

Classification automatique des émotions par exploitation du capteur de mouvement et de la dynamique de frappe sur smartphone

Automatic classification of emotions using motion sensors and keystroke dynamics on smartphones

Nicolas Simonazzi¹, Jean-Marc Salotti^{1,2}, Caroline Dubois³, Philippe Le Goff³

¹ Laboratory, IMS, UMR CNRS 5218, Bordeaux INP, Université de Bordeaux, ENSC, France, nsimonazzi@gmail.com

² Auctus, Inria, France, jean-marc.salotti@ensc.fr

³ Orange Innovation, 23 rue Thomas Edison, 33600 Canéjan, France, caroline.dubois@orange.fr, philippe2.legoff@orange.com

RÉSUMÉ. Nous présentons les résultats d'une étude sur une classification binaire des émotions, à partir de données recueillies grâce aux capteurs de mouvement et aux frappes clavier d'un smartphone et d'un bracelet connecté. À cette fin, nous avons développé une application mobile permettant d'induire des émotions par le biais de vidéos et d'enregistrer les interactions des utilisateurs. Un système d'auto-évaluation numérique spécifique a été développé à partir de la Geneva Emotion Wheel pour aider les participants à exprimer leurs émotions. Les enregistrements des capteurs ont été étiquetés en fonction des déclarations des participants et des conditions vidéo. Une méthode est ainsi proposée pour traiter les données temporelles collectées et classer automatiquement la valence des émotions déclarées grâce à des techniques d'apprentissage automatique. Nous avons testé une classification générale de la valence utilisant toutes les émotions de tous les individus et une classification personnalisée utilisant un sous-ensemble d'émotions d'un seul individu. Le résultat le plus prometteur a été obtenu avec un modèle personnalisé, pour lequel nous avons pu obtenir, en moyenne pour tous les participants, deux tiers de classification correcte de la valence, en utilisant des données fusionnées de différentes modalités.

ABSTRACT. We present the results of a study on a binary classification of emotions, based on data collected through motion sensors and keystrokes of a smartphone and a connected bracelet. To this end, we developed a mobile application to induce emotions through videos and record user interactions. A specific digital self-assessment system was developed based on the Geneva Emotion Wheel to help participants express their emotions. The sensor recordings were labelled according to participants' statements and video conditions. A method is thus proposed to process the collected temporal data and automatically classify the valence of the declared emotions using machine learning techniques. We tested a general valence classification using all emotions from all individuals and a personalized classification using a subset of emotions from a single individual. The most promising result was obtained with a personalized model, for which we were able to obtain, on average across all participants, two-thirds of correct valence classification, using fused data from different modalities.

MOTS-CLÉS. Classification des émotions, détection des émotions, roue des émotions de Genève, usage des smartphones, cognitique, ingénierie cognitive.

KEYWORDS. Emotions classification, emotions detection, Geneva emotion wheel, smartphones usage, cognitics, cognitive engineering.

1. Introduction

L'identification des émotions est particulièrement importante dans le domaine de la relation client lorsque l'utilisateur interagit avec un chatbot [KH 99, STA 21, BRJ 18]. La reconnaissance des émotions représente un enjeu crucial dans l'amélioration des interactions homme-machine, comme le soulignent Calvo et D'Mello [CD 2010]. Selon une méta-analyse récente [KBN 24], l'intégration de capacités de reconnaissance émotionnelle dans les systèmes d'IA conversationnelle peut améliorer significativement la satisfaction utilisateur et l'efficacité des interactions. Que le chatbot soit basé sur la reconnaissance

d'expressions simples et des réponses stéréotypées ou sur un grand modèle de langage (LLM) et des réponses structurées, il arrive souvent que l'utilisateur ne soit pas satisfait des réponses fournies et qu'un sentiment de frustration, voire de colère, émerge et soit renforcé par l'incapacité de l'agent artificiel à prendre en considération les émotions, quand la bascule vers un conseiller humain ne se produit pas. La reconnaissance de certaines émotions pourrait permettre d'évaluer plus finement la qualité des chatbots, voire contribuer à faire basculer de façon plus pertinente et efficace l'interaction vers un opérateur humain. Sa réponse pourrait intégrer des formules plus empathiques comme "Je comprends que vous n'êtes pas satisfait, je vous propose de parler à un conseiller ...". Les LLMs actuels présentent encore des limitations importantes dans la détection des nuances émotionnelles, particulièrement dans les interactions textuelles où les indices non verbaux sont absents [YQ 21]. Cette lacune est d'autant plus problématique que les smartphones sont préférés aux ordinateurs pour les activités sociales telles que l'utilisation de services de messagerie instantanée [BRJ 18].

Notre étude vise précisément à explorer les méthodes permettant de reconnaître en temps réel les émotions, en restreignant le contexte d'usage à celui d'une interaction avec un chatbot au moyen d'un smartphone. Dans ce type d'interaction, les expressions non verbales comme le ton ou l'expression du visage sont absentes. Selon Walther, les utilisateurs pourraient exprimer leurs émotions en utilisant des émoticônes ou en consacrant plus de temps à la rédaction de textes [WAL 96]. Ces stratégies offrent un contrôle supplémentaire mais les expressions émotionnelles spontanées sont difficiles à communiquer. L'approche alternative, développée dans cette étude, consiste à exploiter les capteurs disponibles tels que l'accéléromètre et le gyroscope intégrés et la dynamique de frappe [PIC 95, AGO 13]. Intuitivement, un utilisateur en colère doit agiter son smartphone plus que d'habitude. De manière générale, les émotions impactant directement notre comportement, il est possible que de subtiles variations, comme par exemple un léger tremblement de la main qui tient le smartphone selon une certaine fréquence et amplitude, ou des accélérations plus amples selon une direction donnée, puissent être observées et soient caractéristiques de l'émotion ressentie. Il est bien entendu utopique de prétendre pouvoir identifier les émotions des utilisateurs de cette façon avec un excellent taux de réussite, car il existe de nombreuses variables qui impactent la manipulation du smartphone, comme la façon de tenir l'appareil, la fatigue musculaire, le confort de la position, la luminosité, la taille de l'écran, le bruit ambiant, qui n'ont aucun rapport avec les émotions. D'autre part, de la même manière qu'une personne peut cacher ses émotions ou en simuler une autre, quels que soient les éléments détectés, l'émotion éventuellement associée ne pourrait être que probable et non certaine. Enfin, sachant qu'il existe une grande variabilité comportementale pour exprimer une émotion, la généralisation paraît difficile.

Ces limites étant assumées, nous proposons d'aborder malgré tout le problème de l'identification des émotions à partir des capteurs du smartphone afin de déterminer le potentiel de cette approche, c'est-à-dire le pourcentage de reconnaissance des émotions qu'il est possible d'obtenir. Pour cela, il est nécessaire de travailler sur un jeu de données. Nous avons ainsi mené une expérience exploitant un outil de messagerie instantanée sur smartphone, nous avons collecté les données des capteurs de mouvement et des frappes clavier, ainsi que les sentiments subjectifs fournis par les utilisateurs. De plus, nous avons collecté des données physiologiques à l'aide d'un bracelet Empatica E4 équipé de capteurs physiologiques, activité électro-dermique, activité cardiaque, température de peau [Zwy 18]. Nous avons entraîné des modèles d'apprentissage automatique généraux et personnalisés avec ces données afin de déterminer si certaines émotions peuvent être associées à un comportement humain spécifique.

Les concepts théoriques, les travaux existants sur la reconnaissance des émotions sur smartphones et le modèle choisi sont présentés dans la partie 2. Le protocole expérimental de mesure et d'induction des émotions, ainsi que les résultats obtenus, sont présentés dans la partie 3. La partie 4 est consacrée à l'apprentissage automatique qui a été mis en œuvre pour tenter d'identifier automatiquement la valence émotionnelle à partir des données des capteurs des smartphones. Un résumé des résultats et des perspectives est présenté dans la conclusion.

2. Emotions et données du smartphone

2.1. Emotions et informatique affective

Un enjeu important est de définir ce qui est entendu sous le terme d'émotions. À la fin du XXe siècle, la théorie de l'évaluation était devenue l'une des théories dominantes en sciences des émotions [ARN 60]. Cette évaluation provoque une tendance à l'action, ainsi que des modifications physiologiques et comportementales.

Une émotion peut également être définie comme un phénomène hypothétique, non observable directement [SCH 05]. Cependant, des épisodes émotionnels peuvent être déduits d'indicateurs externes. Dans le prolongement de la théorie de l'évaluation, un modèle multi-composante a été proposé [SCH 05]. Une émotion est ainsi définie comme : « un épisode de changements interdépendants et synchronisés dans les états de la totalité ou de la plupart des cinq sous-systèmes de l'organisme, en réponse à l'évaluation d'un stimulus externe ou interne pertinent pour les préoccupations majeures de l'organisme. »

Le premier et le plus important sous-système organique, essentiel à la théorie de l'évaluation, est la composante cognitive. Elle évalue les événements ou objets internes ou externes. Quatre autres composantes sont décrites dans ce modèle :

- Composante neurophysiologique (symptôme corporel) : régulation systémique ;
- Composante motivationnelle (Tendances à l'action) : préparation et orientation de l'action ;
- Composante d'expression motrice (symptôme corporel) : communication des réactions et du comportement ;
- Composante du sentiment subjectif (expérience émotionnelle) : surveillance de l'état interne et des interactions organisme-environnement.

Chaque mouvement effectué avec un smartphone est impacté par ces composantes. Au niveau respiratoire ou cardiovasculaire, le décalage spatial est de l'ordre du millimètre, le décalage temporel de l'ordre du dixième de seconde, mais il existe également des impacts plus faibles dus à l'activité neuronale ou aux fluctuations hormonales. Par conséquent, même si nous ne pouvons pas observer directement un phénomène émotionnel, nous pouvons utiliser des indices concrets issus de ces composantes pour déduire l'état émotionnel d'un individu. Des modèles informatiques, dont le travail fondateur de Picard, ont été proposés à cette fin et définissent le domaine de l'informatique affective [PIC 95, PCB 17, SKG 20, DKB 20]. Afin de reconnaître automatiquement les états émotionnels, les modalités suivantes doivent être prises en compte : caractéristiques visuelles [MSJ 18, KO 18], audio [SRK 18], textuelles [CBF 17], physiologiques [SXY 18], ainsi que les mouvements des appareils tels que le smartphone et la fréquence de frappe des touches [MK 29, YQ 21]. Une revue récente offre une synthèse des méthodes utilisées dans ce domaine et les difficultés rencontrées [KBN 24]. La méthode d'apprentissage la plus efficace semble être la machine à vecteur de support (SVM) et les recommandations sont notamment de privilégier les modèles permettant l'explicabilité de la classification obtenue et la quantification des incertitudes.

2.2. Catégorisation des émotions à partir des smartphones

Il existe de nombreuses façons de définir des modèles computationnels des émotions et de les catégoriser. Les chercheurs en théorie de l'évaluation ont développé des modèles de catégorisation basés sur l'appréciation des événements et de leurs conséquences. Par exemple, le modèle OCC classe 22 émotions selon la désirabilité de l'événement, ses conséquences pour soi ou pour autrui, et les raisons de l'événement [OT 90].

Deuxièmement, les modèles dimensionnels se basent sur un état émotionnel pouvant être représenté par un point dans un espace multidimensionnel. Chaque dimension est associée à une caractéristique émotionnelle spécifique. L'un des modèles les plus connus est le modèle dit « circumplex de l'affect »,

proposé par Russell en 1980 [RUS 80]. Il s'agit d'un modèle bidimensionnel, composé de valence et d'activation.

- Valence : elle illustre la caractéristique positive ou négative du sentiment ;
- Éveil (ou activation) : elle définit la mobilisation d'énergie provoquée par le processus émotionnel.

Certains auteurs ont proposé des modèles plus complexes à trois dimensions ou plus [MEH 96, BVV 14]. Dans le contexte difficile des interactions avec un smartphone, sans utilisation d'informations sémantiques (pas d'analyse du sens des mots), c'est le modèle de Russell que nous avons retenu dans cette étude. Les deux paramètres considérés sont donc la valence et l'intensité de l'émotion.

2.3. Capteurs de mouvement

Les études d'informatique affective menées sur des appareils tels que les ordinateurs nécessitent souvent des objets connectés et des capteurs invasifs ou encombrants pour enregistrer les composantes physiologiques ou motrices des interactions utilisateur. Cependant, les smartphones, grâce à leurs capteurs intégrés, permettent de collecter facilement des données utilisateurs afin d'inférer leurs émotions en fonction du contexte environnemental [PAP 17]. Kolakowska *et al* détaillent les différents capteurs utilisables pour développer des systèmes de reconnaissance automatique des émotions sur smartphones [KSS 20] : caméra, microphones, GPS, écran tactile et capteurs de mouvement.

Politou et al. ont examiné de nombreux défis dans le domaine de l'informatique affective mobile [PAP 17]. Premièrement, la confidentialité est particulièrement vulnérable lors de l'utilisation des smartphones. De nombreux capteurs peuvent collecter des données sensibles, telles que la localisation (avec les données GPS), les caractéristiques faciales (avec la caméra) ou les particularités vocales (avec le microphone). Pour les capteurs de mouvement, il est certes plus difficile d'identifier une personne, mais cela pourrait être possible si les données peuvent être corrélées à d'autres informations. Le contexte d'utilisation doit donc être contrôlé.

Deuxièmement, les ressources de traitement des appareils numériques, en particulier portables, sont limitées. L'enregistrement, le traitement et l'utilisation des données pour l'apprentissage automatique nécessitent des ressources importantes. D'autre part, la production de systèmes de reconnaissance représente une complexité trop importante pour des appareils limités comme les smartphones. Il est donc apparu pertinent d'explorer des données moins critiques pour la vie privée des utilisateurs et moins consommatrices d'énergie. Une étape préliminaire et essentielle de notre étude a donc consisté à construire des jeux de données représentatifs d'une activité de communication par messagerie sur un smartphone. Dans cette perspective, les données issues des capteurs de mouvement et de la dynamique tactile, en particulier la dynamique de frappe, nous ont semblé prometteuses [MK 20, YQ 21].

Les capteurs de mouvement comprennent l'accéléromètre, le gyroscope et le magnétomètre. L'accéléromètre est un capteur de force qui détermine en temps réel le vecteur d'accélération du smartphone selon les principes fondamentaux de la dynamique. Le gyroscope enregistre les rotations et les mouvements angulaires appliqués au téléphone. A l'instar des capteurs d'un avion, il enregistre les rotations autour des trois axes de l'espace : tangage autour de l'axe x, roulis autour de l'axe y et lacet autour de l'axe z. Enfin, le magnétomètre est une boussole électronique, capable de déterminer la direction de l'appareil dans l'espace à partir des mesures de champs magnétiques. Ces capteurs ont été spécifiquement utilisés pour la reconnaissance d'une partie significative de l'activité humaine [AGO 13]. En effet, ils sont particulièrement appropriés pour les applications mobiles de sport ou de santé, qui peuvent utiliser ces informations pour déterminer la durée et l'intensité des activités physiques entreprises par les utilisateurs. Concernant la reconnaissance des émotions, Cui et al. [CLZ 15] ont montré des vidéos aux participants pour susciter la joie ou la colère, puis leur ont demandé de marcher avec un smartphone attaché à leurs chevilles et à leurs poignets. Après avoir filtré les signaux de l'accéléromètre, les auteurs ont calculé 114 caractéristiques et entraîné des modèles d'apprentissage automatique (SVM, arbres de décision, réseaux de neurones, etc.). Ils ont réussi à identifier les émotions

avec un score de 90 %. Olsen et Torresen ont effectué des travaux comparables, mais dans un cadre non contrôlé [OT 16]. Les participants ont été invités à auto-déclarer leurs émotions lors d'activités de marche quotidiennes, en utilisant les dimensions de valence et d'activation. Bien que le nombre de participants soit plus faible, des modèles d'apprentissage ont été utilisés pour obtenir un taux de classification correcte des émotions allant jusqu'à 75 %.

Des activités plus spécifiques à l'utilisation du smartphone ont également fait l'objet de travaux. Par exemple, Ruensuk *et al* ont étudié le comportement des utilisateurs lors de tâches de navigation et de messagerie, grâce à des capteurs inertIELS (accéléromètre et gyroscope) et à l'écran tactile [ROC 19]. Les utilisateurs ont été invités à annoter leurs émotions sur des degrés de valence et d'activation et à interagir avec un agent conversationnel sur les médias qu'ils ont consultés. Grâce à des modèles d'apprentissage automatique, un taux de reconnaissance de 81 % a été obtenu pour distinguer les valences positives et négatives dans la tâche de navigation. Les résultats pour la tâche de conversation sont inférieurs avec un taux de 69 %, ce qui montre une plus grande difficulté à traiter les informations issues de ce type d'activité.

2.4. Dynamique de frappe au clavier

Clayton Epp suggère que la dynamique de frappe peut être définie comme suit [ELM 11] : « La dynamique de frappe est l'étude des schémas temporels uniques de la frappe d'un individu, et inclut généralement l'extraction de caractéristiques temporelles de frappe telles que la durée d'une pression sur une touche et le temps écoulé entre chaque pression. »

Sur les smartphones, les écrans tactiles permettent d'écrire du texte à l'aide de claviers virtuels. L'étude des frappes sur un smartphone est similaire à celle réalisée sur des claviers d'ordinateur. Plusieurs études ont déjà été menées sur les frappes et les états émotionnels [GGM 17a, GGM 17b, GHG 19, GWB 02, MK 20]. Ghosh définit la distance inter-frappe comme la durée entre deux événements clavier consécutifs, cette distance étant calculée à partir de l'horodatage de ces événements [GGM 17a]. À partir de cette caractéristique, il est possible de calculer des caractéristiques statistiques telles que la moyenne, la médiane, le maximum ou le minimum pour les sessions d'utilisation du clavier. Avec son équipe, il a mené une étude de terrain avec une application spécialement développée pour enregistrer les événements clavier. Pendant trois semaines, les données clavier ont été collectées sur les smartphones des utilisateurs. Un formulaire d'auto-évaluation des émotions a été proposé sous certaines conditions (quantité minimale de texte produit, changement d'application et durée minimale de la session). Les participants pouvaient ensuite choisir de s'annoter selon les quatre catégories : joie, détente, tristesse, stress. Les auteurs ont conçu des modèles d'apprentissage automatique pour chaque participant. Ils ont obtenu une précision moyenne de 73 % pour la classification des quatre états rapportés.

Une difficulté importante, cependant, dans l'utilisation des frappes au clavier, réside dans le fait que les mesures inter-taps et la dynamique en général peuvent dépendre de la taille de l'écran, de la largeur des touches, de facteurs anthropomorphiques, de la sensibilité de l'écran ou de l'utilisation habituelle d'un autre appareil par l'utilisateur.

Dans le cadre de cette recherche, nous avons cherché à étudier de nouvelles sources de données issues des interactions avec les smartphones. Plus précisément, les données des capteurs de mouvement et des claviers virtuels permettent de recueillir des informations pertinentes. Nous avons donc construit une base de données composée de données d'interaction avec les smartphones, qui doivent être annotées avec des étiquettes associées aux états émotionnels. Une fois suffisamment de données collectées, un système de classification automatique des émotions basé sur des techniques d'apprentissage automatique peut être développé. Ces deux étapes sont présentées dans les parties suivantes.

3. Induction d'émotions

3.1. Déclenchement des émotions

Bien que déclencher des émotions quantifiées et précises soit une tâche particulièrement difficile, les modalités suivantes peuvent être utilisées à cet effet : images, vidéos, musique et même odeurs. L'utilisation de stimuli est l'une des approches les plus courantes. Néanmoins, on peut ajouter à ces modalités perceptives l'exécution d'activités telles que des jeux, des tests de logique ou des interactions avec d'autres individus pour provoquer des émotions. Certains auteurs ont comparé les performances des différentes méthodes. Westermann et al. [WSS 96] ont réalisé une méta-analyse de 250 articles afin de comparer l'impact de 11 procédures d'induction d'émotions. Selon cette étude, la présentation d'un film ou d'une histoire s'est avérée la plus efficace pour induire des états d'humeur positifs et négatifs.

Plusieurs bases de données vidéo ont été produites et testées auprès de participants. En 1995, Gross et Levenson ont rassemblé des extraits de films, les ont classés selon huit émotions cibles et ont sélectionné les deux meilleurs extraits annotés pour chaque émotion cible, d'après les annotations des participants [GJR 95]. Cependant, la fiction limite la criticité d'une situation et sa capacité à susciter des émotions. Il peut donc être judicieux d'utiliser des vidéos de situations réelles et d'éviter les personnages et situations fictifs. Samson et al. [SKS 16] ont constitué une base de données à partir de vidéos maison disponibles sur Internet. Ces vidéos facilitent l'accès à ce type de contenu. Environ 200 extraits ont été évalués sur les dimensions de valence, d'activation, ainsi que sur des émotions discrètes telles que l'amusement, le dégoût ou la colère. Enfin, les vidéos ont été classées en quatre catégories générales de valence : positive, négative, mixte et neutre. Nous proposons ici de les utiliser.

3.2. Caractérisation des émotions

De nombreuses méthodes d'auto-évaluation ont été utilisées pour caractériser les émotions dans des contextes expérimentaux. Par exemple, le PANAS [WCT 88] (Positive Affect and Negative Affect Schedule) évalue 20 émotions sur une échelle de Likert à 5 points pour une période spécifique (instant présent, jour, semaine). Le calcul de chaque item donne la valence globale. Une autre méthode consiste à utiliser des échelles picturales, comme le Self-Assessment Manikin conçu par Bradley et Lang [BL 94]. Les dimensions du plaisir, de l'excitation et de la dominance peuvent être évaluées avec une échelle de Likert à 9 points illustrée par des personnages. L'avantage des échelles picturales est leur grande accessibilité à travers les cultures et leur utilisation avec les enfants [GIL 94]. Cependant, les échelles dimensionnelles offrent un faible niveau de granularité pour une compréhension de l'émotion ressentie par le sujet.

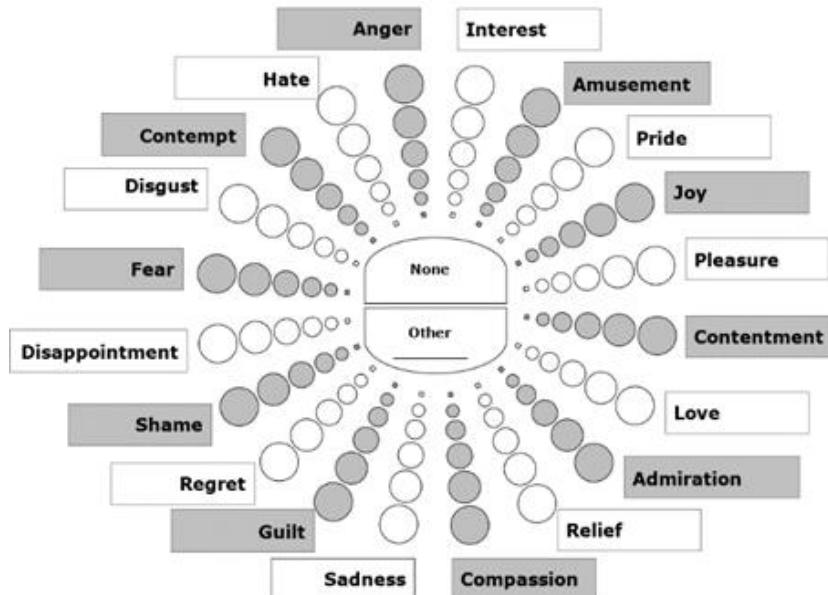


Figure 1. La Geneva Emotion Wheel (GEW)

Pour combiner modèles discrets et dimensionnels, Scherer *et al.* ont développé la Geneva Emotion Wheel (GEW) [SCH 05, SSF 13]. Sur la dernière version de la "roue", 20 émotions sont affichées sur deux dimensions : valence et niveau de contrôle (voir figure 1). De plus, le niveau d'intensité de l'émotion ressentie est indiqué par la taille des cercles sur la roue. Différentes méthodes peuvent être utilisées pour analyser les données. Les annotations des participants peuvent être analysées en fonction des dimensions et des étiquettes. La GEW est donc un outil flexible et adaptable à de nombreuses situations. La conception d'une version numérique a été réalisée dans notre étude pour permettre l'auto-évaluation directement sur smartphone et faciliter l'analyse des résultats. On pense qu'elle pourrait être utilisée dans d'autres applications dans le domaine de l'informatique affective. Selon Kolakowska *et al*, la plupart des études sur l'informatique affective mobile collectent le sentiment subjectif via PANAS ou SAM [KSS 20, WCT 88, BL 94]. Néanmoins, l'utilisation de la GEW sur smartphone pourrait être plus adaptée en offrant des possibilités d'analyse supplémentaires par rapport aux instruments traditionnels.

Une explication détaillée de cet outil, appelé GEWMI, est disponible dans un article précédent, consacré uniquement à sa description [SJM 21]. De plus, en complément des mesures subjectives de la GEW, des données physiologiques ont été fournies par le bracelet Empatica E4, conçu pour les besoins de la recherche [ZEW 20, AKJ 21, ZPL 21]. Il comprend, outre un capteur de volume sanguin, un accéléromètre, un capteur de température cutanée et un capteur électrodermique (voir tableau 1). Ces capteurs sont synchronisés sur les mêmes horodatages. Il a été utilisé, entre autres, pour collecter des données physiologiques lors d'épisodes émotionnels induits, afin de construire des outils de reconnaissance automatique des émotions [ZWY 18]. Bien que moins précis qu'un moniteur Holter cardiaque, il a été démontré qu'il enregistrait des données comparables pendant 85 % du temps expérimental, tout en étant moins invasif [MPR 16].

Capteurs	Fréquence d'échantillonnage	Caractéristiques mesurées
Accéléromètre	32Hz	Accélération sur 3 axes en m.s ⁻²
Capteur de débit sanguin	64 Hz	Puissance nW
Capteur électrodermal	4 Hz	Activité électrodermale µS
Capteur thermique infrarouge	4 Hz	Température de la peau (°C)

Tableau 1. Données enregistrées avec le système Empatica E4.

3.3. Protocole Expérimental

Des expériences ont été menées pour acquérir des données de capteur de mouvement et de frappe de clavier de smartphone tout en induisant des émotions avec des clips vidéo. Une application mobile a été développée dans laquelle un chatbot interagit avec l'utilisateur et lui impose une séquence fixe de tâches. Elle est accessible à l'adresse suivante : <https://github.com/Orange-OpenSource/GEWMI>. Une vidéo est d'abord présentée, puis l'utilisateur est invité à la décrire et à annoter ses émotions ressenties à l'aide du GEWMI (voir Figure 2).

Une sélection de 12 clips vidéo et leurs étiquettes émotionnelles associées ont été choisis dans la base de données de clips conçue par Samson *et al* [SKS 16]. Il y a quatre extraits vidéo de valence positive, quatre de valence neutre et quatre de valence négative. Ces vidéos durent environ 30 secondes et représentent des scènes filmées par des amateurs, collectées en ligne. Les clips vidéo à valence négative présentent des accidents avec des blessures explicites. Les clips vidéo à valence positive représentent des situations amusantes avec des bébés ou des animaux. Les clips vidéo à valence neutre représentent des situations banales telles que la préparation d'un plat dans la cuisine ou la visite d'un musée. Voir le tableau 2 pour la liste complète.

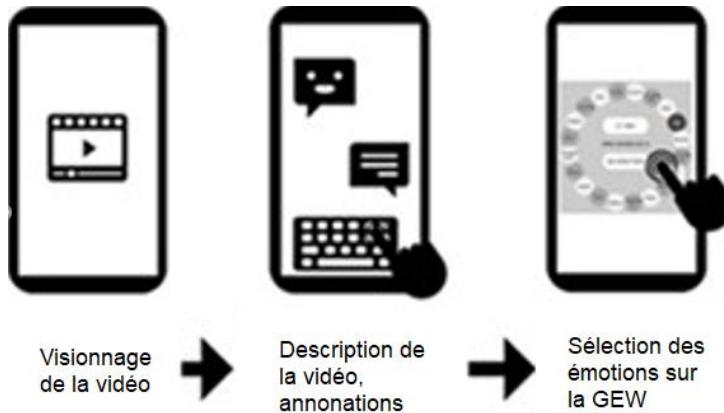


Figure 2. Protocole expérimental. Ces actions sont répétées 12 fois avec des vidéos différentes.

Les sujets peuvent sélectionner jusqu'à deux étiquettes d'émotion et l'intensité correspondante. Il est possible de n'en sélectionner aucune ou de créer une nouvelle étiquette en utilisant un élément appelé « autre ». Les mêmes étapes sont répétées plusieurs fois. Le type de vidéo est alterné pour éviter les biais. De plus, un autre clip vidéo neutre a été sélectionné pour créer une tâche d'entraînement et s'assurer que l'utilisateur comprend toutes les étapes.

Thème de la vidéo	Condition	Valence (1 to 7)
Garçon renversé par une voiture	négative	1,60
Morsure par un crocodile	négative	1,77
Ruade d'un cheval contre un homme	négative	2,35
Patineur qui se casse le bras	négative	1,68
Bébé qui lève les bras	positive	5,32
Bébé qui fait une grimace drôle	positive	5,01
Chat qui sort sa langue	positive	4,85
Panda qui éternue beaucoup	positive	4,68
Ouvrier qui trie des boîtes	neutre	3,04
Des gens sont assis sur une terrasse	neutre	3,29
Personne qui prépare le repas	neutre	3,37
Personnes qui visitent un musée	neutre	3,28
Train qui arrive en gare	entraînement	3,06

Tableau 2. Sélection des clips vidéo dans la base de données de Samson et al [33].

Capteurs	Fréquence d'échantillonnage	Caractéristiques mesurées
Accéléromètre	50 Hz	Accélération 3 axes m/s ²
Gyroscope	50 Hz	Rotation 3 axes (rad/s)
Magnétomètre	50 Hz	champ magnétique 3 axes (G)

Tableau 3. Données enregistrées grâce à l'application mobile.

Au total, 40 participants ont complété notre expérience dans les laboratoires d'Orange Labs, 21 femmes et 19 hommes, âgés de 18 à 64 ans. L'ensemble de l'expérience a été réalisé sur un Huawei P Smart (P8 lite 2018). Il est important de noter que les participants étaient assis loin des tables pour s'assurer qu'ils tenaient le smartphone dans leur main. Au cours de chaque expérience, les données des capteurs inertIELS et de frappe avec les horodatages associés à chaque étape de la tâche ont été collectées et stockées sur un serveur (voir tableau 3). D'autres données ont été enregistrées : âge, sexe, habitudes d'utilisation.

3.4. Résultats de l'expérience avec la GEW

Le rôle des mesures obtenues avec la GEW est de déterminer la valence de chaque vidéo en fonction de chaque participant. On suppose dans cette étude que les participants ne font pas d'erreurs dans le choix des étiquettes qui correspondent à leurs sentiments. Il est important de noter que même si la valence est censée être connue à partir des étiquettes de la base de données, chaque participant peut avoir un sentiment subjectif qui peut être différent de celui de la base de données. Le premier enjeu est donc d'analyser la distribution des labels, leur intensité et leur valence. Les résultats sont présentés dans la Figure 3. Pour rappel, les étiquettes de valence positives sont celles du côté droit de la GEW et les étiquettes de valence négatives sont celles du côté gauche. Plusieurs observations peuvent être faites. Premièrement, de manière générale, l'analyse de la valence montre une expression de sentiments subjectifs cohérents avec les étiquettes des clips vidéo.

Cependant, le nombre d'étiquettes déclarées avec une valence positive est plus important que prévu. Une analyse plus détaillée montre que le choix d'étiquettes positives dans des conditions a priori négatives est largement représenté par l'étiquette « compassion ». Cette étiquette est ambiguë. En effet, la compassion est considérée comme une émotion positive dans le sens où elle correspond à un sentiment de soutien à autrui, alors même que la situation est ressentie comme négative par la personne qui subit. Il y a certainement un biais dans la catégorisation de l'étiquette de compassion, mais comme elle est considérée comme une émotion positive dans les modèles émotionnels, nous avons décidé de la conserver dans cette catégorie.

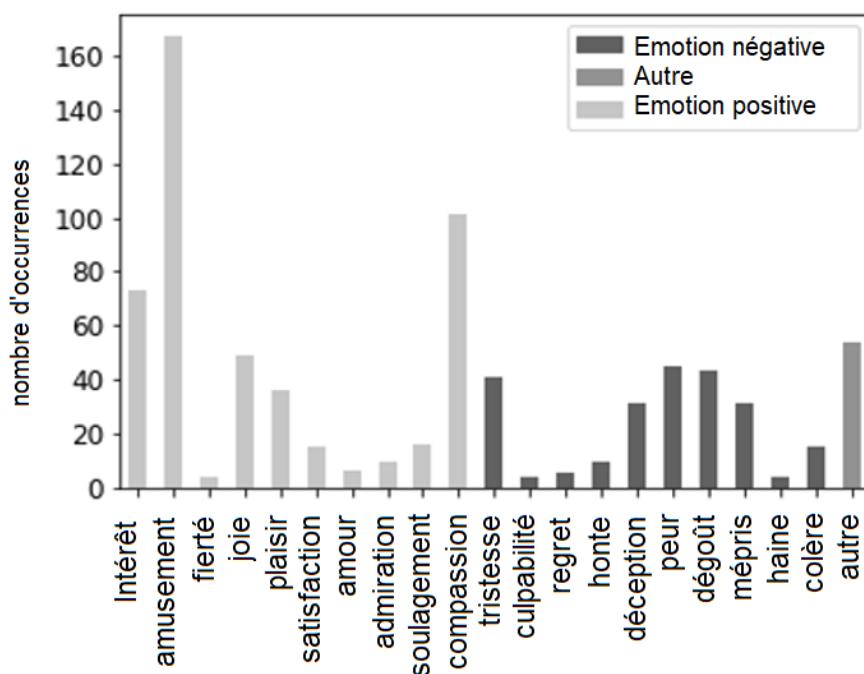


Figure 3. Distribution des étiquettes.

Enfin, « fierté », « amour » et « culpabilité » ont très peu d'occurrences (moins de 10). Cela peut s'expliquer par la nature externe du stimulus, qui n'implique pas directement le sujet. Au contraire,

certaines émotions exprimées comme « intérêt », « amusement », « compassion » ou « tristesse », ont généré de nombreuses occurrences.

4. Classification automatique des émotions

4.1. Méthodologie

Pour l'apprentissage et la classification des émotions, la période la plus intéressante correspond à la tâche de description qui suit le déclenchement de l'émotion par les vidéos. Pendant cette période, les données de mouvement et de frappe sont disponibles. Une question importante est de définir les résultats qui doivent être appris en fonction des étiquettes émotionnelles sélectionnées. Chaque étiquette pourrait être utilisée comme une classe, mais le nombre d'échantillons est trop petit et la variabilité est trop élevée pour une telle approche. De plus, plusieurs étiquettes ont parfois été sélectionnées deux fois (jusqu'à deux étiquettes pour chaque vidéo) et parfois la valence est positive pour la première étiquette et négative pour la seconde. Comme déjà mentionné, l'étiquette « compassion » est considérée comme positive mais il existe une ambiguïté dans son interprétation. Par souci de simplicité, il est proposé ici de ne considérer que deux classes. La première classe est appelée « valence négative ». Ceci est attribué aux vidéos pour lesquelles l'utilisateur a sélectionné au moins une étiquette avec une valence négative selon la roue des émotions de Genève (partie gauche de la roue). La deuxième classe est appelée « valence non négative ». Ceci est attribué à toutes les autres vidéos. Sur la base de cette stratégie d'étiquetage, 6 018 échantillons ont été étiquetés comme non négatifs et 5 502 ont été étiquetés comme négatifs.

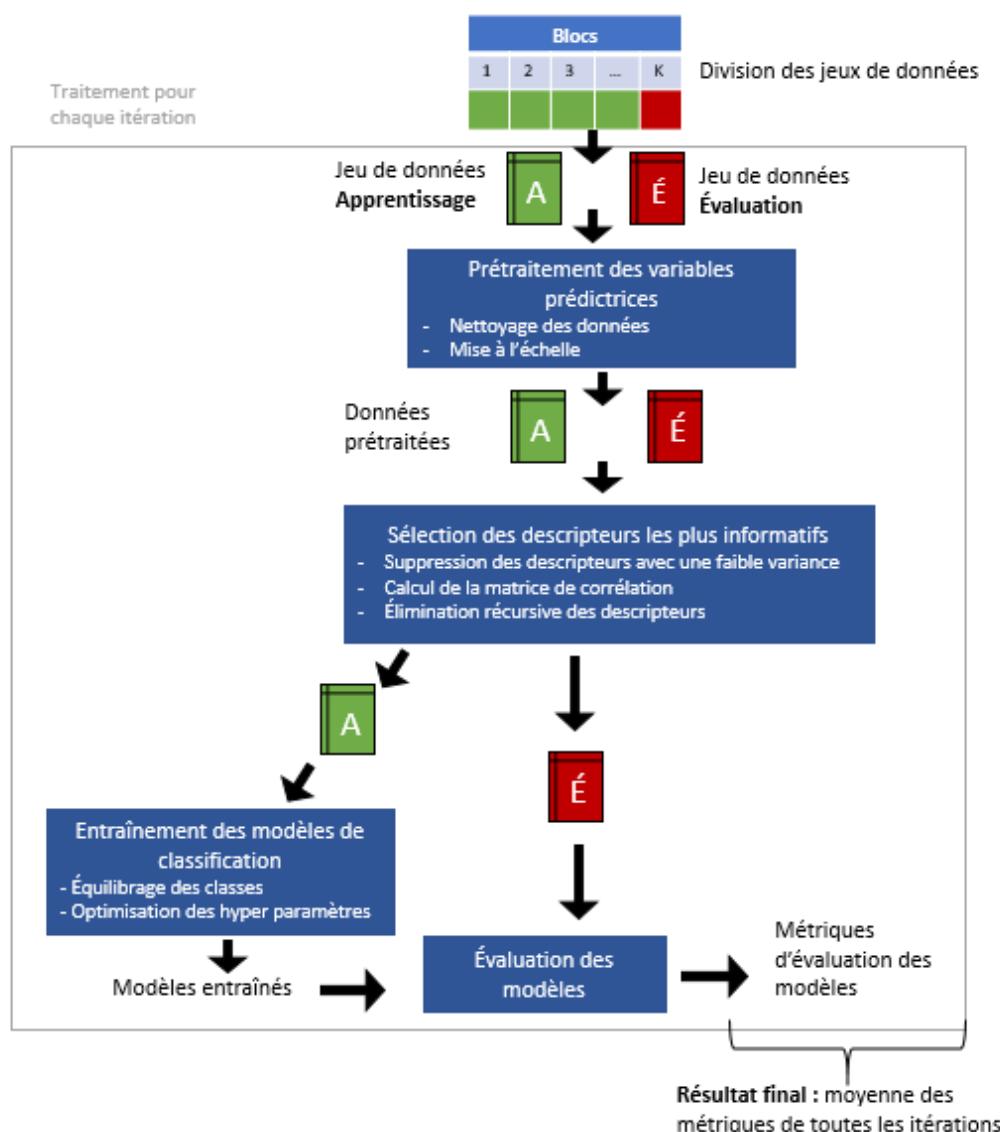


Figure 4. Méthode générale pour l'apprentissage supervisé et la classification.

Deux problèmes distincts peuvent alors être abordés.

- Classification de la valence issue d'un entraînement individuel : les données d'une seule personne peuvent être utilisées pour l'entraînement afin d'adapter la classification au profil de la personne.
- Classification de la valence issue de la formation globale : à condition que la réponse émotionnelle soit similaire pour un grand nombre d'individus, les données de tous les participants peuvent être utilisées pour la formation afin de voir si la classification des émotions peut être généralisée.

La figure 4 illustre l'ensemble du processus d'apprentissage automatique et les opérations appliquées à chaque ensemble de données pour obtenir les performances des classificateurs. L'ensemble de données global est séparé en deux sous-ensembles, un ensemble de données d'entraînement et un ensemble de données d'évaluation (« T » et « E » dans la figure 4). Chaque étape est présentée dans les parties suivantes.

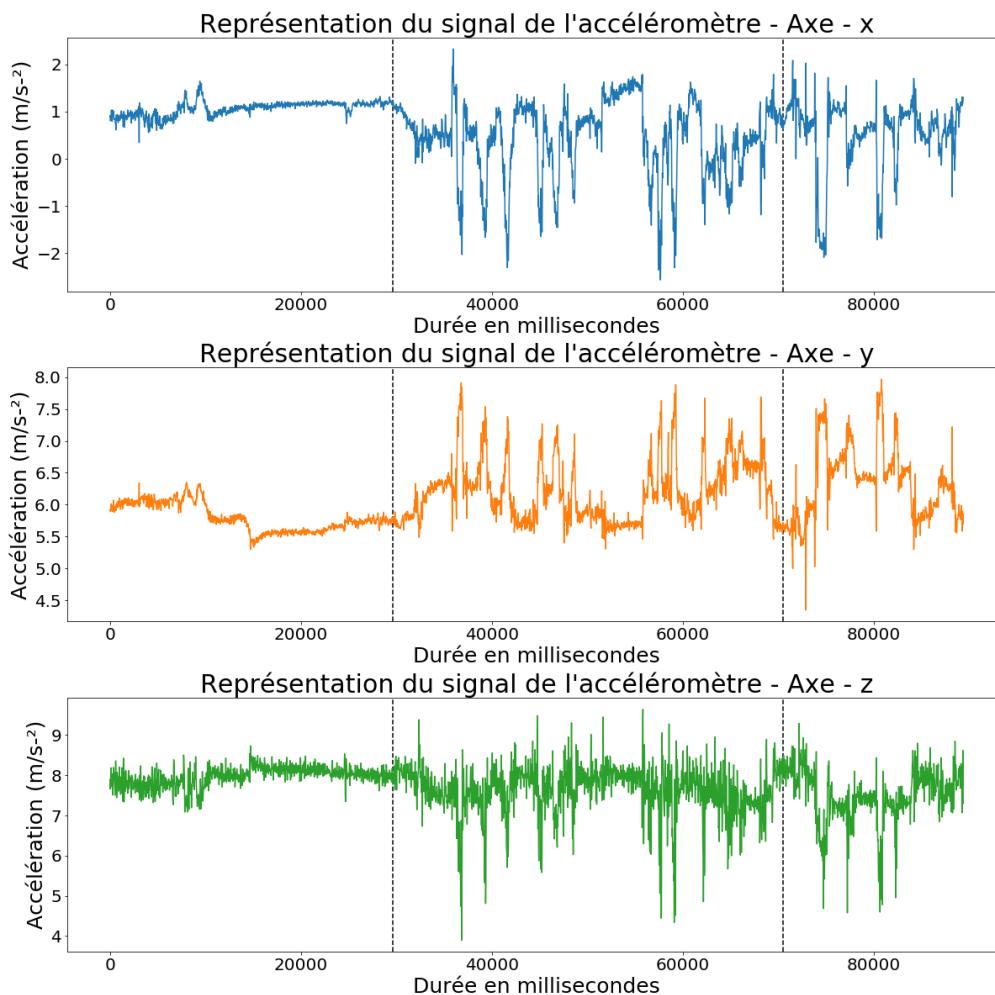


Figure 5. Données de l'accéléromètre du smartphone.

4.2. Traitement du capteur de mouvement

Des échantillons de données d'accéléromètre sont présentés figure 5. Ces séries chronologiques représentent le signal brut de l'accéléromètre selon les trois axes (x, y et z) au cours d'une expérience. Elles comprennent les phases de visualisation, de description et d'auto-évaluation. Visuellement, la différence d'amplitude entre les différentes phases peut être facilement observée. Dans la phase de visualisation, le signal est relativement plat et, dans les phases suivantes, il y a de nombreux pics, qui sont causés par l'écriture des participants et par la sélection des étiquettes sur la GEW. Pour l'apprentissage automatique, les données brutes ne sont pas appropriées et il est préférable d'extraire les fonctionnalités les plus pertinentes. L'ensemble du processus peut être résumé par les étapes suivantes : prétraitement, rééchantillonnage, extraction de caractéristiques.

1) Prétraitement

Trois filtres ont été appliqués aux données du capteur de mouvement à l'aide de la bibliothèque Python SciPy [VGO 20] :

- pour réduire le bruit du signal et supprimer les pics isolés, un filtre médian a été utilisé avec une taille de fenêtre de trois [KSS 20] ;
- deuxièmement, l'application d'un filtre Butterworth passe-haut à 0,3Hz a permis de supprimer l'effet de la pesanteur ;
- enfin, afin de réduire davantage le bruit, un filtre passe-bas Butterworth de troisième ordre à 20 Hz a été appliqué. Selon Anguita et al. [AGO 13], 99 % des informations sur le mouvement humain sont en effet détectables en dessous de 15 Hz.

2) Rééchantillonnage

L'échantillonnage consiste à découper les signaux en intervalles réguliers. Une exigence importante est de déterminer les intervalles de temps les plus appropriés des échantillons qui vont alimenter la méthode d'apprentissage automatique. Notre hypothèse est que les émotions peuvent avoir un impact sur l'intensité et la fréquence des mouvements, mais la variabilité pourrait être très élevée. De grandes fenêtres temporelles pourraient donc être nécessaires pour détecter et identifier les caractéristiques associées à la réponse émotionnelle. Afin de capturer suffisamment d'informations, des échantillons de 10 secondes ont été choisis (500 mesures) avec un décalage de 2,5 secondes (125 mesures) permettant un chevauchement de 75 % entre 2 échantillons consécutifs, ce qui a été recommandé par Yonggang Lu *et al* pour optimiser la quantité de données utilisées [LWL 17].

3) Extraction de caractéristiques

Un ensemble de caractéristiques est calculé pour chaque échantillon afin de constituer un ensemble de données. Les caractéristiques ont été divisées en trois catégories : caractéristiques du domaine temporel, caractéristiques du domaine fréquentiel et caractéristiques des transformées en ondelettes.

Parmi ces caractéristiques, une information importante est le "jerk" (à-coup), qui correspond à la première dérivée du signal mesuré par l'accéléromètre [AGO 13]. Les caractéristiques de temps et de fréquence ont été calculées à l'aide des bibliothèques NumPy et SciPy [VGO 20, WCV 11]. Ces calculs ont été appliqués aux signaux des capteurs accéléromètre, gyroscope et magnétomètre, pour chaque axe. Ces informations sont utilisées dans la recherche sur la reconnaissance de l'activité humaine à partir d'enregistrements de capteurs de mouvement sur les smartphones [AGO 13] et plus récemment dans la reconnaissance des émotions [KSS 20, CLZ 15, MH 16, RCH 20].

De plus, selon Edmond Mitchell *et al.* [MMO 13], la transformée en ondelettes est particulièrement adaptée à l'extraction de caractéristiques informatives à partir de signaux de capteurs de mouvement, car elle s'adapte aux séries temporelles discontinues. Les caractéristiques associées aux ondelettes ont été calculées à l'aide de la bibliothèque PyWavelets [LGW 19]. Suivant la méthodologie proposée par Yufei Chen et Shen [CS 17], nous avons appliqué une transformée en ondelettes avec une forme d'onde Daubechies 4 sur cinq bandes de fréquences : 0,3-1,25 Hz, 1,25-2,5 Hz, 2,5-5 Hz, 5-10 Hz et 10-20 Hz. Pour chaque bande de fréquence, l'énergie de fréquence a été calculée.

4.3. Traitement des informations de frappe au clavier

Les caractéristiques de frappe sont calculées à partir des événements du clavier toutes les 10 secondes, selon la fenêtre d'échantillonnage. Chaque événement de clavier était horodaté à la milliseconde près. Surjya Ghosh *et al* [GGM 17a] définissent la distance "inter-tap" comme l'intervalle de temps entre deux événements clavier consécutifs. Cette mesure temporelle a été utilisée pour calculer un ensemble de paramètres complémentaires permettant de caractériser chaque échantillon de 10 secondes au cours de la phase de description :

- le nombre de caractères dans l'échantillon ;
- la moyenne inter-tap ;
- le maximum inter-tap ;
- le minimum inter-tap ;
- l'écart-type inter-tap.

4.4. Sélection des caractéristiques

1) Prétraitement

Les valeurs des caractéristiques ont été mises à l'échelle entre 0 et 1 à l'aide d'une procédure de normalisation standard.

2) Sélection des caractéristiques

Afin de ne retenir comme variables d'entrées de l'apprentissage que les caractéristiques les plus pertinentes, une sélection a été opérée selon les critères courants du domaine :

- en suivant la méthodologie proposée par Ruensuk *et al*, les caractéristiques à faible variance ont été éliminées [RCH 20] ;
- une matrice de corrélation a été générée en calculant d'abord les corrélations entre les caractéristiques, puis la corrélation entre chaque caractéristique et chaque classe ;
- si deux caractéristiques avaient une corrélation supérieure à 80 %, seule la plus corrélée avec la classe était conservée ;
- enfin, un algorithme d'élimination des caractéristiques de façon récursive (RFE) a été utilisé pour conserver uniquement les caractéristiques les plus pertinentes pour le processus de classification. La fonction de la bibliothèque Python Scikit-learn [PVG 11] est basée sur un algorithme développé par Guyon *et al* [GWB 02]. L'algorithme RFE a été utilisé avec une régression logistique, jusqu'à ce qu'il ne reste plus que la moitié des caractéristiques initiales.

Nom de la caractéristique et description	Coefficient de corrélation
Gyr_jerk_t_xy_corr : Corrélation du jerk entre les mesures X et Y du gyromètre	0.611
Gyr_dwt_y_energy_4 : Energie de fréquence du gyromètre, ondelette du 4 ^{ème} niveau, axe Y	0.594
Gyr_jerk_f_yz_corr : Corrélation du jerk entre les mesures Y et Z du gyromètre	0.587
Acc_dwt_y_energy_3 : Energie de fréquence de l'accéléromètre, ondelette de niveau 3, axe Y	0.517
Acc_f_yz_corr : Fréquence de l'accéléromètre, corrélation Y et Z	0.508
Acc_dwt_y_energy_2 : Energie de fréquence de l'accéléromètre, ondelette de niveau 2, axe Y	0.505
Acc_jerk_f_skewness : Fréquence de l'accéléromètre, asymétrie du jerk	0.504
Gyr_jerk_t_y_mad : Gyromètre, jerk médian, axe Y	0.470
Mag_jerk_f_z_entropy : Magnétomètre, entropie de la fréquence de jerk, axe Z	0.427
Gyr_jerk_t_xz_corr : Gyromètre, corrélation du jerk, axes X et Z	0.420

Tableau 4. Liste des 10 caractéristiques les plus corrélées à la classification recherchée.

Le tableau 4 indique les 10 variables les plus corrélées avec une classe recherchée pour les échantillons associés à une seule personne. Il est difficile d'interpréter les résultats obtenus. Ce qui est espéré, c'est que les mouvements de la main soient corrélés aux émotions. Par exemple, la fréquence des oscillations de la main peut être plus ou moins rapide selon l'émotion, ou la fréquence de ces oscillations peut fluctuer plus ou moins fortement selon l'émotion. Or, précisément, plusieurs caractéristiques associées au jerk, c'est-à-dire la dérivée de l'accélération, en translation ou en rotation, sont dans la liste des 10 variables les plus corrélées avec les classes. Cette première étape est donc prometteuse.

4.5. Apprentissage

Comme le nombre d'échantillons dans chaque classe n'est pas égal, il existe un biais potentiel dans le processus d'entraînement. Une technique SMOTE (Synthetic Minority Over-sampling Technique) a donc été mise en œuvre pour suréchantillonner la classe f minoritaire [CBH 02]. Pour chaque session d'apprentissage, nous avons utilisé la méthode de classification naïve "Scikit-Learn" comme référence pour la comparaison. Cette méthode est simple : le résultat de la classification est déterminé par l'étiquette la plus fréquente dans l'ensemble d'apprentissage. Huit classificateurs d'apprentissage automatique ont ensuite été utilisés sur la base des algorithmes de la bibliothèque Python Scikit-Learn [PVG 11] : Adaboost, Arbre de décision, Gradient Boosting, K Voisins, Support Vector Machine, Perceptron, Régression logistique et Réseau bayésien.

4.6. Classification individualisée

L'apprentissage peut être réalisé avec toutes les données de tous les individus, ou un apprentissage individualisé peut être mis en œuvre. C'est le deuxième cas qui est discuté ici. L'apprentissage a été testé séparément pour chaque personne en exploitant uniquement les données qui lui étaient associées. En entrées, nous avons les caractéristiques retenues et en sortie, nous tentons d'apprendre l'émotion positive ou négative de la personne, qui est supposée être déterminée lors du processus de labellisation à l'aide de la roue des émotions. Comme il y avait 4 vidéos avec une valence négative et 4 vidéos avec une valence non négative présentées à chaque personne, 8 périodes de temps correspondant à la tâche de description et d'étiquetage ont pu être utilisées pour l'apprentissage. Afin d'éviter la corrélation des échantillons (chevauchement), les classificateurs ont été entraînés avec des échantillons associés à 6 périodes et testés à l'aide d'échantillons distincts associés aux 2 périodes restantes. Pour éviter tout biais possible, des échantillons de chaque période ont été utilisés comme ensembles de données d'entraînement et de test, en alternance (validation croisée). Dans certains cas, comme le participant n'a pas choisi d'étiquette avec une valence négative pendant toute la session, les ensembles de données correspondants n'ont pas été utilisés pour l'entraînement. Des ensembles de données provenant de 38 sujets ont été utilisés pour l'identification des émotions à partir de l'entraînement individuel. Ensuite, la précision moyenne des classifications a été calculée (validation croisée). Le tableau 5 indique les résultats de la classification. En moyenne, trois méthodes surpassent la précision de classification naïve de plus de 10 points de pourcentage : le classificateur de gradient boosting avec une précision de 69,89 %, l'Adaboost avec 69,35 % et l'arbre de décision avec 67,73 %. Au niveau individuel, pour 19 personnes sur 38, les classificateurs dépassent la précision de la classification naïve de plus de 10 points de pourcentage.

Méthode de classification	Précision moyenne	Comparaison avec la classification naïve
Classification naïve	56.04 %	0 %
Gradient Boosting	69.89 %	13.85 %
AdaBoost	69.35 %	13.31 %
Arbre de décision	67.73 %	11.69 %
Régression logistique	65.89 %	9.85 %
SVM	65.31 %	8.82 %
Perceptron	65.02 %	9.27 %
Réseau bayésien	63.36 %	7.79 %
K voisins	61.40 %	5.36 %

Tableau 5. Précision moyenne des apprentissages individualisés.

4.7. Classification généralisée

Dans le cas de la classification généralisée, les échantillons de tous les sujets et les labels associés sont exploités pour l'apprentissage des émotions. Pour éviter la dépendance entre les échantillons et ainsi éviter le surapprentissage, une méthode de validation croisée Leave-One-Subject-Out a été mise en œuvre. Tous les ensembles de données des sujets, sauf un, ont été utilisés pour entraîner le classificateur, et l'ensemble de données du sujet restant a été utilisé pour le test. La méthode a été répétée afin que l'ensemble de données de chaque sujet soit utilisé pour l'entraînement et les tests. Par conséquent, 40 sessions d'apprentissage ont été mises en place, une par sujet. Pour chaque session, la précision de chaque classificateur a été calculée. La précision des classificateurs et les différences avec la classification naïve sont présentées dans le tableau 6.

Méthode de classification	Précision moyenne	Comparaison avec la classification naïve
Classification naïve	51.46 %	0 %
K Voisins	56.73 %	5.27 %
Arbre de décision	56.64 %	5.18 %
Perceptron	54.88 %	3.43 %
Gradient boosting	53.93 %	2.47 %
SVM	53.93 %	2.48 %
Adaboost	53.56 %	2.10 %
Logistic regression	53.52 %	2.06 %
Réseau bayésien	52.69 %	1.22 %

Tableau 6. Précision moyenne des apprentissages généralisés.

La classification naïve a une précision de 51,46 %, ce qui correspond à la classification de tous les échantillons dans la classe la plus fréquente (c'est-à-dire la classe non négative). Les autres méthodes ont réussi à atteindre une précision d'au moins 1,22 point de pourcentage supérieure à la précision de la classification naïve.

4.8. Identification des émotions avec le bracelet Empatica E4

4.8.1. Apprentissages individualisés

Afin d'approfondir l'analyse, nous avons effectué des classifications de valence à partir des données du bracelet Empatica E4. Les signaux enregistrés sur cet appareil sont présentés dans le Tableau 1. Les caractéristiques physiologiques ont été calculées à partir du capteur Empatica E4 EDA et du pouls de volume sanguin (BVP), selon la méthode proposée par Schmidt *et al* [SRD 18]. Les caractéristiques de variation de la fréquence cardiaque (VFC) ont été calculées à partir du signal BVP. De plus, les caractéristiques de la réponse de conductance cutanée (SCR) et du niveau de conductance cutanée (SCL) ont été calculées à partir du signal EDA. Les caractéristiques de mouvement du signal d'accélération ont été traitées de la même manière que l'accélération du smartphone.

Nous avons effectué une classification individualisée avec les données de l'Empatica E4 (voir Tableau 7). En raison de problèmes d'enregistrement, seuls 34 sujets disposaient de suffisamment de données pour l'apprentissage. La classification naïve a une précision de 55,77 %. Toutes les autres méthodes obtiennent en moyenne 10 points de pourcentage de plus que la classification naïve, à l'exception de l'arbre de décision. La méthode la plus performante est la régression logistique avec une précision moyenne de 72,89 %. En moyenne, les performances des classifications entraînées avec les données Empatica E4 sont supérieures aux classifications effectuées sur les données du smartphone, mais le résultat peut être biaisé par les ensembles de sujets, qui sont différents.

Sur 34 sujets, 26 avaient au moins un classificateur avec une précision de 10 points de pourcentage supérieure à la classification naïve. En comparant avec les résultats obtenus avec les données d'interaction sur le smartphone, nous avons identifié les profils suivants :

- pour 13 sujets, les méthodes de classification de valence ont une précision de 10% supérieure à la méthode naïve en considérant à la fois les données d'interaction avec le smartphone et les données du bracelet Empatica E4 ;
- pour 13 sujets, les méthodes de classification de valence obtiennent une précision de 10% supérieure à la méthode naïve, lorsque sont considérées uniquement les données Empatica E4 ;
- pour 6 participants, les méthodes de classification de valence obtiennent une précision de 10% supérieure à la méthode naïve, en considérant uniquement les données d'interaction avec le smartphone.

De façon remarquable, ces résultats sont comparables à ceux obtenus dans des études similaires avec un contexte différent [BAR 16, ZEW 20, AKJ 21].

Méthode de classification	Précision moyenne	Comparaison avec la classification naïve
Classification naïve	55.77%	0 %
Régression logistique	72.89%	17.12%
SVM	70.81%	15.04%
Adaboost	70.71%	14.94%
Perceptron	70.45%	14.67%
K voisins	68.88%	13.11%
Gradient boosting	68.51%	12.74%
Réseau bayésien	67.94%	12.16%
Arbre de décision	64.15%	8.38%

Tableau 7. Précision moyenne des apprentissages individualisés avec les données Empatica E4.

4.8.2. Apprentissage individualisé en exploitant les données de tous les capteurs

Afin de tester la capacité du système de reconnaissance des émotions à classer les modalités de valence à partir de sources de données multimodales, nous avons effectué des classifications de valence (négative ou non négative) en fusionnant les données des interactions avec les smartphones avec les mesures physiologiques. Les caractéristiques ont été calculées pour chaque échantillon et fusionnées avec les caractéristiques de mouvement et de frappe. Les données de 34 sujets étaient complètes et disponibles pour un apprentissage individualisé. Les résultats sont présentés dans le tableau 8.

Méthode de classification	Précision moyenne	Comparaison avec la classification naïve
Classification naïve	55.96%	0 %
Adaboost	73.62%	17.66%
Gradient boosting	73.02%	17.06%
Arbre de décision	71.83 %	15.87%
Régression logistique	69.96%	14.01%
SVM	68.82%	12.86%
Perceptron	67.80%	11.85%
Réseau bayésien	67.35%	11.39%
K voisins	63.83%	7.87%

Tableau 8. Précision moyenne des apprentissages individualisés sur les données multi-capteurs.

En moyenne, la classification naïve a une précision de 55,96 %. On peut observer une légère amélioration de la précision pour certaines méthodes, par rapport aux méthodes individualisées basées uniquement sur les données Empatica E4. En effet, en moyenne, le classificateur AdaBoost dépasse la classification naïve de 17,66 points de pourcentage.

4.9. Discussion

La méthode la plus performante dépend du problème considéré, du jeu de données, des paramètres choisis et parfois également d'une part d'aléatoire (par exemple les valeurs affectées initialement aux

poids du perceptron). Dans l'ensemble, il n'y a pas de méthode qui surclasse les autres, les résultats sont du même ordre de grandeur. De plus, la plupart des méthodes sont des boîtes noires qui ne facilitent pas l'explicabilité, d'autant plus que les variables du problème, comme par exemple une corrélation de jerk, sont elles-mêmes complexes et difficiles à appréhender.

De manière générale, apprendre la valence émotionnelle à l'aide d'échantillons de jeux de données de tous les sujets n'est pas très efficace. Le meilleur classificateur n'est que 5,27 % plus performant que la classification naïve, ce qui est à peine significatif. Il y a deux interprétations possibles. La première est que ces caractéristiques sont faiblement corrélées avec la valence émotionnelle. La seconde est que certaines corrélations peuvent exister, mais le profil des caractéristiques dépend fortement de chaque individu et la généralisation à toute la population n'est pas possible. En ce qui concerne l'apprentissage de la valence émotionnelle adapté à chacun, les résultats sont prometteurs. L'avantage de l'apprentissage individualisé par rapport à l'apprentissage général est cohérent avec la méta-analyse proposée par Kołakowska *et al* [KSS 20]. En effet, un apprentissage individualisé s'adapte mieux aux caractéristiques individuelles et réussit donc mieux à distinguer les caractéristiques les plus discriminantes pour chaque participant. Cependant, l'apprentissage individualisé requiert la collecte de données de chaque utilisateur, ce qui est fastidieux. Avec une précision jusqu'à 13 % supérieure à celle de la classification naïve en moyenne, les résultats sont prometteurs. Nous avons malgré tout remarqué une forte variabilité de la performance en fonction des individus. Ainsi, le déclenchement des émotions n'a pas forcément opéré avec la même efficacité pour tout le monde. De manière plus qualitative, nous avons pu observer visuellement le comportement des utilisateurs. Certains d'entre eux s'avèrent plus exubérants dans leurs gestes tandis que d'autres restent réservés tout au long de l'expérimentation. Les caractéristiques les plus importantes semblent être la secousse associée au gyromètre pour les axes X et Y, ainsi que la fréquence dans la bande 4 Hz. Il est difficile de décrire les mouvements associés à de telles caractéristiques. Ils peuvent correspondre à des mouvements rapides utilisant le poignet et le coude. L'analyse d'enregistrements vidéo pour filmer les individus tout au long de la séance pourrait s'avérer pertinente pour mieux comprendre les différences interindividuelles. D'éventuels paramètres importants, qui n'ont pas été enregistrés, pourraient être la taille de la main et la longueur du bras.

Selon R. A. Calvo et D'Mello [CD 10], les méthodes expérimentales pour déclencher des émotions pourraient souffrir d'un contexte insuffisamment engageant, bien que les clips vidéo soient considérés comme des moyens adaptés [WSS 96, FGT 15]. De plus, l'expression des émotions peut être restée limitée et peu détectable lors des interactions avec le smartphone. Un enjeu essentiel est l'évaluation de l'efficacité du déclenchement de l'émotion pour chaque sujet. Nous avons donc inclus des mesures physiologiques avec l'Empatica E4, afin d'augmenter le nombre de caractéristiques considérées. Il serait possible de combiner d'autres modalités pour améliorer la reconnaissance des émotions exprimées. Les expressions faciales ou les postures des participants ont pu être analysées à partir des enregistrements vidéo, afin de mieux identifier l'expression de l'émotion. Cependant, l'intégration de nombreuses modalités implique une complexité méthodologique et technique accrue. De plus, certaines données pourraient soulever des questions d'acceptabilité et de criticité concernant la vie privée.

Il est important de noter, selon Calvo et Mello [CD 10], qu'il n'est pas évident que la mesure de l'évolution d'une caractéristique émotionnelle implique des évolutions mesurables et synchronisées sur les autres caractéristiques. Par exemple, nous avons observé que pour certaines personnes, certaines sources de données semblaient plus efficaces que d'autres. De meilleurs résultats pourraient donc être obtenus si une pré-classification des individus était disponible afin de déterminer quelles sources de données doivent être considérées en priorité et de prédire les performances de la méthode. Néanmoins, une variabilité individuelle peut être observée en fonction de l'humeur ou de la fatigue de la personne. Enfin, nous avons remarqué que la fusion des données du smartphone et du bracelet Empatica E4 a conduit à une légère amélioration des résultats.

Dans notre cas, l'analyse de l'auto-déclaration obtenue avec le GEW a guidé notre étiquetage afin d'obtenir des classes de données pertinentes et équilibrées pour la dimension de valence. On peut

imaginer établir des classifications avec d'autres caractéristiques de la GEW comme le niveau d'intensité ou le nombre d'émotions rapportées, ou encore des étiquettes plus spécifiques.

Sur le plan technique, nous avons choisi un smartphone spécifique avec un écran spécifique. Une certaine variabilité peut être observée si différents appareils sont utilisés, en fonction de la taille de l'écran, de la configuration et de la taille du clavier, ou du type de capteurs. Plusieurs choix de traitement pourraient également influencer les performances des modèles de classification. Cui *et al* ont identifié la taille de la fenêtre d'échantillonnage comme un facteur majeur du résultat de performance, dans la reconnaissance des émotions avec les données des capteurs de mouvement [CLZ 15]. Dans le cas d'une interaction avec un smartphone, il serait alors pertinent de tester différentes fenêtres d'échantillons, dans le but d'établir une taille optimisée pour ce type de données. Le choix des caractéristiques est également un facteur déterminant, notamment au vu de la grande variété de celles qui sont calculables. Il pourrait être possible d'utiliser des modèles d'apprentissage profond, comme les réseaux neuronaux récurrents, pour générer automatiquement des caractéristiques sur de tels enregistrements.

En résumé, la variabilité interindividuelle est supérieure à la sensibilité des outils. Un problème fondamental est la variabilité de la réponse émotionnelle, qui dépend fortement du profil psychologique de chaque personne ainsi que de l'état physiologique et de la posture. Les variabilités intra et inter-individuelle pourraient donc être inévitables et l'identification de la valence émotionnelle resterait probablement toujours incertaine.

5. Conclusion

Une expérience a été menée pour évaluer la faisabilité de l'identification de la valence des émotions à l'aide de capteurs de smartphone et de bracelet. Ce travail est préalable à un autre, plus fourni, comportant une grande masse de données. Plusieurs conclusions peuvent être tirées.

- En ce qui concerne l'induction d'émotions, des valences positives et négatives ont été enregistrées avec succès à l'aide de vidéos et d'une version adaptée de la SME (GEWMI), qui a été mise en œuvre sur les smartphones. Bien que les étiquettes émotionnelles proposées par les utilisateurs étaient en général cohérentes avec la valence a priori des vidéos, plusieurs difficultés ont été rencontrées : a) La valence de l'étiquette « compassion » est ambiguë car elle s'adresse à une autre personne et fait référence à un comportement social ; b) Comme les événements associés aux vidéos n'impliquent pas l'utilisateur lui-même, les sentiments émotionnels étaient principalement dus à l'empathie. Par conséquent, plusieurs étiquettes n'ont jamais été sélectionnées et l'intensité émotionnelle est restée faible.
- En ce qui concerne l'identification automatique de la valence émotionnelle à l'aide de capteurs de smartphones et de bracelets au moyen de techniques d'apprentissage automatique, les résultats montrent que, en général, ce n'est pas possible. Il semble que chaque utilisateur ait sa propre réponse émotionnelle et qu'il ne soit pas possible de généraliser.
- Si l'identification automatique de la valence émotionnelle est limitée à un seul individu et si l'apprentissage automatique est basé uniquement sur les ensembles de données de capteurs associés à cet individu, de meilleurs résultats sont obtenus. Le classificateur AdaBoost dépasse le modèle de référence de 17,66 points de pourcentage, ce qui signifie que le taux de reconnaissance est d'environ deux tiers.
- Dans l'ensemble, nous observons une forte variabilité entre les sujets de l'expérience. Plus en détail, lorsque l'on compare les classifications des données d'interaction avec le smartphone ou le bracelet Empatica E4, certaines personnes obtiennent de meilleurs résultats sur certaines sources de données que sur d'autres. Cela démontre la complexité de la fusion de données multimodales et, par extension, du suivi de plusieurs composantes d'un épisode émotionnel. Cependant, pour une catégorie particulière d'individus, la multimodalité semble être l'une des approches les plus efficaces pour augmenter la précision des modèles de reconnaissance des émotions.

- Nous n'avons pas tenté de combiner les méthodes de classification comme cela est souvent suggéré dans la littérature [KBN 24], car cela n'aurait sans doute pas permis d'améliorer les résultats de façon notable. Lorsqu'aucune méthode ne parvient à identifier une émotion sur un jeu de données, la combinaison de plusieurs méthodes ne peut apporter de gain significatif. Le problème fondamental est que l'information permettant de discriminer les émotions est trop approximative et incertaine.

D'un point de vue méthodologique, plusieurs perspectives sont intéressantes. Tout d'abord, l'utilisation de vidéos pour déclencher des émotions est considérée comme pertinente, mais les sujets sont placés dans une situation passive et ne sont pas très engagés dans la tâche. Nous avons essayé de compenser cela en incluant une tâche de description. Cependant, il serait préférable de définir des protocoles dans lesquels les sujets sont plus actifs et impliqués dans la tâche, la difficulté étant de pouvoir provoquer des états émotionnels de manière contrôlée. Deuxièmement, la réponse émotionnelle dépend fortement de chaque individu, mais il est possible de trouver des catégories d'utilisateurs qui réagissent de la même manière. Il peut s'agir d'informations sur le profil psychologique, l'âge, le sexe, la profession, ainsi que d'informations physiques telles que la taille des membres et la posture. Troisièmement, la capacité de calcul des smartphones augmente chaque année. Dans un avenir proche, il sera peut-être possible d'intégrer dans l'appareil un système d'apprentissage automatique et de classification des émotions de l'utilisateur.

En plus des composantes physiologiques des émotions, il est également possible de prendre en compte les informations sémantiques. Par exemple, si un client se plaint du fonctionnement d'un appareil spécifique qui vient d'être acheté, il y a souvent de la frustration, voire de la colère. Dans un tel cas, les données des capteurs pourraient être utilisées pour corroborer cette interprétation. Combiner les données d'interaction du smartphone avec le traitement du langage naturel pourrait alors être une approche intéressante pour améliorer les systèmes de reconnaissance automatique des émotions et fournir de meilleurs résultats.

Références

- [AGO 13] D. Anguita, A. Ghio, L. Oneto, X. Parra and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," proceedings of the 21st European Symposium on Artificial Neural Networks, Bruges , April 24-26, 2013.
- [AKJ 21] F.A. Alskafi, A.H. Khandoker & H.F. Jelinek, A comparative study of arousal and valence dimensional variations for emotion recognition using peripheral physiological signals acquired from wearable sensors. Proc. of the 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 1104-1107, Nov. 2021.
- [ARN 60] M. B. Arnold, Emotion and personality, Psychological aspects, Vol. 1, Columbia University Press, 1960.
- [BAR 16] L. F. Barrett, "Navigating the science of emotion," in Emotion measurement," in Emotion measurement, Elsevier, p. 31–63, 2016.
- [BL 94] M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," Journal of behavior therapy and experimental psychiatry, vol. 25 (1), p. 49–59, 1994.
- [BRJ 18] C. Bröhl, P. Rasche, J. Jablonski, S. Theis, M. Wille and A. Mertens, "Desktop PC, Tablet PC, or Smartphone? An Analysis of Use Preferences in Daily Activities for Different Technology Generations of a Worldwide Sample," International Conference on Human Aspects of IT for the aged population, p. 3–20., 2018.
- [BVV 14] I. Bakker, T. v. d. Voordt, P. Vink and J. d. Boon, "Pleasure, arousal, dominance: Mehrabian and Russell revisited", Current Psychology, vol. 33, no. 3, p. 405–421, 2014.
- [CBF 17] E. Cambria, D. D. S. Bandyopadhyay and A. Feraco, "A practical guide to sentiment analysis", Springer, 2017.
- [CBH 02] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique", Journal of artificial intelligence research, vol. 16, pp. 321-357, 2002.
- [CD 10] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications", IEEE Transactions on affective computing, vol. 1, pp. 18-37, 2010.

- [CLZ 15] L. Cui, S. Li, W. Zhang, Z. Zhang and T. Zhu, "Identifying emotion from natural walking," arXiv, vol. abs/1508/00413, 2015.
- [CS 17] Y. Chen and C. Shen, "Performance analysis of smartphone-sensor behavior for human activity recognition", IEEE Access, vol. 5, p. 3095–3110, 2017.
- [DKB 20] A. Dzedzickis, A. Kaklauskas and V. Bucinskas, "Human emotion recognition: Review of sensors and methods", Sensors, vol. 20 (3), p. 592, 2020.
- [EF 71] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," Journal of personality and social psychology, vol. 17, no. 2, p. 124, 1971.
- [ELM 11] C. Epp, M. Lippold and R. L. Mandryk, "Identifying emotional states using keystroke dynamics," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, p. 715–724, 2011.
- [FGT 15] R. A. Ferrer, E. G. Grenen et J. M. Taber, « Effectiveness of internet-based affect induction procedures: A systematic review and meta-analysis», Emotion, vol. 15, p. 752, 2015.
- [GIL 94] S. Gil, «Comment étudier les émotions en laboratoire», vol. 4, p. 15–24, 1994.
- [GGM 17a] S. Ghosh, N. Ganguly, B. Mitra and P. De, "Evaluating effectiveness of smartphone typing as an indicator of user emotion", proceedings of the 7th International Affective Computing and Intelligent Interaction (ACII) IEEE, p. 146–151, 2017.
- [GGM 17b] S. Ghosh, N. Ganguly, B. Mitra and P. De, "Tapsense: combining self-report patterns and typing characteristics for smartphone based emotion detection", Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services. ACM, p. 2, 2017.
- [GHG 19] S. Ghosh, K. Hiware, N. Ganguly, B. Mitra and P. De, "Emotion detection from touch interactions during text entry on smartphones", International Journal of Human-Computer Studies, 2019.
- [GJR 95] J. J. Gross, O. John and J. Richards, "Berkeley expressivity questionnaire," Edwin Mellen Press, 1995.
- [GWB 02] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, "Gene selection for cancer classification using support vector machines", Machine learning, vol. 46 (1), p. 389–422, 2002.
- [KBN 24] S.K. Khare, V. Blanes-Vidal, E.S. Nadimin and U.R. Acharya, Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations, Information Fusion, Volume 102, ref. 102019, February 2024.
- [KH 99] D. Keltner and J. Haidt, "Social functions of emotions at four levels of analysis," Cognition & Emotion, vol. 13, no. 5, p. 505–521, 1999.
- [KO 18] B. C. Ko, "A brief review of facial emotion recognition based on visual information," Sensors, vol. 18 (2), p. 401, 2018.
- [KSS 20] A. Kołakowska, W. Szwoch and M. Szwoch, "A review of emotion recognition methods based on data acquired via smartphone sensors", Sensors, vol. 20 (21), p. 6367, 2020.
- [LGW 19] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt and A. O'Leary, "Pywavelets: A python package for wavelet analysis", Journal of Open Source Software, vol. 4 (36), p. 1237, 2019.
- [LWL 17] Y. Lu, Y. Wei, L. Liu, J. Zhong, L. Sun and Y. Liu, «Towards unsupervised physical activity recognition using smartphone accelerometers», Multimedia Tools and Applications, vol. 76 (18), p. 10 701–10 719, 2017.
- [MEH 96] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament", Current Psychology, vol. 14, no. 4, p. 261–292, 1996.
- [MH 16] A. Mottelson and K. Hornbæk, "An affect detection technique using mobile commodity sensors in the wild", Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, p. 781–792, 2016.
- [MK 20] A. Maalej et I. Kallel, Does keystroke dynamics tell us about emotions? A systematic literature review and dataset construction. In 2020 16th International Conference on Intelligent Environments (IE) (pp. 60-67). IEEE, juillet 2020.
- [MMO 13] E. Mitchell, D. Monaghan and N. E. O'Connor, "Classification of sporting activities using smartphone accelerometers," Sensors, vol. 13 (4), p. 5317–5337, 2013.
- [MPR 16] C. McCarthy, N. Pradhan, C. Redpath and A. Adler, "Validation of the empatica e4 wristband", EMBS International Student Conference (ISC) IEEE, p. 1–4, 2016.

- [MSJ 18] D. Mehta, M. F. H. Siddiqui et A. Y. Javaid, «Facial emotion recognition: A survey and real-world user experiences in mixed reality», *Sensors*, vol. 18 (12), p. 416, 2018.
- [OT 90] A. Ortony and T. J. Turner, “What’s basic about basic emotions?”, *Psychological review*, vol. 97, no. 3, p. 315, 1990.
- [OT 16] A. F. Olsen and J. Torresen, "Smartphone accelerometer data used for detecting human emotions," 3rd International Conference on Systems and Informatics (ICSAI) IEEE, p. 410–415, 2016.
- [PAP 17] E. Politou, E. Alepis and C. Patsakis, "A survey on mobile affective computing", *Computer Science Review*, vol. 25, p. 79–100, 2017.
- [PCB 17] S. Poria, E. Cambria, R. Bajpai and A. Hussain, “A review of affective computing: From unimodal analysis to multimodal fusion”, *Information Fusion*, vol. 37, p. 98–125, 2017.
- [PIC 95] R. W. Picard, *Affective computing*, MIT Press, Cambridge, Massachusetts, 1995.
- [PVG 11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss and V. Dubourg, "Scikit-learn: Machine learning in Python", *The Journal of machine Learning research*, vol. 12, pp. 2825-2830, 2011.
- [RCH 20] M. Ruensuk, E. Cheon, H. Hong and I. Oakley, "How do you feel online: Exploiting smartphone sensors to detect transitory emotions during social media use", *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4 (4), p. 1–32, 2020.
- [ROC 19] M. Ruensuk, H. Oh, E. Cheon, I. Oakley and H. Hong, "Detecting negative emotions during social media use on smartphones," *Proceedings of the Asian CHI Symposium Emerging HCI Research Collection*, p. 73–79, 2019.
- [RUS 80] J. A. Russell, “A circumplex model of affect,” *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.
- [SCH 05] K. R. Scherer, “What are emotions? And how can they be measured?,” *Social science information*, vol. 44, no. 4, p. 695–729, 2005.
- [SJM 21] N. Simonazzi, S. Jean-Marc, M. Morelle, C. Dubois et P. Le Goff, «The Geneva Emotion Wheel Mobile Interface: an Instrument to Report Emotions on Android Devices», proceedings of the ERGO'IA conference, Biarritz (France), 2021. Code available in the following repository: <https://github.com/Orange-OpenSource/GEWMI>
- [SKG 20] A. Saxena, A. Khanna and D. Gupta, “Emotion recognition and detection methods: A comprehensive survey”, *Journal of Artificial Intelligence and Systems*, vol. 2, no. 1, p. 53–79, 2020.
- [SKS 16] A. C. Samson, S. D. Kreibig, B. Soderstrom, A. A. Wade and J. J. Gross, "Eliciting positive, negative and mixed emotional states: A film library for affective scientists," *Cognition and Emotion*, vol. 30 (5), pp. 827--856, 2016.
- [SRD 18] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection", *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pp. 400-408, 2018.
- [SRK 18] M. Swain, A. Routray and P. Kabisatpathy, "Databases, features and classifiers for speech emotion recognition: a review", *International Journal of Speech Technology*, vol. 21 (1), p. 93–120, 2018.
- [SSF 13] K. R. Scherer, V. Shuman, J. Fontaine and C. S. Salina, "The grid meets the wheel: Assessing emotional feeling via self-report", *Components of emotional meaning: A sourcebook*, 2013.
- [STA 22] Statcounter, Browser Market Share Worldwide, Oct 2021 – Oct 2022, available at: <https://gs.statcounter.com/>.
- [SXY 18] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu and X. Yang, "A review of emotion recognition using physiological signals", *Sensors*, vol. 18 (7), p. 2074, 2018.
- [VGO 20] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. v. d. Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern and E. Larson, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python", *Nature Method*, 2020.
- [WAL 96] J. B. Walther, “Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction,” *Communication research*, vol. 23, no. 1, p. 3–43, 1996.
- [WCT 88] D. Watson, L. A. Clark. and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the panas scales," *Journal of personality and social psychology*, vol. 54 (6), p. 1063, 1988.
- [WCV 11] S. v. d. Walt, S. C. Colbert and G. Varoquaux, "The NumPy Array: A Structure for Efficient Numerical Computation", *Computing in Science Engineering*, vol. 13 (2), pp. 22-30, 2011.

- [WSS 96] R. Westermann, K. Spies, G. Stahl and F. W. Hesse, "“Relative effectiveness and validity of mood induction procedures: A meta-analysis”, European Journal of social psychology, vol. 26 (4), p. 557–580, 1996.
- [YQ 21] L. Yang et S.F. Qin, A Review of Emotion Recognition Methods From Keystroke, Mouse, and Touchscreen Dynamics. IEEE Access, 9, 162197-162213, 2021.
- [ZEW 20] T. Zhang, A. El Ali, C. Wang, A. Hanjalic & P. Cesar, Corinet: Fine-grained emotion recognition for video watching using wearable physiological sensors. Sensors, 21(1), 52, 2020.
- [ZPL 21] M.S. Zitouni, C.Y. Park, U. Lee, L. Hadjileontiadis, & A. Khandoker, A.. Arousal-valence classification from peripheral physiological signals using long short-term memory networks. Proc. of the 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 686-689, Nov. 2021.
- [ZWY 18] B. Zhao, Z. Wang, Z. Yu and B. Guo, "Emotion sense: Emotion recognition based on wearable wristband", IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation IEEE, pp. 346-355, 2018.