République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieure et de la Recherche Scientifique

Faculté des Sciences Exactes et de Science de la Nature et de la Vie

Mémoire de fin d'étude

Pour l'obtention du diplôme de MASTER

Filière : Informatique

Option : Systèmes d'information

Thème :

Détection de l'engagement académique des étudiants par suivi oculaire

Réalisé par :

FARES Wissal

Devant le jury :

Dr. ACHOURI Mounir	MCB Université Larbi Tebessi	**Président**
Dr. KHELIFA Boudjemaa	MCB Université Larbi Tebessi	**Examinateur**
Dr. TABET KHAOULA	MCB Université Larbi Tebessi	**Encadrante**

Date de soutenance :

09/06/2025

People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

Echahid Cheikh Larbi Tebessi University- Tebessa

Faculty of Exact Sciences and Natural and Life Sciences

A Dissertation Submitted in Partial Fulfillment of the Requirement for Master's Degree

Field: Computer Science

Specialty: Information systems

Subject:

Detecting students' academic engagement through eye tracking

Presented by:

FARES Wissal

In front of the jury:

Dr. ACHOURI Mounir	MCB Larbi Tebessi University	**President**
Dr. KHELIFA Boudjemaa	MCB Larbi Tebessi University	**Examiner**
Dr. TABET KHAOULA	MCB Larbi Tebessi University	**Framer**

College year:

2024/2025



Remerciement

Je tiens à exprimer ma profonde gratitude à Allah, le Tout-Puissant et Miséricordieux, pour m'avoir accordé la force et la patience nécessaires à l'achèvement de ce travail.

Un grand merci à mes parents pour leur soutien infaillible et leurs efforts constants. Leur présence et leurs encouragements ont été une source d'inspiration tout au long de ce projet, et je leur en suis profondément reconnaissante.

Je souhaite également remercier chaleureusement **Dr. Tabet Khaoula** pour ses précieux conseils et son accompagnement pédagogique éclairé. De même, je suis reconnaissante envers toutes les enseignantes qui m'ont transmis leur savoir et m'ont apporté un soutien précieux dans mes études.

J'exprime ma sincère gratitude aux membres du jury pour l'attention portée à mon travail ainsi que pour leurs contributions enrichissantes à cette recherche.

Enfin, un grand merci à toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce travail. Votre soutien et votre collaboration ont été inestimables, et je vous adresse toute ma reconnaissance et mon profond respect.

Merci beaucoup à tous

Dédicace

Avant tout, je rends grâce à Allah le Tout-Puissant pour nous avoir accordé la force et la patience nécessaires à la réalisation de ce projet.

Je dédie ce travail à moi-même, en hommage aux innombrables heures de labeur, de persévérance et de dévouement, aux nuits blanches et aux instants d'épuisement qui ont rendu possible l'aboutissement de ce projet. Qu'il soit le reflet de mon engagement profond envers l'apprentissage, l'effort et l'accomplissement personnel.

Avec l'expression de ma reconnaissance, je dédie ce modeste travail a ceux qui quels que soient les termes embrassés, je n'arriverais jamais à leur exprimer mon amour sincère.

A l'homme de ma vie : mon papa **RIDA**, merci beaucoup pour les ailes que vous m'avez données, pour m'avoir appris à me lever et à élargir mes horizons vers les cieux. Merci de m'avoir accordé toute votre confiance.

À la femme qui a souffert sans me laisser souffrir, qui n'a jamais dit non-amés exigences et qui n'a épargné aucun effort pour me rendre heureuse : mon adorable mère **RAFIKA**.

Àux amoureux du cœur, aux fleurs de notre maison, a ceux qui vivent dans mon cœur, À mon autre moitié, mes chères sœurs **: LINA** et **CHAHED**.

À qui je souhaitais qu'elle soit mon compagnon de graduation, à qui c'était mon bras droit dans la vie, a ceux qui m'ont quitté et qui n'ont pas quitté mon cœur : **MAYMOUNA**, que dieu ait pitié de toi et me ressemble avec toi au paradis, je dédie mon travail à ton âme pure.

À tous mes **Oncles** et **Tantes**, à mes cousins, ainsi qu'à mes voisins et amis que j'ai connus jusqu'à présent, merci pour leur amour et leurs encouragements.

À tous mes collègues….

À tous ma famille.

FARES Wissal

Résumé

L’engagement académique est un facteur clé dans la réussite scolaire des étudiants. Sa mesure, souvent subjective, peut être optimisée grâce aux technologies émergentes comme le suivi oculaire et l’intelligence artificielle. Ce travail propose un modèle intelligent combinant un réseau de neurones convolutif pré-entraîné (VGG16) et un réseau LSTM afin de détecter automatiquement le niveau d’engagement académique à partir de séquences d’images d’yeux. Après avoir étudié les approches existantes, un prétraitement des données a été réalisé sur le dataset Eye Dataset de Kayvan Shah. Le modèle proposé a été évalué en utilisant des métriques telles que la précision, le rappel et le F1-score, montrant des performances supérieures aux méthodes traditionnelles. Cette approche ouvre des perspectives prometteuses pour l'amélioration des dispositifs éducatifs intelligents.

Mots-clés : Engagement académique, Suivi oculaire, Deep Learning, VGG16, LSTM, Détection automatique l'engagement.

Abstract

Academic engagement is a key factor in students' educational success. Its evaluation, often subjective, can be enhanced through emerging technologies such as eye-tracking and artificial intelligence. This work proposes an intelligent model combining a Convolutional Neural Network (VGG16) and a Long Short-Term Memory network (LSTM) to automatically detect the level of academic engagement from sequences of eye images. After reviewing existing approaches, data preprocessing was carried out on the Eye Dataset from Kayvan Shah. The proposed model was evaluated using metrics such as accuracy, recall, and F1-score, demonstrating superior performance compared to other existing models. This approach opens up promising prospects for the development of intelligent educational systems.

Keywords: Academic engagement, Eye-tracking, Deep Learning, VGG16, LSTM, Automatic engagement detection.

ملخص

يُعد الارتباط الأكاديمي عاملاً رئيسياً في نجاح الطلاب في مسارهم التعليمي. وغالباً ما تكون طرق تقييمه تقليدية وذات طابع شخصي. لهذا، يهدف هذا العمل إلى تطوير نموذج ذكي يعتمد على الشبكات العصبية التلافيفية المدربة مسبقاً (VGG16)مع شبكة LSTM للكشف التلقائي عن مستوى الارتباط الأكاديمي انطلاقاً من تسلسلات صور للعين. بعد دراسة المناهج السابقة، تم إجراء معالجة مسبقة لبيانات مجموعةEye Dataset الخاصة ب Kayvan Shah .وقد تم تقييم النموذج باستخدام مقاييس الدقة والاسترجاع ومعامل F1 ، حيث أظهرت النتائج تفوقاً ملحوظاً مقارنة بالطرق التقليدية. تفتح هذا النهج آفاقاً واعدة لتطوير أنظمة تعليمية ذكية.

الكلمات المفتاحية :المشاركة الأكاديمية، تتبع حركة العين، التعلم العميق، VGG16، LSTM، الكشف التلقائي عن المشاركة.

Table Des Matiéres

Remerciement ..... **I**

Dédicace ..... **II**

Résumé ..... **III**

Abstract ..... **IV**

ملخص ..... **V**

Table Des Matiéres ..... **VI**

Liste Des Figures ..... **IX**

Liste Des Tableaux ..... **X**

INTRODUCTION GÉNÉRALE ..... **XI**

Introduction : **1**

Chapitre 01 : Introduction à la détection de l'engagement Académique par suivi oculaire ... **2**

1. Introduction : **2**

2. L'engagement académique: **2**

2.1. Définition : **2**

2.2. L'importance de l'engagement académiques: **2**

2.3. Les différentes dimensions de l'engagement : **3**

2.4. Impact de l'engagement sur la performance académique : **4**

3. Le Suivi oculaire: **5**

3.1. Définition: **5**

3.2. Les métriques de suivi oculaire (ET) : **5**

3.3. Applications du suivi oculaire dans l'éducation : **7**

4. Conclusion : **8**

Chapitre 02 : L'intelligence artificielle dans la détection De l'engagement académique : état de l'art **9**

1. Introduction: **9**

2. Concepts: **9**

2.1. L'intelligence artificielle (IA) : **9**

2.2. Apprentissage automatique « Machine Learning (ML) » : **10**

2.2.1. Définition: 10
2.2.2. Les type de L'Apprentissage automatique : 10
2.2.3. Algorithmes de L'Apprentissage automatique : 12
2.2.4. Les métriques d'évaluations des systèmes de l'Apprentissage automatique : 13
2.3. Apprentissage profonde « Deep Learning (DL) »: 14
2.3.1. Définition: 14
2.3.2. Types de réseaux DL : 14
2.4. Relation entre IA et ML et DL : 16
3. Approches existantes pour la détection de l'engagement académique : 17
3.1. Etat de l'art: 17
3.2. Synthèse: 22
3.3. Justification du choix de CNN(VGG16) - LSTM : 23
4. Conclusion : 24
Chapitre 03 : Détection de l'engagement académique Par suivi oculaire (Conception et évaluation du modèle) 8
1. Introduction: 25
2. Présentation des outils utilisé: 25
3. Description du dataset: 26
4. Chargement du dataset : 27
5. Prétraitement des données: 28
6. Architecture du modèle utilisé: 28
7. Entraînement du modèle: 29
8. Evaluation du modèle: 30
9. Matrice de confusion: 32
10. Analyse comparative avec les travaux existants : 33
11. Interface Utilisateur et Fonctionnalités: 34
11.1. Présentation générale de l'interface : 34
11.2. Fonctionnalités principales : 34
11.2.1. Illustrations de l'interface (Captures d'écran) : 34
12. Conclusion : 42
Conclusion Générale 43

Conclusion ; **44**

Perspectives ; **45**

Bibliographie **29**

Liste Des Figures

Figure1. 1 : Schématisation de l'engagement des étudiants. [4] 4
Figure1. 2: Exemple de détection de mouvement des yeux durant l'apprentissage. 7

Figure2. 1: Les techniques d'Intelligence Artificiel……..10
Figure2. 2: Les techniques de Machine Learning……....11
Figure2. 3: Fonctionnement de Random Forest……...12
Figure2. 4: Arbre de décision……...13
Figure2. 5 : Architecture du VGG16…….... 16
Figure2. 6 : Relation entre IA et ML et DL……..17

Figure3. 1: Exemple d'images extraites du dataset Eye Gaze. 27
Figure3. 2: Chargement du dataset. 28
Figure3. 3 : Architecture du modèle proposé. 29
Figure3.4: Code python pour l'entrainement du modèle vgg16 et LSTM. 29
Figure3. 5: Apprentissage du modèle en 10 époques. 30
Figure3. 6: Courbes d'apprentissage – Accuracy et Loss pour l'entraînement et la validation. 31
Figure3.7: Matrice de Confusion. 32
Figure3.8 : Comparaison des précision (Accuracy) obtenues par différents travaux. 33
Figure3. 9 : Formulaire d'entrée de l'étudiant. 35
Figure3. 10: Interface de classification et interactions. 36
Figure3. 11: Affichage du résultat de classification. 37
Figure3. 12: l'indication de progression affichée lors de la capture d'images. 38
Figure3. 13 : Résultat de classification après capture d'une séquence d'images (Engagé). 39
Figure3. 14 : Résultat de classification après capture d'une séquence d'images (Non Engagé). 40
Figure3. 15: Sauvegarde des résultats dans le fichier Excel. 41
Figure3. 16 : Exemple de fichier PDF généré avec les résultats de prédiction. 41

Liste Des Tableaux

Tabeau 2. 1:Résultats des travaux connexes pour la prédiction du suivi oculaire. 22
Tabeau 2. 2: Résultats obtenus par le modèle CNN (VGG16) + LSTM.....31

INTRODUCTION GÉNÉRALE

Introduction :

L'engagement académique des étudiants joue un rôle fondamental dans le processus d'apprentissage. Il est souvent lié à la motivation, à l'attention, ainsi qu'à l'implication cognitive et comportementale de l'apprenant. Cependant, son évaluation demeure complexe et généralement subjective, s'appuyant sur des questionnaires ou des observations humaines susceptibles de présenter des biais. Avec l'évolution des technologies d'intelligence artificielle et des dispositifs de suivi oculaire (eye-tracking), il devient aujourd'hui envisageable d'exploiter les données visuelles pour analyser de manière plus objective et en temps réel le comportement des apprenants. Dans ce contexte, il devient pertinent d'explorer le potentiel des techniques d'apprentissage profond pour modéliser et prédire l'engagement académique à partir de données issues du eye-tracking.

Ce mémoire s'inscrit dans cette perspective et a pour objectif principal le développement d'un modèle intelligent capable de détecter le niveau d'engagement académique des étudiants à partir de séquences d'images oculaires. Les objectifs spécifiques de cette recherche sont les suivants : étudier les approches existantes de détection d'engagement par eye-tracking ; prétraiter et exploiter un jeu de données oculaires pour entraîner un modèle de classification ; concevoir une architecture hybride combinant un réseau de neurones convolutif pré-entraîné (VGG16) et un réseau LSTM, dans le but de prendre en compte à la fois les dimensions spatiales et temporelles du regard ; évaluer la performance du modèle à l'aide de métriques standards telles que la précision, le rappel et le F1-score ; et enfin, comparer les résultats obtenus à ceux des travaux antérieurs utilisant des approches similaires.

Ce mémoire est structuré en trois chapitres principaux. Le Chapitre 1 introduit la notion d'engagement académique et présente les principes du suivi oculaire dans un contexte éducatif. Le Chapitre 2 propose une revue de l'état de l'art des méthodes traditionnelles et intelligentes de détection de l'engagement. Enfin, le Chapitre 3 est consacré à la conception du modèle proposé, à l'expérimentation menée sur le dataset Eye Dataset de Kayvan Shah, ainsi qu'à l'analyse des résultats obtenus.

Chapitre 01 : Introduction à la détection de l'engagement Académique par suivi oculaire

1. Introduction :

L'engagement académique est un facteur clé influençant la réussite des étudiants et leur expérience d'apprentissage. Il se manifeste à travers plusieurs dimensions, telles que l'engagement comportemental, émotionnel et cognitif, et joue un rôle essentiel dans la performance scolaire. Comprendre et mesurer cet engagement permet aux institutions éducatives d'optimiser leurs stratégies pédagogiques afin d'améliorer l'implication des étudiants.

Dans ce chapitre, nous définirons d'abord le concept d'engagement académique et son importance dans le contexte éducatif. Nous examinerons ensuite ses différentes dimensions ainsi que son impact sur la performance des étudiants. Par la suite, nous introduirons la technologie du suivi oculaire (Eye-tracking), qui permet d'analyser l'attention visuelle des apprenants. Nous présenterons ses principales métriques et explorerons ses applications dans le domaine de l'éducation.

Enfin, ce chapitre mettra en évidence comment l'utilisation du suivi oculaire peut contribuer à une meilleure compréhension de l'engagement académique et à l'amélioration des méthodes d'enseignement.

2. L'engagement académique:

2.1. Définition :

Selon [1],s'engager signifie investir pleinement ses capacités mentales et physiques afin d'apprendre, comprendre et maîtriser des connaissances ou développer des compétences. Bien que souvent associée à la motivation, l'engagement s'en distingue.

La motivation représente un état d'esprit qui incite à agir, constituant ainsi une condition préalable à l'engagement [2] Cependant, l'engagement va au-delà de cette simple impulsion, car il se traduit par des actions concrètes. Il repose à la fois sur une décision mentale et une participation active dans l'activité entreprise, comme le soulignent également [2].

2.2. L'importance de l'engagement académiques:

L'engagement académique est essentiel pour le développement personnel et la réussite scolaire des étudiants. Il influence leur bien-être mental et physique**,** améliore leurs résultats éducatifs et renforce leur sentiment d'appartenance à l'école. Il repose sur des relations positives avec les enseignants et les pairs et peut être vu comme un processus évolutif**.**

Les étudiants peuvent avoir un engagement positif, neutre ou négatif, nécessitant des stratégies adaptées pour renforcer leur attachement à l'école. Un engagement fort mène à une vie plus épanouie et à des expériences éducatives enrichissantes. [3]

2.3. Les différentes dimensions de l'engagement :

Il est important de distinguer les différents types d'engagement dans le contexte de l'apprentissage afin de concevoir des interventions sur mesure visant à enrichir l'expérience des apprenants.

- **L'engagement comportemental :** se traduit par la participation active des élèves aux activités scolaires. Cela inclut l'attention en classe, le respect des règles, l'assiduité ainsi que l'implication dans les activités extrascolaires. Cet engagement joue un rôle clé dans la réussite académique et contribue au bien-être subjectif des élèves. À l'inverse, un faible engagement comportemental est associé à une augmentation des troubles intériorisés (comme la dépression) et extériorisés (tels que la délinquance) à l'adolescence. [4]
- **L'engagement cognitive :** reflète la volonté des élèves à fournir les efforts nécessaires pour comprendre des concepts complexes et maîtriser des compétences difficiles. Les recherches montrent que cet engagement a une influence directe sur la réussite académique : les élèves plus investis cognitivement obtiennent de meilleures notes et de meilleurs résultats aux tests standardisés, tout en bénéficiant d'un bien-être subjectif accru. De plus, une corrélation a été établie entre l'engagement cognitif et le bien-être subjectif des élèves. Plus précisément, il a été démontré que le bien-être subjectif des élèves joue un rôle de médiateur entre leur engagement émotionnel et cognitif, et leur réussite scolaire. [4]

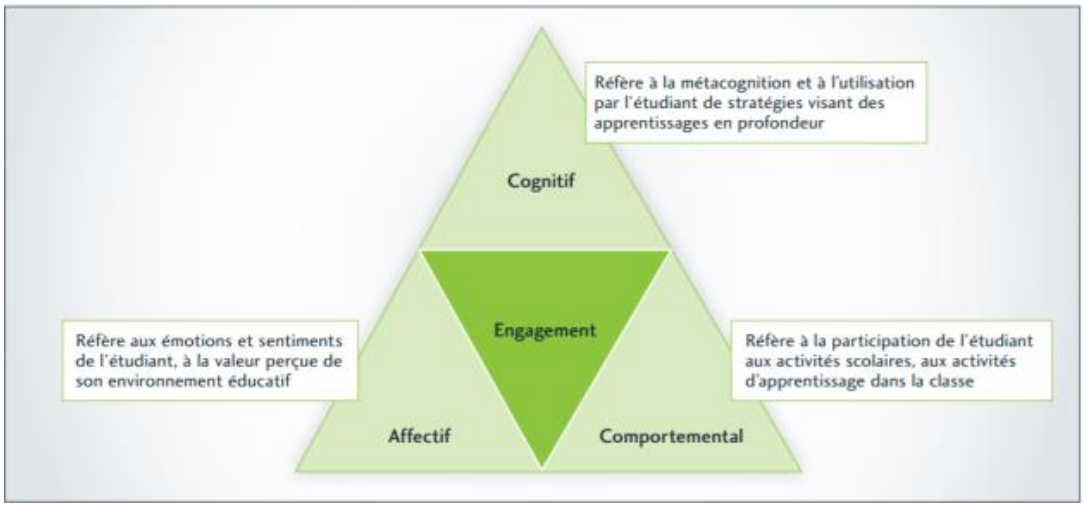


Figure1. 1 : Schématisation de l'engagement des étudiants. *[4]*

- **L'engagement émotionnel :** L'engagement émotionnel fait référence aux réactions positives ou négatives des élèves envers leurs enseignants, leurs camarades et l'école. Il joue un rôle clé dans leur attachement à l'établissement et influence leur motivation à fournir des efforts supplémentaires. Cet engagement englobe divers aspects affectifs, tels que le plaisir d'apprendre, le sentiment d'appartenance, le soutien reçu et les attitudes envers les enseignants, les pairs, l'apprentissage et l'école en général. En plus de sa corrélation avec le bien-être subjectif et la réussite scolaire, l'engagement émotionnel est également influencé par la qualité des relations entre élèves et enseignants ainsi que par les dynamiques de groupe entre pairs. [4]

2.4. Impact de l'engagement sur la performance académique :

L'engagement des étudiants joue un rôle déterminant dans leur réussite académique, en influençant directement leurs résultats scolaires et leur persévérance. Les recherches montrent une corrélation significative entre l'engagement académique et la performance, bien que certains types d'engagement aient un impact plus marqué que d'autres. L'engagement comportemental, qui inclut l'assiduité, la participation active et le respect des règles, est identifié comme le facteur le plus influent sur les résultats académiques. À l'inverse, l'engagement émotionnel et cognitif, bien qu'important pour la motivation et l'intérêt des étudiants, ne suffit pas à garantir de meilleures performances sans une implication concrète dans les activités d'apprentissage. [5]

Comme le confirme l'étude menée par [6] sur l'apprentissage numérique confirme que les étudiants qui s'investissent activement dans des exercices interactifs et des tâches numériques obtiennent de meilleurs résultats aux examens, contrairement à ceux qui se contentent d'assister aux cours ou d'utiliser uniquement des manuels. Cependant, une surcharge de travail répétitif sans réflexion critique peut réduire les bénéfices de

l'apprentissage. En conclusion, un engagement actif et structuré, notamment à travers la participation aux activités interactives et une pratique régulière, est essentiel pour maximiser la réussite académique et limiter le risque d'échec ou d'abandon.

En Résumé, un engagement actif et structuré, qu'il soit en classe ou dans des environnements numériques, est essentiel pour maximiser la réussite académique. La participation aux activités interactives et une pratique régulière sont des éléments clés pour améliorer la performance et réduire le risque d'échec ou d'abandon.

3. Le Suivi oculaire:

3.1. Définition:

Le Suivi oculaire (Eye tracking) est un concept général englobant divers processus de surveillance et de mesure de l'activité oculaire. Il peut inclure le suivi de la position de l'œil par rapport à la tête, l'analyse des mouvements oculaires, l'évaluation du degré de fermeture des paupières, l'observation de la taille et de l'activité des pupilles, ainsi que la détection de la direction du regard. [7]

Selon la définition proposée par [8], « *le eye tracking (ET) est une technologie permettant de déterminer où une personne regarde à un instant donné et comment son regard se déplace d'un point à un autre. L'objectif est de comprendre précisément la fixation visuelle et son évolution en temps réel* ». Autrement dit, cette technologie permet d'analyser les mouvements oculaires afin de mieux comprendre la manière dont les individus perçoivent leur environnement visuel.

En résumé, l'eye tracking, ou suivi oculaire, désigne un ensemble de technologies permettant d'analyser le comportement visuel d'une personne lorsqu'elle explore une image ou un environnement.

3.2. Les métriques de suivi oculaire (ET) :

Les principales mesures utilisées en recherche sur le suivi oculaire sont les fixations et les " saccades", qui sont des mouvements oculaires rapides se produisant entre deux fixations. De nombreuses métriques dérivées en découlent, notamment les mesures du "regard"(gaze) et du "parcours visuel" (scanpath). La taille de la pupille et le taux de clignement des yeux sont également analysés. [9]

- **Fixations :** L'interprétation des fixations varie en fonction du contexte. Lors d'une tâche d'encodage (par exemple, la navigation sur une page web), une fréquence élevée de fixations sur une zone spécifique peut traduire un intérêt accru pour la cible, comme une

photographie dans un article d'actualité. Toutefois, elle peut également indiquer une complexité accrue du stimulus, rendant son encodage plus difficile [10] ; [11]

En revanche, dans une tâche de recherche, l'interprétation peut être inversée : un nombre élevé de fixations uniques ou de regroupements de fixations est souvent le signe d'une plus grande incertitude dans l'identification d'un élément cible [10].

La durée d'une fixation est également liée au temps de traitement accordé à l'objet observé [11]. Il est généralement admis que les représentations externes associées à de longues fixations sont moins significatives pour l'utilisateur que celles associées à des fixations plus brèves [12]

- **Saccades :** Aucun encodage ne se produit durant les saccades, ce qui signifie qu'elles ne fournissent pas d'informations sur la complexité ou la saillance d'un objet dans une interface. Toutefois, les saccades régressives (c'est-à-dire les mouvements oculaires de retour en arrière) peuvent être un indicateur de la difficulté de traitement lors de l'encodage [13]

Bien que la plupart des saccades régressives soient de faible amplitude ne revenant que de deux ou trois lettres dans les tâches de lecture des régressions plus longues, de la taille d'une phrase, peuvent refléter une confusion dans le traitement cognitif avancé du texte [13]. De plus, ces régressions peuvent être utilisées comme une mesure de la valeur de reconnaissance, puisqu'il existerait une relation inverse entre leur fréquence et la saillance d'une phrase.

- **Parcours visuel (*Scanpath*)** : Un scanpath correspond à une séquence complète alternant saccades et fixations. Dans une tâche de recherche, un parcours visuel optimal est généralement représenté par une trajectoire directe vers la cible recherchée, accompagnée d'une durée de fixation relativement courte sur celle-ci [12].
- **Taux de clignement et taille de la pupille (Blink rate and pupil size) :** Ces deux indicateurs sont souvent utilisés pour évaluer la charge cognitive. Un faible taux de clignement est généralement associé à une charge cognitive élevée, tandis qu'un taux de clignement plus élevé peut refléter de la fatigue [14] ; [15]

De même, une dilatation plus importante des pupilles peut indiquer un effort cognitif accru [16] ; [17]. Cependant, ces mesures sont influencées par de nombreux autres facteurs, comme le niveau de luminosité ambiante, ce qui peut introduire des biais dans l'analyse [18]. Pour ces raisons, la taille de la pupille et le taux de clignement sont moins fréquemment utilisés en recherche sur le eye tracking.

3.3. Applications du suivi oculaire dans l'éducation :

- **Analyse des processus d'apprentissage** : L'eye tracking permet d'observer l'attention des apprenants, d'évaluer leur niveau de compréhension et d'analyser leurs stratégies d'apprentissage. Cette technologie aide les enseignants à identifier les points forts et les difficultés de chaque élève, afin d'adapter l'enseignement de manière plus personnalisée [19]
- **Développer des outils pédagogiques plus efficaces :** L'eye tracking permet d'analyser les réactions des apprenants face à divers contenus et supports pédagogiques. Ces analyses facilitent l'amélioration de ces outils en les rendant plus clairs, engageants et mieux adaptés aux besoins spécifiques de chaque élève. [19]

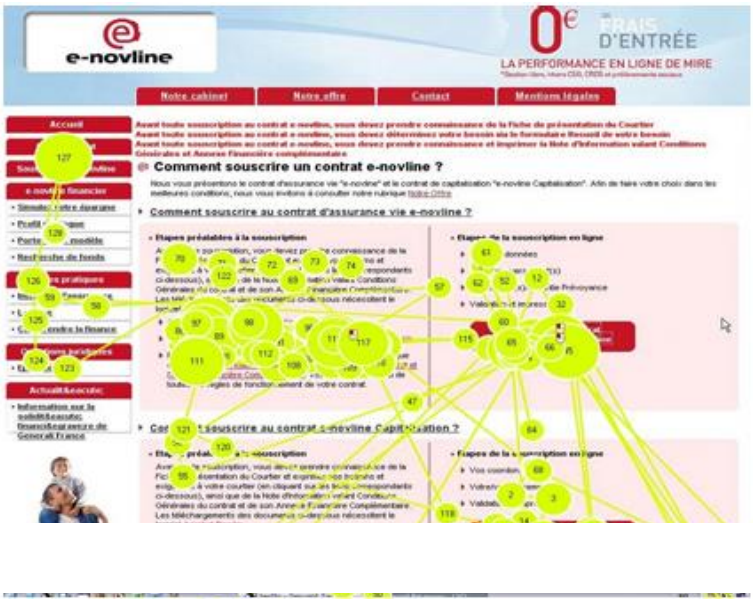


Figure1. 2: Exemple de détection de mouvement des yeux durant l'apprentissage.

4. Conclusion :

Dans ce chapitre, nous avons exploré le concept d'engagement académique, en mettant en avant ses différentes dimensions ainsi que son impact sur la performance des étudiants. L'engagement joue un rôle clé dans la réussite scolaire, influençant la motivation, la persévérance et l'apprentissage.

Nous avons également introduit le suivi oculaire comme une méthode avancée permettant d'évaluer l'engagement des étudiants en analysant leurs mouvements oculaires. Nous avons présenté ses principales métriques et ses applications dans le domaine éducatif, soulignant son potentiel pour améliorer l'expérience d'apprentissage.

En résumé, la combinaison de l'engagement académique et des technologies de suivi oculaire ouvre de nouvelles perspectives pour l'optimisation des stratégies pédagogiques. Cependant, des défis subsistent, notamment en termes de précision des analyses et d'éthique dans l'utilisation des données des apprenants.

Chapitre 02 : L’intelligence artificielle dans la détection De l'engagement académique : état de l’art

1. Introduction:

L'intelligence artificielle (IA) est devenue un outil incontournable dans de nombreux domaines, y compris l'éducation, où elle permet d'analyser et d'améliorer l'engagement académique des étudiants. Grâce aux avancées en Machine Learning (ML) et en Deep Learning (DL), il est désormais possible de concevoir des systèmes intelligents capables de détecter et d'évaluer l'implication des apprenants en temps réel.

Dans ce chapitre, nous commencerons par introduire les concepts généraux de l'IA, en définissant le Machine Learning et le Deep Learning, ainsi que leurs principaux types, algorithmes et métriques d'évaluation. Nous examinerons ensuite la relation entre ces différentes branches de l'IA et leur rôle dans l'analyse de l'engagement académique.

Enfin, nous présenterons les approches existantes pour la détection de l'engagement, en comparant les méthodes traditionnelles (questionnaires, observations) aux techniques avancées basées sur l'IA et le Deep Learning. Nous terminerons par une analyse comparative des méthodes et une justification du choix des modèles CNN-LSTM, qui seront utilisés pour notre approche.

2. Concepts:

2.1. L'intelligence artificielle (IA) :

« L'intelligence artificielle (IA) est une discipline scientifique et technologique dédiée à la conception de machines intelligentes capables de réfléchir de manière autonome et de prendre des décisions, dans le but de reproduire l'intelligence et le comportement humains. » [20] . L'IA est également définie comme un paradigme de programmation dans lequel les programmes traitent des données selon des règles prédéfinies pour atteindre des objectifs. Elle se caractérise aussi comme l'étude des systèmes capables d'agir d'une manière qui, aux yeux d'un observateur, semblerait intelligente. Bien que cette définition soit pertinente, elle ne reflète pas toute l'étendue de l'intelligence artificielle. En effet, les techniques d'IA sont souvent appliquées à la résolution de problèmes simples ou intégrées dans des systèmes plus complexes. Ainsi, une autre approche permet de définir l'intelligence artificielle comme l'utilisation de méthodes inspirées du comportement intelligent des humains et des animaux afin de résoudre des problèmes complexes. [21].

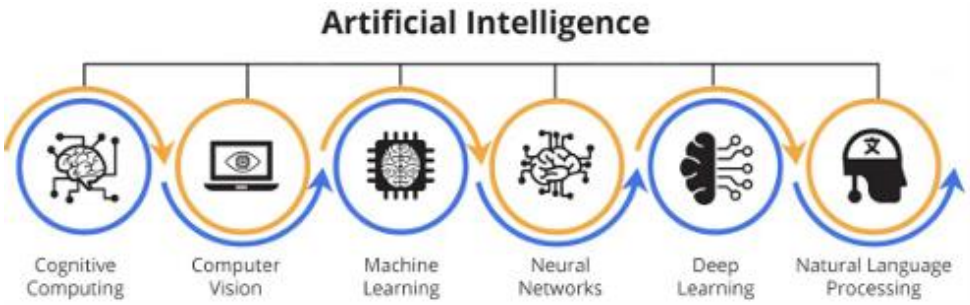


Figure2. 1: Les techniques d'Intelligence Artificiel. *[63]*

2.2. Apprentissage automatique « Machine Learning (ML) » :

2.2.1. Définition:

Le machine Learning (ML) permet aux ordinateurs d'apprendre à partir des données sans instructions explicites. En analysant de vastes ensembles de données, les algorithmes de ML peuvent reconnaître des motifs, faire des prédictions et détecter des anomalies. Cela facilite ainsi la compréhension d'événements complexes, la création de modèles prédictifs et l'identification proactive de problèmes. [22].

2.2.2. Les type de L'Apprentissage automatique :

- **Machine Learning supervisée :**

L'apprentissage supervisé désigne l'ensemble des méthodes permettant d'exploiter une base d'apprentissage afin d'en extraire des informations utiles, facilitant ainsi la manipulation et la classification de nouvelles données. [23]

Voici certains des algorithmes les plus essentiels de l'apprentissage supervisé : [24]

- **Régression** : régression linéaire, régression logistique, méthodes d'ensemble, arbres de décision, forêts aléatoires, et régression par processus gaussien (GPR).
- **Classification** : machine à vecteurs de support (SVM), naïve Bayes, k-plus proches voisins (k-NN).
- **Réseaux neuronaux**.

- **Machine Learning non supervisée :**

Lorsqu'aucune étiquette prédéfinie n'est utilisée, les algorithmes d'apprentissage automatique non supervisés analysent et traitent les données de manière autonome, sans intervention humaine. [25]

Voici certains des principaux algorithmes d'apprentissage non supervisé : [24]

- **Clustering** : K-means, classification hiérarchique, réseaux de neurones, analyse en composantes principales (ACP), ACP à noyau.
- **Analyse d'association** : méthodes catégorielles.

- **Machine Learning par renforcement :**

L'apprentissage par renforcement (Reinforcement Learning) est une approche de Machine Learning de plus en plus répandue. Elle repose sur le principe selon lequel les ordinateurs apprennent par l'expérience, en recevant des récompenses ou des pénalités en fonction de leurs actions. Cette méthode pourrait même représenter une avancée clé vers le développement d'une intelligence artificielle générale, capable de rivaliser avec l'intelligence humaine.

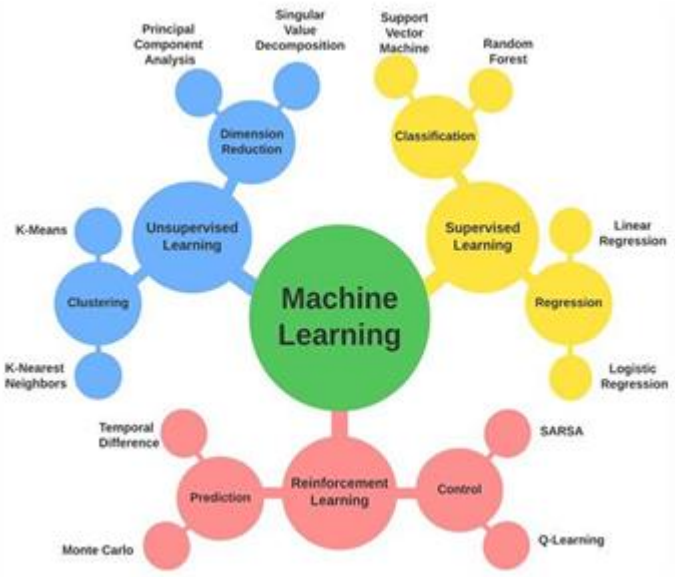


Figure2. 2: Les techniques de Machine Learning. *[64]*

2.2.3. Algorithmes de L'Apprentissage automatique :

Il existe une large gamme d'algorithmes de Machine Learning, dont certains sont plus fréquemment utilisés, notamment pour le traitement de données étiquetées.

- **Random Forest :**

La Random Forest (RF) est un algorithme d'apprentissage supervisé polyvalent, utilisé aussi bien pour la classification que pour la régression. Réputée pour sa flexibilité et sa facilité d'utilisation, cette méthode repose sur un ensemble d'arbres de décision, où une forêt devient d'autant plus robuste qu'elle contient d'arbres. Les forêts aléatoires génèrent ces arbres à partir d'échantillons de données sélectionnés de manière aléatoire, puis agrègent les prédictions de chaque arbre en déterminant la meilleure solution par vote. En outre, elles fournissent un indicateur fiable de l'importance des caractéristiques (features). [26]

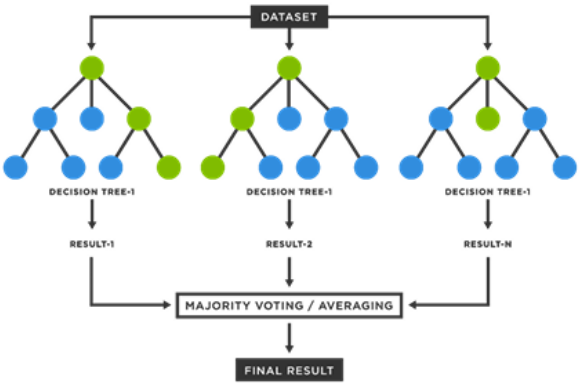


Figure2. 3: Fonctionnement de Random Forest. *[65]*

- **Support Vector Machine (SVM) :**

Cet algorithme offre une précision de classification intéressante. Il est défini comme un espace vectoriel de dimension finie, où chaque dimension correspond à une caractéristique ou un attribut d'un objet. [26]

- **Les arbres de décision:**

Cet algorithme adopte une structure arborescente ou similaire à un organigramme. Il est composé de branches, de feuilles, de nœuds et d'un nœud racine. Les nœuds internes contiennent les attributs, tandis que les branches représentent le résultat de chaque test effectué sur un nœud. Très utilisé pour la classification, l'arbre de décision (DT) se distingue par sa facilité d'utilisation, nécessitant peu de connaissances préalables ou de réglages spécifiques pour fonctionner efficacement. [26]

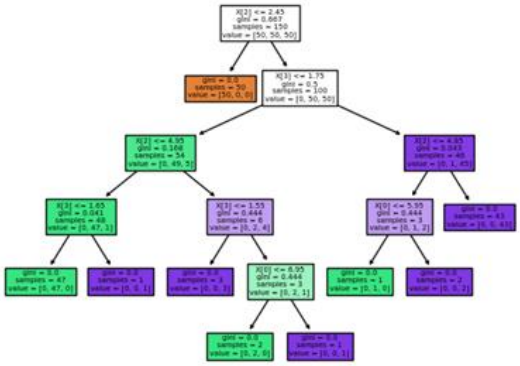


Figure2. 4: Arbre de décision. *[66]*

- **Les k plus proches voisins (KNN) :**

Cet algorithme attribue une classe à une nouvelle instance en se basant sur le vote majoritaire de ses plus proches voisins. Il utilise la distance euclidienne pour mesurer l'éloignement d'un attribut par rapport à ses voisins. [26].

2.2.4. Les métriques d'évaluations des systèmes de l'Apprentissage automatique :

Les métriques ci-dessous sont utilisées par les chercheurs pour évaluer les modèles de prédiction et analyser leurs performances. Nous en proposons une brève définition sans entrer dans les détails approfondis ni les équations mathématiques. [26]

- **Accuracy** : Indique le pourcentage de résultats corrects.

$$\textbf{Accuracy} = \frac{\textbf{TP} + \textbf{TN}}{\textbf{TP} + \textbf{TN} + \textbf{FP} + \textbf{FN}}$$

- **Precision** : Évalue la pertinence des résultats obtenus.

$$\textbf{Precision} = \frac{\textbf{TP}}{\textbf{TP} + \textbf{FP}}$$

- **Recall (ou Sensibilité)** : Mesure la capacité du modèle à identifier les résultats pertinents.

$$\textbf{Recall} = \frac{\textbf{TP}}{\textbf{TP} + \textbf{FN}}$$

- **F-Measure (Le F1-Score)** : Combine la précision et le rappel pour un équilibre entre les deux.

$$\textbf{F1_Score} = \frac{\textbf{2} * \textbf{Precision} * \textbf{Recall}}{\textbf{Precision} + \textbf{Recall}}$$

2.3. Apprentissage profonde « Deep Learning (DL) »:

2.3.1. Définition:

Le Deep Learning (DL), un sous-domaine avancé du machine Learning, exploite des réseaux de neurones complexes pour obtenir des résultats révolutionnaires. Contrairement aux modèles traditionnels, il s'appuie sur des architectures multicouches inspirées du cerveau humain, combinées à des transformations avancées et des technologies basées sur les graphes. Grâce à l'abondance des données et à la puissance du matériel informatique, le DL permet des avancées majeures dans des domaines tels que la reconnaissance d'images et le traitement du langage naturel. [27].

2.3.2. Types de réseaux DL :

Les réseaux de neurones en Deep Learning, également appelés réseaux de neurones artificiels, cherchent à imiter le fonctionnement du cerveau humain en combinant des entrées de données, des poids et des biais. Ces éléments interagissent pour reconnaître, classer et décrire avec précision les objets présents dans les données. Les modèles les plus connus seront abordés dans cette section. [27]

- **Recursive Neural Networks (RvNN) :** Les réseaux de neurones récursifs (RvNNs), inspirés de la mémoire auto-associative, sont particulièrement adaptés au traitement de structures de données hiérarchiques telles que les arbres ou les graphes. Ils peuvent gérer des entrées de tailles variables et produire des sorties de taille fixe grâce à un système d'apprentissage par rétropropagation. Très utilisés en traitement du langage naturel (NLP), ils permettent notamment la classification de phrases et la construction d'arbres

syntaxiques. Leur fonctionnement repose sur la fusion itérative des unités, le calcul de scores de plausibilité et la génération de vecteurs compositionnels représentant l'ensemble de la structure. [27]

- **Recurrent Neural Networks (RNN) :** Les réseaux de neurones récurrents (RNNs) sont des algorithmes de deep learning largement utilisés pour des tâches telles que le traitement de la parole et le traitement du langage naturel (NLP). Contrairement aux réseaux traditionnels, ils sont conçus pour gérer des données séquentielles, ce qui leur permet de capturer le contexte et d'agir comme des unités de mémoire à court terme. Toutefois, ils rencontrent des difficultés avec les dépendances à long terme en raison du problème des gradients qui s'atténuent. [27]
- **Long Short Term Memory (LSTM) :** Les LSTM sont une variante des réseaux de neurones récurrents (RNN), conçus pour apprendre et mémoriser les dépendances à long terme. Cela permet aux RNN de mieux se souvenir des événements passés et d'identifier des motifs temporels pour formuler des prédictions cohérentes. Les LSTM ont établi de nouveaux records en améliorant la traduction automatique, la modélisation du langage et le traitement multilingue du langage naturel. [28]
- **Les réseaux de neurones convolutionnels (CNN ou Convolutional Neural Networks) :** Décomposent un problème en plusieurs sous-parties. Pour chacune d'elles, un groupe de neurones est formé afin d'analyser une portion spécifique des données. Par exemple, dans le cas d'une image en couleur, celle-ci peut être segmentée selon sa largeur, sa hauteur et sa profondeur (représentant les différentes couleurs). [29]

Il existe aujourd'hui plusieurs architectures CNN pré-entraînées largement utilisées pour les tâches de vision par ordinateur, notamment **VGG16**, **ResNet**, **Inception** ou encore **MobileNet**, qui permettent d'accélérer le développement des modèles en tirant parti de connaissances déjà acquises sur de grands ensembles de données comme ImageNet.

- **VGG16 :**

L'architecture VGG-16 (Visual Geometry Group-16) est un modèle d'apprentissage profond développé par des chercheurs de l'Université d'Oxford. Ce réseau de neurones convolutionnels (CNN) a été spécifiquement conçu pour les tâches de reconnaissance et de classification d'images. VGG-16 est largement reconnu pour sa simplicité architecturale ainsi que pour sa complexité computationnelle relativement réduite comparée à d'autres modèles plus complexes tels que ResNet et Inception. L'architecture comprend un total de 16 couches de poids, dont 13 couches convolutionnelles, 3 couches entièrement connectées et 5 couches de max-pooling. [30]

Explorons plus en détail la structure de VGG16 : Ce qui distingue particulièrement VGG16, c'est le choix délibéré de ses concepteurs de réduire le nombre d'hyperparamètres en privilégiant l'utilisation de couches de convolution avec des filtres de taille 3x3, un pas

(stride) de 1 et un rembourrage (padding) constant. À chaque étape, ces couches sont suivies de couches de MaxPooling utilisant des filtres 2x2 avec un pas de 2, selon un schéma strictement répété tout au long de l'architecture. En fin de réseau, on retrouve deux couches entièrement connectées (FC), suivies d'une couche Softmax pour la classification finale. Le chiffre 16 dans VGG16 fait référence aux 16 couches contenant des poids (13 convolutions et 3 FC). Bien que performant, ce modèle reste volumineux avec environ 138 millions de paramètres, ce qui le rend plus lent à entraîner et plus exigeant en ressources comparé à d'autres architectures. [31]

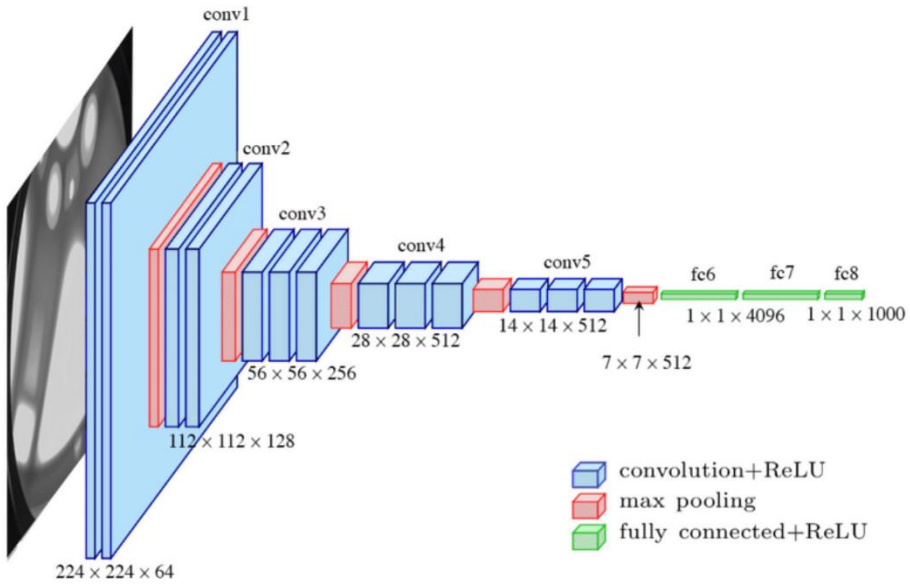


Figure2. 5 : Architecture du VGG16. *[67]*

2.4. Relation entre IA et ML et DL :

- L'IA est l'imitation d'une réponse d'un expert dans un domaine.
- Le ML est de l'IA reposant sur des statistiques (Le ML adapte son algorithme mais a besoin d'un expert pour les données)

Le DL est du ML utilisé avec des réseaux de neurones (Le DL extrait l'information pertinente des données et adapte son algorithme).

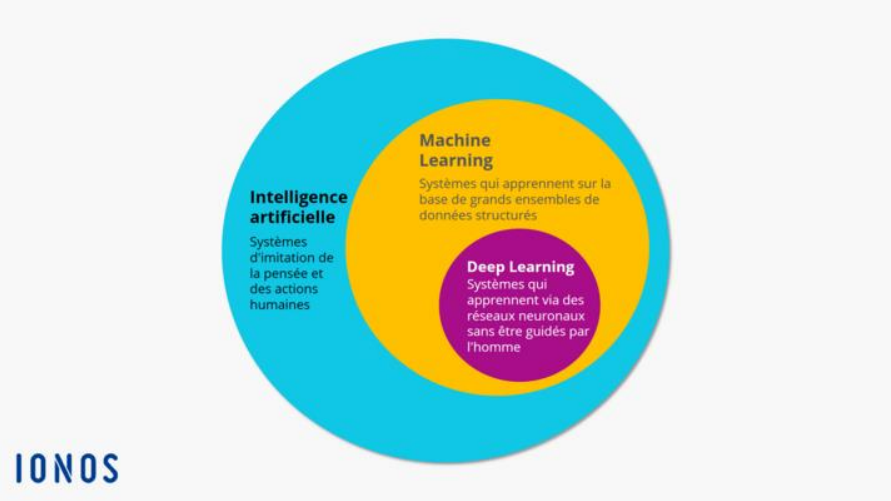


Figure2. 6 : Relation entre IA et ML et DL. *[68]*

3. Approches existantes pour la détection de l'engagement académique :

3.1. Etat de l'art:

Ces dernières années, plusieurs recherches ont été menées sur la prédiction dans le domaine du suivi oculaire. Dans cette section, nous présenterons quelques-unes des solutions récentes explorées. Le tableau 2.1 synthétise les résultats obtenus par ces différentes approches.

- Dans l'étude de [32], une méthode de reconnaissance de l'état des yeux a été proposée en combinant un DNN et un CNN. Le transfert d'apprentissage a été utilisé pour mieux performer sur de petits jeux de données. Ils ont aussi testé une architecture CNN-LSTM sur un dataset de suivi du regard, atteignant 82,27 % de précision. Les expériences sur les datasets CEW, Zhejiang Eyeblink et somnolence au volant montrent que le modèle est robuste. Il est plus performant que les méthodes existantes.
- Par ailleurs, en 2018**,** [33] le modèle Dilated-Net a été proposé pour améliorer l'estimation du regard en intégrant des convolutions dilatées dans les CNN. Testé sur les bases de données Columbia Gaze et MPIIGaze, le modèle a permis une meilleure extraction des caractéristiques visuelles, en tenant compte des variations subtiles dans l'apparence des yeux. Les expérimentations ont montré une amélioration de précision allant jusqu'à 20,8 % par rapport aux CNN classiques, notamment dans la régression du vecteur gaze 3D. Le modèle multi-région Dilated-Net a ainsi démontré des performances supérieures sur les deux datasets.
- De manière similaire, dans l'étude menée par [34]**,** une méthode d'estimation du regard 3D a été proposée avec des réseaux CNN récurrents. Contrairement aux méthodes classiques, elle utilise le visage, les yeux et les points faciaux. Le modèle multi-flux avec module

récurrent capture les infos sur plusieurs images. Testé sur EYEDIAP, il dépasse les anciennes méthodes avec +14,6 % de précision. L'ajout du temps améliore encore les résultats de +4 %, même avec des changements de pose ou de lumière.

- En complément, pour améliorer la précision de la détection de la région du regard, [35] ont proposé un CNN multitâche basé sur la régression des angles de yaw et pitch, permettant une estimation continue de la direction du regard. Le modèle combine cette approche avec l'apprentissage simultané de l'orientation de la tête. Testé sur le Columbia Gaze Dataset, il a atteint une précision de 95,8 % sur l'ensemble de test et 78,2 % en test inter-sujets, démontrant sa robustesse et sa capacité de généralisation.
- En outre, en 2020, [36] une méthode de détection du contact visuel a été proposée en combinant un CNN pré-entraîné avec un SVM. Le CNN a servi d'extracteur de caractéristiques pour la région oculaire. Testée sur le Columbia Gaze Dataset (CAVE-DB), l'approche a montré que l'augmentation des images améliore la précision de la classification binaire. Le modèle a atteint une précision de 91,04 % sans lunettes, contre 86 % avec lunettes, soulignant l'impact des accessoires sur la performance.
- De plus, dans l'étude de [37], un système de détection de la distraction des étudiants en environnement VR éducatif a été développé à partir de données de regard oculaire. Un dataset de 3,4 millions de points a été collecté, et trois classificateurs profonds ont été comparés. Le modèle CNN-LSTM proposé a obtenu une précision de 89,8 % pour classer la distraction en trois niveaux, selon les sections d'activité éducative.
- Par ailleurs, dans l'étude de [38], une méthode de calibration sans contact pour l'estimation du regard a été développée avec un CNN. Elle combine un composant facial et 39 points de repères faciaux pour améliorer l'estimation. Le modèle, après ajustement avec VGG16, atteint 93,36 % de précision. L'utilisation des repères faciaux a renforcé les performances. Cette méthode est prometteuse pour des estimations dans des environnements non contrôlés.
- De la même manière, dans l'étude menée par [39], une méthode de détection du regard, appelée DRNet, a été proposée pour améliorer la robustesse des prédictions en exploitant les différences entre les deux yeux. Le modèle repose sur l'extraction de caractéristiques binoculaires, un module différentiel, et des connexions résiduelles pour optimiser l'estimation. Évalué sur MPIIGaze et Eyediap, DRNet a atteint des erreurs angulaires de 4,57° et 6,14°, surpassant les CNN classiques, même en conditions de bruit visuel.
- En parallèle, dans l'étude de [40], un modèle C-LSTM est proposé pour améliorer l'estimation du regard en exploitant des séquences vidéo. Des couches CNN extraient les traits du visage et des yeux, puis un LSTM capte les variations temporelles pour localiser le point de regard. Le modèle atteint une erreur moyenne de 0.82 cm en horizontal, 0.92 cm en vertical et 6.1° en angulaire. Les tests sur GazeCapture et MPIIGaze confirment sa haute précision.
- En d'autres termes, dans l'étude menée par [41], les auteurs ont utilisé le dataset Eye Dataset de Kayvan pour améliorer la qualité des données à travers un processus de

nettoyage automatique. Leurs étapes incluent la détection des images bruitées, floues ou mal étiquetées, grâce à un modèle CNN appelé VGG_NIN. Ils ont aussi utilisé un algorithme de clustering pour repérer les erreurs de labels et ont répété le processus jusqu'à obtenir un dataset propre. Après nettoyage, ils ont obtenu une précision de plus de 96 %, ce qui montre l'impact direct de la qualité des données sur la performance du modèle. Leur méthode confirme que même un modèle simple peut très bien fonctionner si les données sont bien préparées.

- Dans le même esprit, dans l'étude menée par [42], le dataset Columbia Gaze a été utilisé pour valider un système de prédiction du regard basé sur des CNN. L'approche comprend trois étapes : estimation de la position de la tête, détection des yeux avec Viola-Jones, et estimation du regard avec plusieurs architectures, dont un modèle B-CNN. Le modèle "CoGaz-Net" a atteint 96,88 % de précision, surpassant Inception-ResNet-V2 et VGG-16. Cela montre l'efficacité des CNN et l'importance d'architectures optimisées. Les résultats soulignent la robustesse de cette méthode.
- Enfin, dans l'étude de [43], un modèle nommé Gaze-Swin a été proposé pour l'estimation du regard, combinant un Transformer Swin et un CNN dans une architecture à double branche. Le Transformer capture les dépendances à long terme et les caractéristiques globales, tandis que le CNN gère les détails locaux. Un mécanisme d'attention amélioré, le DA-Attention, a été intégré pour renforcer la robustesse. Le modèle atteint 95 % de précision sur trois jeux de données de référence. Cela surpasse les méthodes existantes et montre la force du modèle hybride CNN-Transformer.

Les résultats des différentes approches étudiées pour la prédiction du suivi oculaire sont également synthétisés dans le tableau2.1.

Référence	Date	Méthodologie	Dataset	Résultats de performance
[32]	2017	CNN-LSTM	Closed Eyes in the Wild (CEW). Eyeblink dataset Dataset personnalisé	Accuracy: 82.27%
[33]	2018	Dilated-Net (CNN avec convolutions dilatées)	Columbia Gaze & MPIIGaze	Amélioration de la précision jusqu'à 20,8 % par rapport aux CNN classiques
[34]	2018	Réseaux de neurones convolutionnels récurrents (Recurrent CNN) exploitant l'apparence et la forme pour une estimation 3D du regard.	EYEDIAP	Réduction de l'erreur de 14,6 % par rapport aux méthodes existantes.
[35]	2020	CNN multitâche : régression pour prédire yaw & pitch + estimation de l'orientation de la tête	Columbia Gaze Dataset (5 880 images, 56 individus)	Accuracy: 95.8% (test), 78,2 % (inter-sujets)

[36]	2020	CNN pré-entraîné + SVM pour la classification du contact visuel	Columbia Gaze Dataset (CAVE-DB)	Accuracy : 91,04 % (sans lunettes), 86 % (avec lunettes)
[37]	2021	CNN-LSTM	Dataset personnalisé de regards (Eye gaze data)	Accuracy: 89.80%
[38]	2022	VGG-16 (pre-trained)	GazeCapture TabletGaze	Accuracy: 93.36%
[39]	2022	Réseau résiduel différentiel profond (DRNet) exploitant la différence d'informations entre les deux yeux.	MPIIGaze, EYEDIAP	Erreur angulaire: 4,57° (MPIIGaze), 6,14° (EYEDIAP).

[40]	2022	CNN-LSTM	MPIIGaze GazeCapture ETH-XGaze	Improved accuracy
[41]	2023	CNN (VGG_NIN)	Eye Dataset	Accuracy :96%
[42]	2023	CNN (VGG16, ResNet50, InceptionV3), Viola-Jones pour la détection des yeux, B-CNN	Columbia Gaze Dataset	Accuracy: 96,88 % (CoGaz-Net), 91,94 % (Inception-ResNet-V2), 81,38 % (VGG-16)
[43]	2024	CNN-Transformer Hybrid (Gaze-Swin)	ETH-XGaze Gaze360 EyeDiap RT-Gene	Superior performance (95%)

Tabeau 2. 1:Résultats des travaux connexes pour la prédiction du suivi oculaire.

3.2. Synthèse:

Plusieurs travaux ont été menés autour de l'estimation du regard et de la détection de l'état d'un individu à partir des mouvements oculaires. Ces recherches s'appuient principalement sur des techniques de Deep Learning, notamment les réseaux de neurones convolutifs (CNN), parfois combinés à d'autres architectures comme les LSTM, les SVM ou

même les Transformers. Les résultats obtenus sont très encourageants, avec des précisions allant de 82 % à plus de 96 % selon les études.

Certaines approches utilisent des séquences d'images pour mieux capter la dynamique du regard dans le temps, en combinant CNN et LSTM. D'autres méthodes s'appuient sur des informations complémentaires comme les points faciaux, la position de la tête ou encore les deux yeux (vision binoculaire), ce qui permet une meilleure estimation du regard, même dans des conditions difficiles.

On retrouve aussi des modèles qui intègrent des mécanismes plus avancés, comme les convolutions dilatées, l'attention ou les architectures hybrides CNN-Transformer. L'utilisation de modèles pré-entraînés ou d'architectures comme VGG16, Inception-ResNet ou encore B-CNN permet également d'améliorer les performances, surtout lorsqu'on travaille avec des jeux de données standards comme Columbia Gaze, MPIIGaze ou EYEDIAP.

Malgré ces bons résultats, quelques défis restent à relever. Par exemple, certains modèles ont du mal à généraliser sur de nouvelles personnes ou dans des environnements différents. Le port de lunettes, les changements de luminosité ou les petits jeux de données peuvent aussi influencer les performances.

Globalement, on remarque que les tendances récentes vont vers des modèles plus flexibles, capables de fonctionner dans des conditions variées, voire en temps réel. L'intégration de séquences, l'exploitation de différentes régions du visage, et l'adaptation à des environnements non contrôlés sont des pistes prometteuses pour les travaux futurs.

3.3. Justification du choix de CNN(VGG16) - LSTM :

Pour détecter l'engagement académique à partir du suivi oculaire, nous avons choisi un modèle hybride puissant : **CNN(VGG16) - LSTM**. Ce choix repose sur les besoins spécifiques de notre tâche et sur les forces complémentaires de ces deux architectures.

Le modèle CNN(VGG16) - LSTM combine deux approches complémentaires. Le CNN, ici **VGG16**, est très efficace pour extraire des détails importants dans une image, comme la position des yeux ou la direction du regard. Ensuite, le LSTM prend le relais pour analyser l'évolution du regard dans le temps, image après image. C'est important, car l'engagement ne se perçoit pas toujours sur une seule image, mais plutôt à travers une séquence.

Grâce à cette combinaison, le modèle peut interpréter à la fois les caractéristiques visuelles instantanées et leur évolution temporelle, ce qui en fait un outil particulièrement adapté à notre problématique de détection d'engagement académique.

4. Conclusion :

Dans ce chapitre, nous avons exploré le rôle de l'intelligence artificielle dans la détection de l'engagement académique. Nous avons défini les concepts clés du Machine Learning (ML) et du Deep Learning (DL), en présentant leurs algorithmes et métriques d'évaluation, tout en mettant en évidence leur complémentarité avec l'IA.

Nous avons ensuite comparé les méthodes traditionnelles de mesure de l'engagement aux approches modernes basées sur l'IA et le Deep Learning. Cette analyse nous a permis de mettre en évidence les avantages des techniques intelligentes pour une évaluation plus précise et objective de l'engagement des étudiants.

Enfin, nous avons justifié le choix des modèles CNN-LSTM, en expliquant leur pertinence pour notre approche. Ces modèles combinent la puissance des réseaux convolutionnels et récurrents, offrant ainsi une solution robuste pour la détection automatique de l'engagement académique.

Chapitre 03 : Détection de l'engagement académique Par suivi oculaire (Conception et évaluation du modèle)

1. Introduction:

Après avoir présenté les fondements théoriques et les approches existantes dans les chapitres précédents, ce chapitre est dédié à la mise en œuvre pratique du modèle proposé pour la détection de l'engagement académique à partir de données de suivi oculaire. Il détaille la méthodologie adoptée, les outils utilisés, le traitement des données, ainsi que la conception, l'entraînement et l'évaluation des modèles CNN (VGG16) et LSTM, ainsi que l'évaluation des performances du modèle.

2. Présentation des outils utilisé:

Dans le cadre de notre projet, le développement du modèle de détection de l'engagement académique a été réalisé à l'aide de **Jupyter**, un environnement interactif largement utilisé pour les projets de machine learning et de deep learning, en raison de sa souplesse et de son interface conviviale. Il permet d'exécuter du code, de visualiser les résultats, et de documenter chaque étape du traitement de manière reproductible.

Le langage **Python** a été choisi pour sa richesse en bibliothèques spécialisées dans l'intelligence artificielle. Parmi les bibliothèques principales, on peut citer :

TensorFlow : un framework puissant pour construire et entraîner des modèles de deep learning, notamment pour gérer les opérations sur GPU [44]

Keras : une API de haut niveau intégrée à TensorFlow qui facilite la création de modèles comme CNN ou LSTM, avec une syntaxe simple et modulaire [45].

NumPy : utilisée pour les opérations numériques et la manipulation de matrices d'images

Matplotlib : permet la visualisation des résultats sous forme de courbes d'apprentissage, de matrices de confusion, etc.

Scikit-learn : utilisée pour le calcul des métriques d'évaluation telles que la accuracy, précision, le rappel et le F1-score [46] .



Tkinter : est une bibliothèque standard de Python dédiée à la création d'interfaces graphiques (GUI). Elle est incluse par défaut avec Python et repose sur Tk, une extension graphique du langage de script Tcl (Tool Command Language). [47]



Pandas : est une bibliothèque Python conçue pour faciliter l'analyse et la manipulation de données. Elle offre des outils puissants comme l'alignement automatique et l'indexation hiérarchique. [48]



OpenCV : qui signifie Open Source Computer Vision, est une bibliothèque de fonctions utilisée pour le développement d'applications de vision par ordinateur en temps réel. La vision par ordinateur désigne l'analyse automatique d'images et de vidéos numériques à l'aide de programmes informatiques. [49]



Pillow : est une bibliothèque de traitement d'images basée sur PIL. Compatible avec tous les systèmes et avec Python 3, Pillow prend en charge de nombreux formats comme JPEG, PNG, BMP, GIF…etc. Elle permet diverses opérations sur les images : filtrage, conversion de couleurs, et bien plus encore. [50]

Ces outils ont permis de développer et de tester efficacement une architecture hybride basée sur CNN (VGG16) et LSTM, en assurant un traitement fluide des données et une analyse rigoureuse des performances du modèle.

3. Description du dataset:

Le dataset utilisé dans ce travail est celui fourni par **Kayvan Shah,** intitulé ***"Eye-dataset"*** (Eye seeing in that direction) [51], il se compose d'environ **14 500 images** extraites de séquences vidéo, représentant les mouvements des yeux d'étudiants en situation d'apprentissage, cependant les différentes directions du regard sont :

- ***Close_look*** : images d'yeux fermés,
- ***Left_look*** : images d'yeux regardant vers la gauche,
- ***Right_look*** : images d'yeux regardant vers la droite,
- ***Forward_look*** : images d'yeux regardant droit devant.

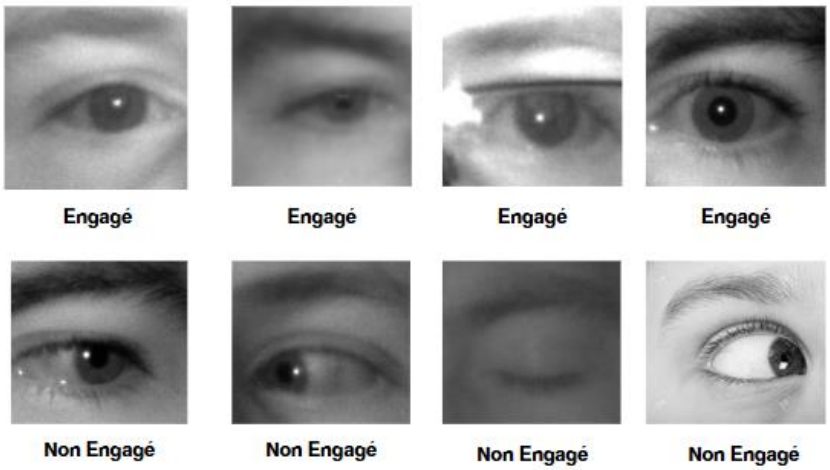


Figure3. 1: Exemple d'images extraites du dataset Eye Gaze. *[51]*

Les images, au format JPG et PNG, ont été redimensionnées pour l'uniformité et sont annotées par un label binaire indiquant l'état d'engagement : 0 pour Engagé "Engaged" et 1 pour non engagé "Not Engaged". Ce dataset permet de capturer des indices visuels comme la direction du regard, les fixations, les clignements ou encore les mouvements saccadiques, qui sont des indicateurs clés de l'attention et de la concentration. L'objectif principal est d'utiliser ces données pour entraîner un modèle de classification binaire capable de prédire automatiquement le niveau d'engagement académique d'un étudiant à partir des images de ses yeux. Le dataset Eye Gaze est accessible via son dépôt sur Kaggle [51].

4. Chargement du dataset :

Ce travail a été réalisé localement en utilisant **Jupyter Notebook** comme environnement de développement. Le dataset utilisé, intitulé **"Eye Dataset"** [51], a été **préalablement traité** et sauvegardé localement dans un fichier compressé au format. npz, nommé dataset.npz. Ce fichier contient trois ensembles principaux : **entraînement (X_train, y_train)**, **validation (X_val, y_val)** et **test (X_test, y_test)**. Le chargement des données s'effectue à l'aide de la bibliothèque **NumPy**, en utilisant la fonction np.load() qui permet de récupérer directement les tableaux de données et leurs étiquettes associées. Cette méthode offre un gain de temps significatif lors de l'entraînement des modèles, car les données sont déjà normalisées, redimensionnées et prêtes à l'emploi.

```
[2]: # Chargement du dataset
dataset_path = r"C:\Users\Boukhatem\Desktop\memoire_m2\archive\Eye_dataset_spliteq\dataset.npz"
data = np.load(dataset_path)

X_train, y_train = data["X_train"], data["y_train"]
X_val, y_val = data["X_val"], data["y_val"]
X_test, y_test = data["X_test"], data["y_test"]

print("Données chargées depuis dataset.npz !")
print(f"Taille initiale de X_train : {X_train.shape}, X_val : {X_val.shape}, X_test : {X_test.shape}")
```

```
Données chargées depuis dataset.npz !
Taille initiale de X_train : (5530, 128, 128, 3), X_val : (690, 128, 128, 3), X_test : (694, 128, 128, 3)
```

Figure3. 2: Chargement du dataset.

5. Prétraitement des données:

Avant l'entraînement du modèle, un prétraitement des images a été nécessaire afin d'assurer l'homogénéité des données et d'améliorer la performance de l'algorithme. Dans un premier temps, toutes les images ont été redimensionnées à 128×128 pixels pour correspondre à l'entrée attendue par le modèle pré-entraîné VGG16. Ensuite, une normalisation des pixels a été réalisée en divisant les valeurs des pixels par 255, afin de les ramener dans l'intervalle [0, 1], ce qui accélère la convergence du modèle durant l'apprentissage. Concernant les étiquettes, Un encodage numérique a été réalisé manuellement (0 pour engagé, 1 pour non_ engagé), puis converti en one-hot encoding avec la fonction to_categorical. Enfin, le dataset a été divisé en trois ensembles distincts : 80% pour l'entraînement, 10% pour la validation, et 10% pour les tests. Cette séparation a été réalisée de manière aléatoire tout en conservant la proportion équilibrée entre les deux classes pour éviter tout biais lors de l'apprentissage.

6. Architecture du modèle utilisé:

Pour définir notre modèle hybride de détection de l'engagement académique, nous avons d'abord utilisé un modèle VGG16 pré-entraîné comme extracteur de caractéristiques. Nous avons chargé VGG16 sans ses couches denses supérieures, en ne conservant que la base convolutive, et défini include_top=False avec une forme d'entrée de 128x128x3. Les caractéristiques spatiales extraites sont ensuite passées à une couche d'aplatissement pour les convertir en un vecteur unidimensionnel.

Nous avons ensuite ajouté une couche de mémoire à long terme (LSTM) de 128 unités afin de capturer les dépendances temporelles entre les séquences d'images de regard. Une couche d'abandon avec un taux de 0,3 a été incluse pour éviter le surapprentissage. Enfin, une couche de sortie dense avec une fonction d'activation sigmoïde a été utilisée pour effectuer une classification binaire (Engagé vs. Non Engagé). Nous avons enfin compilé le modèle à l'aide

de l'optimiseur Adam, avec categorical_crossentropy comme fonction de prétraitement et la précision comme principale mesure d'évaluation.

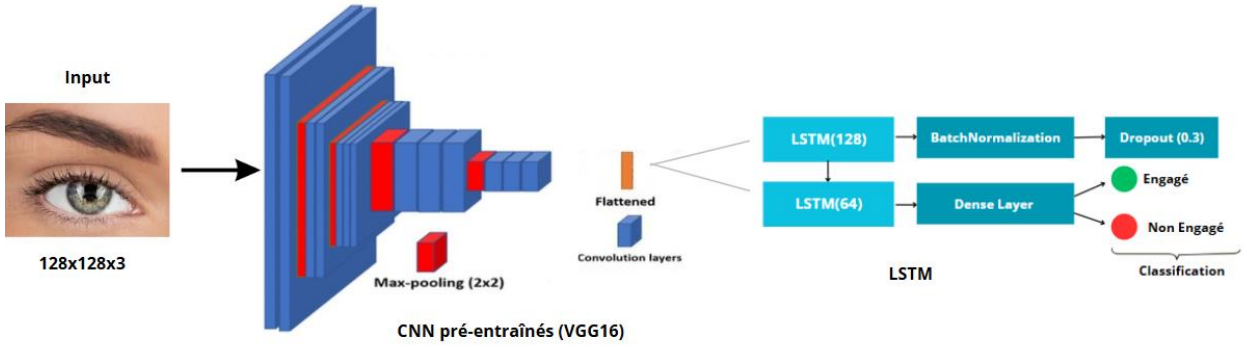


Figure3. 3 : Architecture du modèle proposé.

7. Entraînement du modèle:

Lors de l'entraînement du modèle CNN + LSTM, plusieurs hyperparamètres ont été définis : un taux d'apprentissage de 0.0001, une taille de batch de 32, et un nombre d'époques fixé à 10. Pour la partie LSTM, deux couches ont été utilisées avec respectivement 128 et 64 unités, associées à des techniques de régularisation comme le Dropout (0.4 et 0.3) et la BatchNormalization. La couche de sortie est une Dense avec deux neurones et une activation softmax pour la classification binaire (engagé / non-engagé). L'optimisation a été assurée par l'optimiseur Adam et la fonction de prétraitement to_categorical (). Afin d'améliorer l'apprentissage et éviter le surapprentissage, deux callbacks ont été employés : EarlyStopping et ReduceLROnPlateau. Ces choix ont contribué à stabiliser l'entraînement et à améliorer la performance du modèle.

[6]:
```
# Entraînement du modèle
history = model.fit(
    X_train, y_train,
    epochs=10,
    batch_size=32,
    validation_data=(X_val, y_val),
    callbacks=callbacks
)
```

Figure3.4: Code python pour l'entrainement du modèle vgg16 et LSTM.

8. Evaluation du modèle:

Après l'entraînement du modèle CNN (VGG16) combiné au LSTM, une phase d'évaluation a été réalisée afin de mesurer sa robustesse et sa capacité de généralisation. Cette évaluation s'est faite à partir de l'ensemble de test issu du dataset Eye Dataset [51]. L'évaluation du modèle CNN + LSTM a été réalisée en utilisant différentes métriques, telles que l'accuracy, la précision, le rappel et le score F1. Ces indicateurs permettent de mesurer la performance du modèle sur les données de test en termes de qualité de classification. Une matrice de confusion a également été générée afin d'analyser la répartition des prédictions correctes et incorrectes. Cette étape est essentielle pour vérifier la capacité du modèle à généraliser sur de nouvelles données.

Après la création du modèle, l'étape d'apprentissage a été lancée et répétée plusieurs fois. À chaque fois, le nombre d'époques a été ajusté pour optimiser les résultats. Ce processus a été effectué progressivement en augmentant et en diminuant le nombre d'époques jusqu'à ce qu'on obtienne des résultats valides. Finalement, **10 époques** ont été utilisées pour obtenir des performances satisfaisantes, permettant ainsi au modèle de bien apprendre à partir des données. Les résultats obtenus après 10 époques sont valide, par exemple l'accuracy =0,99 très proche de 1.

La figure 3.5 montre l'apprentissage du modèle en 10 époques.

```
Epoch 1/10
173/173 ━━━━━━━━━━━━━━━━━━━━ 2646s 15s/step - accuracy: 0.8182 - loss: 0.4047 - val_accuracy: 0.9227 - val_loss: 0.2410 - learning_rate: 1.0000e-04
Epoch 2/10
173/173 ━━━━━━━━━━━━━━━━━━━━ 2933s 17s/step - accuracy: 0.9725 - loss: 0.0938 - val_accuracy: 0.9869 - val_loss: 0.0990 - learning_rate: 1.0000e-04
Epoch 3/10
173/173 ━━━━━━━━━━━━━━━━━━━━ 2726s 16s/step - accuracy: 0.9809 - loss: 0.0620 - val_accuracy: 0.9810 - val_loss: 0.0811 - learning_rate: 1.0000e-04
Epoch 4/10
173/173 ━━━━━━━━━━━━━━━━━━━━ 2580s 15s/step - accuracy: 0.9884 - loss: 0.0405 - val_accuracy: 0.9548 - val_loss: 0.0972 - learning_rate: 1.0000e-04
Epoch 5/10
173/173 ━━━━━━━━━━━━━━━━━━━━ 2577s 15s/step - accuracy: 0.9933 - loss: 0.0235 - val_accuracy: 0.9942 - val_loss: 0.0179 - learning_rate: 1.0000e-04
Epoch 6/10
173/173 ━━━━━━━━━━━━━━━━━━━━ 2587s 15s/step - accuracy: 0.9889 - loss: 0.0337 - val_accuracy: 0.9082 - val_loss: 0.1873 - learning_rate: 1.0000e-04
Epoch 7/10
173/173 ━━━━━━━━━━━━━━━━━━━━ 2576s 15s/step - accuracy: 0.9895 - loss: 0.0300 - val_accuracy: 0.5525 - val_loss: 1.6233 - learning_rate: 1.0000e-04
Epoch 8/10
173/173 ━━━━━━━━━━━━━━━━━━━━ 2600s 15s/step - accuracy: 0.9943 - loss: 0.0186 - val_accuracy: 0.7988 - val_loss: 1.0608 - learning_rate: 1.0000e-04
Epoch 9/10
173/173 ━━━━━━━━━━━━━━━━━━━━ 2557s 15s/step - accuracy: 0.9988 - loss: 0.0084 - val_accuracy: 0.8265 - val_loss: 0.5461 - learning_rate: 5.0000e-05
Epoch 10/10
173/173 ━━━━━━━━━━━━━━━━━━━━ 2568s 15s/step - accuracy: 0.9946 - loss: 0.0176 - val_accuracy: 0.9985 - val_loss: 0.0155 - learning_rate: 5.0000e-05
```

Figure3. 5: Apprentissage du modèle en 10 époques.

➢ Après l'entraînement du modèle, on a généré les courbes pour visualiser l'évolution de la performance. Ces graphes montrent comment la perte (loss) et la précision (accuracy) ont évolué pendant les époques, aussi bien pour les données d'entraînement que pour celles de validation. Cela permet de vérifier si le modèle apprend correctement et s'il ne fait pas de surapprentissage.

Les résultats obtenus sont les suivants :

Métrique	Valeur obtenue (%)
Accuracy	99.57 %
Précision	99.43 %
Recall (Rappel)	99.71 %
F1-Score	99.57 %

Tabeau 2. 2: Résultats obtenus par le modèle CNN (VGG16) + LSTM

Le modèle développé montre une **précision parfaite de 99.43 %**, indiquant que presque toutes les instances prédites comme positives correspondent effectivement à des exemples réellement positifs. Par ailleurs, un **recall élevé de 99.71 %** confirme que le modèle a réussi à détecter la grande majorité des cas positifs, avec très peu d'oublis. Enfin, l'**accuracy globale de 99.57 %** démontre la **robustesse générale** du modèle et sa capacité à bien généraliser sur des données inédites. Ces résultats témoignent de l'efficacité de l'architecture CNN (VGG16) combinée à LSTM pour la tâche de détection de l'engagement académique à partir des données de suivi oculaire.

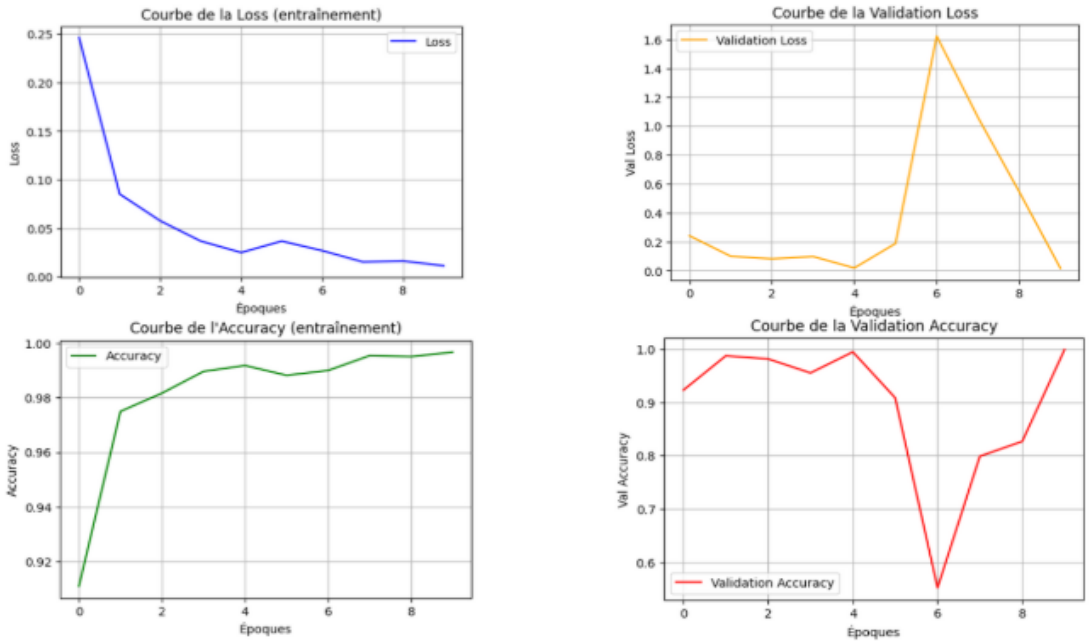


Figure3. 6: Courbes d'apprentissage – Accuracy et Loss pour l'entraînement et la validation.

9. Matrice de confusion:

La figure 3.7 représente la matrice de confusion qui montre que parmi les 343 images des étudiants non engagés, le modèle en détecte correctement 341. De plus, parmi les 347 images des étudiants engagés, il en détecte 346. Cela indique que le modèle fonctionne avec une très haute précision, ce qui le rend prometteur pour être utilisé comme un outil d'évaluation de l'engagement académique des étudiants.

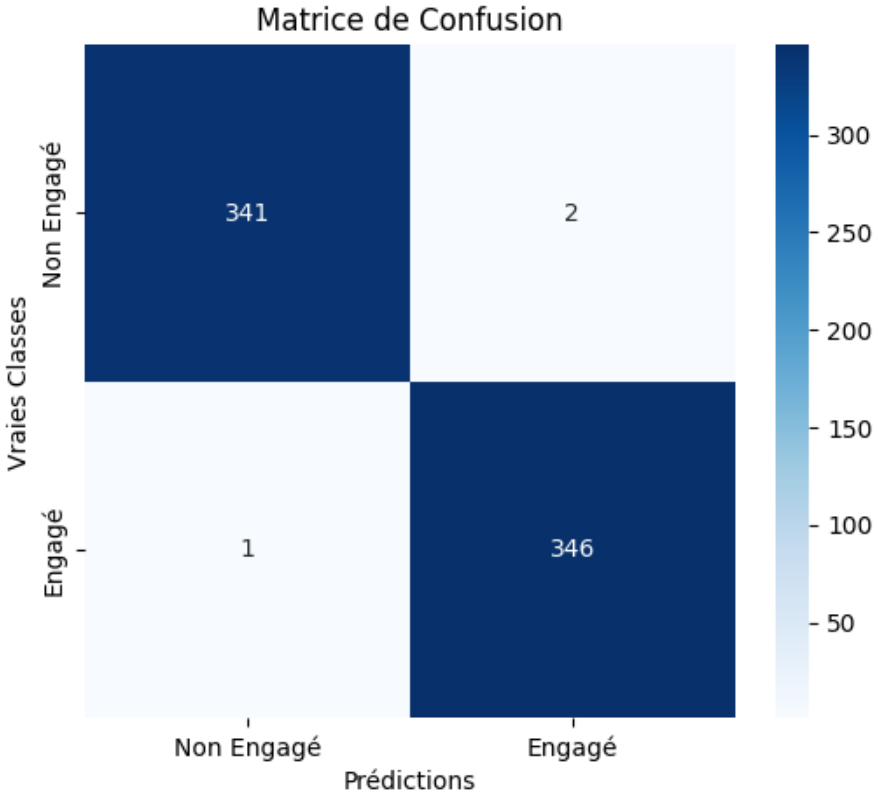


Figure3.7: Matrice de Confusion.

10. Analyse comparative avec les travaux existants :

Le figure 3.8 ci-dessus illustre la comparaison des précisions(accuracy) obtenues par différents travaux qui ont utilisée Eye Dataset [51]. Il apparaît clairement que le modèle proposé dans ce travail, à savoir CNN (VGG16) + LSTM, atteint une accuracy de **99 %**, ce qui constitue la meilleure performance parmi les modèles comparés.

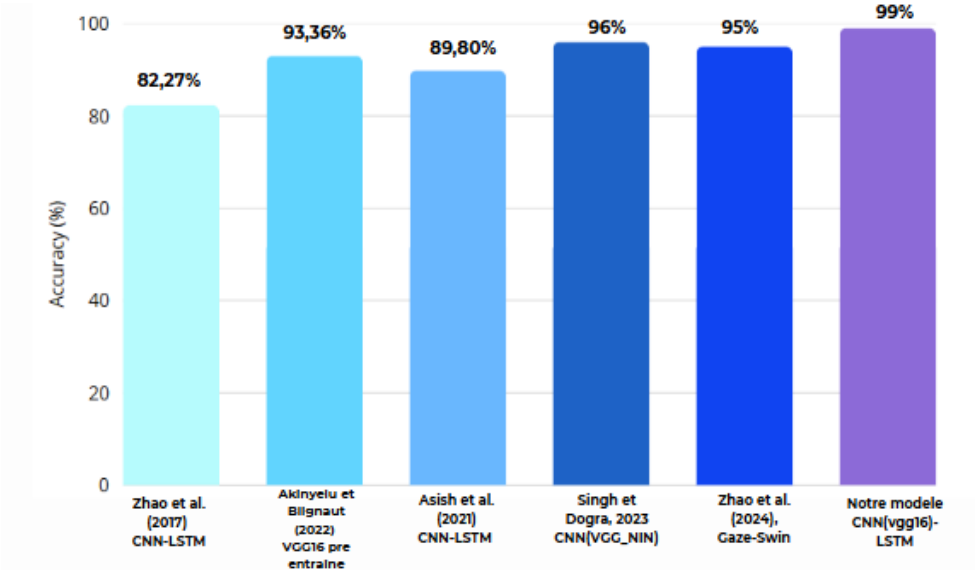


Figure3.8 : Comparaison des précision (Accuracy) obtenues par différents travaux.

- En comparaison:

- Le modèle CNN-LSTM proposé par [32] atteint une précision relativement plus faible de **82.27 %**, ce qui montre l'intérêt d'utiliser un extracteur de caractéristiques plus puissant comme VGG16.
- Le modèle CNN-LSTM [37] atteint **89.8 %**, ce qui reste significativement inférieur au modèle développé dans notre travail.
- Le modèle VGG-16 pré-entraîné de [38] enregistre une précision de **93.36 %,** confirmant que les modèles CNN performants extrayant des caractéristiques visuelles riches apportent un réel avantage.
- Le modèle VGG_NIN utilisé par [41] atteint une précision de 96 %, ce qui met en évidence l'impact d'un bon nettoyage de données sur la performance, même sans recourir à des architectures complexes.
- Enfin, le modèle Gaze-Swin de [43] , utilisant une architecture hybride CNN-Transformer, atteint **95 %**, mais reste tout de même en deçà des performances obtenues par notre approche.

Ainsi, la combinaison judicieuse de VGG16 pour l'extraction spatiale et LSTM pour la modélisation temporelle permet d'exploiter efficacement la richesse des données visuelles et la dynamique du regard, offrant des performances supérieures aux approches traditionnelles et récentes de la littérature.

11. Interface Utilisateur et Fonctionnalités:

11.1. Présentation générale de l'interface :

Afin de rendre le modèle accessible à des utilisateurs, une interface intuitive a été développée. Cette interface permet d'interagir facilement avec le modèle de détection de l'engagement académique, en important des images d'yeux et en recevant une prédiction immédiate sur l'état d'engagement de l'étudiant.

11.2. Fonctionnalités principales :

L'application propose les fonctionnalités suivantes :

- **Chargement d'images** : l'utilisateur peut importer une ou plusieurs images d'yeux depuis son ordinateur ou bien capturer via webcam
- **Classification automatique** : le modèle analyse les images importées et prédit si l'étudiant (personne) est " Engagé " ou "Non Engagé ".
- **Affichage du résultat** : les résultats sont présentés sous forme de texte clair, accompagnés du pourcentage de confiance de la prédiction.

11.2.1. Illustrations de l'interface (Captures d'écran) :

La figure 3.9 montre une interface permettant à l'utilisateur de saisir les informations de base de l'étudiant, telles que le nom, le prénom et la classe. Ces données sont essentielles pour identifier chaque session de classification et structurer les résultats de manière organisée.

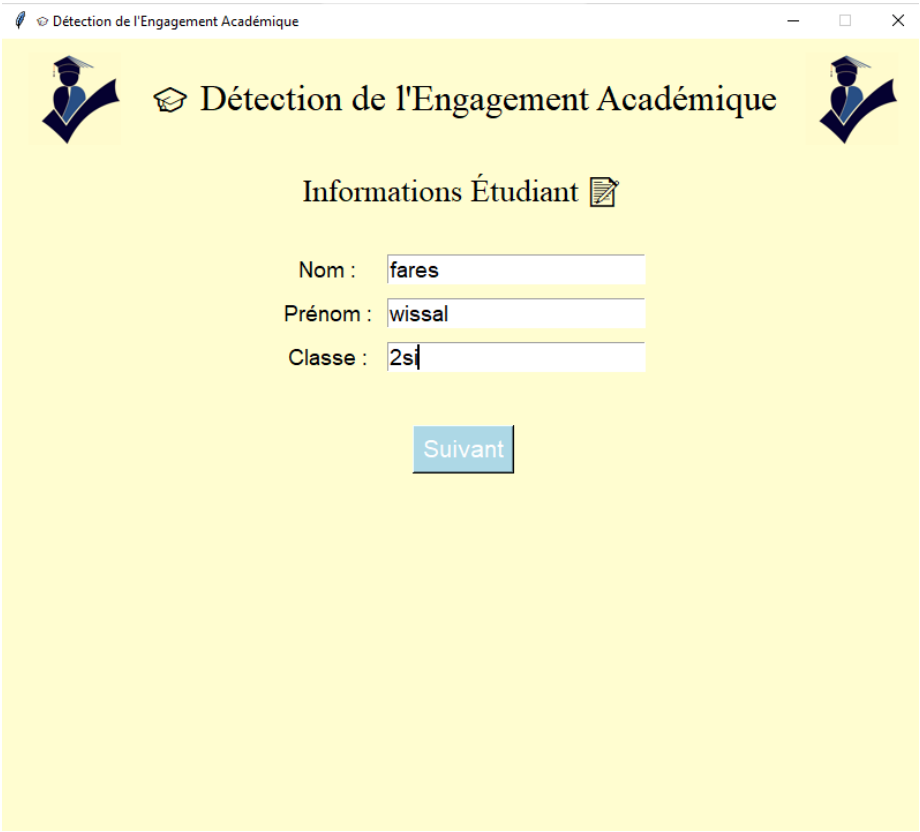


Figure3. 9 : Formulaire d'entrée de l'étudiant.

La figure 3.10 montre l'interface principale à partir de laquelle l'utilisateur peut interagir avec le système. Elle affiche les informations de l'étudiant déjà saisies, et propose plusieurs fonctionnalités :

- **Capture Image Séquence** : pour capturer une séquence d'images à partir de la webcam.
- **Charger une image** : pour importer des images depuis l'ordinateur.
- **Résultat** : une zone prévue pour afficher la prédiction du modèle (engagé ou non engagé).
- **Ouvrir Excel** : pour consulter les résultats enregistrés.

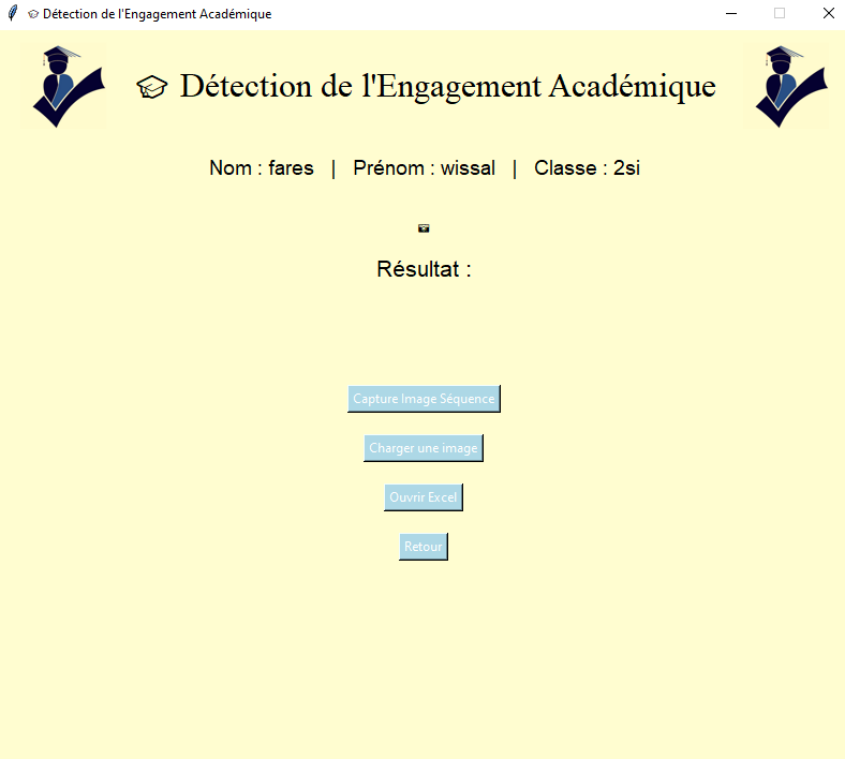


Figure3. 10: Interface de classification et interactions.

La figure 3.11 présente deux exemples de sortie du modèle. L'image de l'œil est affichée au centre de l'interface, suivie du résultat de la classification :

À gauche, le modèle prédit que l'étudiant est Engagé avec une confiance de 99.99%.

À droite, le modèle identifie un cas Non Engagé avec une confiance de 99.52%.

L'utilisation de couleurs (vert pour "Engagé", rouge pour "Non Engagé") facilite la compréhension visuelle du résultat par des utilisateurs non experts.



Figure3. 11: Affichage du résultat de classification.

Lors du clic sur le bouton **"Capture Image Séquence"**, un message temporaire **"Capture en cours..."** s'affiche à l'écran. Cette indication visuelle informe l'utilisateur que le système est en train de capturer la séquence d'images à traiter. Une fois la capture terminée, le message disparaît et le traitement automatique (prédiction) est lancé.

Ce retour visuel améliore l'expérience utilisateur en évitant toute confusion ou attente sans réponse.

La figure 3.12 illustre l'indication de progression affichée lors de la capture d'images.

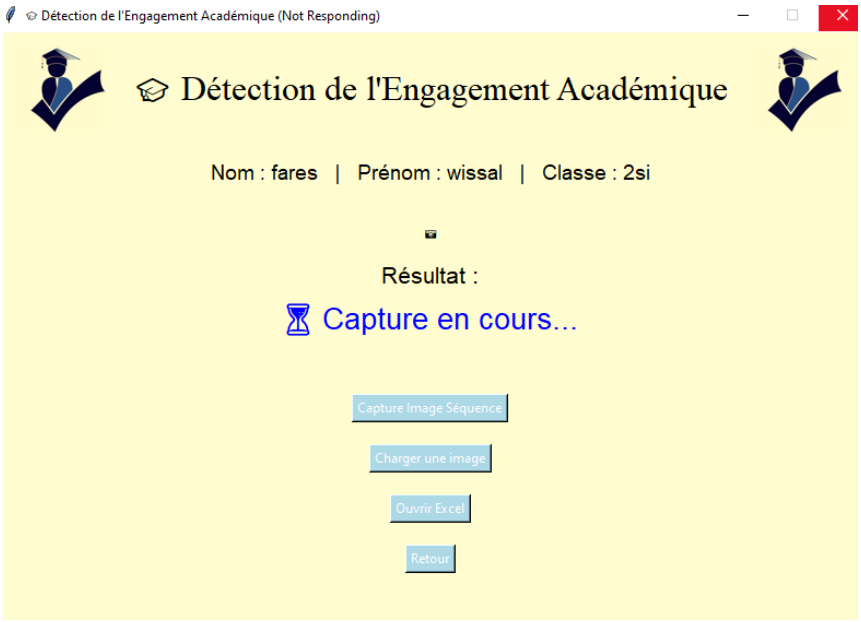


Figure3. 12: l'indication de progression affichée lors de la capture d'images.

La figure 3.13 montre l'interface après la capture complète de la séquence d'images (5 images). Les images sont affichées côte à côte, représentant la séquence analysée par le modèle. Avec l'affichage du résultat **Engagé (99.52%)**.

Le système effectue également une sauvegarde automatique de ces données (nom, prénom, classe, date, résultat, taux de confiance) dans un fichier Excel, ce qui permet d'assurer un suivi structuré et traçable des prédictions effectuées.

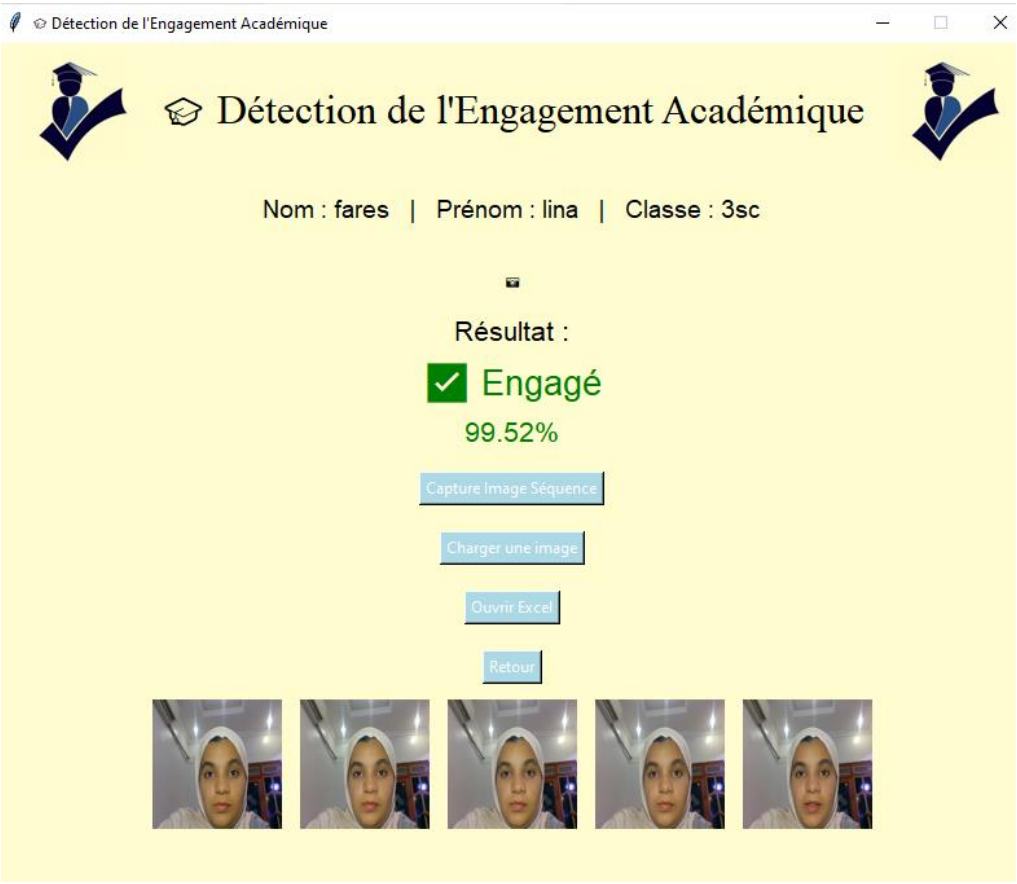


Figure3. 13 : Résultat de classification après capture d'une séquence d'images (Engagé).

La figure 3.14 illustre un cas où le modèle prédit un état de **non-engagement** sur la base d'une séquence capturée.

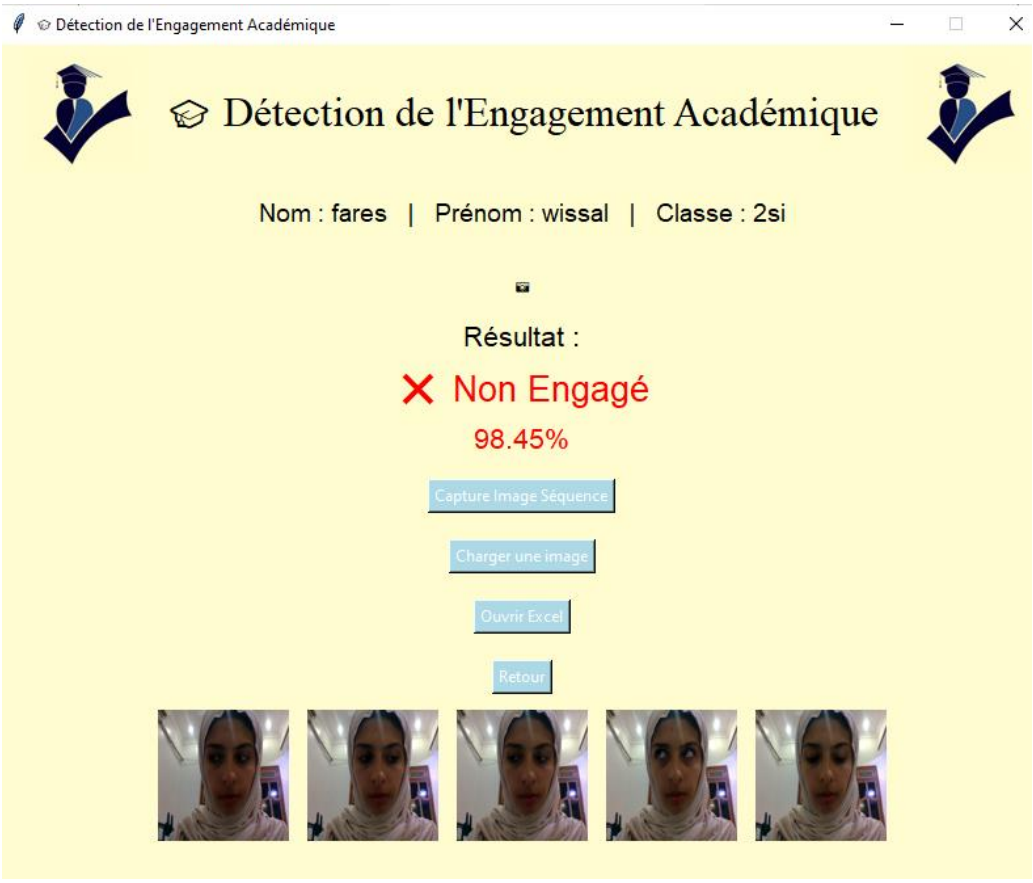


Figure3. 14 : Résultat de classification après capture d'une séquence d'images (Non Engagé).

La figure 3.15 montre que les résultats des prédictions sont automatiquement enregistrés dans un fichier Excel nommé **"résultats_prédictions.xlsx"**. Ce fichier contient les informations suivantes :

- Nom, prénom, classe de l'étudiant
- Résultat de la classification (Engagé / Non Engagé)
- Date et heure de la prédiction

Un bouton dédié **"Ouvrir Fichier Excel"** dans l'interface permet à l'utilisateur de consulter directement ce fichier sans quitter l'application, facilitant ainsi le suivi et l'analyse des résultats enregistrés.

	A	B	C	D	E	F
1	Nom	Prénom	Classe	Résultat	ate et Heure	
2	fares	wissal	2sci	✗ Non E	2025-05-01 21:56:05	
3	fares	wissal	2sci	☑ Engag	2025-05-01 21:56:30	
4	fares	lina	3sci	Engagé	2025-05-01 21:57:36	
5	fares	wissal	2sci	Non Enga	2025-05-01 22:14:45	
6						
7						
8						

Figure3. 15: Sauvegarde des résultats dans le fichier Excel.

Une autre fonctionnalité utile intégrée dans l'application est **l'exportation automatique des résultats au format PDF**. Après chaque prédiction, l'utilisateur peut générer un fichier PDF contenant les informations suivantes :

- Nom, prénom, classe de l'étudiant
- Résultat de la classification

Le fichier est nommé dynamiquement sous la forme : "résultats_nom_prenom_date_heure.pdf "

Par exemple : "résultats_fares_lina_2025-05-01_22-14-45.pdf"

Cela permet d'archiver facilement les résultats individuels et de les partager si nécessaire.

Le PDF généré contient un en-tête "Résultats Prédiction", suivi des détails organisés ligne par ligne.

Résultats Prédiction

Nom: fares

Prénom: lina

Classe: 3sc

Résultat: Engagé

Confiance: 99.52%

Figure3. 16 : Exemple de fichier PDF généré avec les résultats de prédiction.

12. Conclusion :

Ce chapitre a présenté la méthodologie expérimentale adoptée pour la détection de l'engagement académique à partir de données de suivi oculaire. Après avoir décrit le dataset utilisé et le prétraitement appliqué, l'architecture du modèle CNN (VGG16) combiné à LSTM a été détaillée. Les résultats expérimentaux ont démontré l'efficacité de l'approche proposée, avec une précision remarquable de 99.43 %, surpassant plusieurs méthodes existantes de la littérature. Ces performances valident la pertinence du modèle développé pour des applications concrètes dans le domaine de l'éducation intelligente.

Conclusion Générale

Conclusion :

L'objectif principal de notre travail était de développer un système capable de détecter automatiquement l'engagement académique des étudiants à partir de données de suivi oculaire, en combinant des approches de vision par ordinateur et d'apprentissage profond.

Nous avons proposé une architecture basée sur un réseau CNN (VGG16) couplé à un réseau LSTM, permettant de capter à la fois les caractéristiques visuelles des images et la dynamique temporelle du regard. Ce modèle a été entraîné sur un dataset structuré, avec un prétraitement adapté et une séparation équilibrée des données.

L'évaluation du modèle a montré des résultats très satisfaisants, avec une précision globale de 99.43 %, un accuracy de 99.57 %, et une capacité à bien généraliser sur des données inédites. Ces performances surpassent celles de plusieurs méthodes de la littérature, confirmant la pertinence de notre approche.

De plus, une interface utilisateur conviviale a été développée, permettant de charger ou capturer une séquence d'images, d'obtenir un diagnostic clair de l'état d'engagement, et de sauvegarder les résultats. Cela rend l'outil accessible même à des utilisateurs non spécialistes.

En conclusion, cette étude propose une solution fiable et performante pour l'évaluation automatique de l'engagement académique, pouvant être utilisée dans des contextes réels d'enseignement et de suivi pédagogique.

Perspectives :

Plusieurs pistes peuvent être envisagées pour améliorer et étendre ce travail.

Tout d'abord, il serait intéressant de tester le modèle sur des datasets plus variés et plus larges, afin d'enrichir sa capacité de généralisation à différents contextes éducatifs et types de public. Une extension possible est l'inclusion de nouveaux paramètres physiologiques comme les expressions faciales, les mouvements de tête ou encore le clignement des yeux, afin d'affiner davantage la détection de l'engagement.

Ensuite, l'intégration d'architectures plus avancées comme les Transformers visuels (ViT), ou de techniques d'attention, pourrait encore améliorer les performances en capturant plus finement les relations spatiales et temporelles.

Par ailleurs, l'application pourrait être déployée sur des plateformes mobiles ou embarquées pour permettre une utilisation en temps réel dans les salles de classe ou dans des environnements d'apprentissage à distance.

Enfin, une collaboration avec des experts en pédagogie pourrait aider à valider l'outil sur le terrain et à l'adapter aux besoins réels des enseignants et des apprenants. Cela ouvrirait la voie à une éducation plus personnalisée et plus interactive, où le niveau d'engagement pourrait être suivi et renforcé de manière continue.

Bibliographie

Bibliographie:

[1] F. M. Newman, G. G. Wehlage and S. D. Lamborn, " The significance and sources of student engagement," 1992.

[2] L. Pirot and J.-M. De Ketele, "L'engagement académique de l'étudiant comme facteur de réussite à l'université Étude exploratoire menée dans deux facultés contrastées," *Revue des sciences de l'éducation,* vol. 26, p. 367–394, 2000.

[3] M. Furlong and J. Rebelez-Ernst, "School/Student Engagement," in *Metadata of the chapter that will be visualized online*, Springer Science+Business Media Dordrecht, 2013.

[4] V. Tomislava , "Emotional, Behavioural and Cognitive Engagement,Self-Efficacy and Satisfaction of Primary School Students," vol. 3, pp. 1-12, 2024.

[5] M. Çali, L. Lazimi and B. M. L. Ippoliti, "Relationship between student engagement and academic performance," *International Journal of Evaluation and Research in Education (IJERE),* vol. 13, 2024.

[6] B. A. Bertheussen, "Relation between academic performance and students' engagement in digital learning activities," *Journal of Education For Business,* vol. 3, 2016.

[7] R. I. Hammoud, Passive Eye Monitoring: Algorithms, Applications and Experiments, Springer Science & Business Media, 2008.

[8] A. Poole and L. J. Ball, Eye Tracking in HCI and Usability Research, Encyclopedia of Human Computer Interaction, 2006, pp. 211-219.

[9] L. J. Ball, Eye tracking in human-computer interaction and usability research: Current status and future prospects, 2006.

[10] R. J. K. Jacob and S. K. Keith , Eye tracking in human-computer interaction and usability research: Ready to deliver the promises, In H. Deubel & J. R. In Hyönä, The Mind's Eye: Cognitive and Applied Aspects of Eye Movement, 2003.

[11] M. A. Just and P. A. Carpenter, Eye fixations and cognitive processes, Cognitive Psychology, 1976, pp. 441-480.

[12] J. H. Goldberg and X. P. Kotval , "Computer interface evaluation using eye movements: methods and constructs," *International Journal of Industrial Ergonomics,* pp. 631-645, 1999.

[13] K. Rayner and A. Pollatsek, The Psychology of Reading, 1994.

[14] D. Bruneau, M. A. Sasse and J. McCarthy, "The Eyes Never Lie: The Use of Eye Tracking Data in HCI Research," vol. 3 , pp. 542-546, 2002.

[15] J. B. Brookings, . G. F. Wilson and . C. R. Swain, "Psychophysiological responses to changes in workload during simulated air traffic control," *Biological Psychology,* pp. 361-377, 1996.

[16] S. Marshall, "Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity," July 2000.

[17] M. Pomplun and S. Sunkara, "Pupil Dilation as an Indicator of Cognitive Workload in Human-Computer Interaction," vol. 3, pp. 542-546, 2003.

[18] J. H. Goldberg and A. M. Wichansky, "Eye tracking in usability evaluation: A practitioner's guide," in *The Mind's Eye Cognitive and Applied Aspects of Eye Movement Research*, Amsterdam, Elsevier, 2003, pp. 493-516.

[19] P.-L. Liu, "Using eye tracking to understand learners' reading process through the concept-mapping learning strategy," *Computers & Education,* vol. 78, pp. 237-249, 2014.

[20] H. M. SEKKIL and M. MEBROUKI , "Etude comparative entre les différentes architectures des réseaux de neurones convolutifs (CNNs) pour la détection de la fatigue du conducteur," 2021.

[21] W. A. Taylor, What every engineer should know about artificial intelligence, MIT Press, 1988.

[22] J. Sen, Machine Learning., intechopen, 2021.

[23] L. Philippe , F. Gérard , G. Matthieu and S. Chafik , Découverte du Machine Learning - Les outils de l'apprentissage automatique, 2021.

[24] S. Shashidhar, R. Dr. Manjunath and . B. N. Chandrashekar, Essentials of Deep Learning and AI: Experience Unsupervised Learning, Autoencoders, Feature Engineering, and Time Series Analysis with TensorFlow, Keras, and scikit-learn (English Edition), B. Publications, Ed., 25 nov. 2021.

[25] E. B. BELALEM and A. HAFIANE , "Biometric System based on Artificial Neural Networks using Conjugate Gradient Optimization method," 08 07 2019.

[26] M. Aljanabi, M. H. Qutqut and M. Hijjawi, "Machine Learning Classification Techniques for Heart Disease Prediction: A Review," *International Journal of Engineering and Technology,* october 2018.

[27] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, . M. Al-Amidie and . L. Farhan, "Review of deep learning: concepts, CNN architectures, challenges,

applications, future directions," *Journal of Big Data,* 2021.

[28] S. LABS, "Understanding Deep Learning: DNN, RNN, LSTM, CNN and R-CNN," 21 mars 2019. [Online]. Available: https://medium.com/@sprhlabs/understanding-deep-learning-dnn-rnn-lstm-cnn-and-r-cnn-6602ed94dbff.

[29] M. Z. Mokri , "Classification des images avec les réseaux de neurones convolutionnels," p. 21, 03 07 2017.

[30] P. Gayathri, A. Dhavileswarapu, S. Ibrahim and R. Paul, "Exploring the Potential of VGG-16 Architecture for Accurate Brain Tumor Detection Using Deep Learning," *JCMM,* vol. 2, 2023.

[31] K. BOUABDELLI and H. LARBI, "Recherche et classification des images en utilisant les modèles pré-trainés des CNNs," 2020.

[32] L. Zhao, Z. Wang, Z. Guoxin , Q. Yazhou and W. Xiaojin , "Eye state recognition based on deep integrated neural network and transfer learning," *Journal of Ambient Intelligence and Humanized Computing,* vol. 8, p. 635–644, 2017.

[33] Z. Chen and B. E. Shi, "ppearance-based gaze estimation using dilated-convolutions," in *Asian Conference on Computer Vision*, 2018.

[34] . C. Palmero, . J. Selva, M. A. Bagheri and S. Escalera, "Recurrent CNN for 3D Gaze Estimation using Appearance and Shape Cues," in *Conference: British Machine Vision Conference*, 2018.

[35] M. Ewaisha, M. Shawarby, H. Abbas and I. Sobh, "End-to-End Multitask Learning for Driver Gaze and Head Pose Estimation," 2020.

[36] Y. Omori and Y. Shima, "Image Augmentation for Eye Contact Detection Based on Combination of Pre-trained Alex-Net CNN and SVM," *J. Comput,* vol. 15, pp. 85-97, 2020.

[37] "Deep Learning on Eye Gaze Data to Classify Student Distraction Level in an Educational VR Environment," 2021.

[38] P. Blignaut and A. Akinyelu, "Convolutional Neural Network-Based Technique for Gaze Estimation on Mobile Devices," *Front. Artif. Intell,* vol. 4, 2022.

[39] . L. Huang, Y. Li , X. Wang, . H. Wang, . A. Bouridane and . A. Chaddad , "Gaze Estimation Approach Using Deep Differential Residual Network," *Sensors,* vol. 14, 2022.

[40] M.-T. Vo and . S. G. Kong, "Enhanced Gaze Tracking Using Convolutional Long Short-Term Memory Networks," *International Journal of Fuzzy Logic and Intelligent Systems 2022,* vol. 2, pp. 117-127, 2022.

[41] S. Harshit and D. Archit , "Data Cleaning in Eye Image Dataset," Waknaghat, Himachal Pradesh, India, 2023.

[42] R. Karmi, I. Rahmany and N. KHLIFA , "Gaze estimation using Convolutional Neural Networks," 24 February 2023.

[43] R. Zhao, Y. Wang, S. Luo, S. Shou and P. Tang , "Gaze-Swin: Enhancing Gaze Estimation with a Hybrid CNN-Transformer Network and Dropkey Mechanism," *Electronics,* vol. 13(2), p. 328, 2024.

[44] M. Abadi, , P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu and X. Zheng, "TensorFlow: A System for Large-Scale Machine Learning," *OSDI '16,* pp. 265-283), 2016.

[45] F. Chollet, "Keras: Deep learning for humans," 2015.

[46] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research,* vol. 12, p. 2825–2830, 2011.

[47] C. Nguyen, "La GUI tkinter," Universit´e de Toulo.

[48] W. Mckinney, "pandas: a Foundational Python Library for Data Analysis and Statistics," 2011.

[49] "OpenCV with Python," *Tutorials Point,* 2021.

[50] "Python Pillow," *Tutorials Point,* 2020.

[51] S. Kayvan , "Eye-dataset," Kaggle, 2020.

[52] J. Fredricks, W. McColskey, J. Meli and J. Mordica, "Mesurer l'engagement de l'élève de la fin du primaire jusqu'à la fin du secondaire :une description de 21 instruments," 2019.

[53] . R. Voigt, "Gaze Detection with CNNs for Linguistic Research," s.d.

[54] S. Rabba, Y. He, M. Kyan and L. Guan, "Pupillocalization for gaze estimation usingunsupervised graph-based model.," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2017.

[55] . I. GASMI, "Engagement comportemental dans un environnement d'apprentissage par projet," 2023.

[56] l. Bodin, "L'engagement scolaire : manifestations, caractéristiques des élèves engagés et influences des cours à distance.," 2020.

[57] R. Mishra, . M. S. Kiran, M. Maheswaram, . A. Upadhyay and B. Singh, "Investigation of optimal feature for milling chatter identification using supervised machine learning techniques," *Journal of Engineering Research,* vol. 12, pp. 950-962, 2024.

[58] Dvd8719, "Classification tree on iris dataset.png," 22 june 2020. [Online]. Available: https://commons.wikimedia.org/wiki/File:Classification_tree_on_iris_dataset.png.

[59] "Glossaire de l'intelligence artificielle (IA) : Comprendre les 10 termes clés," 1 october 2024. [Online]. Available: https://www.dendreo.com/blog/glossaire-de-lintelligence-artificielle.

[60] A. Pushkar, "What is Artificial Intelligence (AI)?," 2 January 2018. [Online]. Available: https://intellipaat.com/blog/what-is-artificial-intelligence/.

[61] X. Zhang, Y. Sugano, M. Fritz and . A. Bulling, "It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation," 2017.

[62] T. Poitschke, F. Laquai, S. Stamboliev and G. Rigoll, "Gaze-based interaction on multiple displays in an automotive environment," in *Conference2011 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2011*, CityAnchorage, AK, 2011.

[63] digiworld, "Machine Learning Vs Ai," Medium, 2022.

[64] D. Koblah, R. Yu Acharya, O. P. Dizon-Paradis and S. Tajik, "A Survey and Perspective on Artificial Intelligence for Security-Aware Electronic Design Automation," 2022.

[65] . P. Caballero-María, J. Caballero-Villarraso, J. Arenas-Montes , Alberto Díaz-Cáceres , Sofía Castañeda-Nieto, Juan F. Alcalá-Díaz, Javier Delgado-Lista , Fernando Rodríguez-Cantalejo , Pablo Pérez-Martínez , José López-Miranda and Antonio Camargo , "Deep Learning Model Approach to Predict Diabetes Type 2 Based on Clinical, Biochemical, and Gut Microbiota Profiles," *applied sciences,* 2025.

[66] Aleksandr, "Decision Trees and Random Forest," *Medium,* 2021.

[67] Dandi Yang, Cristhian Martinez, Lara Visuña, Hardev Khandhar, Chintan Bhatt and Jesus Carretero, "Detection and analysis of COVID-19 in medical images using deep learning techniques," *open access,* 2021.

[68] "Deep Learning vs Machine Learning : quelle est la différence ?," *IONOS.*

[69] digiworld, "Machine Learning Vs Ai," Medium, 2022.