures.	**Faible ("Small Data").**

Pour valider la robustesse de cette approche, nous établissons deux protocoles expérimentaux dépendant de la complétude de l'observation T_{new} :

- **Analyse Post-Hoc (basée sur la trajectoire complète) :** Le modèle a accès à la trajectoire complète T_{full} de longueur L. Ici, le calcul utilise l'alignement de groupe standard. Cela établit la borne supérieure théorique de performance.
- **Alerte Précoce :** Le modèle observe seulement un préfixe tronqué $T_{partial}$ représentant les premiers 30 % de l'activité.

Dans le scénario d'alerte précoce, appliquer la DTW standard est mathématiquement défectueux. En effet, cette dernière impose une contrainte, obligeant la fin de l'observation partielle (30 % de la chronologie) à s'aligner avec la fin de la signature de référence (100 % de la chronologie). Cela induit des distorsions de déformation temporelle artificielles. Pour atténuer cela, nous employons la DTW à fin ouverte (Open-End DTW ou OE-DTW) [14]. Cette variante relâche la contrainte de point final sur la séquence de référence. Formellement, si Q est la requête étudiante partielle et R est une signature de référence, OE-DTW cherche la longueur de préfixe optimale l au sein de R qui minimise le coût d'alignement :

$$D_{OE-DTW}(Q,R) = min_{l\in[1,|R|]} D_{DTW}(Q, R_{1:l})$$

Cela permet au système de déterminer si le début du comportement de l'étudiant est cohérent avec le début d'une signature de réussite.

3.5 Éthique et Confidentialité : Une Approche « *Privacy-by-Design* »

Le déploiement de systèmes de vision par ordinateur dans les cadres éducatifs soulève des préoccupations légitimes concernant la vie privée et la conformité réglementaire (RGPD en Europe). Pour relever ces défis, cette étude adopte une architecture stricte de *Privacy-by-Design*.
Contrairement aux approches de surveillance traditionnelles basées sur l'enregistrement de flux vidéo bruts, notre pipeline de traitement effectue une extraction immédiate des caractéristiques. Le système ne conserve aucune image du visage de l'étudiant. Seuls des descripteurs numériques anonymisés (points clés du squelette facial - 64 repères 2D, valeurs de valence et d'arousal) sont extraits en temps réel. Cette méthode de minimisation des données assure l'anonymisation de l'identité visuelle de l'étudiant à partir des trajectoires stockées (vecteurs numériques). De plus, l'analyse se limite à la détection de comportement (pose de la tête) et n'opère pas de reconnaissance faciale (identification biométrique), excluant toute utilisation de bases de données d'identité. Cette approche garantit que l'analyse pédagogique bénéficie de la richesse des signaux non verbaux sans compromettre l'anonymat et le droit à l'image des participants, conformément aux directives actuelles sur l'IA de confiance.

4 Résultats Expérimentaux

4.1 Protocole

L'évaluation du cadre SEPT repose sur un corpus de 1849 trajectoires, capturant la dynamique comportementale d'une cohorte d'étudiants en cycle d'ingénierie. Afin de garantir une stricte validité écologique et de simuler les conditions réelles de déploiement d'un système d'alerte précoce, nous avons opté pour un partitionnement temporel des données. Ainsi, l'apprentissage du modèle s'appuie sur un historique constitué de l'intégralité des trois premières évaluations (Quiz 1 à 3), formant un ensemble d'entraînement de 1 279 trajectoires (dont 200 échecs). La capacité de généralisation temporelle et inter-tâches du modèle est ensuite évaluée sur un ensemble de test strictement ultérieur composé des tentatives du dernier quiz (Quiz 4), totalisant 570 trajectoires (dont 80 échecs).

La prévalence de la classe d'intérêt (l'échec) n'étant que de 14,0 % dans l'ensemble de test, ce fort déséquilibre de classe invalide l'utilisation de l'exactitude (accuracy) comme métrique d'évaluation légitime ; en effet, un classificateur majoritaire naïf atteindrait mécaniquement une exactitude de 86 %. Par conséquent, notre protocole d'évaluation s'appuie sur le F1-Score ciblant spécifiquement la classe minoritaire, afin d'évaluer le compromis précision-rappel, ainsi que sur l'Aire Sous la Courbe ROC (AUC) pour quantifier la capacité de discrimination globale et la robustesse du modèle indépendamment de son seuil opérationnel.

Définition de la Trajectoire et Vérité Terrain. Dans le cadre de cette étude, l'unité d'analyse temporelle, désignée comme une « trajectoire », correspond à l'ensemble continu des vecteurs d'état $x_t \in \mathbb{R}^7$ capturés spécifiquement durant le temps de réflexion et de saisie de la réponse pour une seule question. La vérité terrain caractérisant la difficulté a été établie objectivement à l'échelle de l'item cognitif : une trajectoire est étiquetée comme « Échec » (situation de difficulté) si la note obtenue par l'étudiant à la question courante est inférieure à 50 % de la note maximale possible pour cet item. À l'inverse, un score $\geq 50\%$ caractérise une trajectoire de réussite, venant alimenter l'historique de succès individuel ψ_{ind} . Après suppression des séquences inutilisables (occultations du visage, pertes de signal), l'ensemble de données exploitable s'élève à 1849 trajectoires distinctes.

Acquisition et Architecture *Privacy-by-Design*. La capture des flux a été réalisée via un outil de télésurveillance (proctoring) configuré avec un taux d'échantillonnage volontairement bas d'une image par seconde (1 FPS). Cette faible fréquence répond à un double impératif : limiter le bruit à haute fréquence des micro-mouvements non sémantiques. Conformément à notre architecture *Privacy-by-Design*, aucune vidéo brute n'a été stockée. L'inférence des points clés squelettiques et des états affectifs s'est effectuée à la volée, ne persistant que les descripteurs numériques anonymisés.

Conformité Éthique. Le déploiement de ce système a suivi un protocole éthique strict. L'étude a fait l'objet d'une instruction et d'une validation préalable par un comité d'éthique indépendant ainsi que par le Délégué à la Protection des Données de l'établissement. Préalablement à l'expérimentation, l'ensemble des 79 étudiants ont été informés des modalités techniques (absence d'enregistrement vidéo visuel) et ont signés un formulaire de consentement éclairé, en accord avec les directives du RGPD et du futur AI Act.

4.2 Analyse Comparative : Signature de l'étudiant vs. Signature de groupe

Notre hypothèse centrale postule que la difficulté académique se manifeste comme une déviation par rapport aux propres habitudes comportementales d'un étudiant plutôt que comme une déviation par rapport à une signature de groupe. Les résultats valident les Signatures de parcours de réussite Individuelles :

- **Gain de performance :** Le modèle « Étudiant » (régression logistique) atteint un F1-Score de 0,27, surpassant l'approche « groupe » (0,23). Ce gain relatif de +17,4 % confirme que l'historique comportemental personnel est un prédicteur plus granulaire que les moyennes de classe.
- **Pouvoir discriminant :** L'analyse ROC corrobore cette supériorité : alors que l'approche des signatures de parcours de réussite de groupe plafonne à une AUC de 0,61 (SVM), l'approche de signature étudiante monte à 0,67 (SVM). Ce différentiel est critique, car il indique une réduction significative des faux positifs pour le même taux de détection.
- **Stabilité du Modèle :** Les modèles linéaires (SVM, Régression Logistique) surpassent constamment les architectures non linéaires telles que XGBoost ou MLP sur les données de trajectoire basées sur la DTW. Cela suggère que la distance à une « signature personnelle » est une caractéristique linéairement séparable dans le cas des modèles linéaire.

Cette supériorité est corroborée par l'analyse de la courbe ROC présentée en Figure. 4.

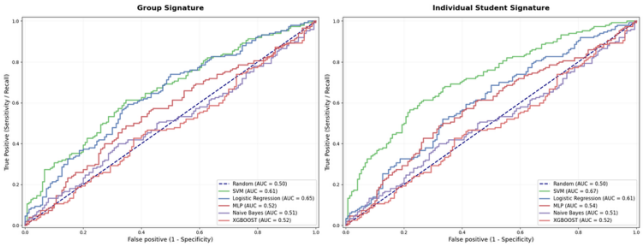


Figure 4 : Courbes ROC. À droite (approche signature Étudiant Individuel), le SVM démontre un pouvoir discriminant significativement meilleur (AUC=0.67). Ce différentiel de +0.06 points d'AUC est critique : il indique que le modèle génère significativement moins de faux positifs pour le même taux de détection.

4.3 Positionnement par rapport aux Études Récentes

Bien que les valeurs absolues (AUC=0.67) puissent paraître modestes, elles doivent être contextualisées par la difficulté intrinsèque de la détection « in-the-wild » (en conditions réelles) utilisant des données conformes à la confidentialité.

Les méthodes utilisant des données démographiques ou des journaux LMS complets atteignent une AUC légèrement plus élevée (0,69-0,71) comme montré dans le Tableau 3. Nous atteignons une performance comparable (seulement 0,04 de différence d'AUC) tout en passant à un paradigme entièrement respectueux de la vie privée. Contrairement à [31], qui dépend du profilage démographique pour le pouvoir prédictif, notre cadre élimine le besoin de données d'attributs personnels.

Tableau 3. Comparaison de SEPT avec des études récentes.

Study	Type de données	Performance (AUC) ↑	Préservation de la confidentialité
SEPT (Our Approach)	Trajectoire émotionnelle et physiques	0.67	**+ +**
Early Warning XGBoost [30]	Socio-Demographique + note du lycée	0.69	- -

MOOC Stacked Ensemble [31]	Journaux complets (LMS) + Données d'information des étudiants	**0.71**	+
EfficientNet Engagement[32]	Vidéo brute (Reconnaissance faciale)	~0.62 (Acc)	- -

4.4 Efficacité de la Détection Précoce (RO2)

Pour qu'un système d'alerte soit utile, il doit identifier les difficultés pendant l'activité. Nous avons testé les modèles sur les premiers 30 % de la trajectoire temporelle.

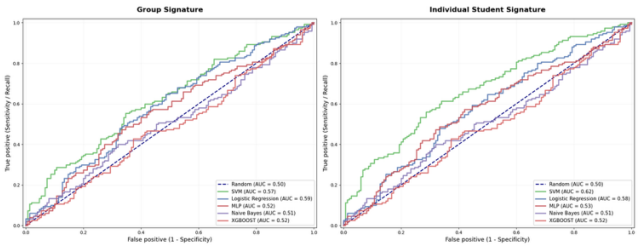


Figure 5 : Performance de détection précoce ($t < 30\%$). La performance de l'approche individuelle reste stable, indiquant que les signaux d'échec sont présents dès le tout début de la tâche.

4.5 Impact Opérationnel : Optimiser les Interventions Ciblées

Au-delà des métriques techniques, la valeur de SEPT réside dans l'optimisation de la « bande passante attentionnelle » limitée d'un instructeur. Dans une simulation d'une cohorte de 70 étudiants avec une prévalence d'échec de 15 % (env. 10 étudiants en difficulté). Pour évaluer cet impact opérationnel, nous analysons la Précision au Rang P (précision *p*) au sein d'un scénario de simulation réaliste. Si un instructeur a la capacité d'effectuer seulement 10 interventions ciblées (p =10), l'efficacité de la stratégie de sélection devient critique.

- Sélection Aléatoire : Un instructeur identifierait seulement 1 ou 2 étudiants en difficulté par hasard.
- Sélection Ciblée SEPT (utilisant les signatures de parcours de réussite Individuelles) : En signalant le Top 10 des étudiants à risque, SEPT identifie correctement 5 étudiants en véritable difficulté.

Cela représente une élévation de performance de 2,5x comparée à une aide humaine non assistée. Dans un cadre « humain dans la boucle » (*human-in-the-loop*), les faux positifs générés par le système ne sont pas des « erreurs » mais des opportunités pour des vérifications bienveillantes avec des étudiants engagés dans une lutte productive. SEPT transforme ainsi une surveillance de masse impossible en un ciblage pédagogique précis et actionnable.

5 Conclusion

Cette étude visait à concevoir et évaluer SEPT, un cadre d'analyse basé sur la trajectoire et intrinsèquement respectueux de la vie privée, pour la détection précoce des difficultés académiques. Au-delà des métriques de performance pures, nos résultats remettent en question l'approche dominante en analyse de données d'apprentissage (*Learning Analytics*), qui tend à évaluer l'engagement étudiant au travers du prisme réducteur d'une norme comportementale globale.

La contribution fondamentale de ce travail réside dans la validation empirique de l'« hypothèse de la signature individuelle ». Nos expériences démontrent que le prédicteur le plus robuste de l'échec académique n'est pas la déviation par rapport à une moyenne de classe, mais la divergence structurelle par rapport à l'historique de succès de l'étudiant lui-même (élévation de performance relative de +17,4 %). Ce résultat s'aligne fortement avec les postulats de la cognition incarnée : l'expression physique de la charge cognitive (ex : faire les cents pas, inclinaisons rythmiques de la tête) est hautement idiosyncratique. Ce constat appelle à un changement de paradigme dans la conception des Systèmes de Tutorat Intelligents (STI) : l'abandon de modèles de classification « *one-size-fits-all* » au profit d'architectures d'auto-étalonnage (*self-benchmarking*), capables de distinguer une "frustration productive" personnelle d'un réel décrochage cognitif.

Nous avons confronté le cadre SEPT aux modèles invasifs de l'état de l'art s'appuyant sur des journaux d'apprentissage complets ou des données socio-démographiques (Tableau 2). Si nous observons un léger compromis prédictif (AUC de 0,67 contre 0,71 pour les méthodes intrusives), nous soutenons que cette différence mathématique marginale masque un avantage opérationnel décisif. Avec l'entrée en vigueur de législations telles que l'AI Act , qui classifie les systèmes d'IA de surveillance éducative et d'évaluation comme présentant un « Haut Risque », les modèles basés sur la reconnaissance faciale biométrique ou le profilage invasif deviennent obsolètes de facto, car légalement et éthiquement impossible à déployer. En s'appuyant exclusivement sur l'extraction éphémère de descripteurs cinématiques et affectifs à basse fréquence (1 FPS), SEPT démontre qu'il est possible de conserver 94 % du pouvoir prédictif de l'état de l'art tout en garantissant un anonymat total des données stockées.

6 Discussion et Travaux Futurs

6.1 Utilité Opérationnelle : L'IA comme aide pédagogique.

La valeur pratique du système a été confirmée dans notre scénario d'Alerte Précoce. En utilisant la méthode de déformation temporelle dynamique ouverte (OE-DTW), SEPT maintient une capacité de discrimination robuste dès les premiers 30 % de l'activité étudiante. En situation de

classe, cela se traduit par une multiplication par 2,5 de la capacité de l'enseignant à cibler les étudiants nécessitant un soutien immédiat. L'objectif de SEPT n'est pas l'automatisation de la notation, mais l'augmentation de la bande passante attentionnelle de l'instructeur. Dans ce cadre « Human-in-the-Loop », les faux positifs générés par le système perdent leur statut d'erreur critique pour devenir de simples opportunités d'interaction pédagogique proactive. L'IA quitte ainsi le rôle de surveillant punitif pour endosser celui d'assistant au diagnostic de la détresse cognitive.

6.2 Limites et Travaux Futurs

Bien que prometteurs, ces résultats comportent plusieurs limites qui orienteront nos travaux futurs.

- **Profil de la cohorte** : Le jeu de données actuel a été collecté auprès d'étudiants en ingénierie face à des tâches convergentes (QCM). Des études ultérieures devront valider la robustesse des trajectoires affectives lors de tâches divergentes ou créatives, notamment en sciences humaines

- **Problème du démarrage à froid (*Cold Start*) :** L'approche par signature individuelle requiert un historique de réussites. Bien que notre système pallie cette lacune transitoire en s'adossant aux "signatures de groupe" lors des premières sessions, la rapidité de convergence vers une signature individuelle stable (nombre minimum de trajectoires requises) reste à modéliser mathématiquement.

- **Résolution temporelle :** L'échantillonnage volontairement parcimonieux à 1 image par seconde, dicté par des contraintes de frugalité et de réduction de l'invasivité, nous prive de l'analyse des micro-expressions faciales (durée < 500 ms). Les travaux futurs pourraient explorer des mécanismes d'attention dynamiques adaptant la fréquence de capture uniquement lors de la détection de transitions affectives critiques.

Divulgation des intérêts. Les auteurs n'ont aucun intérêt concurrent à déclarer qui soit pertinent pour le contenu de cet article.

Références

1. General Data Protection Regulation (GDPR) – Official Legal Text, https://gdpr-info.eu/, last accessed 2024/03/19.
2. Résumé de haut niveau de la loi sur l'intelligence artificielle de l'UE, https://artificialintelligenceact.eu/fr/high-level-summary/, last accessed 2026/01/31.
3. Pekrun, R., Perry, R.P.: Control-value theory of achievement emotions. In: International handbook of emotions in education. pp. 120–141. Routledge/Taylor & Francis Group, New York, NY, US (2014).
4. Camacho-Morles, J., Slemp, G.R., Pekrun, R., Loderer, K., Hou, H., Oades, L.G.: Activity Achievement Emotions and Academic Performance: A Meta-analysis. Educ Psychol Rev. 33, 1051–1095 (2021).
5. Russell, J.: A Circumplex Model of Affect. Journal of Personality and Social Psychology. 39, 1161–1178 (1980).
6. Blumenfeld, P., Paris, A.: School Engagement: Potential of the Concept, State of the Evidence. Review of Educational Research. 74, 59–109 (2004).
7. Wilson, M.: Six views of embodied cognition. Psychonomic Bulletin & Review. 9, 625–636 (2002).
8. Zou, L., Zhang, Z., Mavilidi, M., Chen, Y., Herold, F., Ouwehand, K., Paas, F.: The synergy of embodied cognition and cognitive load theory for optimized learning. Natural Human Behavior. 9, 877–885 (2025).
9. Toward Semantic Gaze Target Detection, https://www.proceedings.com/079017-3858.html, last accessed 2026/01/30.
10. Holstein, K., Aleven, V., Rummel, N.: A Conceptual Framework for Human–AI Hybrid Adaptivity in Education. In: Bittencourt, I.I., Cukurova, M., Muldner, K., Luckin, R., and Millán, E. (eds.) Artificial Intelligence in Education. pp. 240–254. Springer International Publishing, Cham (2020).
11. Zaletelj, J., Košir, A.: Predicting students' attention in the classroom from Kinect facial and body features. Journal of Image Video Processing. 2017, 80 (2017).
12. Mohamad Nezami, O., Dras, M., Hamey, L., Richards, D., Wan, S., Paris, C.: Automatic Recognition of Student Engagement Using Deep Learning and Facial Expression. Proc. Of ECML PKDD. pp 273-289 (2020).
13. Akpanoko, C.E., S., A.T., Cordell, G., Biswas, G.: Investigating the Relations between Students' Affective States and the Coherence in their Activities in Open-Ended Learning Environments. Proc of the 17th International Conference on Educational Data Mining (2024).
14. Grafsgaard, J., Wiggins, J., Boyer, K., Wiebe, E., Lester, J.: Automatically Recognizing Facial Indicators of Frustration: A Learning-Centric Analysis. (2013).
15. Basavaiah, J., Anthony, A.A., N, N.K.H., Mahadevaswamy, Patil, C.M.: Facial Emotion Recognition: A Review on State-of-the-art Techniques. In: 2024 4th International Conference on Data Engineering and Communication Systems (ICDECS). pp. 1–6 (2024).
16. Nadaud, E., Yaacoub, A., Haidar, S., Grand, B., Prevost, L.: Emotion Trajectory and Student Performance in Engineering Education: A Preliminary Study. Int. Conf. on Research Challenges in Information Science, 410-424 (2024).
17. Zambrano, A.F., Ocumpaugh, J., Baker, R.S., Vandenberg, J.: Better to Be Confused or Frustrated Than

Bored: Analyzing Affect Dynamics Across Player Archetypes. In: Carmona, G., Lima, C., Santos, M.J., Benítez, H., Montero-Moguel, L., and Galarza-Tohen, B. (eds.) Advances in Quantitative Ethnography. pp. 384–399. Springer Nature Switzerland, Cham (2026).
18. Baker, R.S.J. d., D'Mello, S.K., Rodrigo, Ma.M.T., Graesser, A.C.: Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. International Journal of Human-Computer Studies. 68, 223–241 (2010).
19. Piot, M., Alabarbe, T., Gonzalez, J., le Bail, C., Prevost, L., Bourdeau, J., Bernard, F.X., Baker, M., Detienne, F.: Joint analysis of verbal and nonverbal interactions in collaborative E-learning. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). pp. 1–5 (2019).
20. D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. Learning and Instruction - LEARN INSTR. 22, (2012).
21. Zhao, J., Itti, L.: shapeDTW: shape Dynamic Time Warping, http://arxiv.org/abs/1606.01601, (2016).
22. Mollahosseini, A., Hasani, B., Mahoor, M.H.: AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. IEEE Trans. Affective Comput. 10, 18–31 (2019).
23. Liu, M., Kollias, D.: Aff-Wild Database and AffWildNet, http://arxiv.org/abs/1910.05318, (2019).
24. Wagner, N., Mätzler, F., Vossberg, S.R., Schneider, H., Pavlitska, S., Zöllner, J.M.: CAGE: Circumplex Affect Guided Expression Inference, http://arxiv.org/abs/2404.14975, (2024).
25. OpenCV: Cascade Classifier, https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html, last accessed 2025/03/26.
26. What is YOLOv5: A deep look into the internal features of the popular object detector, https://arxiv.org/html/2407.20892v1, last accessed 2025/03/26.
27. Nadaud, E., Yaacoub, A., Legrand, B., Prevost, L.: SEPT: Uncovering Student Difficulties through Emotional and Physical Trajectories during Online Assessments. (2025).
28. Kang, X., Nie, Y.: Design and analysis of teaching early warning system based on multimodal data in an intelligent learning environment. PeerJ Comput. Sci. 11, e2692 (2025).
29. Chang, Y.-H., Chen, F.-C., Lee, C.-I.: Developing an Early Warning System with Personalized Interventions to Enhance Academic Outcomes for At-Risk Students in Taiwanese Higher Education. Education Sciences. 15, (2025).
30. Carballo-Mendívil, B., Arellano-González, A., Ríos-Vázquez, N.J., Lizardi-Duarte, M. del P.: Predicting Student Dropout from Day One: XGBoost-Based Early Warning System Using Pre-Enrollment Data. Applied Sciences. 15, 9202 (2025).
31. Ajayi, O.O.: Predicting Student Dropout Risk in Online Learning using Stacked Ensemble Machine Learning and Explainable AI Techniques. International Journal of Computer Applications. 187, 26–29 (2025).
32. Alyuz, N., Okur, E., Genc, U., Aslan, S., Tanriover, C., Esme, A.A.: Unobtrusive and Multimodal Approach for Behavioral Engagement Detection of Students, http://arxiv.org/abs/1901.05835, (2019).