

Remerciements

Tout D'abord et avant tout, Je remercie mon Dieu qui m'a aidé dans mes études et m'a honoré de la grâce de la connaissance et de la compréhension, ainsi pour la capacité et les facilitations qu'il m'a fournies afin d'accomplir mon mémoire.

Je tiens à remercier sincèrement mes parents qui ont fait des efforts d'atteindre à ce que je suis et qui m'ont soutenu et encouragé, en particulier mon père, qui a mis à ma disposition tous les moyens pour que je réussisse dans mes études, Les mêmes grâce à mon mari qui m'a beaucoup soutenu et faites confiance à mes capacités.

Je voudrais adresser toute ma reconnaissance à Madame Benchaïb Yasmine de m'avoir encadré et orienté, et en particulier pour sa patience, sa disponibilité tout au long de la réalisation de ce mémoire et surtout ses conseils qui m'ont beaucoup aidé à accomplir mon projet.

Je tiens à exprimer toute ma gratitude aux membres des jurys qui ont honoré par leur acceptation à évaluer mon mémoire, notamment :

Madame Dalil Youcef Lamia d'avoir accepté de présider le jury de ce mémoire.

Monsieur Lazouni Mohamed Amine d'avoir accepté d'être l'examineur

Je remercie également tous ceux qui m'ont aidé, conseillé et encouragé à fin de réaliser mon mémoire et toute l'équipe pédagogique de département de génie biomédical.

Merci



Dédicace

Je dédie ce modeste travail avec mon grand amour et mon entière gratitude aux mes plus chère gens :

A mon très cher père,

Aucune dédicace ne saurait exprimer la valeur d'amour, le respect, la gratitude et la reconnaissance que je te porte. J'ai pu accomplir ce travail grâce à tous tes sacrifices, tes encouragements, ton amour et ton soutien permanent durant ce parcours. Je te dois ce que je suis aujourd'hui et ce que je serai demain et je ferai toujours de mon mieux pour rester ta fierté et ne jamais te décevoir. Tu es le plus grand père du monde et je suis très chanceux pour toi. Que Dieu le tout puissant te préserve, t'accorde santé, bonheur, quiétude de l'esprit et te protège de tout mal.

A ma mère,

Ce travail exprime ma gratitude d'être aux côtés de moi, tes prières m'ont été d'une aide précieuse et m'ont permis d'atteindre mes buts désirés. Puisse Dieu le tout puissant te donner santé et longue vie.

A mon amoureux mari,

Tu m'as toujours fait sentir votre grand amour, votre tendresse et votre intérêt constant pour moi. Ce travail est une petite expression de mon amour et de mon bonheur d'être à mes côtés tout au long de ma carrière et de m'appuyer moralement. Tu es toute ma vie et j'espère avoir plus de succès ensemble.

Mes chers frères Islam, Mohammed et Youcef pour leur appui et leur encouragement.

Ma cousine décédée Halima qui souhaitait rejoindre l'Université Abou Bakr Belkaïd et qui rêvait d'être diplômée, Je demande à Dieu de t'apporter ses miséricordes et de nous rassembler au paradis.

Ma directrice de mémoire Mme Benchaïb Yasmine, Pour avoir accepté de m'encadrer, pour ses conseils et suggestions et surtout pour sa patience.

Ma famille, pour leur foi en mes capacités, leur désir de mes succès et leur précieux encouragement.

A tous ceux qui veulent me voir réussie et à tous ceux qui font de leur mieux pour réaliser ses rêves.



Table des matières

Remerciements	I
Dédicace	II
Table des matières	III
Table des figures.....	VI
Liste des tableaux	VII
Glossaire	VIII
Introduction générale.....	2
Chapitre I : le trouble de spectre d'autisme.....	3
Introduction.....	4
1. Définition de l'autisme	4
2. Statistique sur l'autisme	5
3. L'étiologie de l'autisme.....	5
3.1 l'Hypothèse Psychologique.....	5
3.2 Hypothèses des facteurs neurologiques.....	6
3.3 Hypothèses génétiques	6
3.4 Hypothèses métaboliques.....	6
3.5 des facteurs environnementaux	7
3.6 Hypothèses biochimiques.....	7
3.7 Hypothèse du virus et vaccination	7
4. Types de trouble du spectre de l'autisme	7
4.1 Le trouble autistique.....	8
4.2 Le syndrome d'Asperger	8
4.3 Le trouble envahissant du développement non spécifié (TED-ns)	8
5. Les signes cliniques de l'autisme	8
5.1 Troubles des interactions sociales	9
5.2 Troubles de la communication verbale et non verbale.....	9
5.3 Comportements stéréotypés et répétitifs	Erreur ! Signet non défini. 0
6. Le diagnostic de l'autisme	Erreur ! Signet non défini. 1
7. Prise en charge.....	13
Conclusion	14
Chapitre II : Revue de la littérature	16
Introduction.....	17
1. Etat de l'art	17
Conclusion	Erreur ! Signet non défini. 0

Chapitre III : Résultats expérimentaux et discussions.....	Erreur ! Signet non défini.1
Introduction.....	Erreur ! Signet non défini.2
1. Méthode et Outils d'étude	Erreur ! Signet non défini.2
1.1 L'algorithme Relieff.....	Erreur ! Signet non défini.2
1.2 Arbre de decision	Erreur ! Signet non défini.2
1.3 ClassificationTree.fit	Erreur ! Signet non défini.4
1.4 MATLAB.....	Erreur ! Signet non défini.4
1.5 langage R.....	Erreur ! Signet non défini.4
1.6 Android Studio	Erreur ! Signet non défini.4
2. Préparation de la base de données	Erreur ! Signet non défini.5
2.1 Description de la base de données.....	Erreur ! Signet non défini.5
2.2 Description des attributs.....	Erreur ! Signet non défini.5
2.3 Exploration de la base de données	Erreur ! Signet non défini.6
2.3.1 Les niveaux des attributs.....	Erreur ! Signet non défini.7
2.3.2 Correction des erreurs existantes dans la base de données	Erreur ! Signet non défini.8
2.3.3 La proportion de l'ensemble de données de la classe (class_ASD)	Erreur ! Signet non défini.8
2.4 Nettoyage des données	Erreur ! Signet non défini.9
2.4.1 Valeurs Manquantes	Erreur ! Signet non défini.9
2.4.2 les doublons	Erreur ! Signet non défini.9
2.5 Visualisation des données	30
2.5.1 La boîte à moustaches – AGE.....	Erreur ! Signet non défini.0
2.5.2 La boîte à moustaches – Result.....	Erreur ! Signet non défini.0
2.5.3 Diagramme en bâton – Ethnicity	Erreur ! Signet non défini.1
2.5.4 Diagramme en bâton – Jaundice.....	Erreur ! Signet non défini.2
2.5.5 Diagramme en bâton - autism.....	Erreur ! Signet non défini.2
2.6 Prétraitement des données.....	Erreur ! Signet non défini.3
2.6.1 Réduction des attributs.....	Erreur ! Signet non défini.3
2.6.2 Codage des données.....	Erreur ! Signet non défini.3
3. Sélection des attributs Par l'algorithme Relieff.....	Erreur ! Signet non défini.4
4. Résultats expérimentaux et discussion	Erreur ! Signet non défini.5
4.1 Classification sur la base de données initiale avec 15 attributs et 248 observations	Erreur ! Signet non défini.5
4.1.1 Résultats d'algorithme ClassificationTree avec validation croisée (10-fold)	Erreur ! Signet non défini.5
4.1.2 Résultats d'algorithme KNN avec validation croisée (10-fold)	Erreur ! Signet non défini.6
4.1.3 Résultats d'algorithme SVM avec validation croisée (10-fold)	Erreur ! Signet non défini.8
4.2 Base de donnée réduit avec 9 attribues et 248 observations	Erreur ! Signet non défini.8

4.2.1	Résultats d’algorithme ClassificationTree avec validation croisée (10-fold)	Erreur ! Signet non défini.8
4.2.2	Résultats d’algorithme KNN avec validation croisée (10-fold)	Erreur ! Signet non défini.9
4.2.3	Résultats d’algorithme SVM avec validation croisée (10-fold)	41
4.3	Comparaison entre les résultats des trois algorithmes avant et après la sélection des attributs par RELIEFF	41
4.4	Discussions	42
	Conclusion	42
Chapitre IV : application mobile « طفلي التوحيدي » 44		
	Introduction	45
1.	Description de l’application	45
2.	La méthode d’apprentissage automatique dans notre application	Erreur ! Signet non défini.6
3.	Les interfaces d’application « طفلي التوحيدي »	Erreur ! Signet non défini.6
3.1	L’interface de « اختبار سمات التوحد »	Erreur ! Signet non défini.7
3.1.1	Les interfaces des questions	Erreur ! Signet non défini.7
3.1.2	L’interface de résultat de test	Erreur ! Signet non défini.8
3.2	L’interface de « عن طيف التوحد »	Erreur ! Signet non défini.9
3.3	L’interface de « تعامل مع سلوكيات طفلك »	Erreur ! Signet non défini.9
3.4	L’interface de « أنشطة »	50
	Conclusion	50
	Conclusion générale	51
	Les Références	54
	Résumé	57

Table des figures

Chapitre III : Résultats expérimentaux et discussions

Figure (1) : Arbre de décision [41].....	23
Figure (2) : les noms des attributs de la base de données.....	27
Figure (3) : statistiques descriptives de base de données	27
Figure (4) : les niveaux des attribues.....	2Erreur ! Signet non défini.
Figure (5) : les noms des attributs après les changements effectués	2Erreur ! Signet non défini.
Figure (6) : la base de données avant et après la suppression des valeurs manquantes.....	29
Figure (7) : Regroupement des valeurs Self et self en un seul niveau de facteur.....	30
Figure (8) : La boîte à moustaches – AGE	30
Figure (9) : La boîte à moustaches – Result	31
Figure (10) : Comparaison entre la valeur de result ≥ 7 et la classe Class_ASD.....	31
Figure (11) : Diagramme d’Ethnicity	31
Figure (12) : Diagramme en bâton – Jaundice	32
Figure (13) : Diagramme en bâton – autism.....	32
Figure (14) : la base de données avec et sans les attributs non significatifs	33
Figure (15) : Arbre de désicion avec un $Tc=96.00\%$, $Sn=92.31\%$ et $Sp=100\%$	36
Figure (16) : Arbre de désicion avec un $Tc=100\%$, $Sn=100\%$ et $Sp=100\%$	39

Chapitre IV : application mobile « طفلي التوحدي »

Figure (17) : Écran de bienvenue pour l’application « طفلي التوحدي ».....	45
Figure (18) : la page d’accueil de l’application « طفلي التوحدي ».....	46
Figure (19) : la page de démarrage de test TSA	47
Figure (20) : les interfaces des questions de test TSA.....	48
Figure (21) : les interfaces des résultats de test TSA	4Erreur ! Signet non défini.
Figure (22) : l’interface des informations sur TSA	4Erreur ! Signet non défini.
Figure (23) : les interfaces des comportements Caractéristiques de l'enfant autiste	Erreur ! Signet non défini.
Figure (24) : les interfaces des activités quotidiennes.....	Erreur ! Signet non défini.

Liste des tableaux

Chapitre III : Résultats expérimentaux et discussions

Table 1 : les attributs et leurs descriptions	25
Table 1 : les attributs et leurs descriptions	26
Table 2 : les cinq premières lignes de l'ensemble de données.....	26
Table 3 : le nombre des valeurs manquantes existent dans chaque attribut	29
Table 4 : les rangs et les poids des attributs en fonction de nombre de K.....	34
Table 5 : Résultats de classificationTree sur la base de données initiale	35
Table 6 : le taux moyen de classificationTree pour chaque essai sur la base de données initiale	35
Table 7 : Résultats de KNN avec distance Euclidienne sur la base de données initiale.....	36
Table 8 : Résultats de KNN avec distance Cityblock sur la base de données initiale.	37
Table 9 : Résultats de KNN avec distance Cosine sur la base de données initiale	37
Table 10 : Résultats de KNN avec distance Correlation sur la base de données initiale.....	37
Table 11 : les meilleurs résultats obtenus par chaque distance sur la base de données initiale	37
Table 12 : Résultats de SVM avec différentes fonction noyau sur la base de données initiale	38
Table 13 : Résultats de classificationTree sur la base de données avec 9 attributs.....	38
Table 14 : le taux moyen de classificationTree pour chaque essai sur la base de données avec 9 attributs	38
Table 15 : Résultats de KNN avec distance Euclidienne sur la base de données avec 9 attributs	39
Table 16 : Résultats de KNN avec distance Cityblock sur la base de données avec 9 attributs	40
Table 17 : Résultats de KNN avec distance Cosine sur la base de données avec 9 attributs	40
Table 18 : Résultats de KNN avec distance Correlation sur la base de données avec 9 attributs	40
Table 19 : les meilleurs résultats obtenus par chaque distance sur la base de données avec 9 attributs	40
Table 20 : Résultats de SVM avec différentes fonction noyau sur la base de données avec 9 attributs	4Erreur ! Signet non défini.
Table 21 : les performances de divers algorithmes d'apprentissage machine sur la base de données de TSA chez les enfants avant et après la sélection des attributs avec l'algorithme RELIEFF	4Erreur ! Signet non défini.

Glossaire

ADI-R : L'interview diagnostique sur l'autisme révisée

ADOS : Echelle d'Observation pour le Diagnostic de l'Autisme

ASD : Autism spectrum disorder

CIM-10 : La 10^{ème} révision classification internationale des maladies

DSM_II : le manuel diagnostique et statistique des troubles mentaux _ version deux

DSM-III : le manuel diagnostique et statistique des troubles mentaux _ version trois

DSM-IV : le manuel diagnostique et statistique des troubles mentaux _ version quatre

DSM-V : le manuel diagnostique et statistique des troubles mentaux _ version cinq

TED : les troubles envahissants du développement

TSA : trouble du spectre de l'autisme

KNN : la méthode des k plus proches voisins

OMS : Organisation mondiale de la Santé

ReliefF : redress pour la régression avec K plus proches voisins

SVM : Les machines à vecteurs de support

TC : contrôles typiques

Introduction générale

Introduction générale

Le spectre de l'autisme est un trouble complexe dans ses causes et très répandu à notre époque qui entrave la croissance normale de nombreux enfants et cause anxiété et souffrance à leurs parents. La différence de ses symptômes d'une personne à l'autre avec la possibilité de disparition de certains et l'émergence d'autres au fil du temps en plus de la similitude de ces symptômes avec les symptômes d'autres troubles rend difficile son diagnostic. Il n'existe pas d'outils ou de test unique pour le diagnostic, mais les spécialistes doivent faire plusieurs tests cliniques et fournir des questionnaires à remplir par les parents afin d'identifier les aptitudes sociales et comportementales de l'enfant et ses capacités linguistiques.

En raison de l'efficacité de l'intelligence artificielle, qui a récemment été démontrée par des études dans le diagnostic de plusieurs maladies telles que la maladie d'Alzheimer et la dépression, et compte tenu de la difficulté de diagnostiquer les troubles du spectre autistique, outre le manque d'études fondées sur l'utilisation de techniques d'apprentissage automatique pour diagnostiquer ces troubles, tant au niveau étranger qu'arabe, nous avons généré la problématique de cette recherche pour détecter les signes qui apparaissent chez l'enfant et indiquent l'incidence des troubles du spectre autistique à l'aide des techniques d'apprentissage automatique, car une détection précoce peut aider à élaborer des plans de traitement et donc améliorer le comportement de l'enfant.

Notre objectif de recherche est la création d'une nouvelle application nommée : «التوحدِي طفلي» pour détecter les caractéristiques de TSA chez les enfants en utilisant une méthode d'apprentissage automatique la plus précise possible. Dans ce contexte, notre travail sera réparti en quatre chapitres, défini comme suit :

- Le premier chapitre présente les notions de base et des définitions générales relatives à l'autisme au cours de l'histoire, ses types et les différentes hypothèses sur ses causes ainsi que ses principaux symptômes, ses méthodes de diagnostic et de prise en charge existantes.
- Le deuxième chapitre est consacré à la présentation des travaux existants dans la littérature liée au diagnostic de l'autisme par des techniques d'apprentissage automatisé.
- Le troisième chapitre qui est le chapitre le plus important dans notre mémoire. Dans lequel, nous allons présenter les différents outils qui vont servir à l'implémentation de notre étude et décrire également la base de données utilisée pour le dépistage des troubles du spectre autistique (TSA) pour les enfants, puis nous présentons les traitements utilisés pour nettoyer la base de données et traiter les valeurs manquantes, réduction des variables et codage. Pour finir ainsi nous allons détailler les résultats obtenus par les différents modèles de classifications réalisés sur notre base de données et nous discutons ces résultats.
- Le quatrième et dernier chapitre sera consacré à la définition de notre application mobile «التوحدِي طفلي» et son mécanisme de travail et affiche ses interfaces et son mode d'emploi.

Enfin, nous allons terminer par une conclusion générale résumant les résultats de notre travail et nous allons également nous attarder sur quelques des perspectives envisagées pour la continuité de ce travail.

Chapitre I : le trouble de spectre d'autisme

Introduction

L'autisme est un trouble dont la prévalence augmente considérablement de plus en plus dans le monde actuel. Comme nous savons que ce trouble est l'un des troubles du développement les plus difficiles et complexes du fait de son impact considérable sur les diverses manifestations du développement de l'enfant, qu'il s'agisse des aspects linguistiques, sociaux et communicatifs, ainsi que des processus d'attention et de perception. De plus, l'impact du trouble de l'autisme ne dépend pas seulement de l'enfant et des différents aspects de sa vie, mais aussi de la famille qui tente de s'adapter à la situation et cherche à fournir toutes les possibilités dont cet enfant a besoin.

Dans ce chapitre, nous tenterons tout d'abord de donner une définition des troubles d'autisme au paragraphe (1) et des statistiques connexes ces dernières années au paragraphe (2), en mentionnant ses principales causes selon différentes hypothèses et présentant ses types au paragraphe (3 et 4). Nous citerons les principaux signes de ce trouble au paragraphe (5) et nous définirons les outils de diagnostic les plus importants actuellement adoptés en plus des méthodes de traitement aux paragraphes (6 et 7), pour donner une idée globale du trouble d'autisme et aider à mieux le comprendre.

1. Définition de l'autisme

L'autisme est un mot qui vient de terme du grec «autos » qui signifie « soi-même», ce terme a été proposé pour la première fois par le psychiatre suisse Eugen Bleuler dans son ouvrage « Dementia preacox oder Gruppe der Scizophrenien » (« Démence précoce ou le groupe des schizophrènes ») en 1911 pour décrire la symptomatologie schizophrénique chez des patients adultes et parmi ces symptômes fondamentaux sont le repli sur soi-même et le retrait social [26]. et en 1943 le médecin pédopsychiatre américain Leo Kanner a été utilisé le terme d'autisme infantile pour décrire le repli sur soi-même dans son article princeps « Autistic disturbances of affective contact » (Les troubles autistiques du contact affectif) ou il a défini les enfants autistes comme sont « les enfants qui ont une inaptitude à établir des relations normales avec autrui et à réagir normalement aux situations » [27]. Un an plus tard, le pédiatre autrichien Hans Asperger a écrit un article intitulé "les psychopathies autistiques pendant l'enfance" dans lequel a utilisé le terme « Psychopathie autistique » [28]. En 1967, Bruno Bettelheim a repris le terme de "mère réfrigérateur" dans son ouvrage « La Forteresse vide » pour expliquer l'autisme chez les enfants [29].

La compréhension des troubles de l'autisme par la communauté scientifique a énormément évolué depuis la première définition de l'autisme par Kanner, De nombreuses appellations ont été successivement utilisées pour nommer le «troubles de l'autisme» :

_ Selon DSM II en 1968, le terme « autisme » a été classé comme « une schizophrénie infantile », En 1980, dans la DSM III, l'autisme a été reconnaître en tant que diagnostic distinct et qu'il a apparaît sous la dénomination « autisme infantile » dans le groupe « trouble globale de développement ». Dans DSM-IV en 1994, le trouble autistique (ou autisme de haut niveau) a été officiellement intégré pour la première fois parmi les troubles envahissants du développement (TED) avec 4 autres troubles (le syndrome d'Asperger, le trouble envahissant du développement – non spécifié (TED-ns), le syndrome de Rett (SR) et le trouble désintégratif de l'enfance). Les TED sont définis comme un groupe de troubles sévères et précoces, caractérisés par des retards et des déformations du développement des habiletés sociales, cognitives et du développement de la communication, en 2013 dans le DSM-V le trouble autistique, le syndrome d'Asperger et le TED-ns sont regroupés sous le terme « trouble du spectre de l'autisme » (TSA) et le syndrome de

Rett et le trouble désintégratif de l'enfance sont exclus des TSA, Il n'y a maintenant que cette appellation pour ces troubles [1].

_ Selon la CIM-10 diffusée par l'OMS a défini l'autisme infantile en 1992 comme : " un trouble envahissant du développement, caractérisé par un développement anormal ou déficient, qui se manifeste avant l'âge de trois ans avec une perturbation caractéristique du fonctionnement dans chacun des trois domaines suivants : interactions sociales, communication, comportements au caractère restreint et répétitif " [2].

L'autisme est défini comme un trouble neurodéveloppemental sévère et précoce qui apparaît au cours des trois premières années de l'âge de l'enfant en raison d'un dysfonctionnement du cerveau et des nerfs, causant des problèmes et des difficultés au niveau du développement et dont le diagnostic se fait sur la base de caractéristiques du comportement et du développement chez l'enfant car Il n'existe pas d'examen médical spécial permettant de détecter l'autisme jusqu'à maintenant.

L'autisme est définitif et perdure toute la vie. Il est caractérisé par trois symptômes principaux : des perturbations dans les interactions sociales, un trouble de Langage et de communication, un comportement restrictif et répétitif, avec des manifestations qui peuvent varier d'un enfant à l'autre et dans le même enfant dans le temps.

2. Statistique sur l'autisme

D'après les études épidémiologiques effectuées ces cinquante dernières années, il semble que la prévalence des troubles du spectre autistique augmente à l'échelle mondiale. Cette augmentation apparente peut s'expliquer de nombreuses façons, notamment par une plus grande sensibilisation au problème de l'autisme, par l'élargissement des critères de diagnostic, le perfectionnement des outils diagnostiques et une notification plus rigoureuse. La prévalence de l'autisme est probablement sous-évaluée dans certaines régions du monde (Afrique, Amérique du sud, pays latins,...) et surévaluée dans les sociétés nord-américaines, occidentales et certains pays asiatiques comme la Corée, pays basés sur des modèles sociétaux sur-stressés

Selon l'OMS en 2017, Les études épidémiologiques estiment que, dans le monde, 1 enfant sur 160 présente un trouble du spectre autistique. Ce chiffre correspond toutefois à une moyenne et la prévalence notifiée varie notablement d'une étude à l'autre. Les études les plus récentes estiment que le taux de prévalence des personnes autistes varierait entre 90 et 120 Individus sur 10 000 soit environ 1% de la population. La prévalence de l'autisme est probablement sous-évaluée dans certaines régions du monde (Afrique, Amérique du sud, pays latins,...) et surévaluée dans les sociétés nord-américaines, occidentales et certains pays asiatiques comme la Corée, pays basés sur des modèles sociétaux sur-stressés [3]. Ce trouble survient quatre fois plus souvent chez les garçons que chez les filles.

En Algérie, le nombre de cas d'autistes en 2018 est situé entre 400 000 et 500 000 enfants adolescents et adultes avec autisme, dont la majorité a été diagnostiquée tardivement, à partir de l'âge de cinq ans et sans aucune prise en charge institutionnelle sérieuse [4].

3. L'étiologie de l'autisme

Les causes de l'autisme n'ont pas été clairement élucidées avec précision et interdépendantes parce qu'il n'y a pas de symptôme spécifique, mais une gamme de symptômes varie en intensité et en qualité d'un enfant à l'autre, Mais des hypothèses multiples ont été émises, tel que :

3.1 l'Hypothèse Psychologique : proposée par les partisans de l'École de psychanalytique, y compris Canner qui soutenait que les pratiques parentales étaient non coordonnées pour leurs enfants et la froideur émotionnelle qui Leur causant de l'autisme, Puis il approuva En 1967 par Bettelheim qui a suggéré que l'autisme se développait en réponse à une dysharmonie dans les interactions précoces entre l'enfant et les parents qui sont accusés de froideur dans leurs émotions envers eux ; Surtout la mère (Mère froide) car elle n'apportait pas l'apaisement dont l'enfant a besoin, et répondait par le rejet aux demandes de l'enfant, Ce qui lui fait regarder le monde comme froid et hostile et répond par un repli sur soi-même, Mais cette hypothèse n'a pas été prouvée et considérée comme infondée [5]. Lorsque les scientifiques ont transféré ces enfants affectés vers d'autres familles, aucune amélioration n'a été observée chez ces enfants et en plus leurs frères sont normaux dans la plupart des cas. Il est également noté que Ce trouble peut parfois commencer à partir de la naissance mais que les comportements familiaux peuvent provoquer des troubles de la personnalité chez les autistes en plus de leur handicap initial.

3.2 Hypothèses des facteurs neurologiques : dans laquelle les chercheurs se sont penchés sur le dysfonctionnement du cerveau par l'utilisation des examens médicaux des radiations ou de la résonance magnétique montrant des preuves que l'organisation du cerveau d'une personne présentant des troubles du spectre autistique diffère de celle d'autres personnes. Ce dysfonctionnement interfère avec le développement normal du cerveau au cours de la grossesse et des premières années de vie.

Selon une étude de Garreau, la stimulation auditive augmente le débit sanguin cérébral dans l'hémisphère cérébral gauche qui contrôle le langage chez l'enfant normal, et dans l'hémisphère droit du cerveau chez l'enfant autiste, ce qui rend souvent le langage déficient chez lui expliquant partiellement les problèmes de socialisation et de communication [5]. Une autre étude sur la zone frontale du cerveau qui implique la reconnaissance des visages et des fonctions sociales, cognitives et affectives a montré que cette zone pouvait être affectée spécifiquement. Où le débit sanguin cérébral a permis de constater un hypodébit frontal dans cette zone chez les enfants autistes âgés de 3 ans. Certains scientifiques ont constaté une atrophie de la taille du cervelet qui intègre les informations tactiles et la sensibilité profonde, assure l'orientation et régule le tonus, la coordination et la force musculaires, Cette atrophie apparaît en particulier des lobes périodiques numéros six et sept, pouvant être réduite à (13%) de la taille du cervelet chez les enfants atteints d'autisme. Et une observation montre que la plupart des enfants autistes souffrent d'un élargissement de l'épaisseur du cortex cérébral (une augmentation du périmètre crânien) et de la présence de défaut dans le mouvement des cellules cérébrales au cours des six mois Premières de grossesse [24].

3.3 Hypothèses génétiques : proposé par Certains chercheurs qui soulignent que les anomalies chromosomiques et génétiques à un stade précoce de la vie d'un fœtus mènent à l'autisme. La plupart des études basant sur des jumeaux monozygotes et dizygotes ont recherché un ou des gènes responsables du trouble autistique, ces études faites ont montré une corrélation génétique importante Chez Les jumeaux monozygotes de 9% lorsque l'un est autiste l'autre l'est aussi, alors que la corrélation est très minime chez les jumeaux dizygotes de 65% si l'un est autiste, l'autre l'est aussi. Cela confirme que L'héritabilité de l'autisme est certaine et toutes les études à ce jour convergent sur l'idée que les facteurs génétiques sont responsables d'au moins 50% des cas d'autisme, mais les chercheurs ne peuvent aujourd'hui pas affirmer clairement le gène responsable de l'autisme [5].

3.4 Hypothèses métaboliques : indique que Certains enfants souffrent de troubles gastro-intestinaux et digestifs, suggérant des hypothèses sur la relation entre les protéines du lait ou le gluten et l'autisme. Ces hypothèses indiquent que les enfants autistes ne peuvent pas en particulier digérer les protéines, notamment les protéines du gluten présentes dans le blé, l'orge

et leurs dérivés (comme le pain et les pâtes). En plus des protéines de caséine dans le lait, l'émergence de peptides non digérés entraîne un effet anesthésique similaire à l'effet de l'opium et de la morphine. Les enfants qui ne digèrent pas le gluten (l'intolérance au gluten) sont sujets à la diarrhée chronique et ils ne se sentent pas bien et leur contact peut devenir bizarre [30].

3.5 des facteurs environnementaux : qui peuvent causer l'autisme chez l'enfant, cela inclut notamment les facteurs prénataux tels que l'exposition de la mère à l'infection par la rougeole allemande, une infection virale pendant la grossesse, exposition à des substances toxiques dans le ventre, carences nutritionnelles avant ou durant la grossesse, le diabète de grossesse, des réactions du système immunitaire durant la grossesse et les facteurs périnataux tel que le prématurité, manque d'oxygène ou hémorragie cérébrale à la naissance. De nombreux médecins insistent sur les effets des complications, notamment celles survenant au cours des trois premiers mois de la grossesse sur l'autisme [10].

D'autre part, Certains chercheurs supposent que l'exposition de l'enfant aux stades critiques de son développement à la pollution de l'environnement (mercure et agent de conservation pour le caoutchouc, le plomb et le monoxyde de carbone) provoque des lésions cérébrales et une intoxication du sang Ce qui entraîne l'autisme chez l'enfant.

3.6 Hypothèses biochimiques : dans lequel la biochimie joue un rôle important dans le travail du corps, en particulier dans le cas de l'autisme car certains étudient et assument que les personnes atteintes du syndrome autistique présentent des perturbations au niveau des concentrations des neurotransmetteurs tels que la dopamine, l'adrénaline, la noradrénaline et la sérotonine, ce dernier présentant une augmentation de 40% de la concentration dans le sang d'enfants autistes et lorsque cette concentration a été réduite par l'utilisation de fluvoxamine, une amélioration et une diminution de certains symptômes chez ces enfants a été observée [10]. En plus d'une amélioration de leurs performances langagières. En outre, les neuropeptides et les bêta-endorphines et l'acide iodophénique se retrouvent généralement à un taux anormalement plus élevé dans le sang des individus autistes que chez les enfants naturels.

3.7 Hypothèse du virus et vaccination : avancée par certains scientifiques ont trouvé un lien entre l'autisme et l'infection maternelle à travers certaines infections virales qui sont soupçonnées d'être nocifs pour le cerveau des enfants, comme la rougeole allemande et l'hypertrophie des cellules virales, Infection cellulaire virale, le cytomégalovirus (CMV), les virus de l'herpès, et peut-être même le virus de l'influenza, et Certains d'autres pensent que la vaccination peut conduire à l'autisme car le système immunitaire ne produit pas assez d'anticorps pour éliminer les virus, de plus il existe des théories sur la relation entre l'autisme et la présence de métaux lourds et de conservateurs dans les vaccins, car ces métaux lourds (tels que le mercure, le plomb et le cadmium) sont dangereux pour le cerveau, en particulier lors de son développement. Dans certains anciens vaccins, des composés du mercure ont été utilisés comme conservateurs (le thimérosal). Bien qu'il s'agisse d'infimes quantités, ce composé puisse s'accumuler après plusieurs vaccinations et le corps ne peut plus se débarrasser facilement du mercure ce qui menace le cerveau. Cette théorie était acceptable, mais cette hypothèse n'a pas été retenue par les centres scientifiques [30].

4. Types de trouble du spectre de l'autisme

Le terme spectre indique que le trouble peut se manifester de différentes manières, ces différences peuvent se présenter par Le nombre de symptômes qui apparaissent, la gravité des

symptômes, l'âge du début des symptômes, et le niveau de fonctionnement intellectuel. Les types de ce spectre sont :

4.1 Le trouble autistique

Le trouble autistique est également appelé autisme, autisme infantile, autisme infantile précoce, syndrome de Kanner ou psychose infantile.

La CIM-10 publiée en 2010, stipule que «l'autisme infantile» est un type de trouble envahissant du développement qui est défini par la présence d'un développement anormal ou altéré qui se manifeste avant l'âge de trois ans, et le type caractéristique de fonctionnement anormal dans les trois domaines de la psychopathologie: interaction sociale réciproque, la communication et comportement restreint, stéréotypé et répétitif. Outre ces caractéristiques de diagnostic spécifiques, une série d'autres problèmes non spécifiques sont courants, tels que les phobies, les troubles du sommeil et de l'alimentation, les crises de colère et des gestes auto-agressifs [2].

4.2 Le syndrome d'Asperger

Le syndrome d'Asperger est défini cliniquement en 1981 par Lorna Wing à partir de la psychopathie autistique décrite en 1944 par Hans Asperger. Comme tous les TSA, le syndrome d'Asperger se caractérise par le même type d'anomalies qualitatives d'interactions sociales réciproques qui caractérisent l'autisme, associées à des intérêts spécifiques ou des comportements répétitifs mais est caractérisée aussi par une intelligence normale ou supérieure à la moyenne. Il se distingue principalement de l'autisme par le fait qu'il n'y a pas de retard général ni de retard dans le développement du langage ou de la cognition. Il s'associe souvent à une maladresse physique et une utilisation atypique de la parole [6].

4.3 Le trouble envahissant du développement non spécifié (TED-ns)

Le TED-ns doit être utilisé en cas de développement anormal et altéré présent seulement après trois ans et d'absence d'anomalies suffisantes et démontrables dans un ou deux des trois domaines de la psychopathologie requis pour le diagnostic de l'autisme, il est souvent appelé autisme atypique car les critères de trouble autistique ne sont pas remplis par exemple en raison de l'âge tardif, de la symptomatologie atypique ou de la symptomatologie sous-seuil, ou de tous ces facteurs. Même si TED-ns est considéré comme plus doux que l'autisme typique, ce n'est pas toujours vrai car certaines caractéristiques peuvent être moins graves et d'autres plus graves [7]. L'autisme atypique survient le plus souvent chez les personnes souffrant de retard mental grave et chez les personnes présentant un trouble du développement spécifique grave du langage réceptif.

Selon DSM-V, TED-ns est un diagnostic utilisé pour altération grave et généralisée du développement d'interactions sociales réciproques ou d'aptitudes à la communication verbale et non verbale, ou lorsqu'un comportement, des intérêts et des activités stéréotypés sont présents, et selon la CIM-10 publiée en 2010, "l'autisme atypique" est "un type de trouble envahissant du développement qui diffère de l'autisme infantile soit par son âge

d'apparition, soit par son incapacité à satisfaire les trois critères des critères de diagnostic [7].

5. Les signes cliniques de l'autisme

L'autisme chez l'enfant se caractérise par trois symptômes principaux :

5.1 Troubles des interactions sociales

De nombreux enfants autistes présentent souvent une faible interaction sociale, et décrite comme vivant dans leur propre monde, ils sont isolés du monde dans lequel ils vivent. Ce manque d'interaction sociale est la caractéristique prédominante de ces enfants et parmi ses manifestations [8] :

- _ Un manque de compréhension des relations et des obligations sociales.
- _ Les enfants autistes ignorent les autres et ne sont pas affectés par leur présence ou de les approcher, ne vont pas spontanément échanger des conversations avec eux et ils trouvent difficile de devenir ce qu'une personne pense ou comment il se sent en regardant son visage fois (ceci présente un attribut utilisé dans notre algorithme pour la classification des enfants ont un TSA).
- _ Ils ont du mal à exprimer leurs sentiments et ils trouvent difficile de déterminer les intentions ou les sentiments des autres (ceci présente un attribut utilisé dans notre algorithme pour la classification des enfants ont un TSA).
- _ Ils préfèrent l'isolement de l'existence avec les autres et préfèrent jouer seul et ne cherchent pas automatiquement à engager les autres dans des activités de divertissement, d'intérêt ou d'accomplissement et ils ont la difficulté à se faire de nouveaux relations (ceci présente un attribut utilisé dans notre algorithme pour la classification des enfants ont un TSA).
- _ Un rejeté de contact physique (le portage et la prise dans les bras) et un désir de communiquer émotionnellement
- _ L'absence des comportements d'anticipation qui se mettent habituellement en place dès le très jeune âge.
- _ L'expression gestuelle est appauvrie (l'enfant ne pointe pas du doigt ou pointe dans le but d'obtenir quelque chose) ou exagérées.
- _ Ils n'aiment pas attirer les attentions des gens, comme ils ont une absence de partage d'intérêts ou de demande d'aide.
- _ Ils ont un manque de communication visuelle et évitent de regarder le visage de quelqu'un d'autre.
- _ Les enfants à l'âge de 0 à 6 mois ont un Absence ou rareté du sourire social qui déclenché par un stimulus lié aux personnes (notamment aux parents) et de geste d'anticipation (ne tend pas les bras quand on va le chercher dans son berceau) et il ont impression d'indifférence au monde sonore, et pas de suivi visuel des mouvements dans l'environnement du bébé (par exemple le mouvement d'une main qui bouge), et à l'âge de 6 à 12 mois, l'enfant ne répond pas à son prénom quand les parents l'appelle et ne manifeste pas de réaction [9].

5.2 Troubles de la communication verbale et non verbale

Les troubles du langage et de la communication chez les enfants autistes tant dans le registre verbal et non verbal apparaissent massivement entre 2 et 5 ans, ce qui les fait souffrir de retard mental complet ou d'absence de développement du langage et de difficultés de communication

qui affectent leur capacité à comprendre les signes et symboles sociaux et leur capacité à les utiliser. Parmi Les troubles de la communication remarquable [8] :

_ Un Manque de compréhension de nombreuses significations des mots que les enfants autistes reçoivent des autres et de la difficulté à généraliser leurs concepts et l'utilisation turbulente de ces mots ce qui les mettant mal à l'aise dans les conversations sociale (ceci présente un attribut utilisé dans notre algorithme pour la classification des enfants ont un TSA).

_ Un manque évident de capacité d'expression, de sorte que ce qu'ils disent est étrange ou n'est pas lié à un sujet sérieux et utilisent aussi un Langage abstrait mal compris.

_ les enfants autistes ne pas pouvoir contrôler leur conversation avec les autres (ne savent pas quand parler et quand ils arrêtent d'écouter) et suivre facilement plusieurs conversations à la fois (ceci présente un attribut utilisé dans notre algorithme pour la classification des enfants ont un TSA).

_ Pour les enfants ont un langage développé et capables de parler, ils ne pourront à entamer ou à poursuivre une conversation avec une autre personne ou avec ses amis (ceci présente un attribut utilisé dans notre algorithme pour la classification des enfants ont un TSA).

_ ne pas utiliser d'autres modes de communication comme L'expression gestuelle et lorsqu'elle apparaît, elle est rarement coordonnée avec le regard.

_ Ils ont souvent fréquemment d'une mémoire auditive pour cela ils répètent souvent des mots qui seront entendus dans un passé lointain.

_ Ils ont peu ou pas de réponse appropriée aux instructions qui leur sont données, et ils n'utilisent pas les pronoms.

_ l'enfant autiste manifeste généralement par la suite un phénomène d'inversion pronominale (il exprime ses besoins en « tu » ou en « il » au lieu de « je » par exemple).

_ l'enfant autiste ne pouvait interpréter les émotions correspondantes telles la peur, la colère, la joie ou la tristesse.

_ Ils ont une répétition involontaire ou automatique des mots, de passages ou de sons qui semblent résonner avec eux.

_ L'absence de jeu de fiction automatique sous ses différentes formes chez l'enfant autiste et il n'aime pas jouer aux jeux impliquant de (faire semblant) avec les autres enfants (ceci présente un attribut utilisé dans notre algorithme pour la classification des enfants ont un TSA).

_ chez les enfants de 0 à 6 mois, Les troubles de communication sont caractérisés par l'apathie ,l'impression d'anormalité (bébé trop calme ou excité),et un peu de vocalisation ou absence de babillage qui est la forme de langage utilisée par les tout petits et est composé d'onomatopées et de sons simples. Chez les enfants de 6 à 12 mois, il y a une absence de geste d'anticipation, et d'imitation dans la communication gestuelle (faire « coucou », « au revoir », « bravo »...) [9].

5.3 Comportements stéréotypés et répétitifs

Les enfants autistes présentent de nombreux comportements qui peuvent être décrits comme des anomalies caractéristiques de l'autisme dont ses causes sont très complexes dans leur compréhension, ces enfants ont des intérêts limités et sont également très intéressés par certaines choses et ont une personnalité anormale en plus qu'ils passent une grande partie de ses vies à pratiquer des comportements typiques et routines répétitive anormaux qui se manifestent à travers que [8]:

_ Ces enfants peuvent passer leur temps dans la répétition automatique des activités (le scintillement, flottement, flexion des mains, doigts ou mouvements complexes de tout le corps).

_ Ils adhèrent aussi à une routine et à un rituel anormaux qui ne sont pas pratiques.

_ Ils ont une mauvaise adaptation lors de la connexion à des objets.

_ Ils souffrent de crises de colère, et de dépression nerveuse et d'automutilation.

_ Ils ont un comportement agressif soit verbal ou non verbal (physique).

_ Ils n'utilisent pas les jouets d'une façon fonctionnelle, l'attention étant parfois dirigée vers une seule partie de l'objet manipulée qui est utilisé dans des activités répétitives et excluse les autre,

ces activités stéréotypées peuvent être produites sans l'aide d'un objet et seulement par l'utilisation d'une partie du corps.

_ Ils ont une sensibilité exacerbée aux modifications de l'environnement (changement de place des objets ou des meubles dans la chambre, introduction d'un nouveau jouet, changement de trajet pour aller au parc), et le changement le plus minime dans son environnement l'amène souvent à une grande intolérance et à une forte angoisse.

_ L'enfant autiste ont une réaction paradoxale au bruit où il peut ne pas réagir à un bruit fort juste à côté de lui mais au contraire sembler sensible à des bruits à peine audibles quand d'autres ne le remarquent pas (ceci présente un attribut utilisé dans notre algorithme pour la classification des enfants ont un TSA) [9].

_ Il peut adopter des postures étranges dans lesquelles il peut se figer pendant de longs moments. En plus de ces 3 caractéristiques il y a d'autres pathologies associées appelées aussi comorbidités sont des pathologies secondaires souvent liées à l'autisme, les personnes autistes sont particulièrement affectées par une ou plusieurs d'entre elles. La déficience mentale est une condition qui accompagne fréquemment d'autisme à côté d'autre manifestations tel que les troubles psychiatriques, les troubles sensoriels, les troubles du sommeil, les troubles de l'alimentation et les difficultés motrices et de coordination

6. Le diagnostic de l'autisme

L'autisme est souvent diagnostiqué tardivement. Pourtant la nécessité de détecter l'autisme chez les jeunes enfants est impérative en le référant au diagnostic effectué par une équipe multidisciplinaire expérimentés et formée pour examiner son développement, comprendre un orthophoniste, un psychomotricien, un psychologue pédiatre spécialement formé aux bilans diagnostic de l'autisme (ADI-R, ADOS, CARS...) et un pédopsychiatre ou un neuropédiatre capable d'effectuer la synthèse et les éventuels examens complémentaires nécessaires, et poser enfin le diagnostic.

Les enfants diagnostiqués précocement peuvent bénéficier d'une intervention précoce ce qui se traduit par l'élaboration d'un plan de traitement comportemental intégré pour réduit ainsi le problème des maladies liées à ce trouble. Mais le diagnostic de l'autisme reste l'un des problèmes les plus importants auxquels sont confrontés les chercheurs dans le domaine des problèmes de l'enfance, cela peut être dû au fait que les symptômes de l'autisme partagent les mêmes symptômes avec d'autres troubles tels que le retard mental, les handicaps émotionnels, en plus les recherches sur l'autisme sont relativement récentes car les connaissances de la communauté scientifique ont commencé à être limitées dans les années 1950 et plus précisément dans les années 1990, et des recherches sont toujours en cours pour mieux comprendre l'autisme.

Le diagnostic de l'autisme repose sur le repérage d'un certain nombre de signes comportementaux. Lorsqu'un nombre minimum de symptômes est présent, on parlera alors d'autisme. S'il existe des symptômes mais que leur nombre n'atteint pas un certain seuil, d'autres diagnostics pourront être portés, le diagnostic doit être basé aussi sur la collecte d'un certain nombre de données auprès des parents en particulier l'histoire très précise du développement et la description des comportements, aptitudes et difficultés présentés par l'enfant, ainsi que sur l'observation directe de l'enfant. Il existe plusieurs critères fiables pour diagnostiquer l'autisme, tels que :

Critères de diagnostic selon la CIM – 10 qui comprend la présence, avant l'âge de 3 ans, d'anomalies ou d'altérations du développement, dans au moins un des domaines suivants [10] :

- (1) Langage (type réceptif ou expressif) utilisé dans la communication sociale
- (2) Développement des attachements sociaux sélectifs ou des interactions sociales réciproques
- (3) Jeu fonctionnel ou symbolique

Critères de diagnostic selon DSM-IV suggère que les symptômes autistique comprennent les anomalies des interactions sociales, de la communication et des activités pour diagnostiquer l'autisme chez un enfant [10].

le **CHAT** (Check-list for Autism in Toddlers) est un outil de dépistage qui permet de détecter les premiers signes de l'autisme chez les enfants âgés de 16 à 30 mois, il a été initialement mis au point et validé dans les années 90 par une équipe de chercheurs dont Simon Baron-Cohen. Il est constitué de deux questionnaires, l'un réservé aux parents compte 9 items qui testent plusieurs domaines de développement et l'autre au médecin ou travailleur social compte les 5 items spécifiques à l'autisme. Une nouvelle étude a révélé que trois items du CHAT peuvent prédire de façon précise un diagnostic d'autisme :

Absence de pointage proto déclaratif.

Absence de comportement d'attention conjointe.

Absence de jeu de « faire semblant ».

L'analyse des résultats suggère que si un enfant de 18 mois échoue systématiquement à ces trois items, il y a 83% de chances qu'il sera par la suite diagnostiqué autiste [31].

Le **M-CHAT** (Modified Check-list for Autism in Toddlers) ou CHAT modifié qui a été publié en avril 2001 par Diana Robins, il est actuellement considéré comme l'un des rares outils pouvant permettre un dépistage valable des signes précoces d'autisme parmi différentes populations d'enfants non sélectionnés [40].

Les six items clé qui ont été identifiés comme pouvant prédire l'autisme sont :

Item 2 : Intérêt pour les autres enfants

Item 7 : Pointage proto-déclaratif (pointer du doigt pour indiquer un intérêt dans quelque chose)

Item 9 : Amener des objets aux parents

Item 13 : Imitation

Item 14 : Réponse de l'enfant à son prénom quand il est appelé

Item 15 : Capacité à suivre le pointage de l'adulte

Si l'enfant échoue à 2 ces items ou à 3 de l'ensemble ces items, un suivi et une évaluation doivent être entrepris. Cependant, le M-Chat n'est pas conçu pour être interprété par les parents, qui effectuent les observations et notent les réponses aux items. Les parents concernés doivent consulter un service de psychologie ou de neuropédiatrie spécialisé pour l'interprétation des résultats, et notamment un "Centre de référence pour l'autisme".

ESAT (Early Screening for Autistic Traits, 2006) : est un outil de dépistage destiné aux parents constitué de 19 items portant sur les comportements précurseurs du TSA chez les enfants âgés de 14 et 15 mois. Les items se répondent par « oui » ou « non » et abordent entre autres aspects, les intérêts de l'enfant, l'expression des émotions, le contact visuel, le sourire social et l'attention conjointe. Selon Oosterling et ses collaborateurs (2009), l'ESAT présente une valeur prédictive positive de 68 % lorsque l'outil est utilisé chez des enfants âgés de 8 à 44 mois. Ces valeurs psychométriques augmentent lorsque l'outil est testé auprès d'enfants âgés de 8 à 24 mois, avec une valeur prédictive positive de 75 %. L'ESAT est publié en anglais uniquement [11].

Le **PDDST-II** (Pervasive Developmental Disorders Screening Test-II, Siegel, 2004) est un outil de dépistage développé pour identifier les enfants âgés de 18 à 48 mois à risque de présenter un TSA et de les distinguer de ceux ayant un autre trouble développemental. L'outil propose trois formes de questionnaires, à sélectionner selon le milieu clinique dans lequel il est employé. Le niveau 1 est destiné aux cliniciens des services de première ligne (médecins de famille et pédiatres) pour déterminer la pertinence d'une évaluation plus approfondie du TSA, ce niveau présente une sensibilité de 92 % et une spécificité de 91 %, ce qui en fait un outil de dépistage

intéressant pour les cliniciens qui ne sont pas spécialisés en TSA. Le niveau 2 est destiné aux spécialistes de la petite enfance et il vise à discriminer les enfants à risque de présenter un TSA de ceux à risque de souffrir d'un trouble de langage, d'un retard global de développement ou d'un retard intellectuel, niveau présente une sensibilité de 73 % et une spécificité de 49 %, et apparaît donc moins recommandé comme outil pour distinguer le TSA d'autres troubles. Le niveau 3 est considéré comme un outil diagnostique destiné aux cliniciens aptes à conclure à la présence d'un TSA (médecins et psychologues formés à cet effet), ce niveau présente une sensibilité de 58 % et une spécificité de 60 %, ce qui semble insuffisant pour estimer adéquatement le degré de sévérité du TSA. De plus, Le PDDST-II permet de déterminer, selon les signes précoces, si l'enfant est susceptible de présenter une forme légère ou sévère du trouble, il est publié uniquement en anglais [11].

Afin d'aider à l'établissement ou la confirmation du diagnostic nosologique, des questionnaires et des procédures d'observation standardisés ont été mis au point. Les plus utilisés sont :

L'ADI-R (Autism Diagnostic Interview-Revised) de Lord (1994) L'ADI-R est un entretien semi-structuré mené avec les parents, ou une personne qui s'occupe de l'enfant, permettant de compléter un premier diagnostic de l'autisme. Il s'agit d'une entrevue semi-structurée faite par un clinicien avec les parents ou tuteurs de l'enfant. Cet instrument prend environ 1½ à 2 heures à administrer, et peut être utilisé avec des enfants de 24 mois ou plus (ayant un âge mental d'au moins 18 mois). Il s'agit de recueillir le plus d'informations possible dans Les interactions sociales réciproques, la communication et le langage et les comportements stéréotypés et répétitifs [12].

L'ADOS (Autism Diagnostic Observation Schedule) est un outil fréquemment utilisé en ce qui concerne les comportements de l'enfant de 12 mois, de l'adolescent ou de l'adulte. L'ADOS est une échelle d'observation pour le diagnostic de l'autisme et se positionne comme la référence internationale dans ce domaine et pour évaluer les habiletés de communication, de socialisation et de jeu de l'enfant et l'utilisation de l'imagination par les personnes soupçonnées d'un TSA [12].

ASRS (L'Autism Spectrum Rating Scales ; Goldstein et Naglieri, 2010) est un questionnaire de dépistage des comportements autistiques pour déterminer la présence d'un TSA chez les enfants et les adolescents âgés de 2 à 18 ans. L'ASRS consiste en quatre protocoles ; une version pour les parents et une pour l'enseignant, disponible pour deux tranches d'âge, soit pour les enfants âgés de 2 à 5 ans et une pour ceux âgés de 6 à 18 ans. Chaque version est disponible en format complet et en format écourté. Les versions longues comprennent 71 questions et les formes courtes comprennent 15 items. La somme des scores obtenus réfère à trois niveaux de sévérité soit un niveau léger, un niveau élevé et un niveau très élevé. Selon les auteurs de l'outil, la fidélité test-retest se situe entre 78 % et 92 % pour les trois formes de questionnaire mais il est disponible uniquement en anglais [11].

SCQ (Social Communication Questionnaire ; Rutter, Bailey et Lord, 2003) est un questionnaire de dépistage utilisé auprès d'enfants de 4 ans et plus, présentant un âge mental d'au moins 2 ans. Il consiste en 40 items que les parents remplissent en notant la présence ou l'absence de comportements ciblés. Les items comportementaux ciblent les symptômes spécifiques au TSA, soit les anomalies de la communication, les particularités sur le plan des interactions sociales et la présence de comportements restreints, répétitifs et stéréotypés [6]. Le SCQ a été conçu pour précéder l'utilisation de l'ADI-R et, selon Corsello et ses collaborateurs (2007), lorsqu'ils sont utilisés ensemble ils identifient adéquatement les enfants avec un TSA [11].

7. Prise en charge

Il n'existe pas de traitement efficace et définitif pour l'autisme, mais il existe différentes façons d'améliorer le comportement des autistes avec des taux de réussite variables. L'intervention doit être mise en route plus tôt car un dépistage et un diagnostic précoces, ainsi que des programmes de traitement et de réadaptation appropriés, offrent des moyens plus efficaces de réduire les symptômes. En revanche, le retard dans l'intervention thérapeutique entraîne une détérioration accrue et une exacerbation des symptômes ou l'émergence d'autres symptômes, ce qui limite considérablement l'efficacité des interventions, les recherches suggérant que les interventions avant l'âge de cinq ans sont plus efficaces et ont un impact plus important sur le développement de l'enfant autiste.

Parmi les interventions, la pharmacothérapie qui est utilisée pour réguler et modifier le système nerveux à l'origine du comportement anormal et aider à réduire les taux élevés de l'excitation et de l'anxiété et réduit le comportement subversif, mais n'affecte pas les carences fondamentales et comme peut conduire à des problèmes plus graves, il convient donc de l'utiliser avec prudence. Il convient de noter que le type de médicament, la quantité de dose et la durée de prise du médicament sont la spécialité du pédiatre tel que la sécrétine pour faciliter la digestion des protéines et donc réduire le peptide opiacé qui a en relation avec l'autisme [25], le Fenfluramine a été utilisé pour diminuer le taux de sérotonine dans le sang des autistes [25], Naltrexone et Lithium. En outre, des scientifiques conseillent de donner à l'enfant une quantité supplémentaire de vitamines, en particulier la vitamine B6 qui aide certains cas d'autisme qui ont beaucoup d'activité, améliore leur comportement, les troubles du langage et du sommeil, réduit l'irritabilité et augmente l'attention [25]. Il y a aussi le régime sans gluten et Caséine car l'enfant autistes est incapable à digérer ces protéines à savoir la caséine qui est une protéine présente dans lait et de ses dérivés, tandis que le gluten se trouve dans le blé, l'orge, l'avoine, et tous les dérivés de blé, car ils souffrent d'un dysfonctionnement intestinal et gastrique, Cette digestion incomplète conduit à l'émergence des peptides [25].

De plus il y a des programmes comme la méthode de Lovaas (ABA) est l'une des méthodes les plus courantes de thérapie comportementale. Cela dépend du fait que le comportement peut être contrôlé en examinant l'environnement dans lequel il se produit et en contrôlant les facteurs qui l'induisent. Chaque comportement répond à l'effet, cette méthode comporte plusieurs séries de programmes plus faciles à réaliser et visant à développer des comportements positifs, à réduire les comportements inappropriés, à récompenser l'enfant pour tout bon comportement ou non, et à punir pour inconduite [32]. Le programme TEACCH (Traitement et éducation¹ des enfants autistes ou atteints de troubles de la communication associés) a été créée en 1966 et a développé par Eric Schupler en 1971 est une méthode d'apprentissage complète qui ne traite pas d'un seul aspect, tel que le langage ou le comportement, mais assure une formation complète des enfants, cette méthode de traitement est conçue individuellement en fonction des besoins de chaque enfant [32]. Programme d'utilisation d'images en communication (PECS) Ce programme a été créé par Andrew Bondy et Lori Frost en 1994, il est basé sur l'utilisation de la personne autiste des images des objets comme alternative à la parole. Il convient donc à la personne qui souffre de déficits linguistiques où la communication est initiée par l'échange des images qui représentent ce qu'il veut avec d'autre personne (père, mère, enseignant) [32].

L'un des traitements les plus efficaces pour les personnes autistes est l'exercice vigoureux qui signifie un entraînement aérobique de 20 minutes ou plus, 3 à 4 jours par semaine car il est associé à une diminution des comportements stéréotypiques (auto-stimulants), de l'hyperactivité, de l'agressivité, de l'automutilation et de la destructivité. Il y a aussi La Musicothérapie et la Thérapie par le jeu pour améliorer les aptitudes sociales et se débarrasser de certains comportements inacceptables des autistes [25].

Conclusion

L'autisme est l'un des troubles les plus graves car il affecte tous les aspects de la personnalité cognitive, sociale et linguistique de l'enfant et la proportion de sa prévalence augmente constamment. Les causes de l'autisme sont inconnues et peuvent être psychologiques, génétiques, neurologiques, métabolique, biochimiques ou environnementales. En plus de ses symptômes caractérisés par un trouble de relations sociales et de communication et des Comportements stéréotypés et répétitifs, et ce qui rend Le trouble plus difficile et le processus de diagnostic en raison du grand nombre de symptômes et de leur similitude avec d'autres troubles tels que le handicap mental, la surdité... etc. Ce qui nécessite de recourir à des outils de diagnostic précis et intégrés pour la détection précoce de ce trouble puis l'élaboration d'un programme de traitement et d'éducation, car il n'existe aucun moyen efficace de guérir l'autisme.

Après la présentation des informations sur l'autisme dans ce chapitre, nous aborderons dans le chapitre suivant les études les plus importantes menées pour détecter l'autisme chez les enfants à l'aide des techniques d'apprentissage automatiques.

Chapitre II : Revue de la littérature

Introduction

Bien qu'il existe très peu de travaux de recherche sur le diagnostic des TSA basé sur l'apprentissage automatique en raison de l'absence des bases de données accessibles au public, nous présenterons les principales études menées sur ce sujet dans ce chapitre. A travers ces études nous allons nous inspirer de leurs procédures et aussi les comparer avec nos résultats.

1. Etat de l'art

En 2012, Dennis P. Wall et al [13] ont utilisé des techniques d'apprentissage automatique pour étudier l'ensemble complet des réponses à l'ADI-R disponible à AGRE (Autism Genetic Research Exchange) pour 891 personnes atteintes d'autisme et 75 personnes ne répondant pas aux critères d'un diagnostic d'autisme.

Une comparaison a été réalisée entre les performances de 15 algorithmes différents d'apprentissage automatique sur l'item 93 de l'ADI-R complet. Les résultats ont montré que l'arbre de décision alternatif (ADTree) donnait les meilleurs résultats en termes de sensibilité et de spécificité de classification, avec une sensibilité parfaite de 1,0, un taux de faux positif (FPR) de 0,013 et une précision globale de 99,90%. Le classifieur ADTree a montré que 7 des 93 éléments de l'ADI-R 93% de moins que la totalité étaient suffisants pour classer l'autisme avec cette précision. La performance de ce classificateur à 7 questions a été testée sur un ensemble de 1654 personnes autistes de la Simons Foundation (SSC) et une collection de 322 personnes autistes du Boston Autism Consortium (AC). Dans les deux cas, le classificateur a effectué une analyse statistique avec une précision statistique de près de 100%.

Les données utilisées contenaient également une abondance d'enfants plus âgés, avec une densité la plus élevée entre 5 et 17 ans. Cependant, une performance de classification presque parfaite a été obtenue pour les enfants de 4 ans et moins, le plus jeune individu ayant 13 mois, mais la taille des échantillons d'enfants plus jeunes était petite.

En 2015, J A Kosmicki et al [14] ont utilisé un apprentissage automatique pour évaluer l'un des meilleurs instruments d'évaluation clinique des TSA, le plus largement utilisé, le programme d'observation diagnostique de l'autisme (ADOS), qui permet de vérifier si seul un sous-ensemble de comportements peut différencier les enfants du spectre de l'autisme.

ADOS repose sur l'observation comportementale en milieu clinique et comprend quatre modules, le module 2 étant réservé aux personnes possédant un vocabulaire limité et le module 3 aux niveaux supérieurs de fonctionnement cognitif. Huit algorithmes d'apprentissage automatique ont été exécutés en utilisant la sélection progressive des fonctionnalités sur des feuilles de résultats des modules 2 et 3 de 4540 personnes. Le résultat indique que 9 des 28 comportements capturés par les éléments du module 2, et que 12 des 28 comportements capturés par le module 3 sont suffisants pour détecter le risque de TSA avec une précision de 98,27% et 97,66%, respectivement.

Une réduction de plus de 55% du nombre de comportements avec une perte de précision négligeable dans les deux modules suggère que les méthodes informatiques et statistiques ont un rôle à jouer dans la rationalisation de la détection et du dépistage du risque de TSA.

En 2016, Anibal Sólón Heinsfel et al [18] ont fait un étude qui a pour but d'appliquer des algorithmes d'apprentissage en profondeur pour identifier les patients présentant un trouble du spectre de l'autisme à partir d'un grand ensemble de données d'imagerie cérébrale des patients atteints de TSA rs-IRMf, de la base de données mondiale multisites connue sous le nom de ABIDE qui incluent des images rs-IRMf, des images cérébrales structurales et des informations phénotypiques pour chaque patient [28]. L'étude a utilisé des données provenant de 505

individus TSA et de 530 contrôles appariés (contrôles typiques, TC), et s'est basée uniquement sur les modèles de connectivité fonctionnelle du cerveau des patients qui identifient objectivement les participants atteints de TSA et classer les sujets en TSA et TC. La connectivité fonctionnelle fournit un index du niveau de co-activation des régions du cerveau sur la base rs-IRMf puis révéler les modèles neuronaux issus de la classification.

Les données ont été sélectionnées à partir du pipeline de prétraitement C-PAC. Les données IRMf ont été corrigées sur le temps de coupe, le mouvement et l'intensité du voxel ont été normalisés. Les données fonctionnelles ont été filtrées par passe-bande (0,01–0,1 Hz) et enregistrées dans l'espace à l'aide d'une méthode non linéaire.

Le réseau neuronal profond (deep learning) a obtenu une précision de classification moyenne de 70% (sensibilité de 74%, spécificité de 63%) à partir des plis de validation croisée dans l'identification des TSA par rapport aux patients témoins dans l'ensemble de données. Les résultats montrent également que l'algorithme a surperformé les autres méthodes supervisées utilisées pour la comparaison.

En 2017, Halim Abbas et al [16] ont appliqué les techniques de Machine Learning (ML) aux données cliniques de référence obtenues auprès de milliers d'enfants à risque de trouble du spectre autistique afin de créer un outil de dépistage de l'autisme économique, rapide et facile à appliquer.

Deux algorithmes sont formés pour identifier l'autisme, l'un basé sur des questionnaires courts et structurés déclarés par les parents, et l'autre sur l'identification des comportements clés à partir de courtes vidéos maison semi-structurées d'enfants.

Nous démontrons une amélioration significative de la précision par rapport aux outils de dépistage standard dans un échantillon d'étude clinique de 162 enfants. Nous discutons ensuite du défi que représente l'extension des algorithmes d'apprentissage automatique à des conditions autres que l'autisme, et nous proposons un cadre général permettant d'utiliser des algorithmes d'apprentissage automatique pour rechercher simultanément la présence de nombreuses conditions différentes.

Une étude clinique multicentrique de $n = 162$ enfants est réalisée pour déterminer les performances de ces algorithmes et par rapport aux outils de dépistage standard. Les résultats démontrent une amélioration significative de la précision par rapport leurs combinaisons dans les mesures de l'ASC, de la sensibilité et de la spécificité.

En 2018, Mofleh Al-diabat [17] a mené une étude qui a pour objectif d'évaluer l'utilisation de l'exploration de données basée sur des règles floues qui est une approche prometteuse permettant d'automatiser le processus de dépistage des TSA et d'améliorer la précision et l'efficacité de la détection de TSA afin de réduire les coûts de soins d'autisme

Dans l'étude, les performances de l'algorithme d'exploration de données floues appelé FURIA (Fuzzy Unordered Rule Induction) ont été examinées dans la détection des traits de TSA chez les enfants. FURIA construit des modèles de dépistage de manière automatisée à partir de contrôles et de cas historiques, puis utilise ces modèles pour détecter la possibilité de traits autistiques chez de nouveaux individus, il gère les règles If-Then. Sa principale qualité est qu'il contient d'importantes parties de connaissances (règles floues) que non seulement les cliniciens et le personnel médical peuvent interpréter, mais également les membres de la famille, les enseignants et les soignants.

Puis une comparaison a été effectuée sur les performances d'algorithme FURIA par rapport à d'autres algorithmes d'exploration de données afin de révéler les avantages et les inconvénients de FURIA. Les algorithmes JRIP, RIDOR et PRISM ont été utilisés car ils génèrent des règles sous la forme d'If-Then, comme le fait FURIA, pour une comparaison équitable. De plus, il s'agit d'algorithmes d'exploration de données basés sur des règles qui ont fait leurs preuves dans différentes applications de classification,

L'ensemble de données utilisé dans l'article comprend plus de 24 variables recueillies récemment chez des enfants âgés de 4 à 11 ans à l'aide d'une application mobile appelée ASDTests, développée

en 2017 avec 292 instances dans le référentiel de données de l'UCI, mais le travail a été effectué sur une base de données mise à jour avec 227 instances enfant.

Des résultats empiriques ont révélé que les règles floues de FURIA permettaient de détecter des traits de TSA avec une précision de classification pouvant aller jusqu'à 91,35%. Et un taux de sensibilité de 91,40%; ces résultats étaient supérieurs aux autres techniques. Bien que FURIA produise un taux de spécificité acceptable, à savoir 88,09%, d'autres techniques d'exploration de données ont généré de meilleurs résultats de spécificité.

Vaishali et Sasikala ont réalisé un travail en août 2018 [15] qui visait à concevoir un modèle de prédiction TSA automatisé avec des ensembles de comportement minimum, sélectionnés dans une base de données de diagnostic TSA avec l'utilisation de wrapper, de sélection de caractéristiques Firefly binaire à objectif unique basé sur l'intelligence en essaim pour une réduction de la dimensionnalité. L'hypothèse de cette étude est de déterminer si les modèles d'apprentissage automatique formés avec des ensembles de comportements minimaux sont capables d'améliorer les performances ou non.

Dans cette étude, un ensemble de données de diagnostic TSA avec 21 fonctions obtenues à partir du référentiel d'apprentissage automatique UCI (c'est la même base de données que nous avons utilisée dans notre étude). La base de données enfants TSA est formée à 8 algorithmes différents d'apprentissage automatique utilisant une validation croisée 10 fois. Les résultats obtenus sont comparés aux résultats des modèles d'apprentissage automatique obtenus après la sélection de caractéristiques avec l'algorithme Firefly binaire.

La comparaison des résultats des modèles d'apprentissage avant et après la sélection des fonctions a montré que les 3/5 modèles d'apprentissage améliorent considérablement les performances avec des ensembles de comportements optimaux. Ce qui montre que les ensembles de comportements minimum peuvent conserver la structure de l'ensemble de données dans l'apprentissage automatique.

L'algorithme Firefly binaire a démontré que 10 caractéristiques parmi 21 caractéristiques d'TSA données suffisent à distinguer les patients atteints de TSA des patients non-TSA. Les résultats obtenus produisant une précision moyenne de 92,12% à 97,95% avec des sous-ensembles de caractéristiques optimaux approximativement égaux à la précision moyenne générée par l'ensemble de données de diagnostic de TSA.

De plus en 2018, Qandeel Tariq et al [19] ont émis une hypothèse qui stipule l'utilisation de l'analyse d'apprentissage automatique sur une vidéo enregistrée à domicile de 3 minutes, montrant des enfants avec et sans TSA, afin de parvenir à une classification rapide et précise de l'autisme par une méthode automatique.

Dans l'étude, un test a été effectué de manière prospective pour déterminer si les fonctionnalités des modèles d'apprentissage automatique optimisés peuvent être extraites par des évaluateurs sur une base de données anonymisées qui sont disponibles dans le référentiel github et concerne des courtes vidéos d'enfants autistes âgés entre 1 à 11 ans.

un portail Web mobile a été créé pour évaluer les vidéos afin d'évaluer 30 caractéristiques comportementales (contact visuel, sourire social) utilisés par 8 modèles d'apprentissage automatique indépendants pour l'identification des TSA, les résultats ont démontré que plusieurs modèles (arbres de décision en alternance, SVM, régression logistique, un noyau radial et SVM linéaire) se sont bien comportés, un classifieur à cinq caractéristiques LR (LR5) a fourni la plus grande précision 92% pour tous les âges testés, ce qui indique que le marquage des vidéos personnelles pour la classification de l'autisme par apprentissage automatique peut produire des résultats précis dans des délais très courts, à l'aide d'appareils mobiles.

En 2018 aussi, Fatiha Nur Büyükoflaz et Ali Öztürk [20] ont réalisé une étude dans son but est de mettre en place des méthodes de dépistage simples et efficaces. Dans cette étude, les performances ont été comparées à l'aide de trois méthodes de classification différentes, Naive Bayes, IBk (k-voisins les plus proches), RBFN (réseau à fonction de base radiale) et Random Forest, dans la base de données de dépistage des troubles du spectre autistique UCI 2017 pour enfants. À la suite de l'expérience, la méthode Random Forest s'est avérée plus efficace que les méthodes Naive Bayes, IBk et RBFN.

En Novembre 2018, Mirac Baris Usta et al [21] ont fait une étude, son objectif est d'examiner les facteurs prédictifs des résultats (par exemple, le QI avant traitement, les troubles psychiatriques concomitants, les compétences adaptatives et langagières, etc.) à l'aide des méthodes d'apprentissage automatique, qui sont de nouvelles méthodes de calcul, notamment l'estimation statistique, les théories de l'information et l'apprentissage mathématique, découvrant automatiquement des modèles utiles dans de grandes quantités de données.

L'étude comprenait 433 enfants (âge moyen : $72,3 \pm 45,9$ mois) diagnostiqués avec un TSA et permet de tester les performances des algorithmes d'apprentissage automatique (Naïve Bayes, Modèle linéaire généralisé, Régression logistique, Arbre de décision) sur les données, y compris les 254 éléments des formulaires de base.

Les auteurs trouvent un diagnostic précoce, une rééducation précoce, la sévérité des symptômes du TSA lors de l'évaluation initiale des résultats prévus. En outre, ils ont constaté que des diagnostics psychiatriques concomitants affectaient l'issue des symptômes de TSA en observation clinique. Les modèles d'apprentissage automatique révèlent que plusieurs autres sont plus significatifs (par exemple, l'âge des parents, le poids à la naissance, les variables sociodémographiques, etc.) en termes d'informations pronostiques et de planification du traitement des enfants atteints de TSA.

Conclusion

D'après ce chapitre nous remarquons que durant ces dix dernières années plusieurs chercheurs ont étudié plusieurs modèles d'apprentissage automatique avec des précisions différentes mais importantes et proches du diagnostic classique pour classer les enfants ayant un TSA. La plupart de ces études ont porté sur le dépistage des TSA avec des ensembles de comportement minimum et dans un délai plutôt court. Dans L'objectif principal est d'effectuer un dépistage précoce des symptômes de l'autisme pour réduire les coûts de traitement et poser un diagnostic d'autisme.

Il est donc possible de considérer les études précédentes comme un point de départ à partir duquel la recherche actuelle vise à créer un modèle de prédiction TSA automatisé avec le plus petit ensemble de comportement et une bonne précision. À partir de là, nous lancerons le chapitre suivant, qui contient nos expérimentations et les résultats obtenus, puis nous allons les comparer avec ces études précédentes pour évaluer la précision de nos algorithmes dans la détection de l'autisme chez les enfants.

Chapitre III : Résultats expérimentaux et discussions

Introduction

Après avoir présenté le trouble du spectre d'autisme avec ses types, ses causes, ses symptômes et les méthodes utilisées pour le diagnostiquer. Dans ce chapitre, nous présentons nos expériences visant à détecter le trouble de spectre de l'autisme chez les enfants par l'algorithme d'apprentissage automatique supervisé « arbre de décision » sous le langage de programmation MATLAB, en utilisant une base de données de dépistage des TSA pour les enfants avec 21 attributs et 292 observations puis nous comparons ces résultats avec les résultats d'autres algorithmes (KNN et SVM), en plus de prédire le minimum des attributs cibles pour détecter TSA à partir de cet ensemble d'attributs.

Donc, dans la première partie, nous allons fournir les méthodes et les outils que nous avons utilisés, puis dans la deuxième partie nous allons faire un prétraitement et un nettoyage des données pour préparer la base de données à la classification, par la correction des erreurs, le traitement des valeurs manquantes, la suppression des valeurs aberrantes et le codage. Ensuite, dans la troisième partie nous allons utiliser l'algorithme RELIEFF pour sélectionner les variables cibles pour les utiliser dans notre modèle de classification, et en quatrième partie nous présentons nos expérimentations avec les résultats obtenus et les discussions.

1. Méthode et Outils d'étude

Dans notre étude, nous utilisons l'arbre de décision pour la création des modèles d'apprentissage automatique afin de détecter le TSA chez les enfants, puis nous comparons ses résultats avec celles de l'algorithme de SVM et KNN pour que nous puissions les évaluer, et nous nous basons sur l'algorithme RELIEFF pour la sélection des attributs pour la détection de ce trouble avec un nombre minimum de comportements qui caractérisent TSA et la performance la plus précise. R et l'outil Matlab ont été utilisés pour construire nos modèles d'apprentissage automatique pour la classification des TSA.

1.1 L'algorithme Relief

Relief est un algorithme développé par Kira et Rendell [23] en 1992 qui utilise une méthode de filtrage pour sélectionner les entités particulièrement sensibles aux interactions entre elles. Il calcule un score pour chaque fonctionnalité, qui peut ensuite être appliqué pour classer et sélectionner les fonctionnalités les plus performantes. Ces scores peuvent également être appliqués en tant que poids caractéristiques. Mais il était limité aux problèmes de classification avec deux classes. Puis en 1994, Kononenko [23] a amélioré Relief sous le nom de ReliefF (redress pour la régression avec K plus proches voisins) qui peut traiter des problèmes de multiclassés et il est plus robuste et peut également traiter des données incomplètes et bruitées. Sous MATLAB, RANKED sont des index de colonnes dans la matrice de données en entrée X classés par importance d'attribut, ce qui signifie que RANKED (1) est l'indice du prédicteur le plus important. WEIGHT sont des poids d'attributs allant de -1 à 1 avec des poids positifs importants attribués aux attributs importants. Les rangs et les poids des attributs calculés par relèvent dépendent généralement de K.

1.2 Arbre de décision

L'arbre de décision (*Classification Tree*) est une technique d'apprentissage supervisé d'aide à la décision représentant un ensemble de choix sous la forme graphique d'un arbre, il est une méthode qui permette d'obtenir des modèles à la fois explicatifs et prédictifs. Un ensemble de données étiquetées a été utilisé pour lesquelles la valeur de la variable-cible est connue afin de construire l'arbre, puis les résultats ont été extrapolés à l'ensemble des données de test. Dans ces structures d'arbre, les feuilles représentent les valeurs de la variable-cible et les embranchements

correspondent à des combinaisons de variables d'entrée qui mènent à ces valeurs comme le montre la figure (1). En apprentissage et en fouille de données, un arbre de décision décrit les données et serait utilisé comme point de départ au processus de décision [37].

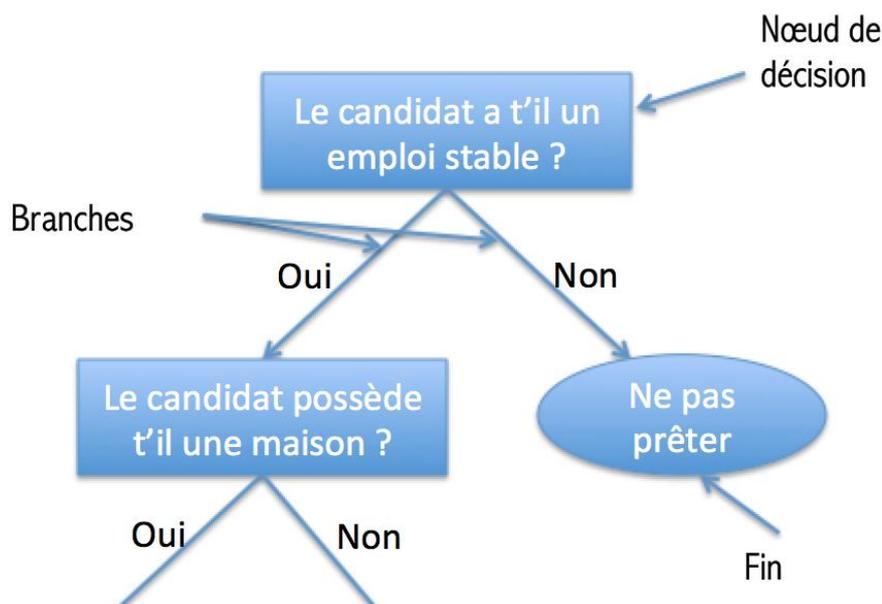


Figure (1) : Arbre de décision [41]

Parmi les avantages d'arbre décision est d'être lisible et rapide à exécuter. Il s'agit de plus d'une représentation peuvent être calculé automatiquement à partir de bases de données par des algorithmes d'apprentissage supervisé. Ces algorithmes sélectionnent automatiquement les variables discriminantes à partir de données non-structurées et potentiellement volumineuses. Ils peuvent ainsi permettre d'extraire des règles logiques de cause à effet qui n'apparaissaient pas initialement dans les données brutes, et il a la possibilité d'obtenir des règles en langage naturel [38].

Les arbres de décision sont généralement construits de haut en bas, Fondamentalement, étant donné un ensemble de données, un nœud est créé et un test T est sélectionné pour ce nœud. Chaque test induit une partition dans la base de données, chaque sous-ensemble de la partition correspondant à un seul résultat de test et contenant les éléments de données pour lesquels le test fournit ce résultat. En règle générale, le test pour lequel les sous-ensembles de la partition sont homogènes au maximum par rapport à un attribut cible (la classe) est sélectionné. Pour chaque sous-ensemble P_j de la partition P induite par T , la procédure est répétée et les nœuds créés deviennent les enfants du nœud actuel. La procédure s'arrête lorsque le critère d'arrêt réussit, c'est généralement le cas lorsqu'aucun bon test ne peut être trouvé ou lorsque la base de données est déjà suffisamment homogène. Dans ce cas, le sous-ensemble devient une feuille de l'arbre et dans cette feuille sont stockées des informations sur le sous-ensemble (telles que la classe majoritaire). Le résultat de l'appel initial de l'algorithme est l'arbre de décision complet [37].

Nous utilisons l'arbre de décision dans notre travail pour créer des modèles de classifications qui vont nous aider à visualiser nos résultats et à extraire les règles qui nous permettent de détecter les TSA chez les enfants.

1.3 ClassificationTree.fit

La fonction `ClassificationTree.fit ()` renvoie un arbre de classification basé sur les variables d'entrée (appelés prédicteurs, entités ou attributs) `X` et la sortie (réponse) `Y`. `tree` est un arbre binaire, où chaque nœud de branchement est fractionné en fonction des valeurs d'une colonne de `X` [36].

Elle effectue les étapes suivantes pour créer des arbres de décision :

- _ Commencez avec toutes les données d'entrée et examinez toutes les divisions binaires possibles sur chaque attribut.
- _ Sélectionnez une division avec le meilleur critère d'optimisation. _ Imposer la scission.
- _ Répétez de manière récursive pour les deux nœuds enfants.

L'explication nécessite deux éléments supplémentaires sont : la description du critère d'optimisation et une règle d'arrêt tel que Le nœud est pur (nœud ne contient que des observations d'une classe) [36].

Nous utilisons la fonction `ClassificationTree.fit ()` pour prédire les classes des données de test.

1.4 MATLAB

MATLAB est langage de programmation conçu par Cleve Moler à la fin des années 1970, il a émulé par un environnement de développement du même nom pour mettre en œuvre des algorithmes, l'analyse des données, leur visualisation et le calcul numérique. Il a été développé par la société The MathWorks qui est reconnu mondialement comme le principal éditeur de logiciels de calcul scientifique et technique [34]. Nous utilisons la version de matlab R2012a développé en 9 février 2012 et ses algorithmes pour programmer nos modèles de classification.

1.5 langage R

R est un langage de Programmation et un environnement pour l'informatique statistique et les graphiques, crée au courant des années 1990 par Ross Ihaka et Robert Gentleman. Il s'agit d'un projet GNU similaire au langage et environnement S qui est souvent le véhicule de choix pour la recherche en méthodologie statistique, et R fournit une moyenne open source de participer à cette activité et disponible en tant que logiciel libre [35]. Il possède de nombreuses fonctionnalités pour la modélisation, le reporting, les tests statistiques classiques, analyse de séries chronologiques, classification, regroupement et la visualisation graphique des données. Nous utilisons la version R 3.5.2 développé en 20 décembre 2018 pour la description statique, la visualisation et le prétraitement de nos données.

1.6 Android Studio

Android Studio est un nouvel environnement de développement pour développer des applications mobiles Android conseillé par Google. Il est basé sur IntelliJ IDEA et utilise le moteur de production Gradle. Il est annoncé le 15 mai 2013 lors du Google I/O et le 8 décembre 2014, il passe de version bêta à version stable 1.0 [39].

Android Studio permet principalement d'éditer les fichiers Java/Kotlin et les fichiers de configuration XML d'une application Android. Il propose des outils pour gérer le développement d'applications multilingues et permet de visualiser rapidement la mise en page des écrans sur des

écrans de résolutions variées simultanément. De plus, Il intègre par ailleurs un émulateur permettant de faire tourner un système Android virtuel sur un ordinateur [39].

Android Studio permet de voir chacun des changements visuels que les développeurs effectuez sur leurs applications et en temps réel, ils pourront voir aussi son effet sur différents appareils Android. En outre, il fournit des assistants et des modèles pour les éléments communs trouvés sur tous les programmeurs Android, et il présente un éditeur complet avec une panoplie d'outils pour accélérer le développement de votre application.

Nous utilisons la version Android Studio 3.0.1 développé en 9 novembre 2017 pour la programmation de notre application mobile qui permet de faire la détection des TSA chez les enfants autistes, en basant sur les résultats de nos modèles de classifications.

2. Préparation de la base de données

2.1 Description de la base de données

L'ensemble de données que nous avons choisi, intitulé « Autistic Spectrum Disorder Screening Data for Children », est un ensemble de données open source disponible dans le référentiel UCI Machine Learning. Il a été donné au référentiel le 24 décembre 2017 par M. Fadi Fayez Thabtah, département de la technologie numérique, Manukau Institute of Technology, Auckland, Nouvelle-Zélande [33].

L'ensemble de données brutes comprend 292 observations de 21 variables, y compris la classe cible. Ces données décrivent les résultats du dépistage des TSA, dont certains semblent avoir été récoltés à partir d'une application développée par Dr.Fadi Fayez appelée «ASDTests», disponible pour les appareils Android et iOS, et il a été développé en 2017 pour accélérer le dépistage des TSA chez différents groupes cibles, y compris les tout-petits, les enfants, les adultes et les adolescents. Dans notre étude, l'accent est mis sur les instances liées à la catégorie des enfants âgés entre 4 à 11 ans.

Parmi les 21 attributs il y a dix attributs binaires représentant les questions de dépistage (A1_Score à A10_Score) générées par l'outil AQ-10-child, ainsi neuf variables catégorielles (gender, ethnicity, jaundice, austim, country_of_res, used_app_before, age_desc, relation et Class / ASD). Il existe également deux variables numériques nommées age et result.

2.2 Description des attributs

La description officielle des attributs est au-dessous selon [22]. Nous avons inclus les questions réelles associées aux attributs A1_Score à A10_Score qui demandent une réponse binaire simple, soit d'accord ou pas.

Attribute	Type	Description
Age	Number	Age en années
Gender	String	Masculine ou feminine
Ethnicity	String	Liste des ethnies communes au format text
Jaundice	Boolean	Si le cas est né avec la jaunisse
Austim	Boolean	Si un membre de la famille immédiate a été diagnostiqué avec un autisme
contry_of_res	String	Liste des pays de résidence au format text
used_app_before	Boolean	Si l'utilisateur a utilisé l'application de dépistage avant le test pour le même enfant
Result	Integer	Score final obtenu sur la base de l'algorithme de notation de la méthode de dépistage utilisée
age_desc	String	Type de méthode de dépistage choisi en fonction de

		la catégorie d'âge
Relation	String	qui remplit le test
Class/ASD	Boolean	Le résultat final est si la personne a un TSA ou no
Question 1 Answer	Binaire	Il/elle remarque souvent de petits sons quand d'autres ne le remarquent pas
Question 2 Answer	Binaire	Il/elle concentre généralement plus sur l'ensemble de l'image plutôt que sur les petits détails
Question 3 Answer	Binaire	Il/elle peut facilement suivre plusieurs conversations à la fois
Question 4 Answer	Binaire	Il/elle trouve facile d'entreprendre plusieurs activités à la fois
Question 5 Answer	Binaire	Il/elle est incapable de maintenir une conversation avec ses ami(e)s
Question 6 Answer	Binaire	Il/elle est à l'aise dans une conversation sociale (bavardage)
Question 7 Answer	Binaire	Quand il/elle lit une histoire, il/elle trouve difficile de déterminer les intentions ou les sentiments des personnages
Question 8 Answer	Binaire	À l'école maternelle, il/elle aimait jouer aux jeux impliquant de (faire semblant) avec les autres enfants
Question 9 Answer	Binaire	Il/elle trouve facile de devenir ce qu'une personne pense ou comment il se sent en regardant simplement son visage
Question 10 Answer	Binaire	Il/elle a la difficulté à se faire de nouveaux amis

Table 1 : les attributs et leurs descriptions

2.3 Exploration de la base de données

La première étape de l'analyse de la base de données brute consiste à explorer. Nous pouvons peut-être découvrir les modifications à apporter qui faciliteront l'utilisation du base de données et pourraient potentiellement accroître la précision des résultats.

Ci-dessous, il y a les cinq premières lignes de l'ensemble de données. Cela donne un bon aperçu de ce à quoi ressemble une observation typique à partir des données.

A1_Score	A2_Score	A3_Score	A4_Score	A5_Score	A6_Score	A7_Score	A8_Score	A9_Score	A10_Score
1	1	0	0	1	1	0	1	0	0
1	1	0	0	1	1	0	1	0	0
1	1	0	0	0	1	1	1	0	0
0	1	0	0	1	1	0	0	0	1
1	1	1	1	1	1	1	1	1	1

age	gender	Ethnicity	jundice	austim	contry_of_res	used_app_before	result	age_desc	relation	Class/ASD
6	M	Others	No	No	Jordan	No	5	'4-11 years'	Parent	NO
6	M	'Middle Eastern '	No	No	Jordan	No	5	'4-11 years'	Parent	NO
6	M	?	No	No	Jordan	Yes	5	'4-11 years'	?	NO
5	F	?	Yes	No	Jordan	No	4	'4-11 years'	?	NO
5	M	Others	Yes	No	'United States'	No	10	'4-11 years'	Parent	YES

Table 2 : les cinq premières lignes de l'ensemble de données

On a extrait les noms des attributs par l'expression suivante :

```
> names(dataASD)
[1] "A1_Score"      "A2_Score"      "A3_Score"      "A4_Score"
[5] "A5_Score"      "A6_Score"      "A7_Score"      "A8_Score"
[9] "A9_Score"      "A10_Score"     "age"           "gender"
[13] "ethnicity"     "jundice"       "austim"        "contry_of_res"
[17] "used_app_before" "result"       "age_desc"      "relation"
[21] "Class.ASD"
```

Figure (2) : les noms des attributs de la base de données

En utilisant la fonction (summary), on peut voir quelques statistiques descriptives de base pour les attributs, ou on peut remarquer qu'il y a un certain nombre de valeurs manquantes dans age, ethnicity et relation. En outre il existe une catégorie dupliquée dans l'attribue relation comme le montre l'image au-dessous :

```
> summary(dataASD)
  A1_Score   A2_Score   A3_Score   A4_Score   A5_Score   A6_Score   A7_Score
Min.   :0.0000  Min.   :0.0000  Min.   :0.0000  Min.   :0.0000  Min.   :0.0000  Min.   :0.0000  Min.   :0.0000
1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000
Median :1.0000  Median :1.0000  Median :1.0000  Median :1.0000  Median :1.0000  Median :1.0000  Median :1.0000
Mean   :0.6336  Mean   :0.5342  Mean   :0.7432  Mean   :0.5514  Mean   :0.7432  Mean   :0.7123  Mean   :0.6062
3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:1.0000
Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000

  A8_Score   A9_Score   A10_Score   age   gender   ethnicity   jundice   austim
Min.   :0.0000  Min.   :0.0000  Min.   :0.0000  4     :92  f: 84  White-European  :108  no :212  no :243
1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000  5     :45  m:208  Asian           : 46  yes: 80  yes: 49
Median :0.0000  Median :0.0000  Median :1.0000  6     :39  ?      ?              : 43
Mean   :0.4966  Mean   :0.4932  Mean   :0.726  7     :27  'Middle Eastern ' : 27
3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:1.0000  11    :26  'South Asian'    : 21
Max.   :1.0000  Max.   :1.0000  Max.   :1.0000  8     :21  Black           : 14
                                     (Other):42  (Other)         : 33

  contry_of_res  used_app_before  result  age_desc  relation  Class.ASD
'United Kingdom': 49  no :281  Min.   : 0.00  '4-11 years':292  'Health care professional': 13  NO :151
'United States' : 42  yes: 11  1st Qu.: 5.00  ?              ?              : 43  YES:141
India           : 42  ?      Median : 6.00  Parent         :214
Australia       : 23  ?      Mean   : 6.24  Relative       : 17
Jordan          : 20  ?      3rd Qu.: 8.00  self           : 1
'New Zealand'  : 13  ?      Max.   :10.00  Self           : 4
(Other)        :103
```

Figure (3) : statistiques descriptives de base de données

2.3.1 Les niveaux des attributs

Les niveaux de chaque attribut sont exprimés ci-dessous :

```
> str(dataASD)
'data.frame': 292 obs. of 21 variables:
 $ A1      : int  1 1 1 0 1 0 1 1 1 0 ...
 $ A2      : int  1 1 1 1 1 0 0 1 1 0 ...
 $ A3      : int  0 0 0 0 1 1 1 1 1 1 ...
 $ A4      : int  0 0 0 0 1 0 1 1 1 1 ...
 $ A5      : int  1 1 0 1 1 1 1 1 1 1 ...
 $ A6      : int  1 1 1 1 1 1 1 1 1 0 ...
 $ A7      : int  0 0 1 0 1 0 0 1 1 1 ...
 $ A8      : int  1 1 1 0 1 1 1 1 0 1 ...
 $ A9      : int  0 0 0 0 1 0 0 0 0 0 ...
 $ A10     : int  0 0 0 1 1 1 1 0 0 0 ...
 $ age     : Factor w/ 9 levels "?", "10", "11", ...: 6 6 6 5 5 4 5 5 3 3 ...
 $ gender  : Factor w/ 2 levels "f", "m": 2 2 2 1 2 2 2 1 1 1 ...
 $ ethnicity : Factor w/ 11 levels "'Middle Eastern '", ...: 8 1 3 3 8 3 11 1 1 3 ...
 $ jaundice : Factor w/ 2 levels "no", "yes": 1 1 1 2 2 1 1 1 1 1 ...
 $ autism  : Factor w/ 2 levels "no", "yes": 1 1 1 1 1 2 1 1 1 2 ...
 $ contry  : Factor w/ 52 levels "'Costa Rica'", ...: 33 33 33 33 10 23 9 16 16 15 ...
 $ used_app_before: Factor w/ 2 levels "no", "yes": 1 1 2 1 1 1 1 1 1 1 ...
 $ result  : int  5 5 5 4 10 5 7 8 7 5 ...
 $ age_desc : Factor w/ 1 level "'4-11 years'": 1 1 1 1 1 1 1 1 1 1 ...
 $ relation : Factor w/ 5 levels "'Health care professional'", ...: 3 3 2 2 3 2 3 3 3 2 ...
 $ Class_ASD : Factor w/ 2 levels "NO", "YES": 1 1 1 1 2 1 2 2 2 1 ...
```

Figure (4) : les niveaux des attribues

La vérification des niveaux de nos attributs catégorielles montre que les attributs relation, ethnicity et contry_of_res ont plus de 2 niveaux. Et les valeurs manquantes symbolisent par « ? »

2.3.2 Correction des erreurs existantes dans la base de données

La vérification de structure de l'ensemble de données montre que les noms des attributs contiennent des erreurs d'orthographe dans (jundice, austim, contry_of_res) et que l'attribue Class / ASD causera des problèmes en raison du caractère de la barre oblique illégal.

Nous allons d'abord corriger certains noms des attributs pour améliorer la lisibilité. Nous pouvons également corriger les erreurs d'orthographe / caractères non autorisés dans les attributs (austim à autism, jundice à jaundice, contry_of_res à contry et Class / ASD à Class _ASD) et nous changeons les noms des attributs A1_Score à A10_Score en A1 à A10

```
> names(dataASD)
 [1] "A1"      "A2"      "A3"      "A4"      "A5"
 [6] "A6"      "A7"      "A8"      "A9"      "A10"
[11] "age"     "gender"  "ethnicity" "jaundice" "autism"
[16] "contry"  "used_app_before" "result"    "age_desc" "relation"
[21] "Class_ASD"
```

Figure (5) : les noms des attributs après les changements effectués

2.3.3 La proportion de l'ensemble de données de la classe (class_ASD)

Comme nous allons essayer de classer si un individu peut ou non être atteint de TSA, il serait utile de connaître la proportion de l'ensemble de données classée Yes ou No. Nous pouvons voir ici que 48.29% (141 cas) des observations présentes ont été classées comme potentiellement atteintes de TSA, et 51.71% (151 cas) étant classées comme n'ayant pas de TSA. Cela indique qu'il existe peu de biais en faveur de la classe No d'attribut cible dans cet ensemble de données.

2.4 Nettoyage des données

Maintenant que nous avons transformé la base de données en un état beaucoup plus facilement lisible, nous pouvons commencer à nettoyer les données des informations indésirables pour les préparer à une utilisation avec l'algorithme d'apprentissage automatique que nous avons choisis.

2.4.1 Valeurs Manquantes

La sortie ci-dessous affiche un nombre d'observations avec des valeurs manquantes pour chaque attribut.

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
0	0	0	0	0	0	0	0	0	0

age	gende	ethnicit	Jaundic	Autis	contr	used_app_befor	resul	age_des	relatio	Class_AS
4	0	43	0	0	0	0	0	0	43	0

Table 3 : le nombre des valeurs manquantes existent dans chaque attribut

Il y en a 43 valeurs manquantes pour les attributs ethnicity et relation, et 4 valeurs manquantes pour l'attribut age. Ces valeurs manquantes sont principalement catégoriques. Cela rend difficile la création de valeurs de remplacement car on ne peut pas remplacer la moyenne ou la médiane pour des attributs non numériques.

La meilleure solution consiste à supprimer toutes les observations contenant des valeurs manquantes.

Les observations qui contiennent des valeurs manquantes sont les mêmes dans les attributs ethnicity et relation et 3 observations d'attribue age, mais il y a une observation contient une valeur manquants seulement dans age et parce que c'est la seule observation, nous pouvons le supprimer aussi car n'affecte pas sur l'apprentissage, donc on a 43 observations plus une observation qu'indique 44 observations à supprimer

La sortie ci-dessous indique le nombre de valeurs manquantes dans la base de données et confirme qu'il n'y a plus d'NAN.

```
> dim(dataASD)
[1] 292 21
> dataASD1=na.omit(dataASD)
> dim(dataASD1)
[1] 248 21
```

Figure (6) : la base de données avant et après la suppression des valeurs manquantes

Le comptage des observations présentes dans la nouvelle base de données sans valeurs manquantes révèle qu'il y a maintenant 248 observations, confirmant que nous avons effectivement perdu 44 observations.

2.4.2 les doublons

Il existe également une catégorie dupliquée dans l'attribut relation. Les catégories Self et self sont évidemment destinées à être une valeur.

La meilleure solution consiste à regrouper les deux valeurs en un seul niveau de facteur comme le montre la sortie ci-dessous

```

> levels(dataASD$relation)
[1] "'Health care professional'" "?"           "Parent"
[4] "Relative"                "self"      "Self"
> levels(dataASD$relation)=gsub("self","Self",levels(dataASD$relation))
> levels(dataASD$relation)
[1] "'Health care professional'" "?"           "Parent"
[4] "Relative"                "Self"

```

Figure (7) : Regroupement des valeurs Self et self en un seul niveau de facteur

Donc après les modifications et les nettoyages des données manquantes mise en œuvre, nous avons eu à la fin une base de données contenant 248 observations de 21 attribues.

2.5 Visualisation des données

Après le traitement des valeurs manquantes et les doublons, nous pouvons maintenant commencer à explorer visuellement les données avec lesquelles nous allons travailler.

2.5.1 La boîte à moustaches – AGE

La boîte à moustaches ci-dessous représente la répartition par âge des mâles et des femelles et leur classe (ASD YES / ASD NO). Nous pouvons voir que l'âge varie de 4 ans au 11 ans pour les deux sexes. Les femelles semblent avoir une distribution plus large d'individus classés comme ayant un TSA.

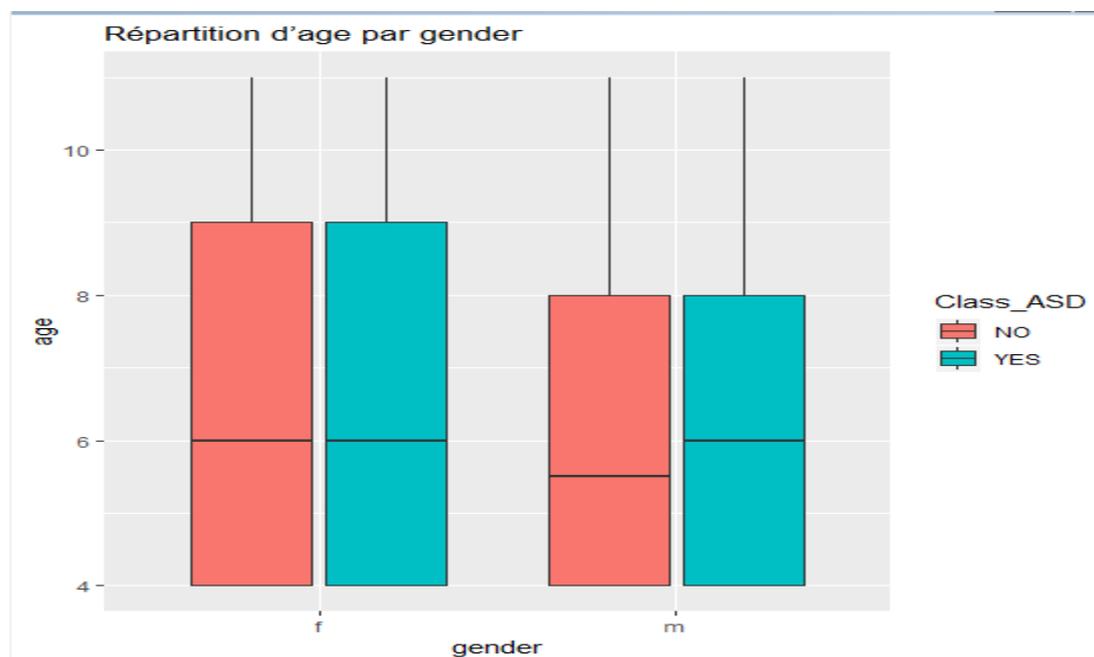


Figure (8) : La boîte à moustaches – AGE

2.5.2 La boîte à moustaches - Result

La boîte à moustaches ci-dessous représente la distribution des scores de résultats de dépistage pour les mâles et les femelles et leur classe (ASD YES / ASD NO). Nous pouvons constater qu'un score d'environ 7 correspond à un individu considéré comme ayant un TSA pour les deux sexes.

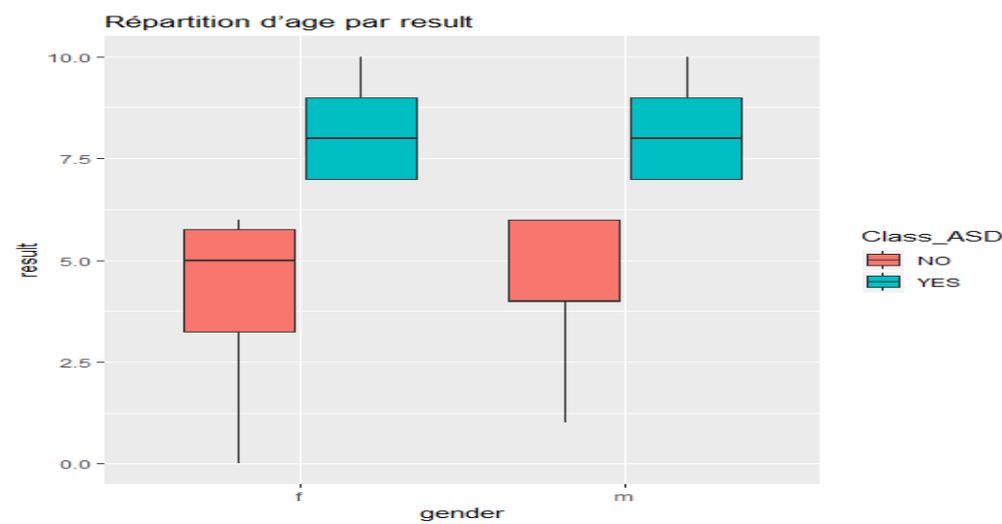


Figure (9) : La boîte à moustaches – Result

L'interrogation des données confirme qu'un score de 7 ou plus a pour résultat qu'un individu est classé comme ayant un TSA. Nous calculons le nombre de result égal ou supérieur à 7 et le nombre d'observations avec Class_ASD = YES et Class_ASD =NO. Le résultat ci-dessous confirme qu'ils sont exactement les mêmes.

```
> summary(dataASD$Class_ASD)
NO YES
151 141
> summary(dataASD$result>=7)
Mode FALSE TRUE
logical 151 141
```

Figure (10) : Comparaison entre la valeur de result >= 7 et la classe Class_ASD

2.5.3 Diagramme en bâton - Ethnicity

Le graphe au-dessous représente les proportions de chaque ethnie présente dans les données. Les Européens blancs représentent environ un tiers des données, suivis des Asiatiques et des peuples du Moyen-Orientale.

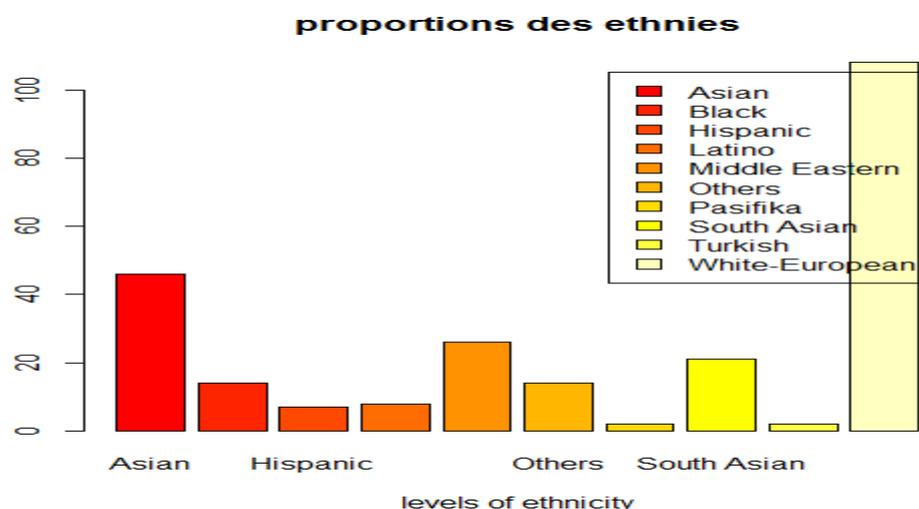


Figure (11) : Diagramme d’Ethnicity

2.5.4 Diagramme en bâton - Jaundice

Le graphe au-dessous représente la relation entre un diagnostic de jaunisse à la naissance et le fait d’être un enfant classé comme ayant un TSA. Nous remarquons qu’il n’y a pas de relation significative entre les deux présents dans cet ensemble de données parce que le nombre des personnes ont diagnostiqué une jaunisse à la naissance est approximativement le même pour une classification ASD = YES et ASD = NO.

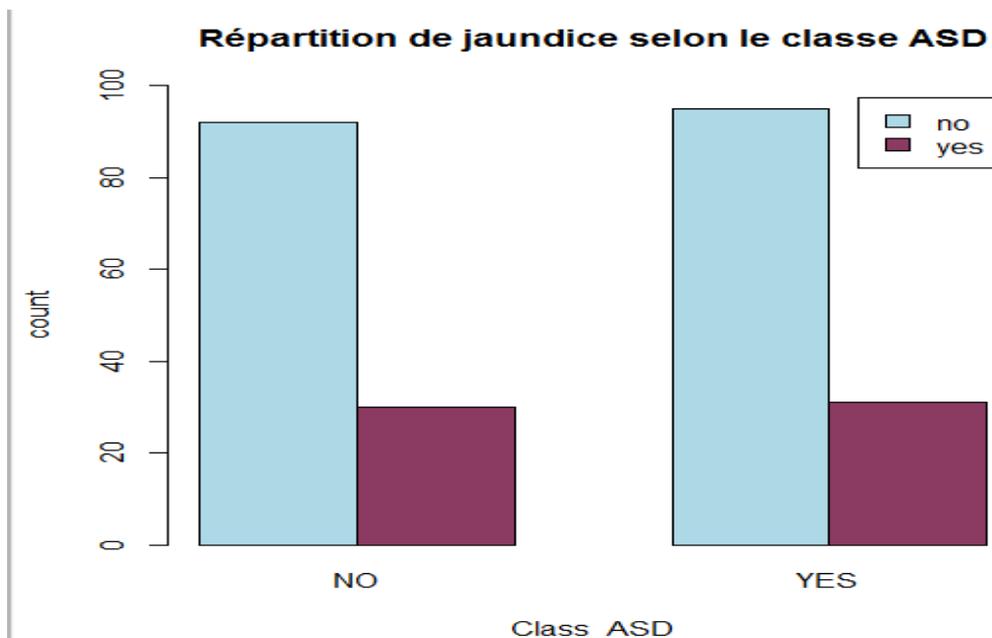


Figure (12) : Diagramme en bâton – Jaundice

2.5.5 Diagramme en bâton - autism

Le graphe au-dessous représente la relation entre les enfants dont un membre de ses familles est atteint de TSA et classés comme tels. Il ne semble pas exister de relation significative entre les deux attributs présentes dans cet ensemble de données, car le nombre des enfants ayant un membre de ses familles ayant un TSA est approximativement le même pour une classification ASD = YES et ASD = NO.

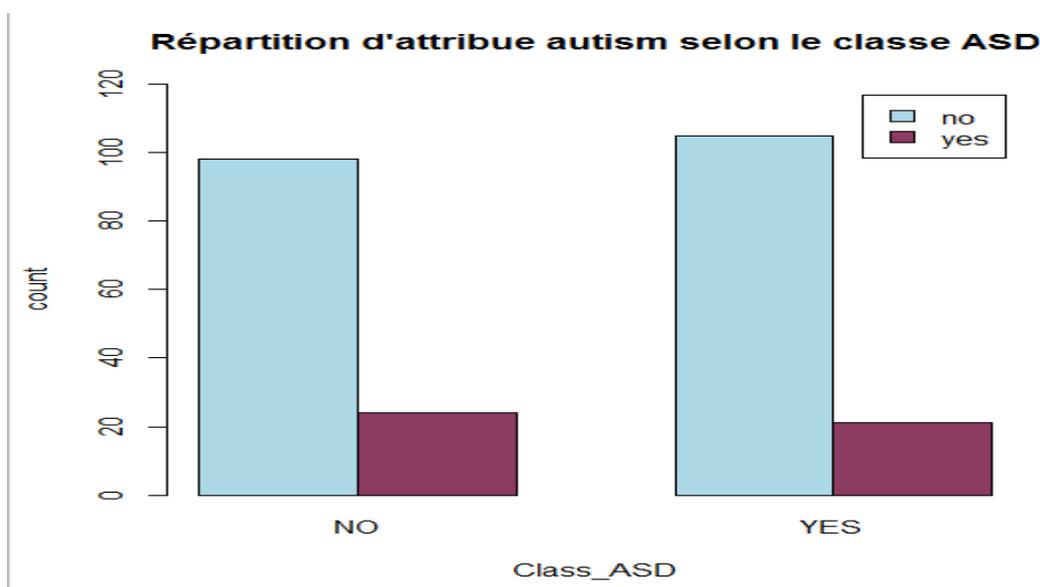


Figure (13) : Diagramme en bâton – autism

2.6 Prétraitement des données

2.6.1 Réduction des attributs

Un certain nombre des attributs présents dans l'ensemble de données n'offrent aucun avantage à notre analyse.

contry_of_res : Ce attributs a 52 niveaux, Ce qui est trop pour que l'algorithme d'apprentissage puisse le gérer et il n'affiche aucune importance significative pour l'attribut cible Class_ASD. Donc, pour assurer la compatibilité, il sera supprimé.

used_app_before : Cet attributs indique si un enfant a utilisé l'application de dépistage ou non. Ce n'est pas un indicateur significatif de notre attribut cible, donc cela sera également supprimé.

age_desc : Cet attribut permet de classer la tranche d'âge d'un individu. Toute personne de plus de 4 ans et moins de 11 ans est classée comme enfant. Nous remarquons que tous les observations dans la base de données âgés entre 4 et 11 ans, ce qui apparait que ce facteur ne comporte qu'un seul niveau et ne peut donc offrir aucun avantage significatif pour l'analyse.

Result : comme indiqué La boîte à moustaches - Result, une valeur de résultat de 7 ou plus sera toujours classée dans Class_ASD = YES. Par conséquent, l'inclusion de cet attribut signifie que les algorithmes d'apprentissage automatique auraient essentiellement déjà le résultat du attribut cible, Pour cela, il sera également supprimé.

Relation : Cet attribut montre qui remplit le test de dépistage. Ce n'est pas un indicateur significatif de notre variable cible, donc cela sera également supprimé.

Après la suppression des attributs non significatifs, il reste que les suivant :

```
> names(dataASD)
 [1] "A1"          "A2"          "A3"          "A4"          "A5"
 [6] "A6"          "A7"          "A8"          "A9"          "A10"
[11] "age"         "gender"      "ethnicity"   "jaundice"    "autism"
[16] "contry"     "used_app_before" "result"     "age_desc"    "relation"
[21] "Class_ASD"
> dataASD1=subset(dataASD, select=-c(contry,used_app_before,result,age_desc,relation))
> names(dataASD1)
 [1] "A1"          "A2"          "A3"          "A4"          "A5"          "A6"          "A7"          "A8"
 [9] "A9"          "A10"         "age"         "gender"     "ethnicity"   "jaundice"   "autism"     "Class_ASD"
> dim(dataASD1)
 [1] 248 16
> |
```

Figure (14) : la base de données avec et sans les attributs non significatifs

Nous avons eu à la fin une base de données contenant 248 observations de 16 attributs avec la classe.

2.6.2 Codage des données

Le codage des valeurs des attributs permet de convertir des attributs qualitatifs non ordinaux en données numériques. Cela peut être une étape de prétraitement nécessaire en fonction du type de classification à effectuer parce que certains algorithmes ne pouvant pas gérer de données catégorielles. Ce codage au début du processus d'apprentissage peut donner plus d'options d'analyse si on en avait besoin.

Le codage est fait par prendre les niveaux d'un attribut et donner à chacun d'eux son propre attribut numérique où Les auteurs de la base de données ont attribué la valeur 0 ou 1 en fonction

de la réponse donnée lors du test de dépistage par le participant. En particulier, pour les questions 1, 5, 7, 10, la valeur «1» est attribuée à la fonction lorsque le participant répond «définitivement d'accord», tandis que la valeur «1» sera attribué à «définitivement en désaccord» pour les questions 2, 3, 4, 6, 8 et 9. De plus, un 2 indique une réponse positive et un 1 une réponse négative (oui / non). Et ce qui concerne un attribut catégorique composé de plusieurs niveaux, un codage donne à chaque niveau son propre attribut.

Dans le cas des attributs binaires telles que le genre, nous pouvons le coder par donne la valeur 1 pour indiquer que l'observation est une femme et un 2 pour indique qu'il s'agit d'un homme.

3. Sélection des attributs Par l'algorithme Relieff

En utilisant l'algorithme Relieff avec k voisins les plus proches par classe, pour classer les attributs selon leurs importances, Relieff renvoie les rangs et les poids des prédicteurs pour les données d'entrée X qui représentent les 14 attributs (de A1 à autism) et le vecteur de classe y qui représente l'attribut Class_ASD, nous changeons à chaque fois le nombre de K et nous obtenons les résultats dans le tableau ci-dessous :

Nbr de K plus proches voisins par classe	les rangs (les attributs classés du plus important au moins important)	le poids de chaque attribut
10	4 10 8 1 5 9 3 7 11 6 13 14 2 15 12	0.1004 -0.0235 0.0366 0.1596 0.0963 0.0011 0.0149 0.1021 0.0399 0.1200 0.0029 -0.0715 -0.0074 -0.0181 - 0.0470
20	4 10 5 8 1 9 3 7 6 2 13 11 15 14 12	0.0948 -0.0085 0.0660 0.1747 0.1209 -0.0065 0.0440 0.1078 0.0671 0.1291 -0.0235 -0.0510 -0.0119 -0.0492 - 0.0454
40	4 10 5 9 8 1 3 7 6 2 13 11 15 14 12	0.0888 -0.0016 0.0669 0.2129 0.1163 0.0244 0.0477 0.1016 0.1083 0.1328 -0.0208 -0.0656 -0.0061 -0.0539 - 0.0530
80	4 10 9 5 8 1 3 7 6 2 13 11 14 15 12	0.0894 0.0048 0.0718 0.2355 0.1157 0.0442 0.0551 0.1117 0.1191 0.1300 -0.0231 -0.0662 -0.0095 -0.0516 - 0.0538
140	4 10 9 5 8 1 3 7 6 2 13 11 14 15 12	0.0903 0.0055 0.0742 0.2389 0.1161 0.0474 0.0575 0.1142 0.1230 0.1306 -0.0226 -0.0648 -0.0091 -0.0507 - 0.0528

Table 4 : les rangs et les poids des attributs en fonction de nombre de K

Nous observons qu'il existe des cas sélectionnent 8 attributs avec un poids positives et le reste avec un poids négatives, et des cas sélectionnent 9 attributs avec un poids positives et le reste avec un poids négative, et des autres cas sélectionnent 10 attributs avec un poids positives et le reste avec un poids négative. Mais dans la plus part des cas il y a 9 attributs qui ont toujours des poids positives et élevé, ils sont les attributs A4, A10, A9, A5, A8, A1, A3, A7 et A6 dans l'ordre décroissante des poids.

Nous avons discuté avec la docteure en psychopathologie de l'enfant Madame BENDIOUIS SARAH, elle nous a confirmé la validité de nos résultats obtenus et l'importance de ces 9 attributs et elle nous a informés qu'ils sont suffisants pour la détection de TSA chez les enfants. Donc, nous pouvons garder seulement ces 9 attributs pour les utilisés dans notre modèle de classification avec un minimum de sous-ensembles de caractéristiques pour détecter TSA chez les enfants plus rapidement que les méthodes de diagnostic traditionnelles. De plus, nous allons nous appuyer seulement sur ces attributs pour créer notre propre application mobile de dépistage de TSA chez les enfants.

4. Résultats expérimentaux et discussion

Dans cette partie, nous allons travailler avec trois algorithmes d'apprentissage automatique pour la classification : KNN, SVM et arbre de décision sur la base de données initiale avec 15 attributs puis nous comparons leurs résultats obtenus sur la base de données après la sélection des attributs par RELIEFF avec 9 attributs

4.1 Classification sur la base de données initiale avec 15 attributs et 248 observations

4.1.1 Résultats d'algorithme ClassificationTree avec validation croisée (10-fold)

Nous avons faire 3 essais et dans chaque essai nous avons obtenus 10 arbres, chacun a sa propre taux, puis nous comparons entre les taux moyennes des essais :

L'arbre	1er essai			2 ème essai			3 ème essai		
	Tc (%)	Sn (%)	Sp (%)	Tc (%)	Sn (%)	Sp (%)	Tc (%)	Sn (%)	Sp (%)
1	80.00	88.89	75.00	88.00	84.62	91.67	92.31	86.67	100
2	92.00	85.71	100	95.83	100	92.31	84.00	83.33	84.62
3	91.67	100	85.71	88.00	100	81.25	87.50	84.62	90.91
4	95.83	92.31	100	88.00	91.67	84.62	84.00	90.00	80.00
5	76.92	73.33	81.82	88.00	84.62	91.67	84.00	78.57	90.91
6	88.00	90.91	85.71	92.00	85.71	100	95.83	92.31	100
7	91.67	91.67	91.67	95.83	92.31	100	88.00	84.62	91.67
8	88.00	84.62	91.67	84.00	84.62	83.33	80.00	83.33	76.92
9	80.00	73.33	90.00	96.00	92.31	100	83.33	83.33	83.33
10	84.00	84.62	83.33	88.00	90.91	85.71	95.83	100	92.31

Table 5 : Résultats de classificationTree sur la base de données initiale

L'essai	1 ère essai	2 ème essai	3 ème essai
Tc moy (%)	86.81	90.37	87.50
Sn moy (%)	86.54	90.68	86.68
Sp moy (%)	88.49	91.06	89.07
Temps de calcul (s)	5.14943	4.53096	3.06954

Table 6 : le taux moyen de classificationTree pour chaque essai sur la base de données initiale

D'après les tableaux précédents montrant nos 3 essais, le taux moyen d'algorithme classificationTree varie de 86.82% à 90.37 pour la base de données initiale, et dans chaque essai nous avons des arbres avec des taux différents. L'arbre 9 dans le deuxième essai a donné le taux de précision le plus élevé de 96.00%, Sn=92.31% et Sp=100%. Parmi ses règles :

- _ Si l'enfant trouve facile d'entreprendre plusieurs activités à la fois est certainement non.
- _ Et Si l'enfant a la difficulté à se faire de nouveaux amis est certainement oui.
- _ Et Si quand l'enfant lit une histoire, il/elle trouve difficile de déterminer les intentions ou les sentiments des personnages est certainement oui.
- _ Alors le diagnostic est TSA.

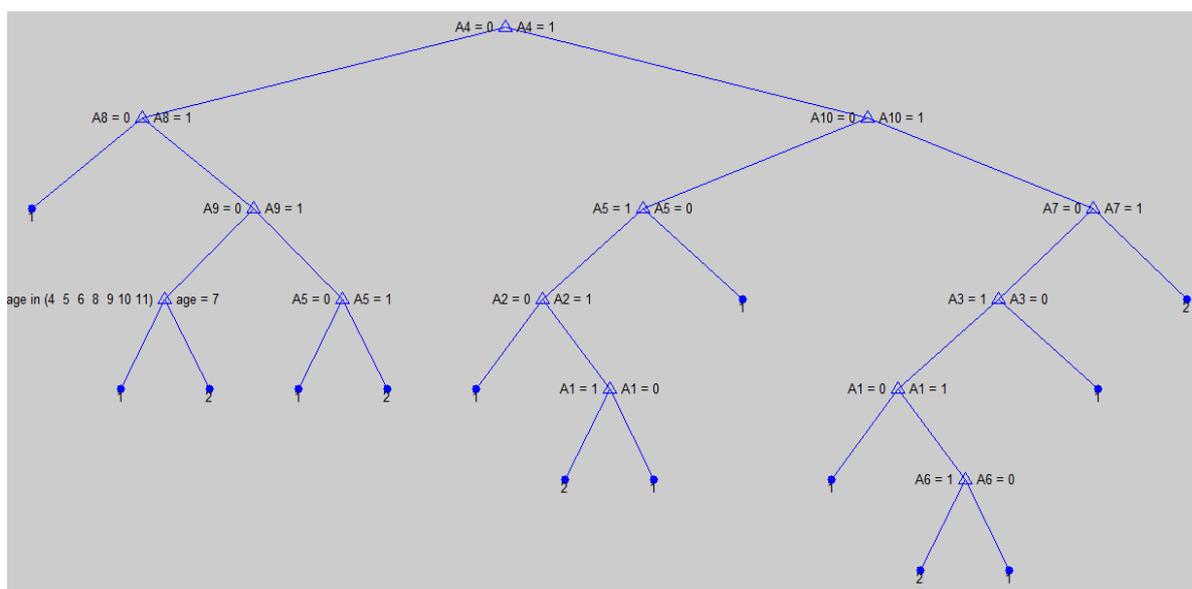


Figure (15) : Arbre de décision avec un Tc=96.00%, Sn=92.31% et Sp=100%.

4.1.2 Résultats d'algorithme KNN avec validation croisée (10-fold)

Nous avons faire 3 essais pour chaque distance (Euclidienne, Cityblock, Cosine, Correlation) avec différentes nombre de voisins K, et nous avons eu les résultats suivants :

Euclidienne k	1 er essai			2 ème essai			3 ème essai		
	Tc moy (%)	Se moy (%)	Sp moy (%)	Tc moy (%)	Se moy (%)	Sp moy (%)	Tc moy (%)	Se moy (%)	Sp moy (%)
3	82.94	71.68	93.82	82.27	69.00	95.37	83.54	72.14	94.48
4	81.92	68.74	94.69	82.14	69.27	94.41	81.03	70.00	91.77
5	81.80	68.61	94.53	84.87	72.92	69.28	81.96	66.30	96.74
6	79.06	62.55	94.98	82.94	70.67	94.91	82.83	66.14	98.61
8	82.65	65.81	99.16	79.84	63.15	69.11	77.69	59.56	95.06
10	77.11	56.15	97.43	78.12	59.85	95.80	80.22	63.44	96.22
20	80.94	62.00	99.33	78.60	57.97	98.45	84.93	70.09	99.14

Table 7 : Résultats de KNN avec distance Euclidienne sur la base de données initiale

Cityblock	1 er essai			2 ème essai			3 ème essai		
Nbr voisins k	Tc moy (%)	Se moy (%)	Sp moy (%)	Tc moy (%)	Se moy (%)	Sp moy (%)	Tc moy (%)	Se moy (%)	Sp moy (%)
3	85.60	72.79	98.07	85.25	76.14	94.16	85.55	76.15	94.74
4	86.30	76.39	95.92	86.31	73.57	98.95	83.84	71.59	95.69
5	81.83	68.88	94.44	81.45	65.00	97.50	77.28	57.87	95.95
6	81.56	65.45	97.10	87.00	74.81	98.79	82.71	66.16	98.97
8	83.06	67.19	98.46	84.93	69.22	100	85.31	70.34	100
10	84.11	67.75	100	86.12	71.69	100	82.88	65.83	100
20	83.53	66.00	100	86.74	73.15	100	83.94	67.52	100

Table 8 : Résultats de KNN avec distance Cityblock sur la base de données initiale

Cosine	1 er essai			2 ème essai			3 ème essai		
Nbr voisins k	Tc moy (%)	Se moy (%)	Sp moy (%)	Tc moy (%)	Se moy (%)	Sp moy (%)	Tc moy (%)	Se moy (%)	Sp moy (%)
3	83.91	72.29	95.35	82.33	69.53	94.51	84.73	71.60	97.49
4	83.67	70.99	96.15	82.88	68.99	96.25	83.13	72.40	93.64
5	81.37	68.36	94.13	75.96	55.33	96.09	80.66	65.81	95.08
6	82.38	67.56	96.81	85.18	71.01	99.06	80.62	63.28	97.26
8	76.66	57.08	95.47	79.52	59.51	98.59	79.72	62.13	97.09
10	83.73	68.91	98.08	80.11	60.32	98.87	74.71	51.87	96.80
20	77.51	55.57	98.44	74.15	50.36	97.00	76.91	55.74	96.97

Table 9 : Résultats de KNN avec distance Cosine sur la base de données initiale

Correlation	1 er essai			2 ème essai			3 ème essai		
Nbr voisins k	Tc moy (%)	Se moy (%)	Sp moy (%)	Tc moy (%)	Se moy (%)	Sp moy (%)	Tc moy (%)	Se moy (%)	Sp moy (%)
3	75.13	61.39	88.14	75.54	66.45	84.14	73.87	64.42	82.76
4	73.06	65.74	80.20	77.47	71.72	83.10	77.58	71.53	83.37
5	70.73	59.36	81.71	63.66	51.89	74.88	68.00	52.81	82.91
6	74.45	63.54	85.07	70.17	54.79	85.05	72.69	62.62	82.24
8	68.11	54.62	81.00	71.89	63.53	80.16	71.45	58.61	83.93
10	69.52	55.42	83.23	69.14	53.16	84.95	69.76	55.69	83.35
20	64.92	53.65	75.80	67.46	51.49	82.75	65.39	45.00	85.23

Table 10 : Résultats de KNN avec distance Correlation sur la base de données initiale

La distance	Euclidienne	Cityblock	Cosine	Correlation
Nbr voisins k	20	6	6	4
Tc moy (%)	84.93	87.00	85.18	77.58
Se moy (%)	70.09	74.81	71.01	71.53
Sp moy (%)	99.14	98.79	99.06	83.37
Temps de calcule (s)	1.259367	1.021804	1.079523	1.15602

Table 11 : les meilleurs résultats obtenus par chaque distance sur la base de données initiale

D'après les tableaux précédents montrant nos 3 essais, le nombre de voisins K qui donne le meilleur taux varie d'une distance à l'autre dans l'algorithme KNN, Où la distance Correlation avec K=4 a donné la précision le plus moins 77.58%, et la distance Cityblock avec K=6 a donné la précision le plus élevé de 87.00%.

4.1.3 Résultats d'algorithme SVM avec validation croisée (10-fold)

Nous allons effectuer 3 essais pour chaque fonction noyau (lineaire, mlp, Radial) et nous avons eu les résultats suivants :

fonction noyau	1er essai			2 ème essai			3 ème essai		
	Tc moy (%)	Se moy (%)	Sp moy (%)	Tc moy (%)	Se moy (%)	Sp moy (%)	Tc moy (%)	Se moy (%)	Sp moy (%)
Lineaire	97.81	91.60	100	98.12	91.36	96.73	98.93	100	97.66
mlp	85.48	86.89	84.13	87.50	87.70	87.30	90.32	90.98	89.68
Radial	72.18	95.36	46.83	72.98	98.36	48.41	72.18	99.18	46.03

Table 12 : Résultats de SVM avec différentes fonction noyau sur la base de données initiale

D'après nos 3 essais, la fonction noyau Radial a toujours donné la précision le plus moins 72.18%, et la fonction noyau lineaire a toujours donné une précision le plus élevé de 97.81% à 98.93.

4.2 Classification sur la base de donnée réduit avec 9 attribues et 248 observations

4.2.1 Résultats d'algorithme ClassificationTree avec validation croisée (10-fold)

Dans cette partie, nous allons suivre le même protocole avec 3 essais et dans chaque essai nous avons obtenus 10 arbres, chacun a son propre taux, puis nous allons comparer les résultats obtenus entre les taux moyens des essais :

L'arbre	1 ère essai			2 ème essai			3 ème essai		
	Tc (%)	Sn (%)	Sp (%)	Tc (%)	Sn (%)	Sp (%)	Tc (%)	Sn (%)	Sp (%)
1	92.00	91.67	92.31	96.00	92.31	100	92.31	86.67	100
2	92.00	85.71	100	100	100	100	80.00	73.33	90.00
3	91.67	100	85.71	88.00	90.91	85.71	91.67	85.71	100
4	91.67	85.71	100	84.00	84.62	83.33	88.00	90.91	85.71
5	92.31	92.31	92.31	88.00	88.00	100	88.00	80.00	100
6	80.00	81.82	78.57	88.00	84.62	91.67	96.00	92.31	100
7	91.67	91.67	91.67	95.83	92.31	100	88.00	84.62	91.67
8	92.00	85.71	100	88.00	81.25	100	88.00	100	80.00
9	80.00	73.33	90.00	96.00	92.31	100	87.50	84.62	90.91
10	84.00	84.62	83.33	84.00	83.33	84.62	95.83	100	92.31

Table 13 : Résultats de classificationTree sur la base de données avec 9 attributs

L'essai	1 ère essai	2 ème essai	3 ème essai
Tc moy (%)	88.73	90.78	89.53
Sn moy (%)	87.26	88.16	87.82
Sp moy (%)	91.39	94.53	93.06

Table 14 : le taux moyen de classificationTree pour chaque essai sur la base de données avec 9 attributs

D'après les tableaux précédents montrant nos 3 essais, le taux moyen d'algorithme classificationTree varie de 88.73% à 90.78% pour la base de données avec 9 attributs, et dans chaque essai nous avons des arbres avec des taux différents. L'arbre 2 dans le deuxième essai a donné un taux de précision parfait de 100%, Sn=100% et Sp=100%. Parmi ses règles :

- _ Si l'enfant trouve facile d'entreprendre plusieurs activités à la fois est certainement non.
- _ Et Si l'enfant a la difficulté à se faire de nouveaux amis est certainement non.
- _ Et Si l'enfant est incapable de maintenir une conversation avec ses ami(e)s est certainement oui.
- _ Et Si Il/elle remarque souvent de petits sons quand d'autres ne le remarquent pas est certainement non.
- _ Alors le diagnostic est non TSA.

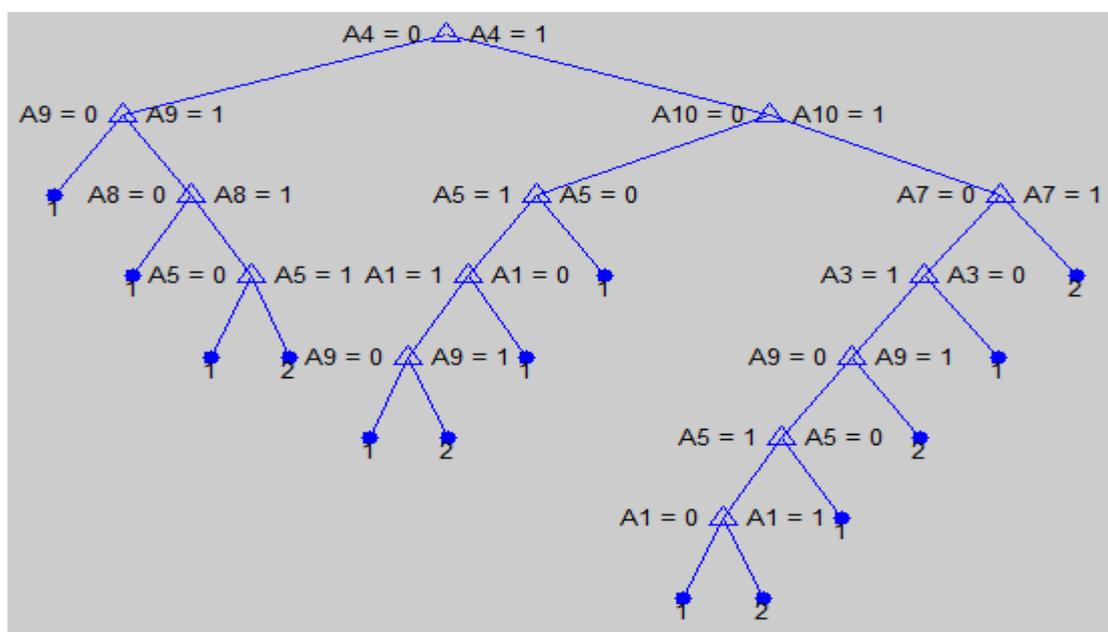


Figure (16) : Arbre de décision avec un Tc=100%, Sn=100% et Sp=100%.

4.2.2 Résultats d'algorithme KNN avec validation croisée (10-fold)

Nous avons fait 3 essais pour chaque distance (Euclidienne, Cityblock, Cosine, Correlation) avec différents nombre de K voisins, et nous avons eu les résultats suivants :

Euclidienne Nbr voisins k	1 er essai			2 ème essai			3 ème essai		
	Tc moy (%)	Sn moy (%)	Sp moy (%)	Tc moy (%)	Sn moy (%)	Sp moy (%)	Tc moy (%)	Sn moy (%)	Sp moy (%)
3	91.15	86.01	96.14	89.42	80.57	98.14	91.36	86.82	95.69
4	89.82	84.44	94.90	89.77	83.44	95.85	89.83	84.74	94.78
5	91.16	83.32	98.62	91.96	86.45	97.33	90.07	81.24	98.72
6	87.94	80.57	94.95	90.01	83.91	95.94	89.89	82.56	96.84
8	89.51	79.53	99.24	85.05	71.84	97.69	90.03	80.53	99.49
10	85.62	71.97	98.94	86.40	75.03	97.33	85.29	73.96	96.47
20	88.56	76.84	100	89.48	78.29	100	87.38	74.44	100

Table 15 : Résultats de KNN avec distance Euclidienne sur la base de données avec 9 attributs

Cityblock	1 er essai			2 ème essai			3 ème essai		
	Tc moy (%)	Sn moy (%)	Sp moy (%)	Tc moy (%)	Sn moy (%)	Sp moy (%)	Tc moy (%)	Sn moy (%)	Sp moy (%)
3	87.79	79.05	96.05	87.35	80.02	94.48	87.35	81.40	93.10
4	92.34	88.63	95.92	93.41	88.52	98.16	91.37	86.17	96.27
5	87.25	76.30	97.92	89.70	80.02	98.87	89.45	79.70	98.87
6	87.99	76.75	98.86	86.79	76.78	96.41	86.43	77.68	94.72
8	90.37	80.49	100	89.06	80.06	97.56	89.31	78.72	99.62
10	87.70	79.09	96.14	89.47	79.98	98.47	86.40	73.71	98.94
20	87.53	74.59	100	87.12	74.01	100	87.29	74.42	100

Table 16 : Résultats de KNN avec distance Cityblock sur la base de données avec 9 attributs

Cosine	1 er essai			2 ème essai			3 ème essai		
	Tc moy (%)	Sn moy (%)	Sp moy (%)	Tc moy (%)	Sn moy (%)	Sp moy (%)	Tc moy (%)	Sn moy (%)	Sp moy (%)
3	88.51	78.47	98.43	89.01	79.45	98.44	85.51	71.22	99.33
4	82.02	66.10	97.63	89.06	77.89	99.92	83.79	68.62	98.47
5	80.48	60.27	100	80.98	61.81	100	84.04	68.15	100
6	82.01	63.18	100	82.56	64.41	100	82.04	63.53	99.92
8	81.66	62.69	100	82.30	64.34	100	83.33	66.11	100
10	78.95	57.03	100	77.78	54.64	100	78.08	55.64	100
20	74.74	48.31	100	70.58	40.42	100	70.51	40.53	100

Table 17 : Résultats de KNN avec distance Cosine sur la base de données avec 9 attributs

Correlation	1 er essai			2 ème essai			3 ème essai		
	Tc moy (%)	Sn moy (%)	Sp moy (%)	Tc moy (%)	Sn moy (%)	Sp moy (%)	Tc moy (%)	Sn moy (%)	Sp moy (%)
3	74.38	88.17	60.53	76.76	86.41	67.35	71.90	86.01	58.27
4	79.19	89.79	68.61	79.80	89.51	70.21	73.92	81.36	66.76
5	81.46	72.60	89.95	83.41	79.06	87.66	78.98	73.80	84.06
6	84.38	80.17	88.47	83.84	75.70	91.77	87.10	84.73	89.38
8	84.04	76.61	84.60	71.89	82.28	86.85	82.05	70.00	93.44
10	83.73	72.00	86.40	69.14	82.52	90.20	84.54	77.31	91.78
20	69.72	68.38	76.33	67.46	73.86	78.78	70.48	71.55	69.44

Table 18 : Résultats de KNN avec distance Correlation sur la base de données avec 9 attributs

La distance	Euclidienne	Cityblock	Cosine	Correlation
Nbr voisins k	5	4	4	6
Tc moy (%)	91.96	93.41	89.06	87.10
Sn moy (%)	86.45	88.52	77.89	84.73
Sp moy (%)	97.33	98.16	99.92	89.38

Table 19 : les meilleurs résultats obtenus par chaque distance sur la base de données avec 9 attributs

D'après les tableaux précédents montrant nos 3 essais, le nombre de voisins K qui donne le meilleur taux varie d'une distance à l'autre dans l'algorithme KNN, Où la distance Euclidienne avec K=5 a donné la précision le plus moins 91.96 %, et la distance Cityblock avec K=4 a donné la précision le plus élevé de 93.41 %.

4.2.3 Résultats d'algorithme SVM avec validation croisée (10-fold)

Nous avons faire 3 essais pour chaque fonction noyau (lineaire, mlp, Radial) et nous avons eu les résultats suivants :

Fonction noyau	1 er essai			2 ème essai			3 ème essai		
	Tc moy (%)	Sn moy (%)	Sp moy (%)	Tc moy (%)	Sn moy (%)	Sp moy (%)	Tc moy (%)	Sn moy (%)	Sp moy (%)
Lineaire	95.97	100	92.06	97.58	100	95.24	97.18	98.36	96.03
mlp	79.84	80.33	79.37	83.06	85.25	80.95	83.87	84.43	83.33
Radial	87.50	99.18	76.19	90.32	99.18	81.75	90.32	99.18	81.75

Table 20 : Résultats de SVM avec différentes fonction noyau sur la base de données avec 9 attributs

D'après nos 3 essais, la fonction noyau mlp a toujours donné la précision le plus moins de 79.84% à 83.87%, et la fonction noyau lineaire a toujours donné une précision le plus élevé de classification de 95.97% à 97.58%.

4.3 Comparaison entre les résultats des trois algorithmes avant et après la sélection des attributs par RELIEFF

L'algorithme	KNN		SVM		ClassificationTree	
	Avant	Après	Avant	Après	Avant	Après
Méthode	Distance Cityblock Avec K = 6	Distance Cityblock Avec K = 4	Fonction noyau lineaire	Fonction noyau lineaire		
Tc moy (%)	87.00	93.41	98.93	97.58	90.37	90.78
Se moy (%)	74.81	88.52	100	100	90.68	88.16
Sp moy (%)	98.79	98.16	97.66	95.24	91.06	94.53

Table 21 : les performances de divers algorithmes d'apprentissage machine sur la base de données de TSA chez les enfants avant et après la sélection des attributs avec l'algorithme RELIEFF

L'évaluation de divers modèles d'apprentissage automatique sur notre base de données de diagnostic des enfants atteints de TSA a permis d'observer une précision varie de 87,00% à 98.93% sur la base de données initial avec 15 attributs. D'où le classifieur Knn avec K = 6 et une distance Cityblock a produit la plus faible précision de 87,00% et le classificateur SVM a produit la précision de prédiction le plus élevé de 98.93%, même pour la sensibilité 100% et la spécificité 97.66%. L'arbre de décision a montré une précision moyenne de 90.37%.

Après la sélection des 9 attributs avec RELIEFF qui sont A4, A10, A9, A5, A8, A1, A3, A7 et A6, les modèles d'apprentissage automatique d'apprentissage avec ces attributs sélectionnées ont donné une précision comprise entre 90.78% et 97,58%. Le modèle Knn a produit une précision de 93,41% avec K=4 et la même distance, ce qui montre une amélioration de 6,41% par rapport

au modèle Knn formé avec l'ensemble de données initial. Nous remarquons cette amélioration aussi pour le classifieur l'arbre de décision qui a produit une précision de 90.78%. Mais le SVM a eu une précision 97.58% représentant une diminution 1.35%.

4.4 Discussions

La comparaison des résultats des modèles d'apprentissage automatiques avant et après la sélection des attributs a montré que 2/3 des modèles d'apprentissage automatique améliorent considérablement les performances avec un minimum de comportements qui caractérisent TSA et dans 1/3 des cas, il a affiché des performances décentes. Ce qui nous montre que les ensembles de comportements minimum qui caractérisent TSA peuvent conserver la structure de l'ensemble de données dans l'apprentissage automatique.

En outre, D'après nos résultats, nous remarquons que le SVM a donné les meilleurs résultats pour notre base de données et donc il est le plus précis pour la détection de TSA chez les enfants, mais il nous manque l'interprétabilité des résultats obtenus. Au contraire, l'arbre de décision nous donne une précision de 90.78% une sensibilité de 88.16% et une spécificité de 94.53%, considéré comme bons résultats par rapport à ceux obtenus dans l'Etat de l'art tel que L'algorithme Firefly binaire qui a produise une précision moyenne de 92,12% à 97,95% avec 10 caractéristiques parmi 21 caractéristiques d'TSA de même notre base de données dans l'étude effectué par Vaishali et Sasikala en août 2018 [15]. Et FURIA qui a détecté les traits de TSA avec une précision de classification pouvant aller jusqu'à 91,35% et un taux de sensibilité de 91,40%; dans l'étude réalisé par Mofleh Al-diabat [17] en 2018.

De plus l'arbre de décision produit des règles logiques qui nous procurent l'interprétabilité des résultats pour pouvoir discuter avec le médecin afin de valider notre travail.

Etant donné que nos résultats obtenus sont très satisfaisants et approuvés par un médecin spécialiste du domaine, nous choisissons l'arbre de décision comme un modèle de classification automatique pour la détection de TSA chez les enfants de 4 à 11 ans, et nous allons nous baser sur les règles de l'arbre qui a donné le meilleur résultat de classification avec une précision de 100% pour programmer notre application mobile pour faire le test de détection de TSA.

Conclusion

La plupart des méthodes TSA existantes consistent en un vaste ensemble de questions sur les comportements de communication, sociaux et répétitifs, et reposent sur des règles expertes du domaine dotées d'une fonction de notation de base permettant de détecter les traits autistiques.

Nous avons opté pour l'utilisation des algorithmes d'apprentissage automatique tels que l'arbre de décision, SVM et KNN pour automatiser le processus de dépistage des TSA chez les enfants. Pour cela nous avons fait un prétraitement sur notre base de données qui comprend 292 observations de 21 variables, comprenait un nettoyage des valeurs manquantes et élimination des attributs qui n'offrent aucun avantage à notre analyse, pour que nous obtenons notre base de données initiale 248 observations de 16 attributs, y compris la classe cible. Puis, nous utilisons l'algorithme RELIEFF pour la sélection des attributs qu'il nous sélectionne 9 attributs parmi les 15 attributs de la base initiale, donc nous avons une base de données réduites avec 248 observations de 10 attributs, y compris la classe cible. Nous avons appliqué nos algorithmes sur la base de données initiale et sur la base avec les 9 attributs sélectionnés, puis nous avons comparé nos résultats obtenus. Nous avons remarqué que les performances de certains algorithmes ont augmentés après la sélection des comportements. L'algorithme SVM a donné les meilleurs taux de classification par rapport aux autres, et l'arbre de décision avec 9 attributs a donné un taux de classification de 90.78% avec une sensibilité de 88.16% et une bonne

spécificité de 94.53%. L'avantage en plus des arbres de décision réside dans les représentations graphiques et les règles de décision qui expliquent les résultats de classification obtenus.

Nos résultats obtenus sont très satisfaisants, et après discussion avec un médecin spécialiste, il a validé nos résultats et les règles de décision extraites à partir des arbres de décision.

Pour cette raison nous nous sommes basés sur les résultats de classification par arbres de décision pour créer notre application mobile pour la détection de TSA chez les enfants que nous allons présenter dans le chapitre suivant et nous allons afficher ses interfaces et montrer son mode d'utilisation.

Chapitre IV : application mobile « طفلي التوحيدي »

Chapitre IV : application mobile « طفلي التوحدي »

Introduction

Dans notre projet de fin d'étude, nous visons à créer une application mobile nommée « طفلي التوحدي » permettant de détecter les signes des troubles de spectre d'autisme chez les enfants par l'utilisation de technique d'apprentissage automatique.

Dans ce chapitre, Nous allons présenter notre application et la méthode d'apprentissage automatique sélectionnée pour faire notre test, puis Nous allons afficher ses interfaces et le mode de son utilisation.

1. Description de l'application

Notre application « طفلي التوحدي » est un moyen simple, facile à utiliser et compréhensible dédié aux parents pour les aider à détecter les signes des troubles de spectre d'autisme chez leurs enfants âgés de 4 à 11 ans , en effectuant un test sur leurs comportements quand ils remarquent des actes anormaux sur eux. Mais l'application ne fournit pas un diagnostic de ce trouble, donc les parents devraient être soumis à des spécialistes pour les résultats finaux. De plus, Notre application offre des suggestions et des conseils pour faire face à ces enfants diagnostiqués autistes. Il est seulement en arabe.



Figure (17) : Écran de bienvenue pour l'application « طفلي التوحدي »

2. La méthode d'apprentissage automatique dans notre application

Pour réaliser le test de dépistage de TSA dans notre application « طفلي التوحيدي », nous nous basons sur les résultats que nous avons obtenus dans le chapitre de Résultats expérimentaux et discussions. Nous avons utilisé les résultats de l'arbre de décision avec validation croisée (10-fold) après la sélection de 9 attributs par la méthode RELIEFF, qui nous a donné un taux de classification correcte de 90.78%, une sensibilité de 88.16% et une spécificité de 94.53% pour la classification de TSA chez les enfants. Puis nous avons sélectionné parmi les 10 arbres que nous avons obtenu, un deuxième arbre qui a donné une précision parfaite de 100%, la sensibilité 100% et la spécificité 100% et nous allons utiliser ses règles logiques qui utilise 9 attributs pour les formuler afin de former notre test de dépistage.

3. Les interfaces d'application « طفلي التوحيدي »

Après que l'écran de bienvenue apparaisse quelques secondes, nous allons passer à la page d'accueil de notre application qui contient quatre interfaces :



Figure (18) : la page d'accueil de l'application « طفلي التوحيدي »

3.1 L'interface de « اختبار سمات التوحد »

Cette partie est consacrée pour faire le test de dépistage de TSA, les parents peuvent commencer ce test en appuyant sur le bouton « بدء الاختبار », puis ils se passeront vers des pages contenant des questions liées au comportement de leur enfant.

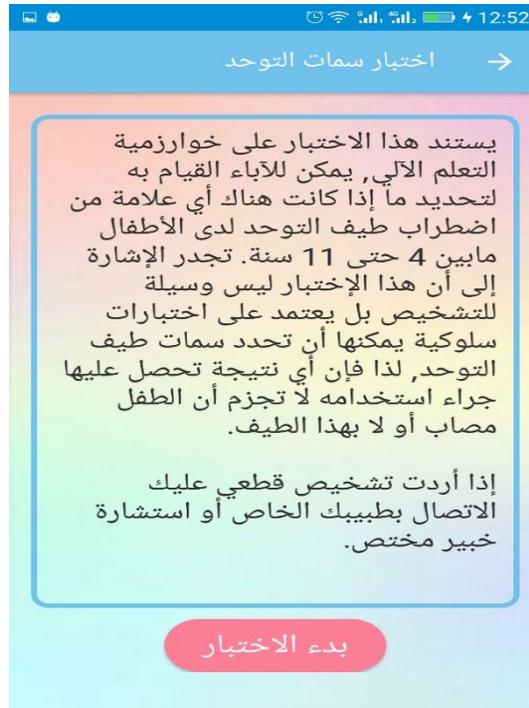


Figure (19) : la page de démarrage de test TSA

3.1.1 Les interfaces des questions

Elle contient les questions liées au comportement des enfants qui représentent nos 9 attributs de la base de données qui ont été adoptés par l'arbre de décision en établissant leurs règles. Les parents doivent compléter le test en répondant à toutes les questions par oui ou non en appuyant sur le bouton « نعم » ou « لا »



Figure (20) : les interfaces des questions de test TSA

3.1.2 L'interface de résultat de test

Après avoir répondu à toutes les questions, l'application transférera les parents vers cette interface afin de visualiser les résultats de test en fonction des règles de notre arbre de décision. Un résultat positif ou négatif de TSA s'affichera en plus d'un bouton « البدء من جديد » permettant de relancer le test.

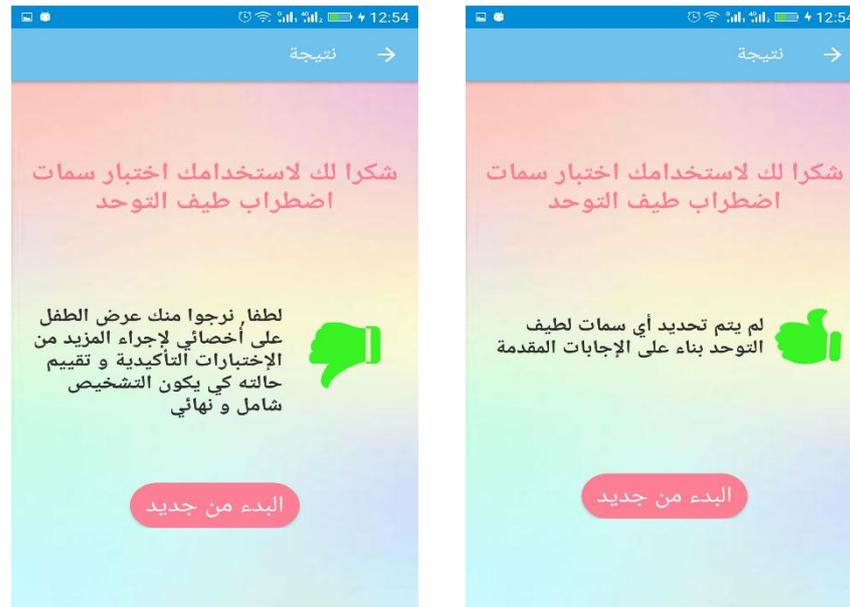


Figure (21) : les interfaces des résultats de test TSA

3.2 L'interface de « عن طيف التوحد »

Elle contient des informations sur le spectre de l'autisme et des conseils et astuces pour les parents ayant un enfant autiste. Vous pouvez naviguer entre leurs pages en cliquant sur un des titres.



Figure (22) : l'interface des informations sur TSA

3.3 L'interface de « تعامل مع سلوكيات طفلك »

Cette partie définit les mauvais comportements qui caractérisent l'enfant autiste et posent des difficultés aux parents. Vous pouvez appuyer sur n'importe quel bouton pour accéder à son interface afin d'obtenir des conseils pour comprendre votre enfant et gérer ses comportements.

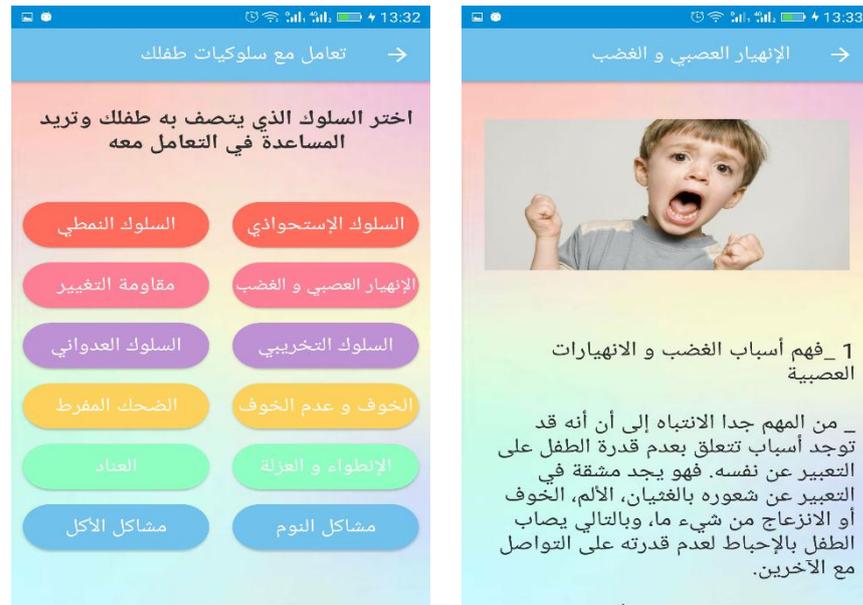


Figure (23) : les interfaces des comportements Caractéristiques de l'enfant autiste

3.4 L'interface de « أنشطة »

Cette partie procure des suggestions d'activités que les parents peuvent faire avec leur enfant autiste dans leur vie quotidienne. Le mode d'activité peut être obtenu en cliquant sur une suggestion.



Figure (24) : les interfaces des activités quotidiennes

Conclusion

Dans ce chapitre nous avons présenté notre application mobile que nous avons créé pour faciliter et aider les parents à découvrir l'autisme chez leur enfant et à mieux comprendre ce trouble et à les guider dans leur gestion des comportements difficiles de leur enfant autiste. Le mode d'utilisation de l'application a été clairement présenté dans les différentes figures.

Conclusion générale

Conclusion générale

Dans notre projet de fin d'études, nous avons pour objectif de développer une application mobile nommée « طفلي التوحدي » qui se base sur l'apprentissage automatique permettant la détection de TSA chez les enfants.

Dans ce contexte, toute une partie a été consacrée sur l'étude théorique de l'autisme et ses comportements distinctifs. Cette partie a été suivie par une autre sur les différentes études menées pour détecter la TSA chez les enfants de différentes manières.

Ensuite, Nous nous sommes concentrés à concevoir un modèle de prédiction de TSA automatique avec des ensembles de comportement minimum sélectionnés dans un jeu de données de diagnostic TSA avec l'algorithme RELIEFF pour la sélection des attributs et déterminer si les modèles d'apprentissage automatiques formés avec des ensembles de comportements minimaux sont capables d'améliorer les performances ou non. Nous avons évalué les algorithmes SVM, KNN et arbre de décision, et nous avons remarqué que le SVM a donné les meilleurs résultats sur notre base de données bien qu'il n'a pas amélioré sa performance après la sélection des attributs.

Au contraire les classifieurs KNN et arbre de décision ont amélioré leur performance après la sélection d'attributs. De là, nous avons adopté les résultats de l'arbre de décision avec validation croisée 10-fold qui nous a donné une précision de 90.78%, une sensibilité de 88.16% et une spécificité de 94.53% avec 9 attributs sélectionnés par RELIEFF et approuvée par un médecin spécialiste.

Le choix du classifieur arbre de décision a été dicté par son principe de fonctionnement qui réside dans le fait qu'il présente des règles que non seulement les cliniciens et le personnel médical peuvent interpréter, mais également les membres de la famille, les enseignants et les aidants naturels. Ces règles sont une source d'information qui peut aider les différentes parties prenantes à comprendre les principaux facteurs d'influence des TSA. Puis nous avons basé notre choix sur les règles de l'arbre de décision qui ont obtenu une précision élevée pour développer le test de TSA dans notre application.

Nous envisageons comme perspectives d'utiliser d'autres bases de données collectées sur des groupes d'enfants de plus jeune âge pour aider la détection précoce de TSA, car notre étude a été réalisée sur 248 observations âgées de 4 ans à 11 ans, et d'essayer d'autre algorithmes d'apprentissage automatiques pour la détection de TSA.

Les Références

Bibliographie :

- [1] Baudouin Forgeot d'Arc, L'autisme d'un DSM à l'autre, Afis_Association française pour l'information scientifique 2013.
- [2] CIM-10 Version 2008, Chapitre V, « Troubles mentaux et du comportement (F84.0) ».
- [3] Organisation mondiale de la santé, « Trouble du spectre autistique », 4 avril 2017
- [4] F.Abdelkrim Journée internationale de l'autisme, 2 avril 2018. [5] Patrick Juignet, Les autismes, In : Philosophie, science et société. 31 juillet 2017, Mis à jour : 17 novembre 2018.
- [6] Asperger Syndrome, Research Autism information service (department within the National Autistic Society), 27 Jun 2018.
- [7] Pervasive Developmental Disorder (Not Otherwise Specified), Research Autism information service (department within the National Autistic Society), 27 Jun 2018.
- [8] Trouble du Spectre Autistique TSA1, Francine Lussier, M.Ed., M.Ps., Ph.D., directrice et neuropsychologue au CENOP, 2016.
- [9] Asperger's syndrome, the complete guide, Tony Attwood, 2007, Jessica Kingsley Publisher, site de Comprendre l'autisme.
- [10] Psychopathologie de l'enfant et de l'adolescent, Dumas E. Jean, et 4e édition revue et augmentée, 2014.
- [11] Revue de psychoéducation, Volume 46, numéro 1, 2017, p. 1-253, La démarche et les outils d'évaluation clinique du trouble du spectre de l'autisme à l'ère du DSM-5.
- [12] LISTE DES TESTS DISPONIBLES, TESTOTHÈQUE, Université Paris Descartes, Paris V, 2018.
- [13] Dennis P. Wall, Rebecca Dally, Rhiannon Luyster, Jae-Yoon Jung, Todd F. DeLuca, « Use of Artificial Intelligence to Shorten the Behavioral Diagnosis of Autism », Reçu le 15 juillet 2011 ; Accepté le 26 juillet 2012 ; Publié le 27 août 2012.
- [14] J A Kosmicki, V Sochat, M Duda & D P Wall, « Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning », États-Unis (2015).
- [15] Vaishali, Sasikala « A machine learning based approach to classify autism with optimum behavior sets », School of Computer Science and Engineering, VIT University, Vellore, India, 25 août 2018, INFORMATICS FOR HEALTH & SOCIAL CARE 2017, VOL. 00, NO. 00, 1–20
- [16] Halim Abbas, Ford Garberson, Eric Glover, Dennis P. Wall, « Machine learning for early detection of autism (and other conditions) using a parental questionnaire and home video screening », publiée en IEEE, Date de la conférence: 11-14 décembre 2017, Date d'ajout à IEEE Xplore: 15 janvier 2018.
- [17] Mofleh Al-diabat « Fuzzy Data Mining for Autism Classification of Children » Department of Computer Science, Al Albayt University, Jordan, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 7, 2018.
- [18] Anibal Sólton Heinsfel, Alexandre Rosa Franco, Augusto Buchweitz, Felipe Meneguzzi de Rio Grande do Sul, Brazil et R. Cameron Craddock de New York, New York USA, « Identification of autism spectrum disorder using deep learning and the ABIDE dataset », Reçu le 1 novembre 2016 ; Reçu sous forme révisée le 30 juin 2017 ; Accepté le 22 août 2017, Publié par Elsevier Inc., NeuroImage : Clinical 17 (2018) 16–23.

- [19] Qandeel Tariq, Jena Daniels, Jessey Nicole Schwartz, Peter Washington, Haik Kalantarian, Dennis Paul Wall « Mobile detection of autism through machine learning on home video: A development and prospective validation study », eçu le 8 juin 2018; Accepté: 25 octobre 2018; Publié: 27 novembre 2018, California, United States of America
- [20] Fatiha Nur Büyükoflaz, Ali Öztürk, Havelsan A.Ş, « Early autism diagnosis of children with machine learning algorithms », Bilgisayar Mühendisliği Bölümü, Izmir, KTO Karatay Üniversitesi, Konya, Türkiye 2018, Éditeur IEEE.
- [21] Mirac Baris Usta ,Koray Karabekiroglu ,Berkan Sahin ,Muazzez Aydin ,Abdullah Bozkurt,Tolga Karaosman,Armagan Aral ,Cansu Cobanoglu,Aysegül Duman Kurt,Neriman Kesim,İrem Sahin &Emre Ürer, « Use of machine learning methods in prediction of short-term outcome in autism spectrum disorders », Nov 2018.
- [22] Thabtah, F. (2017). ASDTests. A mobile app for ASD screening. accessed March 10th, 2017.
- [23] Marko Robnik Sikonja, Igor Kononenko, Theoretical and Empirical Analysis of Relief and RRelief, Machine Learning Journal (2003) 53:23-69

- [24] اضطراب التوحد (الاعراض- الأسباب – العلاج), ا.د. اروة محمد ربيع الخيري 2018. جامعة بغداد. مجلة الآداب / العدد 32
- [25] التوحد كيف نفهمه ونتعامل معه؟, د.فهد بن حمد المغلوث, الطبعة الأولى 2006, مؤسسة ملك الخيرية

Webographie :

- [26] https://fr.wikipedia.org/wiki/Eugen_Bleuler#Recherches_et_th%C3%A9ories
- [27] Andgélina Andrieux, Psychologue.fr,<http://www.memoires-theses.com/psychologie/memoire-these/08-09-2005/autisme.html>
- [28] https://fr.wikipedia.org/wiki/Psychopathie_autistique
- [29]https://fr.wikipedia.org/wiki/M%C3%A8re_r%C3%A9frig%C3%A9rateur#Bruno_Bettelheim
- [30] Jean Steyaert, Comprendre l'autisme, <https://www.participer-autisme.be/go/fr/comprendre-l-autisme/comprendre-le-diagnostic/les-causes-de-l-autisme/autres-causes-de-l-autisme/le-systeme-immunitaire.cfm>
- [31] Dépistage, Cinzia Tolfo, 2005, inforautisme,http://www.inforautisme.com/02quoi/depistage_chat.htm.
- [32] Association rennaisse loi 1901 de parents et de personnes porteuses de troubles du spectre autistique, <https://www.autismaide35.com/les-methodes>
- [33] Thabtah, F. Fayez. (2017). Autistic Spectrum Disorder Screening Data for Children Data Set.<https://archive.ics.uci.edu/ml/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++>.
- [34] <https://fr.mathworks.com/company/aboutus.html>
- [35] <https://www.r-project.org/about.html>
- [36] http://cda.psych.uiuc.edu/statistical_learning_course/supervised_learning_matlab.pdf

[37] [https://fr.wikipedia.org/wiki/Arbre_de_d%C3%A9cision_\(apprentissage\)#G%C3%A9n%C3%A9ralit%C3%A9s](https://fr.wikipedia.org/wiki/Arbre_de_d%C3%A9cision_(apprentissage)#G%C3%A9n%C3%A9ralit%C3%A9s)

[38] https://fr.wikipedia.org/wiki/Arbre_de_d%C3%A9cision

[39] https://fr.wikipedia.org/wiki/Android_Studio

[40] Dépistage, Cinzia Tolfo, 2005, http://www.inforautisme.com/02quoi/depistage_M-chat.htm

[41] Machine Learning : classification à l'aide des arbres de décisions : fonctionnement et application en NodeJS, 17/10/2016 , <https://maximilienandile.github.io/2016/10/17/Machine-Learning-classification-a-l-aide-des-arbres-de-decisions-fonctionnement-et-application-en-NodeJS/>

Résumé

Les techniques d'apprentissage automatiques entre aujourd'hui dans une phase de croissance rapide dans le domaine de la santé pour aider les médecins dans les décisions cliniques notamment dans le diagnostic des maladies grâce à sa capacité d'analyse d'une quantité importante de données. Nous visons à travers notre projet de fin d'études à utiliser un algorithmes d'apprentissage automatique pour détecter les signes de l'autisme chez les enfants en nous basant sur le moins nombre de comportement possible qui indiquent ce trouble et avec la plus grande précision possible. Nous avons également utilisé nos résultats obtenus pour créer une application mobile nommée « طفلي التوحدي » qui aide les parents à effectuer le test d'autisme pour leurs enfants.

Mots clés : l'autisme, TSA, apprentissage automatique, arbre de décision, enfant autiste.

Abstract

Machine Learning techniques enter now a phase of rapid growth in the field of health to help doctors in clinical decisions, particularly in the diagnosis of diseases through its ability to analyze a large amount of data . We aim through our end-of-studies project to use a machine learning algorithm to detect the signs of autism in children based on the least number possible of behaviors that indicate this disorder and with the greatest precision possible. We also used our results to create a mobile app called "طفلي التوحدي" that helps parents to perform the autism test for their children.

Key words : autism, ASD, machine learning, decision tree, autistic child.

الملخص

تدخل تقنيات التعلم الآلي الآن مرحلة من النمو السريع في مجال الصحة لمساعدة الأطباء في اتخاذ القرارات السريرية، وخاصة في تشخيص الأمراض من خلال قدرتها على تحليل كمية كبيرة من البيانات. نحن نهدف من خلال مشروع نهاية الدراسات خاصتنا إلى استخدام خوارزمية التعلم الآلي للكشف عن سمات التوحد عند الأطفال استنادًا على أقل عدد ممكن من السلوكيات التي تشير إلى هذا الاضطراب وبأقصى قدر ممكن من الدقة. استخدمنا أيضًا نتائجنا التي تحصلنا عليها لإنشاء تطبيق للهاتف النقال يسمى "طفلي التوحدي" يساعد الآباء على إجراء اختبار التوحد لأطفالهم.

الكلمات المفتاحية: التوحد، اضطراب طيف التوحد، التعلم الآلي، شجرة القرار، طفل توحدي.