UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

MIAGE RÉSEAU DES MIAGE DE FRANCE
Sorbonne

Machine Learning : Analyse des expressions faciales pour la détection de mensonge

Mémoire présenté par :

Monica Sen

Pour l'obtention du Master 1 MIAGE

De l'université Paris 1 Panthéon - Sorbonne

Année universitaire : 2023-2024

Date de soutenance : 03/04/2024

Année universitaire : Rébecca DENECKERE

Membre du jury : Camile SALINESI

REMERCIEMENTS

Je tiens à remercier tout d'abord Madame Rebecca DENECKERE, ma directrice de mémoire, pour son aide et ses indications qui m'ont permis de réaliser ce mémoire état de l'art.

Et, je voudrais par-dessus tout remercier Imane MIHOUBI, ma camarade et amie de l'université Panthéon Sorbonne, pour son soutien moral et ses encouragements tout au long de la rédaction de ce mémoire.

Table des matières

Résumé

Existe-t-il des comportements observables ou des indices capables de différencier un menteur d'une personne qui dit la vérité ? La question intrigue depuis des siècles, reflétant la nature inséparable du mensonge de nos interactions sociales. Face à l'incapacité humaine à discerner le vrai du faux, et aux limites du polygraphe, l'avènement des technologies de reconnaissance faciale se profile comme une alternative prometteuse. Bien que l'utilisation des expressions faciales, y compris les micro et macro-expressions, dans la détection de mensonge suscite le débat, certains chercheurs soutiennent que ces indices non verbaux pourraient être révélateurs.

Cet état de l'art explore l'analyse des expressions faciales à travers le machine learning pour la détection du mensonge, interrogeant la qualité des données requises pour développer des modèles efficaces.

Mots clés : Détection de mensonge, Machine Learning, CNN, Micro-expression, FACS

Glossaire

Mot	Définition
IA	Intelligence Artificielle
ML	Machine Learning
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short Term Memory
FACS	Facial Action Coding System
AU(s)	Action Unit(s)
Frame	Image ou cadre dans une vidéo
Landmark	Point de repère (facial)

1. Introduction

1.1 Motivation & Contexte

1.1.1 Evolution des techniques de détection de mensonge

Presque tous les chercheurs dans le domaine de la détection de la tromperie s'accordent à dire qu'il n'existe pas de "nez de Pinocchio" qui puisse servir d'indicateur permettant de repérer facilement un mensonge [27]. Le mensonge est un phénomène omniprésent dans nos sociétés, se produisant quotidiennement à de multiples reprises [28]. Malgré cette fréquence, la faculté des êtres humains à le détecter ne semble pas aller au-delà du pur hasard. En moyenne, un individu identifie correctement un mensonge dans seulement 47 % des cas [29]. Étonnamment, les adultes ne montrent pas une meilleure capacité à discerner le mensonge chez les enfants, avec un taux de réussite de 47,5 % [30]. Bien que les chiffres puissent varier d'une étude à l'autre, ils se situent généralement autour des 50 %, ce qui équivaut à la probabilité aléatoire d'un « pile ou face ».

Une détection du mensonge efficace peut prévenir et éviter des dangers et préjudices potentiels. Parmi les méthodes traditionnellement employées, le polygraphe est le plus répandu. Toutefois, sa nature intrusive qui nécessite une connexion physique au corps de l'individu durant l'interrogatoire [16] et exige le consentement d'un suspect, souvent peu disposé à passer le test [17], soulève des préoccupations. Les individus, conscients qu'ils sont surveillés, peuvent élaborer des stratégies pour déjouer l'appareil. Avec un entraînement adéquat, les suspects peuvent feindre l'innocence en utilisant des techniques spécifiques telles que mentir lors des questions préliminaires, contracter leurs muscles ou se mordre la langue [24]. Les capteurs peuvent également affecter la stabilité psychologique du suspect et rendent la détection de mensonges plus complexe [15].

Le polygraphe s'est régulièrement révélé faillible [23], incriminant des innocents tout en disculpant des coupables. De plus, le risque de partialité inhérent à l'intervention humaine nécessaire pour réaliser les tests [29] ainsi que la dépendance aux réponses physiologiques telles que la pression sanguine, le rythme cardiaque, la conductivité de la peau, les tremblements musculaires et la respiration pendant les interrogatoires [18] rendent son application à grande échelle infaisable.

1.1.2 Le rôle du Machine Learning dans l'analyse comportementale

L'avancée des technologies d'intelligence artificielle marque un tournant dans notre capacité à comprendre et analyser le comportement humain. Bien que l'IA vise à imiter le cerveau humain, comprendre le comportement humain s'avère difficile, étant donné la nature intrinsèquement complexe de la science du comportement.

Le comportement humain émerge d'une interaction entre trois éléments fondamentaux : les actions, la cognition et les émotions. Les actions, qui englobent tout ce que nous pouvons observer et mesurer directement, se juxtaposent aux cognitions — nos pensées, images mentales, compétences, connaissances et expériences. Parmi ces facettes du comportement, le mensonge est identifié comme un comportements cognitif [24]. La cognition, telle qu'explorée en philosophie et en psychologie politique, fait référence aux

processus conscients et intentionnels qui sous-tendent la pensée et la connaissance tels que la perception, l'attention, la mémoire, le langage, l'apprentissage, le raisonnement, le jugement et la pensée d'ordre supérieur qui peuvent être délibérément contrôlés [25]. Une émotion est un état mental temporaire caractérisé par une activité cognitive intense et un sentiment qui n'est pas considéré comme résultant de la connaissance ou du raisonnement [26]. Le fondement de la psychologie humaine se façonne à travers les réactions aux comportements guidés par les émotions. Des émotions telles que la joie, la peur, la tristesse, la colère, la surprise, l'excitation, la culpabilité, le regret, la haine et la curiosité, bien qu'elles soient distinctes des comportements, jouent un rôle dans la motivation de certaines actions [24].

Dans le domaine de la recherche en santé mentale, l'IA offre une approche qui abordent les problématiques sous un angle statistique pour optimiser la prédiction. Au départ, des canaux unimodaux tels que l'audio, les images, le texte, et des canaux physiologiques comme l'ECG, l'EEG, le GSR, le BVP, etc., été utilisés pour prédire et classer le comportement cognitif avec des modèles d'IA/ML. Plus tard des modèles multimodaux ont été démontré de fortes capacités de compréhension multimodale et de généralisation dans plusieurs tâches cognitives en aval. Les encodeurs visuels et linguistiques entraînés de manière multimodale, qui se rapprochent plus du cerveau que ceux entrainés de manière unimodale du point de vue de l'encodage neuronal, constituent des ressources pour les neuroscientifiques [32].

Grâce à l'accessibilité à d'énormes volumes de données et les progrès des capacités de calcul, le machine learning se révèle particulièrement adapté pour aborder des défis complexes et hétérogènes, comme anticiper les individus à risque de troubles mentaux, prédire les émotions, repérer le stress, identifier les intentions, ou encore détecter le mensonge [24].

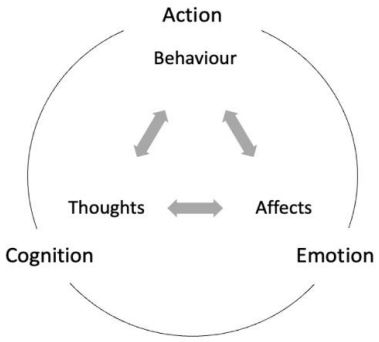

Figure 1: Relation entre Cognition, Emotion et Action
Source: [24]

1.1.3 Enjeux éthiques et sociétaux de l'automatisation de la détection de mensonge

Un comportement trompeur perçu à tort comme bénin peut entraîner des pertes financières, des sentiments de stupidité, de méfiance, et la rupture de relations personnelles [31]. Malgré le caractère inoffensif de nombreux mensonges, ses conséquences revêtent une importance majeure dans divers contextes, incluant les enquêtes criminelles, la sécurité aéroportuaire, ou les procédures judiciaires [24]. Le mensonge fait l'objet de nombreux champs de recherche en psychologie, ce qui a mené à une multitude d'études sur la détection du comportement trompeur dans le domaine des technologies appliquées, impliquant les domaines linguistiques, comportementaux et physiologiques [12].

"... des indices peu clairs et fiables, nous questionnons la validité de l'utilisation de l'intelligence artificielle incluant des indices de mensonge, qui n'ont actuellement aucun soutien empirique" [11]

L'utilisation des systèmes d'IA dans la détection de mensonge et notamment son automatisation dans le monde réel pourrait avoir des impacts sociétaux négatifs et soulever des préoccupations éthiques.

1.2 Définition de la problématique

Compte tenu des limites associées à l'utilisation du polygraphe et à la nécessité d'identifier le mensonge dans une variété de contextes sans recourir à l'intervention humaine, les chercheurs se tournent vers des solutions basées sur le machine learning. Les êtres humains ayant pour habitude d'observer et de comprendre les indices en les regardant, l'approche la plus naturelle pour analyser ces signes serait d'enregistrer des vidéos ou de prendre des photos. Dans ce cadre, la problématique qui guide ce mémoire est :

Est-ce que le machine learning peut aider à détecter le mensonge sur le visage des interlocuteurs ?

Cette interrogation ouvre sur plusieurs axes de recherche :

- L'efficacité des approches de machine learning dépend intrinsèquement de la qualité et de la disponibilité des données sur lesquelles les modèles sont formés. Face à la rareté des données de haute qualité, quelles techniques sont utilisées pour améliorer leur apprentissage ?
- Les micro-expressions semblent être des indices intéressants pour détecter le mensonge, peuvent-elles à elles seules fournir une indication fiable d'un comportement trompeur ?
- Parmi les modèles et algorithmes disponibles, lesquels sont privilégiés pour la reconnaissance des expressions faciales ?
- Enfin, considérant l'existence de modèles de détection de mensonge, quelle est la probabilité que ces systèmes s'avèrent fonctionnels dans des contextes réels, compte tenu des variations individuelles et culturelles dans les expressions du mensonge ?

À noter que dans ce mémoire, le terme de « mensonge » ne se limite pas strictement au fait d'exprimer des choses fausses. À l'instar du terme anglophone « deception », il inclut les comportements trompeurs tels que l'omission, la feinte et tout autre action visant à induire en erreur.

2. Méthodologie de recherche

Ayant un attrait pour les sciences comportementales, mon sujet de mémoire s'est porté sur l'analyse des indices révélateurs de mensonge. L'intégration du machine learning découle des avancées dans le domaine de la reconnaissance faciale, et permet de répondre aux exigences de notre cursus en Master 1 MIAGE, en associant une thématique comportementale avec une dimension informatique.

Dans la recherche sur la détection du mensonge, on distingue les indices verbaux et non-verbaux. Face au risque de dispersion dû à l'ampleur des recherches et à la diversité des méthodes analyser chaque indice, il était nécessaire d'affiner mon champ d'étude. J'ai décidé de me concentrer sur les indices non-verbaux, en particulier les expressions faciales.

Les mots-clés repris dans le tableau ci-dessous me permettent de définir mon périmètre, d'identifier les concepts clé ainsi que leurs synonymes, et de faciliter la recherche des articles pertinents. Mes recherches ont essentiellement été effectuées à l'aide de l'outil MIAGE Scholar, un moteur de recherche Scopus récupérant majoritairement des sources fiables, conçu pour appuyer nos travaux de recherche.

Mots-clés	Machine learning	Detect	Lie	Facial expression
Synonymes	Deep learning	Detect	Lying	Facial analysis
	-	Prediction	Liar	Face recognition
	-	Predict	Deception	Micro expression

***Tableau 1** : Définition des mots-clés de recherche et de leurs synonymes*

Ces termes me permettent d'établir la chaîne de recherche, où chaque mot-clé est regroupé avec ses synonymes par un « ou exclusif » et chaque groupe de mots liés par un opérateur « et ».

```
(TITLE-ABS-KEY("detection") OR TITLE-ABS-KEY("detect")
OR TITLE-ABS-KEY("detector") OR TITLE-ABS-KEY("detecting")) AND
(TITLE-ABS-KEY("lie") OR TITLE-ABS-KEY("lying")
OR TITLE-ABS-KEY("liar") OR TITLE-ABS-KEY("deception")) AND
(TITLE-ABS-KEY("micro expression") OR TITLE-ABS-KEY("facial expression")
OR TITLE-ABS-KEY("face recognition") OR TITLE-ABS-KEY("facial analysis")) AND
(TITLE-ABS-KEY("machine learning") OR TITLE-ABS-KEY("deep learning"))
```

La phase initiale de sélection a permis de recueillir 108 résultats grâce à la chaîne de recherche. Un premier examen rapide du titre et de l'abstract de chaque article m'a permis de conserver 35 articles. L'utilisation de l'outil Parsifal a marqué la seconde phase de filtrage. Une analyse approfondie des introductions et la prise de notes préliminaires m'ont permis de conserver 26 articles. L'étape suivante consistait en une lecture détaillée des articles restants, en prenant en compte leur accessibilité et de leur date de publication pour privilégier les travaux les plus récents. Cette analyse m'a permis d'évaluer et de réévaluer la pertinence de chaque article au regard de la problématique de mon mémoire.

Un article est jugé pertinent s'il aborde un ou plusieurs des sujets suivants :

- Reconnaissance des expressions faciales,
- Reconnaissance des émotions,
- Multimodalité incluant les expressions faciales, spécifiquement dans le contexte de la détection de mensonge au moyen de techniques de machine learning.

Au terme de ce processus, 9 articles ont été retenus pour constituer le corpus de ce mémoire:

[1] Nam, Borum, Joo Young Kim, Beomjun Bark, Yeongmyeong Kim, Jiyoon Kim, Soon Won So, Hyung Youn Choi, and In Young Kim. "FacialCueNet: unmasking deception-an interpretable model for criminal interrogation using facial expressions." Applied Intelligence 53, no. 22 (2023): 27413-27427.

[2] Yildirim, Suleyman, Meshack Sandra Chimeumanu, and Zeeshan A. Rana. "The influence of micro-expressions on deception detection." Multimedia Tools and Applications 82, no. 19 (2023): 29115-29133.

[3] Monaro, Merylin, Stéphanie Maldera, Cristina Scarpazza, Giuseppe Sartori, and Nicolò Navarin. "Detecting deception through facial expressions in a dataset of videotaped interviews: A comparison between human judges and machine learning models." Computers in Human Behavior 127 (2022): 107063.

[4] Zhou, Xingchen, Rob Jenkins, and Lei Zhu. "An Honest Joker reveals stereotypical beliefs about the face of deception." Scientific reports 13, no. 1 (2023): 16649.

[5] Islam, Siam, Popin Saha, Touhidul Chowdhury, Asif Sorowar, and Raqeebir Rab. "Non-invasive deception detection in videos using machine learning techniques." In 2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), pp. 1-6. IEEE, 2021.

[6] Shen, Xunbing, Gaojie Fan, Caoyuan Niu, and Zhencai Chen. "Catching a liar through facial expression of fear." Frontiers in Psychology 12 (2021): 675097.

[7] Bruer, Kaila C., Sarah Zanette, Xiao Pan Ding, Thomas D. Lyon, and Kang Lee. "Identifying liars through automatic decoding of children's facial expressions." Child development 91, no. 4 (2020): e995-e1011.

[8] Alaskar, Haya. "Hybrid Metaheuristics with Deep Learning Enabled Automated Deception Detection and Classification of Facial Expressions." Computers, Materials & Continua 75, no. 3 (2023).

[9] Belavadi, Vibha, Yan Zhou, Jonathan Z. Bakdash, Murat Kantarcioglu, Daniel C. Krawczyk, Linda Nguyen, Jelena Rakic, and Bhavani Thuriasingham. "MultiModal deception detection: Accuracy, applicability and generalizability." In 2020 Second IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), pp. 99-106. IEEE, 2020.

Article	Date de parution	Type
[1]	2023	Journal scientifique
[2]	2023	Journal scientifique
[3]	2022	Journal scientifique
[4]	2023	Journal scientifique
[5]	2021	Conférence
[6]	2021	Journal scientifique
[7]	2020	Journal scientifique
[8]	2023	Journal scientifique
[9]	2020	Conférence

***Tableau 2** : Récapitulatif des articles avec leur date de parution et leur type*

Le développement de l'état de l'art suit un ordre logique en 3 parties :

1. La protocoles expérimentaux pour créer des situations où le mensonge est induit en laboratoire.
2. Les indices de mensonge étudiés pour la prédiction.
3. Les techniques de ML utilisées pour la classification.

3. Background

3.1 Concepts clés

Pour la suite de ce mémoire, il est nécessaire de poser les notions ainsi que les outils, algorithmes et modèles souvent utilisés dans la tâche de la reconnaissance faciale.

Expressions faciales

Les expressions faciales sont un aspect du comportement humain reconnu comme l'aspect le plus saillant et influent de la communication humaine. Le terme "expression facial" est utilisé par les chercheurs pour définir certains mouvements récurrents des muscles du visage qui transmettent des pensées, des émotions ou des comportements. [33] [source 2]

Macro et micro-expressions

Les expressions faciales peuvent être catégorisées en deux types : les macro-expressions et les micro-expressions. C'est la durée de l'expression faciale, et non son intensité, qui différencie les macro-expressions des micro-expressions. Les macro-expressions durent entre 0.75 et 2 secondes et sont facilement perçues par les humains, tandis que les micro-expressions, plus brèves, durent moins de 0.5 seconde. [34] [source 2]

Caractéristiques non-verbales

Les caractéristiques non-verbales incluent les expressions faciales, mais peuvent comprendre d'autres signes tels que le mouvement des yeux ou de la bouche, le temps de réponse, la fréquence de déglutition ou la symétrie du visage [source 1, 2, 11].

Action Unit (AU)

Les Unités d'Action (AUs) sont les actions fondamentales de muscles individuels ou des groupes de muscles, qui sont activées lors d'expressions faciales. Chaque micro-expression est générée par un motif spécifique d'activation des AUs. [35] [source 3]

Action Unit(AU)	*Description*	*Action Unit(AU)*	*Description*
1	Inner Brow Raiser	17	Chin Raiser
2	Outer Brow Raiser	18	Lip Puckerer
4	Brow Lowerer	20	Lip stretcher
5	Upper Lid Raiser	22	Lip Funneler
6	Cheek Raiser	23	Lip Tightener
7	Lid Tightener	24	Lip Pressor
9	Nose Wrinkler	25	Lips par
10	Upper Lip Raiser	26	Jaw Drop
11	Nasolabial Deepener	27	Mouth Stretch
12	Lip Corner Puller	28	Lip Suck
13	Cheek Puffer	41	Lid droop
14	Dimpler	42	Slit
15	Lip Corner Depressor	44	Squint
16	Lower Lip Depressor	45	Blink

Figure 1 : Description des unités d'action
Source : [5]

Facial Action Coding System (FACS)

Le FACS, développé par Paul Ekman et Wallace V. Friesen, est un système de codage visant à décrire toutes les expressions faciales possibles du visage en utilisant les unités d'action. Les mouvements faciaux étaient à l'origine classifiés à la main résultant en la définition de 46 AUs pour identifier jusqu'à 7 000 combinaisons d'AUs différentes [35]. Cette nomenclature a largement été utilisée dans les recherches sur la détection de mensonges pour identifier les différences dans l'activation des AUs entre les déclarations véridiques et mensongères [source 3].

3.2 Algorithmes et modèles de reconnaissance faciale

3.2.1 FACET

FACET est un logiciel FACS automatisé qui permet de coder les expressions faciales en temps réel. Il analyse image par image l'activation des AUs et les traduit en émotions. FACET fournit des scores normalisés indiquant la probabilité qu'un codeur humain qualifié identifie un ensemble de mouvements musculaires comme une émotion. FACET fournit des scores pour 10 expressions émotionnelles : joie, tristesse, colère, dégoût, surprise, peur, confusion, frustration, mépris, et neutre. [36] [source 7]

3.2.2 OpenFace

OpenFace est un outil avancé d'analyse comportementale conçu pour extraire et analyser des caractéristiques faciales à partir de vidéos. Il fournit des données détaillées sur l'emplacement des landmarks faciaux, la confiance de l'algorithme de reconnaissance faciale, la direction du regard, la posture de la tête, ainsi que la présence et l'intensité des AUs, facilitant l'étude automatisée des expressions faciales. [37]

3.2.3 Random Forest

Le Random Forest est un algorithme de classification composé d'arbres de prédiction individuel qui forment une forêt décisionnelle. Chaque arbre dans le modèle est un arbre de classification qui prédit une sortie. La forêt classe les individus en examinant le vote majoritaire des prédictions des arbres individuels. Plusieurs couches d'aléa sont introduites dans le modèle :

- La première couche est induite lors du « bagging », une approche d'agrégation qui consiste à sélectionner un ensemble d'entraînement à partir d'une sous-section aléatoire des données. Ces informations sont utilisées pour développer chaque arbre.
- La seconde couche est introduite par le fait que le Random Forest sélectionne aléatoirement un sous-ensemble de caractéristiques pour développer chaque arbre. Contrairement à d'autres modèles qui utilisent toutes les caractéristiques pour construire un arbre.

3.2.4 Convultional Neural Network (CNN)

Les réseaux de neurones convolutionnels appartiennent au deep learning, un sous-domaine du machine learning inspiré par la structure des réseaux de neurones. Ils ont particulièrement contribué à la vision par ordinateur plus que tout autre algorithme et sont utilisés dans diverses tâches tels que la segmentation d'images, la reconnaissance de vidéos et d'images et la classification d'images. La structure d'un CNN se composent de plusieurs couches :

- *Couche de convolution* : constitue la première couche du réseau qui traite l'image d'entrée. Elle utilise un filtre de taille MxM pixels pour parcourir l'image et créer une « feature map » qui nous donne des informations générales sur l'image, telles que les bords et les coins.
- *Couche de pooling* : réduit la taille de la feature map obtenues à partir de la couche de convolution précédente.
- *Couche « Fully-connected »* : forme les dernières couches et est placée avant les couches denses. Chaque neurone de cette couche est connecté à tous les neurones de la couche précédente.
- *Couche dense* : est placée à la fin de l'architecture et sert à classifier ou à prendre des décisions basées sur les caractéristiques extraites par les couches précédentes.
- *Poids* : détermine l'importance de chaque entrée pour la production de la valeur de sortie.
- *« Dropout »* : méthode de régularisation pour réduire le surajustement (« overfitting ») qui consistent à ignorer aléatoirement certaines unités du réseau de neurones pendant l'entraînement.
- *Softmax* : fonction qui transforme les valeurs d'entrée en probabilités, ce qui facilite l'interprétation des résultats. Elle est souvent utilisée dans la dernière couche pour la classification.

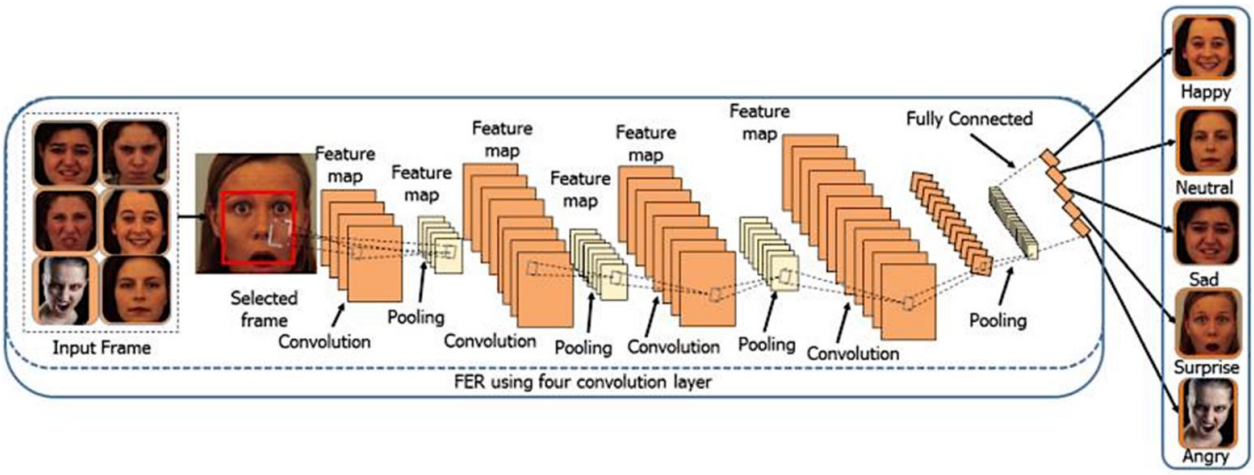

Figure 2 : Processus d'apprentissage du CNN pour la reconnaissance des émotions faciales (FER)
Source : [39]

3.2.5 3D CNN (C3D)

Le réseau de neurones convultionnel 3D est un CNN conçu pour traiter l'information spatio-temporelle en améliorant l'identification d'images en mouvement et d'images 3D. Ce réseau extrait automatiquement des caractéristiques des clips vidéo sans nécessité de réaliser une extraction de caractéristiques préalable. Il considère les deux dimensions spatiales de chaque image et la dimension temporelle. [3]

3.2.6 Long Short Term Memory (LSTM)

Les LSTM sont un type de réseau de neurones qui peut également classer et traitées des données séquentielles. Ils ont été conçus pour pallier aux problématiques de disparition ou d'explosion des gradients qui se manifeste lors de l'apprentissage de séquences longues et qui rendent le réseau incapable de retenir des informations des étapes précédentes. L'avantage du LSTM réside dans son architecture qui intègre des unités de mémoire permettant de se « souvenir » ou « d'oublier » des informations de manière sélective. [40]

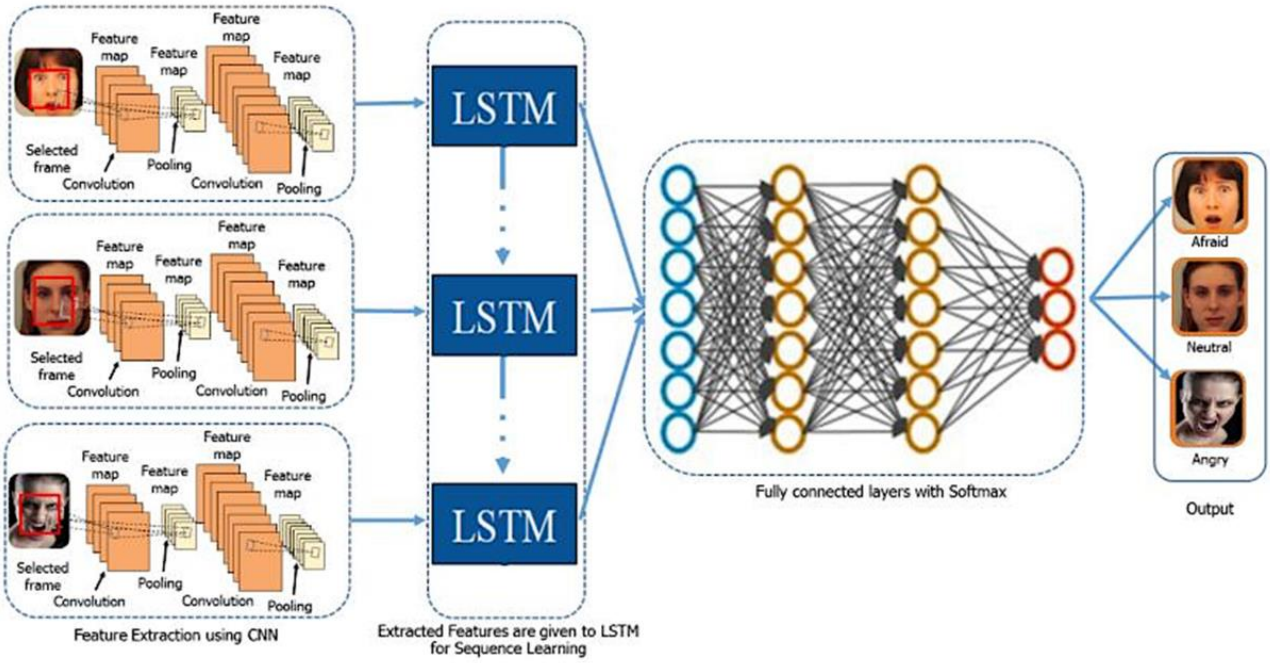

Figure 3 : Utilisation du CNN et LSTM dans la reconnaissance faciale des émotions
Source : [39]

4. Machine learning pour la détection de mensonge

L'ensemble des articles du corpus proposent d'étudier la détection du mensonge à travers l'analyse vidéo et l'utilisation d'un modèle de prédiction. Les étapes peuvent être résumées tel que :

1. **Préparation du modèle** : le modèle de prédiction de ML/DL est entraîné sur une base de données existante ou sur un jeu de données issue de l'expérience de l'étude.
2. **Expérience** : une expérience est menée en laboratoire pour reproduire des situations où le sujet peut mentir ou adopter des comportements trompeurs.
3. **Jeu de données** : les sujets sont filmés durant l'expérience, constituant un ensemble de vidéos sur lequel le modèle peut travailler.
4. **Extractions des indices** : différents indices sont extraits des vidéos via des algorithmes ou modèles de ML/DL pour la prédiction pour identifier le mensonge.
5. **Classification** : le modèle de prédiction est appliqué et classe les vidéos en fonction de la présence de mensonge.
6. **Évaluation des modèles** : la performance du modèle à prédire le mensonge est analysée à travers l'utilisation de différentes métriques statistiques.

4.1 Protocoles expérimentaux

Cette partie a pour but de synthétiser les protocoles expérimentaux conduits dans l'**étape 2**, afin de décrire les scénarios et mécanismes utilisés pour étudier le mensonge en laboratoire. Cela permet également de poser le contexte pour l'analyse des indices de mensonge.

Les indices proviennent soit :

- D'un ensemble de jeu de données publiques, auquel cas les chercheurs [1, 5] se concentrent sur la capacité du modèle à prédire le mensonge.
- D'un jeu de données que constitue eux-mêmes à travers la réalisation d'une expérience.

Les méthodes utilisées pour provoquer le mensonge varient d'une étude à l'autre. Pour une meilleure lecture, le nom des différents *Dataset* utilisés sera souligné. On peut les résumer en 3 catégories, reflétant le niveau d'enjeu associé au mensonge.

4.1.1 Situation à enjeu élevé

Les situations à enjeu élevé correspondent aux contextes où les conséquences d'être pris en mensonge sont importantes. Aucun des études du corpus n'a reproduit une situation où les enjeux sont élevés. Cependant [1, 9] ont utilisé le *Real life trial dataset*, contenant des vidéos de procès judiciaire où le verdict détermine si une déclaration est véridique ou mensongère.

L'étude [1], dans *DDCIT dataset*, utilise des techniques polygraphiques pour mesurer la réponse galvanique de la peau (GSR) du sujet dans le but de simuler un environnement proche des interrogatoires criminelles. Le fait que le sujet soit conscient que ses réactions physiologiques sont enregistrées peut engendrer un sentiment d'immersion qui peut accentuer la perception d'enjeu dans l'expérience. Malgré la présence d'une équipe de professionnelle dans le domaine de la détection du comportement trompeur, les situations

provoquées dans *DDCIT dataset* ne peuvent pas être d'enjeu comparable à de réel cas d'infractions et de crimes.

4.1.2 Situation à enjeu modéré

Les situations à enjeu modéré concernent les contextes où les conséquences du mensonge sont modérées, souvent liées à des interactions sociales ou professionnelles courantes. Les méthodes comprennent les faux crimes, les jeux de cartes et les interrogatoires.

Dans [9], il est demandé au sujet de voler ou non 50$ *(Crime dataset)* puis de convaincre l'expérimentateur de son innocence. Dans [7], on suggère à un enfant de mentir par omission sur des jouets qu'il avait prétendument cassés *(Toys dataset)* sous peine d'avoir des ennuis.

Dans [1, 4], une récompense financière est attribuée au sujet s'il réussissait à tromper l'expérimentateur, ce qui permettait de stimuler la motivation des participants à se montrer persuasifs et à continuer de participer à l'expérience. Un score de 6/7 en moyenne témoigne que les participants dans [4] avaient bien compris les consignes et désiraient gagner, ce qui permet d'enregistrer des réactions plus authentiques.

L'expérience menée dans [4] pour le *Joker dataset*, permet de reproduire une situation où le sujet peut choisir sa stratégie de jeu et par conséquent le type de mensonge qu'il va utiliser. Dans la recherche, le mensonge simple (« simple deception ») est souvent étudié, au détriment d'un type de mensonge plus subtile, le mensonge sophistiqué (« sophisticated deception ») courant dans des contextes de compétition tels que la rivalité politique, la guerre, le sport, le jeu (comme au poker), les affaires et la diplomatie, et qui vise à induire en erreur.

Are they presenting the truth?	Do they intend to deceive?	Do they expect to be believed?	Classification
Yes	No	Yes	Plain Truth
No	Yes	Yes	Simple Deception
Yes	Yes	No	Sophisticated Deception

Figure 4 : Classification des stratégies de mensonge
Source : [4]

4.1.3 Situation à enjeu faible

Les situations à enjeu faibles représentent les contextes où les répercussions du mensonge sont minimes. Les méthodes utilisées par [2, 3, 9] comprennent le jeu de rôle et l'invention d'histoire, le sujet est assigné à un rôle, soit en personne honnête (« truth-teller ») soit en menteur (« liar ») et répond à des questions ouvertes sur des sujets banals.

Dans [9], le sujet donne son avis sur un film choisi aléatoirement *(Opinion dataset)*. Dans [3], le participant est interrogé sur des vacances passées, fictives ou non *(Holiday dataset)*. L'utilisation de souvenirs de vacances comme sujet de mensonge a déjà été adoptée dans des recherches antérieures [41], étant donné que se rappeler des vacances mobilise les mêmes processus cognitifs que le fait de fournir un alibi durant une enquête criminelle, mais dans un contexte à faible risque.

Dans les expériences menées dans [3, 9], les participants étaient libres de s'exprimer librement pendant une durée impartie tandis que dans [2], le sujet ne devait répondre que par « Oui » ou « Non ».

4.1.4 Tableau récapitulatif

Les paramètres tels que l'enjeu perçu dans une situation donnée ainsi que la stratégie de mensonge adoptée par le sujet sont à prendre en compte pour contextualiser les résultats obtenus à l'issue des prédictions par le modèle de ML. Les expérimentations en laboratoire ne peuvent pas toujours refléter les contextes réels de mensonge. Et s'il existe des marqueurs faciaux qui diffèrent d'une stratégie de mensonge à une autre, les confondre pourraient conduire à interprétations erronées.

Le tableau ci-dessous fait la synthèse des données sur lesquels les modèles seront testés. Le champ « labellisation » correspond à la manière dont les vidéos sont étiquetées (vérité/mensonge), il permet de comprendre comment la vérité objective est connue. Pour *Bag of lies*, *Crime*, *Opinion* et *Holiday*, les participants révèlent à l'expérimentateur s'ils avaient choisi de mentir ou pas. Pour *Mood*, les instructions à être sincère, mentir à moitié ou mentir complètement sont données chaque session de questions. Pour *Joker* et *DDCIT*, la vérité est obtenue lorsque l'expérimentateur vérifie les cartes du participant, ce qui ne peut pas être remis en question. Pour *Real life trial*, les vidéos sont labellisés selon le verdict judiciaire : coupable sera labellisé mensonge, non-coupable et exonéré sera labellisé vérité. Le niveau d'enjeu n'est pas donné dans chaque article mais est déterminé arbitrairement selon les informations fournies.

Nom dataset	Real life trial	Bag of lies	Crime	Opinion	Joker	Holiday	DDCIT	Mood	Toys
Réf article	[1, 9]	[5]	[9]	[9]	[4]	[3]	[1]	[2]	[7]
Nombre participants	56	35	-	-	40	62	105	15	158
Détails participants	▪ 21 femmes ▪ 35 hommes 16 à 60 ans	▪ 10 femmes ▪ 25 hommes			▪ 20 femmes ▪ 20 hommes 18 à 23 ans	▪ 43 femmes ▪ 19 hommes 20 à 29 ans	▪ 51 femmes ▪ 54 hommes 20 à 30 ans	▪ 7 femmes ▪ 8 hommes 25 à 33 ans	▪ 7 filles ▪ 8 garçons 4 à 9 ans
Origine	-	-	-	-	Chinois	Italien	Coréen	Coréen	Latino, africain américain
Méthodologie	Cas réel de procès judiciaire	Raconter une histoire à partir de photos	Faux crime commis ou non (vol de 50$)	Discours libre (avis d'un film)	Jeu de cartes (Joker)	Discours libre (vacances)	Interrogatoire sur une information cachée (Oui/Non)	Question ouverte (Oui/Non)	Faux crime non commis (jouets cassés)
Niveau d'enjeu	Élevé	Faible	Modéré	Faible	Modéré	Faible	Modéré	Faible	Modéré
Labellisation	Verdict judiciaire	Les participants	Les participants	Les participants	Les cartes	Les participants	Les cartes	Les instructions données	
Taille des données	121 vidéos ▪ 61 mensonge ▪ 60 vérité	325 vidéos ▪ 162 mensonge ▪ 163 vérité	-	-	120 images 240 vidéos ▪ 32 mensonge ▪ 30 vérité	62 vidéos ▪ 32 mensonge ▪ 30 vérité	630 vidéos ▪ 210 mensonge ▪ 420 vérité	-	Proportion élevée de classe négative
Dataset existant	Oui	Oui	Non	Non	Non	Non	Non	Non	Non

***Tableau** : Synthèse des données utilisées pour la prédiction*

4.2 Indices de mensonge

4.2.1 Macro et micro-expressions

L'utilisation des expressions faciales et, en particulier, les micro-expressions reposent sur la théorie de la « fuite » involontaire d'une expression (« leakage theory »). Les micro-expressions seraient d'après Ekman, des indicateurs universels des émotions et pourraient révéler nos véritables intentions [35].

Les études du corpus [1, 3, 4, 5, 9], utilisent l'outil OpenFace pour extraire les expressions faciales, les résultats obtenus permettent de mesurer la présence de certaines AUs et/ou d'évaluer l'intensité d'une unité d'action.

Par exemple dans [1], les occurrences de micro-expressions sont comptabilisées en se basant sur la durée des AUs. En considérant qu'une micro-expression dure jusqu'à 0.5 seconde, les AUs qui apparaissent pour une durée inférieure à cette limite sont comptées. Les expressions qui n'apparaissent que sur une seule frame sont considérées comme des erreurs de détection et exclues.

Pour [5], les caractéristiques obtenues, ne sont pas utilisées tel quel. Les données sont pré-traitées en passant par un certain nombre d'étapes :

- « *Apex Frame Selection* » : sélection des images les plus expressives d'une séquence vidéo.
- « *Feature Selection* » : sélection des caractéristiques les plus pertinente pour la classification. Dans l'étude, les AUs les plus pertinents pour identifier le mensonge sont : AU14 (plissement externe des lèvres), AU23 (tension refermante des lèvres) et AU12 (étirement du coin des lèvres).

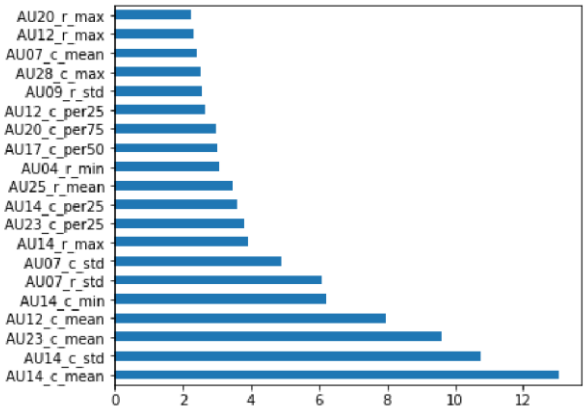

Figure 5 : AUs les plus pertinentes déterminées lors de la Feature Selection
Source : [5]

- « *Data Discretization* » : transformation des caractéristiques continues en valeurs discrètes, ce qui permet de réduire les effets de valeurs aberrantes et simplifier la structure des données.
- « *Feature Scaling* » : normalisation des valeurs des caractéristiques, cette étape s'assure que toutes les caractéristiques contribuent équitablement au processus d'apprentissage.

Pour [9], un traitement préalable des données est également nécessaire pour former le modèle CNN. Les chercheurs ont ajusté toutes les vidéos pour qu'elles aient la même taille d'image et le même nombre d'images par seconde (24 fps) et utilisent outil de PyTorch pour créer de plus petite vidéos de 12 images.

Dans [2], le modèle CNN de prédiction n'inclue pas l'utilisation d'OpenFace pour l'extraction des caractéristiques, en revanche il est entraîné à partir des micro-expressions qu'il identifie. Ce dataset *FER-2013* est un jeu de données qui contient environ 35 890 images, labellisé avec 7 émotions : la colère, la peur, le dégoût, la joie, la surprise, la tristesse, et la neutralité.

4.2.2 Émotions spécifiques

Si les micro-expressions ne peuvent être réprimés lorsqu'une personne ment, alors certaines émotions pourraient être spécifiques au mensonge. Les expressions faciales peuvent être traduites en émotions dans [2, 6, 7]. Dans [7], les expressions faciales des enfants sont traduites en émotions grâce à FACET. Deux moments clés de la vidéo sont analysés sous l'hypothèse que la charge cognitive peut varier et donc que les émotions exprimées sont différentes :

- La première phase correspond à la période où l'enfant entend la question posée. Il est conscient qu'on va lui demander de parler de l'évènement qui s'est produit, en l'occurrence sur une faute qu'il n'a pas commise. Il sera plus curieux d'apprendre sur quoi il va être interrogé, et va évaluer la situation plutôt que chercher à surveiller ses expressions. Durant cette phase, les expressions faciales peuvent être particulièrement informatives, car les capacités d'inhibitions des enfants pourraient ne pas être encore activées.
- La seconde phase correspond à la période qui suit immédiatement la première, une fois que la question est posée dans son intégralité. L'enfant peut se trouver dans un processus de décision quant au fait de savoir quoi dire à l'expérimentateur et comment formuler sa réponse. Cette phase est intéressante pour évaluer les expressions des enfants qui décident de mentir par omission.

Les émotions de surprise et de peur apparaissent dans [7] comme des indicateurs de mensonge puisque ce sont les expressions qui sont le plus présente en comparant les émotions qui apparaissent dans une situation où l'enfant ment et une où il ne ment pas.

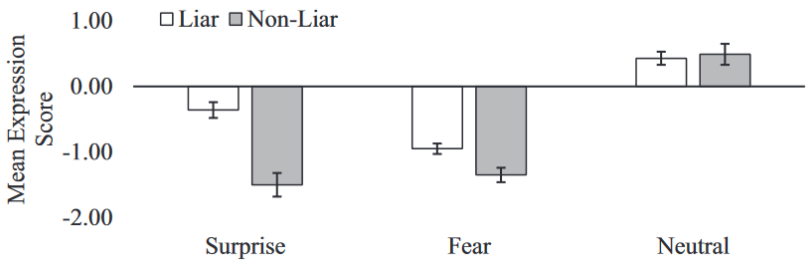

Figure 6 : Scores moyens d'expression des émotions
Source [7]

Des indices verbaux peuvent également être analysés pour le mensonge incluant la vitesse du discours, la fréquence des pauses, les incohérences [1]. Cependant l'analyse faciale est plus

pertinente chez les enfants car ils manifestent moins d'indices verbaux que les adultes, ce qui limiterait les analyses.

Dans [2, 7] ce n'est pas la présence d'une émotion en particulier qui permettait d'étiqueter une vidéo mais plutôt la variation soudaine d'une émotion à une autre et la différence des émotions survenues entre différentes phases. Une émotion de baseline est identifiée dans [2] afin de pouvoir comparer.

La micro-expression dominante identifiée est l'expression neutre, elle correspond aux moments où le sujet n'interagit pas avec l'expérimentateur et où il est sincère, cette expression traduit une confiance et un confort lorsqu'il répond aux questions, ses niveaux d'émotion sont stables et il ne présente pas de variation soudaine d'humeur. Dans une phase suivante, le sujet devait à la fois mentir et dire la vérité, ce qui peut générer de l'inconfort et se traduire en des variations d'émotions (Figure 12). L'expression neutre est comparée à l'expression dominante identifiée lors de cette phase, en l'occurrence la joie. Le modèle comprend donc que la joie et le changement d'humeur de neutre à joyeux sont des indicateurs de mensonge.

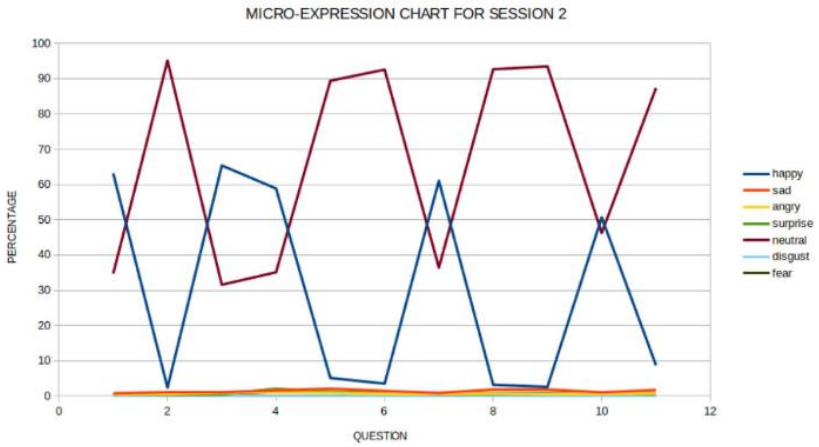

Figure 7 : Echantillon de micro-expression
Source [2]

L'expression de peur est particulièrement étudiée dans [6], sous l'hypothèse que, dans des situations à enjeux élevés, le sujet craindrait de se faire prendre. Même si la peur peut être retrouvée chez une personne honnête et peut donc être mal interprété, le degré de répression sera différent entre un menteur et un non-menteur.

4.1.3 Approche multimodale

Il existe plusieurs approches concernant la multimodalité, l'une consiste à étudier plusieurs indices séparément, l'autre qui consiste à considérer tous les indices pour les étudier comme un tout.

FacialCutNet

Le modèle FacialCueNet de [1] utilise un ensemble d'indicateurs :

- Fréquence des unités d'actions : certains AUs sont identifiées comme permettant de distinguer le mensonge tel que AU15 (Abaissement des coins externes des lèvres) AU17 (élévateur du menton), AU20 (Étirement externe des lèvres), AU25 (Séparation légère des lèvres), AU45 (Clignotement). Ces AUs sont donc étudiés à travers leur présence dans les vidéos et traduites en valeurs binaires.
- Symétrie faciale : les changements dans les distances euclidiennes des landmarks sont évalués. Cette approche utilise la symétrie gauche/droite du visage comme indice et calcule la corrélation entre les distances pour déduire la symétrie faciale.
- Direction du regard : la moyenne, l'écart-type, la skewness (asymétrie), la kurtosis (aplatissement), les valeurs minimales et maximales pour chaque œil sont extraites et analysées.

Le modèle est composé de plusieurs composants combinant CNN, LSTM et un module d'attention. Une fois que les indices sont extraits, FacialCutNet les concatènent et prédit une valeur de sortie.

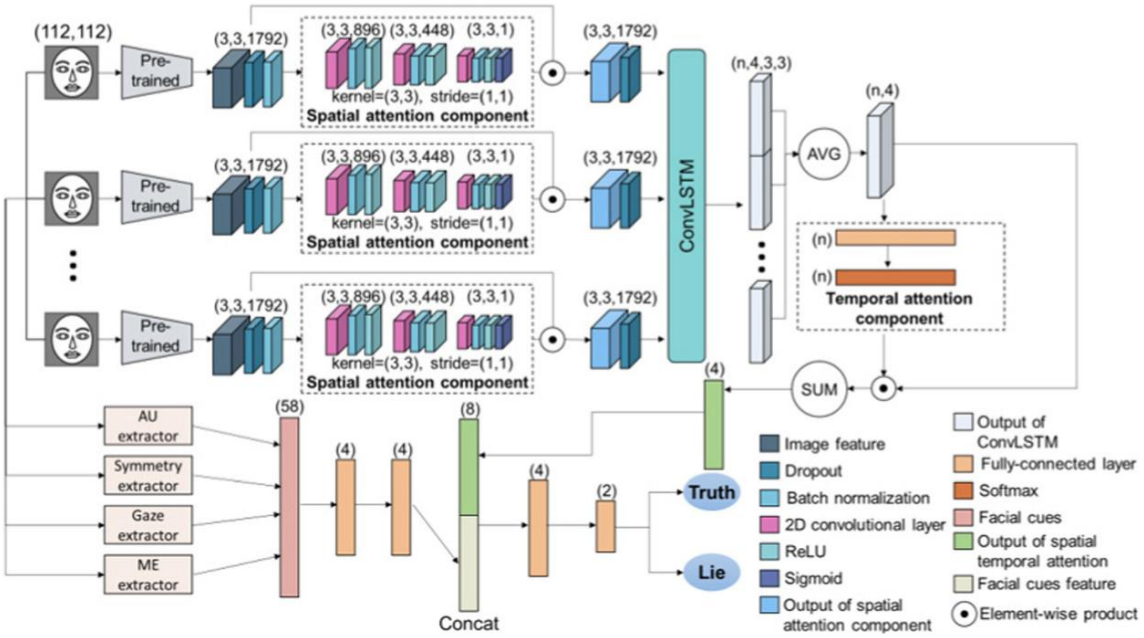

Figure 8 : Méthode de classification du mensonge avec FacialCueNet
Source : [1]

Dynamique oculaire & clignement des yeux

Un élément clé de la dynamique faciale est le mouvement des yeux, c'est-à-dire l'ouverture et la fermeture des yeux dans chaque image vidéo. Dans [9], la dynamique oculaire, mesurée par le Eye Aspect Ratio (EAR) qui représente la relation entre la largeur et la hauteur de l'œil, et la fréquence de clignement des yeux, analysée à travers d'histogrammes de pourcentage de clignement, sont étudiés.

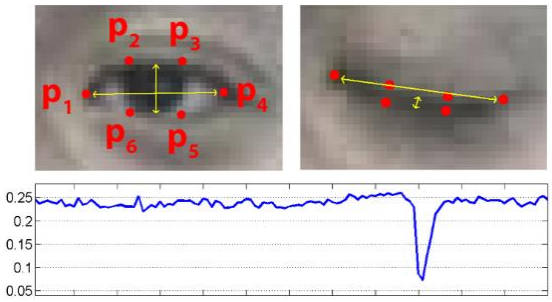

Figure 9 : Landmarks faciaux 2D des yeux
Source : [44]

Charge cognitive

En plus d'analyser les expressions faciales, [4] propose une analyse cognitive en mesurant la durée de leur performance des participants. Cela repose sur l'hypothèse que la charge cognitive est plus élevée dans les situations de mensonge, car le sujet fait plus d'effort pour gérer cette situation, ce qui engagerait des processus mentaux plus importants que les situations où l'individu est honnête.

4.3 Techniques de classification

Après avoir extraits les indices de mensonge, les chercheurs se retrouvent face à une problématique de classification commune. Étant donné un ensemble de vidéos d'entrée (input) pour lesquelles la vérité est connue, le problème de la détection de mensonge peut se présenter comme suit :

Input : Un ensemble de vidéos $V = \{v_1, \ldots, v_N\}$ et les étiquettes correspondantes $Y = \{y_1, \ldots, y_N\}$ où $y_i \in \{\text{mensonge}, \text{vérité}\}$.

Question : Existe-t-il un modèle de marchine learing f capable d'étiqueter correctement v_i pour tout $v_i \in V$, c'est-à-dire $P(f(v_i) \neq y_i) < \epsilon$, pour un ϵ (seuil d'erreur) arbitrairement petit mais supérieur à 0 ?

Figure 10 : Formulation mathématique du problème
Source : [9]

Les modèles prédiction utilisent différentes méthodes pour classifier les vidéos comme trompeuse ou véridiques. À noter que les modèles de classification peuvent être les mêmes que ceux utilisés pour l'extraction des caractéristiques faciales. Pour la suite de l'analyse, nous étudierons les méthodes de ML/DL qui ont été le plus utilisés pour l'ensemble des études du corpus.

4.3.1 Classificateur (ML)

Random Forest

Dans [7], l'utilisation du Random Forest est motivé par sa fiabilité et stabilité sur de petits échantillons, tandis que dans [9], il est utilisé à titre de comparaison. Ce modèle est utilisé avec les caractéristiques obtenues par OpenFace comme données d'entrée pour prédire le mensonge dans [9], alors que dans [7], Random Forest est utilisée à la fois pour quelles caractéristiques ou combinaisons de caractéristiques étaient pertinentes et pour la classification.

SVM

Le SVM est sélectionné dans [3] car il s'est révélé performant pour de petites quantités de données. Deux approches ont été utilisées pour ce classificateur : la première repose sur l'extractions des caractéristiques faciales à partir de caractéristiques définies par des experts (« handcrafted features ») tandis que la seconde se sert d'OpenFace pour extraire les caractéristiques faciales. Les données obtenues servent ensuite d'input au SVM pour la prédiction.

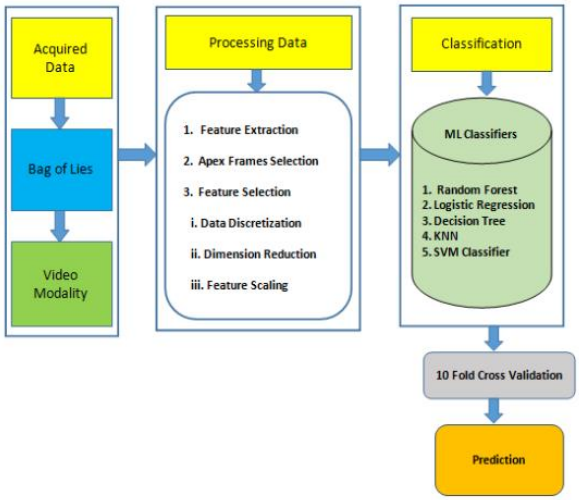

Figure 11 : Modèle de prédiction avec des classificateurs
Source : [5]

4.3.2 Réseau de neurones (DL)

CNN

Dans [2], le modèle est entraîné à partir de *FER-2013*. En comparant les différences des émotions présentes entre les phases où il est sincère et celles où il ne l'est pas.

C3D

Dans [3], bien que le C3D puisse extraire les caractéristiques faciales lui-même, étant donné qu'il reçoit en entrée des images brutes de la vidéo, OpenFace a été utilisé de sorte à cibler uniquement le visage dans les vidéos. Les dimensions de chaque image sont ensuite normalisées à 112 x 112 pour alimenter le C3D. Ce pré-traitement de la donnée permet de surmonter les contraintes de mémoire GPU.

LSTM

Dans [3], LSTM est utilisé sur des caractéristiques faciales également extraites par OpenFace. Contrairement aux autres modèles, LSTM peut traiter directement la séquence d'activations des AUs sans avoir besoin de faire un calcul d'agrégation. Le modèle base ses prédictions sur toutes les images des vidéos et leurs positions réciproques. Comme l'entraînement de LSTM nécessite beaucoup de temps, le modèle a été validé uniquement sur les données obtenues lors de la phase interrogatoire, car les auteurs estiment qu'elle contient plus d'indices évidents au mensonge qu'une phase de discours libre où le sujet n'est pas confronté directement.

Contraintes techniques

Pour surmonter les contraintes de mémoire GPU, rencontrées dans [3], la vidéo est divisée en bloc de 16 images pour prédire la vidéo entière en agrégeant les prédictions des blocs précédents. Cette méthode permet d'estimer l'intensité moyenne des indices de mensonge dans l'ensemble de la vidéo.

La problématique des classes déséquilibrés peut également être rencontrée. Bien que les auteurs dans [7], ont corrigé leur échantillon pour tenir compte du fait qu'il y avait plus de classes menteurs que de classes non-menteur durant le développement du modèle, celui-ci reste significativement biaisé pour classer des enfants comme menteur plutôt que non-menteur.

4.3.3 Tableaux récapitulatifs

Réf article	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[9]	[11]
OpenFace	X		X	X	X	X	X	X	
Handcrafted features			X						
CNN		X							
Custom dataset								X	

***Tableau** : Synthèse des méthodes d'extractions des indices utilisées*

Réf article	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[9]	[11]
Random Forest					X	X	X	X	
Decision Tree					X				
SVM			X		X				
LSTM			X						
CNN			X					X	
Logistic Regression					X				
KNN					X				
Multiple Instance Learning								X	
FacialCueNet	X								

***Tableau** : Synthèse des méthodes de classification utilisées*

4.3 Analyse des résultats

4.3.1 Métriques

Pour mesurer l'efficacité des modèles de machine learning, la plupart des études évoquées se basent sur les métriques suivantes. Les termes « Accuracy » et « Precision » ne seront pas traduits en français pour ne pas les confondre.

Area Under the Curve (AUC)

La mesure AUC est la plus populaire et est largement adoptée en machine learning [43]. Elle équivaut à l'Accuracy ce qui traduit la capacité du modèle à faire des prédictions correctes lors de la phase d'entraînement.

Accuracy

Représente le rapport entre le nombre de prédictions correctes et le nombre total de prédictions. Elle mesure la fréquence à laquelle un modèle prédit correctement le résultat. [2]

Precision

Représente la nombre de prédictions positives bien effectuées par le modèle. Elle mesure la fréquence à laquelle un modèle prédit correctement la classe positive. [2]

Recall

Représente la proportion de positifs bien prédit par le modèle. Elle mesure la fréquence à laquelle un modèle identifie correctement les instances positives (vrais positifs) parmi tous les échantillons positifs réels de l'ensemble de données. [2]

F1-Score

Représente la moyenne harmonique des valeurs de précision et de rappel. Il s'agit d'un bon indicateur pour évaluer la performance d'un modèle. Un F1-Score proche de 1 indique que le modèle est performant. [2]

4.3.1 Évaluation des modèles

Expression faciale

Le modèle SVM couplé à OpenFace utilisée dans [3, 5] semblent démontrer de meilleures performances comparés aux autres modèles de prédiction pour leur dataset respectif. *Bag of lies*, obtient une accuracy de 61.54%. Les autres modèles réussissaient néanmoins à classifier correctement les vidéos, avec un résultat supérieur à 53% pour les métriques : accuracy, precision, recall et F1-score, pour la majorité des modèles utilisés. *Holiday* présente un AUC de 0.72 pour les vidéos sans charge cognitive et un AUC de 0.78 pour les vidéos avec charge cognitive. De manière générale, les classificateurs testés dans [3] montrent de meilleurs performances sur les vidéos avec charge cognitive, ce qui suggère que mentir demande plus d'effort que dire la vérité, en particulier quand des questions inattendues sont posées.

Dans une expérience menées dans [3] pour déterminer la performance de l'humain, les résultats montrent que le genre ou le niveau d'éducation n'a pas d'incidence sur la capacité à prédire le mensonge, ce qui rejoint [7] sur le fait que le sexe n'est pas un facteur qui fait varier

l'aptitude à discerner le mensonge. De plus, la performance de l'humain comparée aux machines diminue, peut-être dû à un « effet de fatigue », où l'attention diminue au fil du temps. Malgré le fait que l'humain se basent se plusieurs critères tels que les détails dans la narration, le contact visuel et les expressions faciales, il reste moins bon que les machines avec un AUC de 0.57.

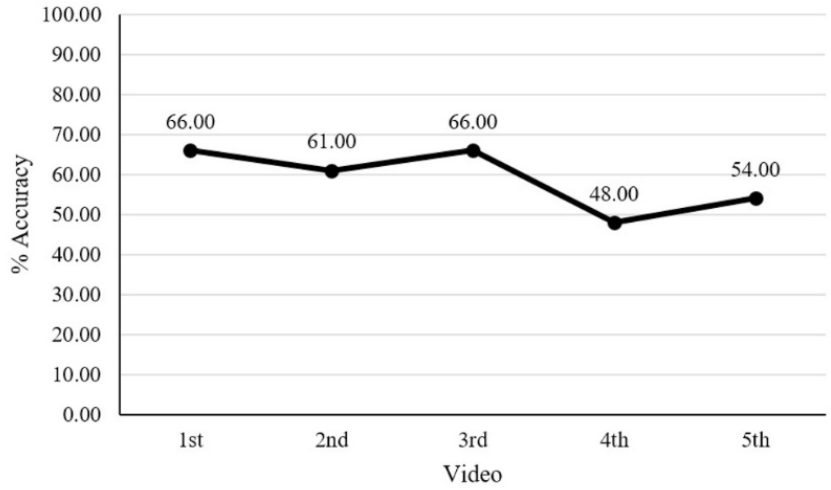

Figure 12 : Accuracy moyenne des humains en fonction de l'ordre des vidéos regardées
Source : [3]

En revanche, [9] se montre moins positif quant aux capacités des modèles à identifier le mensonge. Avec Random forest, Les résultats montrent un surapprentissage sur les données d'entraînement et une incapacité à généraliser sur l'ensemble de validation. *Real life trial* a obtenu de meilleures performances, avec un AUC de 0.58, comparé à *Crime* et *Opinion*, ce qui peut être dû au fait que les caractéristiques pour ce dataset étaient manuellement annotées, tandis que les caractéristiques étaient apprises par le modèle pour les autres dataset. Lorsque le modèle est entraîné sur *Opinion* et testé sur *Crime*, on constate que sa performance n'est pas significativement plus élevée pour justifier la généralisation du modèle sur un dataset à l'autre. Cela s'explique par le fait que les modèles de prédiction sont hautement spécifiques à leur dataset et ne peuvent pas être testés sur un dataset qui varie beaucoup du dataset original. Enfin, pour le CNN, sa précision sur ces datasets est comparable à un « pile ou face ».

Émotions

L'expression de surprise accrue chez les menteurs être interprétée comme une excitation générale lorsqu'ils sont invités à réfléchir à la transgression (contrairement aux non-menteurs qui ne l'ont pas vécu), elle peut être plus marquée chez les menteurs car c'est la première qu'on les interroge dessus. L'expression de peur peut indiquer que l'enfant perçoit le mensonge comme une violation morale et rencontre un conflit entre deux obligations sociales, l'une étant d'avouer à une personne que les jouets sont cassés, l'autre étant de garder le silence tel que l'avait suggéré une autre personne.

Dans [2], le modèle de prédiction atteint une précision de classification de 74,17 %, ce qui est considéré comme très performant par rapport à d'autres modèles conçus à l'aide de CNN et de l'ensemble de données FER-2013. Cependant cette métrique seule ne permet pas de déterminer la performance réelle du modèle. Il faudrait plutôt évaluer la F1 score qui n'est pas donné dans l'étude, qui plus est, le nombre de participants ne permet pas d'affirmer que le modèle est bon.

Dans [7], le modèle prédit une le mensonge avec une précision constante indépendamment de l'âge, du sexe ou des antécédents de maltraitance des enfants. Cependant, on constate que leur modèle est plus performant chez les enfants que chez les adultes. Les métriques utilisées, tel que la moyenne du score d'expression ou l'amplitude de l'expression pour résumer données de FACS automatiques, peut influencer sur la capacité du modèle à détecter efficacement le mensonge.

Approche multimodale

Les résultats de [4] ont montré que les actes trompeurs duraient plus longtemps que les actes véridiques. Notamment, les vidéos impliquant la stratégie du Mensonge Sophistiqué présentaient des durées significativement plus longues que celles du Mensonge Simple ou de la Vérité pure, ce qui souligne une augmentation de la charge cognitive associée au Mensonge Sophistiqué.

Cette différence est particulièrement marquée dans les conditions dynamiques sans parole, où le temps d'action pour le Mensonge Sophistiqué excédait celui de la situation Dynamique avec parole. Cela suggère que la richesse des médias influence la durée de la performance. Ces résultats corroborent la théorie de la Richesse des Médias, montrant que les stratégies de tromperie plus complexes demandent une charge cognitive accrue, reflétée par des durées d'actions prolongées, spécialement dans des contextes où la communication verbale est absente.

Le mouvement des yeux et le clignement des yeux, indices étudiés dans [9] montrent que les menteurs ont tendance à présenter des schémas de dynamique oculaire plus cohérents, tandis que les personnes qui disent la vérité sont plus détendues et enregistrent souvent une plus grande amplitude de mouvement des yeux. Les performances du réseau de neurone à détecter le mensonge basé sur la dynamique oculaire sont meilleures sur *Opinion* dataset ce qui contredit l'hypothèse de la charge cognitive.

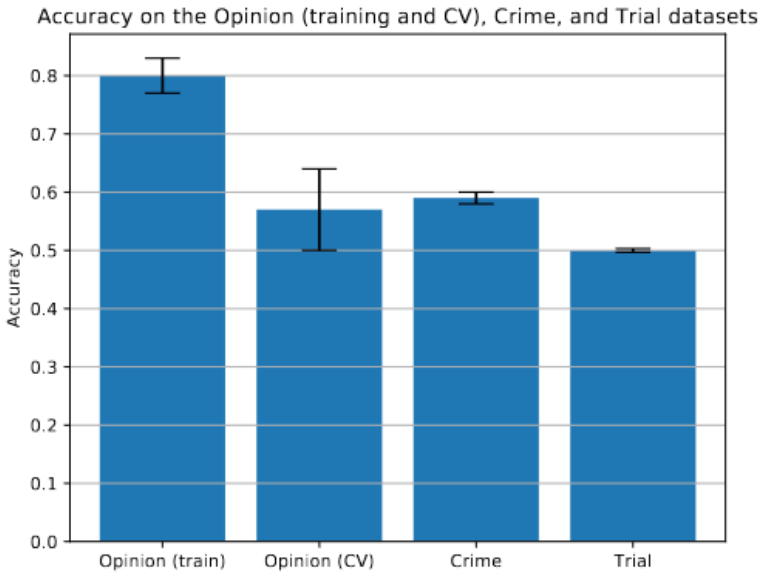

Figure 13 : Accuracy sur Opinion, Crime et Real life trial
Source [9]

Concernant le clignement des yeux, les prédictions des séquences de données du même clip vidéo sont agrégées et l'ensemble du clip est prédit comme « mensonge » si le résultat dépasse un seuil. Ce seuil est sélectionné pour maximiser l'exactitude de l'entraînement avec des taux de faux positifs souhaités. En se basant sur le rapport entre les pourcentages de clignements dans les situations de vérité et de mensonge, on obtient une accuracy de 56.2% sur Crime, ce qui est comparable à la performance humaine. Les résultats montrent également que les menteurs ont tendance à cligner moins des yeux (pourcentage de clignements ≤ 2 %) comparés aux non-menteurs qui clignent des yeux plus fréquemment (pourcentage de clignements ≥ 4 %). Cependant, aucun seuil précis ou motif probabiliste général permet de distinguer clairement les groupes menteurs des groupes de contrôle.

5. Conclusion

L'ensemble des études du corpus s'appuient sur des données limitées, le nombre de participant dans les expériences menés n'excédant pas 200. Le mensonge induit en laboratoire ne peut pas refléter toujours refléter la complexité des mensonges rencontrés dans les contextes réels. Même si les chiffres se montrent positifs quant à la précision des modèles de prédiction, il existe un paradoxe selon lequel un modèle peut présenter une précision très élevée mais qui, finalement, n'est pas applicable sur d'autres contextes car il est trop spécifique aux données sur lesquelles il a été entraîné.

La performance des modèles à prédire le mensonge est équivalente voire légèrement supérieure à la capacité humaine. Les caractéristiques précises liées au mensonge varient d'une étude à l'autre mais peuvent dans l'ensemble, aider le modèle à apprendre les indices de mensonge et à le détecter.

Néanmoins, la question de la subjectivité et du biais se pose toujours. Peut-on dire que le modèle est fiable lorsqu'il est entraîné sur un dataset où la vérité fondamentale n'est pas toujours connue ? Beaucoup de recherches sur la détection de mensonge utilise le Real life trial, alors qu'il peut comporter des erreurs judiciaires et d'interprétation de preuves. Les résultats ne sont pas suffisants justifier l'utilisation de ces modèles ML/DL dans des contextes réels.

Annexes

[1] Nam, Borum, Joo Young Kim, Beomjun Bark, Yeongmyeong Kim, Jiyoon Kim, Soon Won So, Hyung Youn Choi, and In Young Kim. "FacialCueNet: unmasking deception-an interpretable model for criminal interrogation using facial expressions." Applied Intelligence 53, no. 22 (2023): 27413-27427.

[2] Yildirim, Suleyman, Meshack Sandra Chimeumanu, and Zeeshan A. Rana. "The influence of micro-expressions on deception detection." Multimedia Tools and Applications 82, no. 19 (2023): 29115-29133.

[3] Monaro, Merylin, Stéphanie Maldera, Cristina Scarpazza, Giuseppe Sartori, and Nicolò Navarin. "Detecting deception through facial expressions in a dataset of videotaped interviews: A comparison between human judges and machine learning models." Computers in Human Behavior 127 (2022): 107063.

[4] Zhou, Xingchen, Rob Jenkins, and Lei Zhu. "An Honest Joker reveals stereotypical beliefs about the face of deception." Scientific reports 13, no. 1 (2023): 16649.

[5] Islam, Siam, Popin Saha, Touhidul Chowdhury, Asif Sorowar, and Raqeebir Rab. "Non-invasive deception detection in videos using machine learning techniques." In 2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), pp. 1-6. IEEE, 2021.

[6] Shen, Xunbing, Gaojie Fan, Caoyuan Niu, and Zhencai Chen. "Catching a liar through facial expression of fear." Frontiers in Psychology 12 (2021): 675097.

[7] Bruer, Kaila C., Sarah Zanette, Xiao Pan Ding, Thomas D. Lyon, and Kang Lee. "Identifying liars through automatic decoding of children's facial expressions." Child development 91, no. 4 (2020): e995-e1011.

[8] Alaskar, Haya. "Hybrid Metaheuristics with Deep Learning Enabled Automated Deception Detection and Classification of Facial Expressions." Computers, Materials & Continua 75, no. 3 (2023).

[9] Belavadi, Vibha, Yan Zhou, Jonathan Z. Bakdash, Murat Kantarcioglu, Daniel C. Krawczyk, Linda Nguyen, Jelena Rakic, and Bhavani Thuriasingham. "MultiModal deception detection: Accuracy, applicability and generalizability." In 2020 Second IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), pp. 99-106. IEEE, 2020.

[11] Jupe, Louise Marie, and David Adam Keatley. "Airport artificial intelligence can detect deception: or am i lying?." Security Journal 33, no. 4 (2020): 622-635.

[12] Speth, Jeremy, Nathan Vance, Adam Czajka, Kevin W. Bowyer, Diane Wright, and Patrick Flynn. "Deception detection and remote physiological monitoring: A dataset and baseline experimental results." In 2021 IEEE International Joint Conference on Biometrics (IJCB), pp. 1-8. IEEE, 2021.

[13] Bhamare, Ameya Rajendra, Srinivas Katharguppe, and J. Silviya Nancy. "Deep neural networks for lie detection with attention on bio-signals." In 2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI), pp. 143-147. IEEE, 2020.

[14] Kircher, John C., and David C. Raskin. "Polygraph techniques: History, controversies, and prospects." Psychology and social policy (2019): 295-308.

[15] Lewis, Jerry A., and Michelle Cuppari. "The polygraph: The truth lies within." The journal of psychiatry & law 37, no. 1 (2009): 85-92.

[16] Lajevardi, Seyed Mehdi, and Zahir M. Hussain. "Automatic facial expression recognition: feature extraction and selection." Signal, Image and video processing 6 (2012): 159-169.

[17] Porter, Stephen, and Leanne ten Brinke. "The truth about lies: What works in detecting high-stakes deception?." Legal and criminological Psychology 15, no. 1 (2010): 57-75.

[18] Vrij, Aldert, Katherine Edward, Kim P. Roberts, and Ray Bull. "Detecting deceit via analysis of verbal and nonverbal behavior." Journal of Nonverbal behavior 24 (2000): 239-263.

[19] Chebbi, Safa, and Sofia Ben Jebara. "Deception detection using multimodal fusion approaches." Multimedia Tools and Applications 82, no. 9 (2023): 13073-13102.

[20] Feng, Song, Ritwik Banerjee, and Yejin Choi. "Syntactic stylometry for deception detection." In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 171-175. 2012.

[21] Hirschberg, Julia Bell, Stefan Benus, Jason M. Brenier, Frank Enos, Sarah Friedman, Sarah Gilman, Cynthia Girand et al. "Distinguishing deceptive from non-deceptive speech." (2005).

[22] Newman, Matthew L., James W. Pennebaker, Diane S. Berry, and Jane M. Richards. "Lying words: Predicting deception from linguistic styles." Personality and social psychology bulletin 29, no. 5 (2003): 665-675.

[23] Burzo, Mihai, Mohamed Abouelenien, Veronica Perez-Rosas, and Rada Mihalcea. "Multimodal deception detection." In The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition-Volume 2, pp. 419-453. 2018.

[24] Bhatt, Priya, Amanrose Sethi, Vaibhav Tasgaonkar, Jugal Shroff, Isha Pendharkar, Aditya Desai, Pratyush Sinha et al. "Machine learning for cognitive behavioral analysis: datasets, methods, paradigms, and research directions." Brain informatics 10, no. 1 (2023): 18.

[25] Rojas-Barahona, Lina, Bo-Hsiang Tseng, Yinpei Dai, Clare Mansfield, Osman Ramadan, Stefan Ultes, Michael Crawford, and Milica Gasic. "Deep learning for language understanding of mental health concepts derived from cognitive behavioural therapy." arXiv preprint arXiv:1809.00640 (2018).

[26] Spezio, Michael L., and Ralph Adolphs. "Emotional processing and political judgment: Toward integrating political psychology and decision neuroscience." The affect effect: Dynamics of emotion in political thinking and behavior (2007): 71-95.

[27] DePaulo, Bella M., James J. Lindsay, Brian E. Malone, Laura Muhlenbruck, Kelly Charlton, and Harris Cooper. "Cues to deception." Psychological bulletin 129, no. 1 (2003): 74.

[28] DePaulo, Bella M., Deborah A. Kashy, Susan E. Kirkendol, Melissa M. Wyer, and Jennifer A. Epstein. "Lying in everyday life." Journal of personality and social psychology 70, no. 5 (1996): 979.

[29] Bond Jr, Charles F., and Bella M. DePaulo. "Accuracy of deception judgments." Personality and social psychology Review 10, no. 3 (2006): 214-234.

[30] Gongola, Jennifer, Nicholas Scurich, and Jodi A. Quas. "Detecting deception in children: A meta-analysis." Law and human behavior 41, no. 1 (2017): 44.

[31] Lloyd, E. Paige, Jason C. Deska, Kurt Hugenberg, Allen R. McConnell, Brandon T. Humphrey, and Jonathan W. Kunstman. "Miami University deception detection database." Behavior research methods 51 (2019): 429-439.

[32] Lu, Haoyu, Qiongyi Zhou, Nanyi Fei, Zhiwu Lu, Mingyu Ding, Jingyuan Wen, Changde Du et al. "Multimodal foundation models are better simulators of the human brain." arXiv preprint arXiv:2208.08263 (2022).

[33] Frank, Mark G., and Janine Stennett. "The forced-choice paradigm and the perception of facial expressions of emotion." Journal of personality and social psychology 80, no. 1 (2001): 75.

[34] Lu, Guanming, Xiaonan Li, and Haibo Li. "Facial expression recognition for neonatal pain assessment." In 2008 International conference on neural networks and signal processing, pp. 456-460. IEEE, 2008.

[35] Ekman, Paul, and Wallace V. Friesen. "Facial action coding system." Environmental Psychology & Nonverbal Behavior (1978).

[36] Bartlett, Marian Stewart, Gwen Littlewort, Mark G. Frank, Claudia Lainscsek, Ian R. Fasel, and Javier R. Movellan. "Automatic recognition of facial actions in spontaneous expressions." J. Multim. 1, no. 6 (2006): 22-35.

[37] Baltrusaitis, Tadas, A. Zadeh, Y. C. Lim, and L. P. Morency. "Openface 2.0: Facial behavior analysis toolkit. En 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)." (2018): 59-66.

[38] Breiman, Leo, and A. Cutler. "Setting Up." Using, And Understanding Random Forests 4 (2003).

[39] Khan, Amjad Rehman. "Facial emotion recognition using conventional machine learning and deep learning methods: Current achievements, analysis and remaining challenges." Information 13, no. 6 (2022): 268.

[40] Wang, Su-Jing, Bing-Jun Li, Yong-Jin Liu, Wen-Jing Yan, Xinyu Ou, Xiaohua Huang, Feng Xu, and Xiaolan Fu. "Micro-expression recognition with small sample size by transferring long-term convolutional neural network." Neurocomputing 312 (2018): 251-262.

[41] Sartori, Giuseppe, Sara Agosta, Cristina Zogmaister, Santo Davide Ferrara, and Umberto Castiello. "How to accurately detect autobiographical events." Psychological science 19, no. 8 (2008): 772-780.

[42] Gupta, Viresh, Mohit Agarwal, Manik Arora, Tanmoy Chakraborty, Richa Singh, and Mayank Vatsa. "Bag-of-lies: A multimodal dataset for deception detection." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 0-0. 2019.

[43] Ling, Charles X., Jin Huang, and Harry Zhang. "AUC: a better measure than accuracy in comparing learning algorithms." In Advances in Artificial Intelligence: 16th Conference of the Canadian Society for Computational Studies of Intelligence, AI 2003, Halifax, Canada, June 11–13, 2003, Proceedings 16, pp. 329-341. Springer Berlin Heidelberg, 2003.

[44] Soukupova, Tereza, and Jan Cech. "Eye blink detection using facial landmarks." In 21st computer vision winter workshop, Rimske Toplice, Slovenia, vol. 2. 2016.