Dédicace

A ceux qui ont toujours été à mes cotés, mes parents. Tous les mots du monde ne sauraient exprimer l'immense et profonde gratitude que je vous témoigne pour tous les efforts et sacrifices que vous n'avez jamais cessé de consentir pour mon instruction. J'espère avoir répondu à vos espoirs.

Remerciements

Avant tout développement sur cette expérience professionnelle, il me parait indispensable de remercier tous ceux qui m'ont soutenu, et fait de mon stage un moment favorable.

En premier lieu, je tiens à remercier mon encadrant de stage, **Mme Jmal Marwa**, Manager à TELNET HOLDING. Un grand merci pour son accueil chaleureux, ainsi que pour sa patience et ses precis conseils.

Je saisis également cette opportunité pour adresser mes profonds remerciements aux personnels de l'équipe de TELNET INNOVATION LABS. Je tiens à remercier **Mme Othmen Farah**, doctorante à Telnet, qui a été toujours à mon écoute et qui a pu me prodiguer beaucoup de conseils.

Je désire aussi remercier mon encadrant pédagogique **Mr Rebai Chiheb**, pour ses conseils au niveau de l'execution du projet ainsi que sur la rédaction du rapport de fin de stage.

Résumé

Le projet consiste à concevoir et implémenter une solution pour l'authentification des personnes à l'aide du signal électrocardiogramme(ECG). La solution est un système embarqué comportant plusieurs parties consistantes. Ce système a pour application principale le verrouillage/déverrouillage à distance des appareils(un smartphone). Le produit final est une application Android capable d'authentifier son utilisateur.

Plus précisément, l'utilisateur n'a qu'à positionner ses doigts sur un capteur d'acquisition spécifique pour enregistrer son signal instantané. Ensuite, une phase de prétraitement est réalisée. Puis, plusieurs descripteurs, ayant pour rôle de confirmer l'unicité du signal, sont extraits. Les descripteurs sont fournis pour un modèle de classification Machine Learning, qui permet d'identifier une corrélation entre les signaux de la personne concernée et par suite de l'authentifier.

Les résultats de ce travail affirment que la biométrie basée sur le signal électrocardiogramme est fiable et peut être utilisée pour l'authentification des individus. Néanmoins, plusieurs recherches doivent être réalisées afin d'améliorer la performance des systèmes biométriques à base d'ECG.

Table des matières

Table des figures

Introduction générale

Le signal ECG (ElectroCardioGram) est une caractéristique universelle. Il est utilisé depuis plusieurs décennies comme un outil de diagnostic efficace et fiable dans les applications médicales.

Récemment, la possibilité d'utiliser ce signal ECG comme outil biométrique a été suggérée. Sa validité est bien étayée par le fait que les différences physiologiques et géométriques du coeur chez de différents sujets révèlent une certaine unicité dans les caractéristiques du signal.

En effet, le signal ECG de chaque individu contient un motif unique provenant des différences existantes en morphologie chez les individus.

Ces résultats des recherche ont été un bon motif pour les experts de Telnet pour proposer notre sujet intitulé «Authentification par signal électrocardiogramme» dans le cadre d'un stage de fin d'études à l'École supérieure des Communications de Tunis visant à l'obtention du diplôme d'ingénieur en télécommunications pour l'année universitaire 2018/2019.

L'objectif principal du projet est de concevoir et implémenter une solution pour l'authentification des individus par signal ECG qui sera utilisé comme moyen d'authentification.

Ce rapport est décomposé comme suit :

Le premier chapitre est une introduction du cadre général du projet comportant une présentation de l'organisme d'accueil, la problématique, et les aspects théoriques sur lesquels on s'est basés.

Le second chapitre est une description de l'architecture de la solution conçue et de différents algorithmes utilisés.

Ensuite dans le troisième chapitre, on passera à l'implémentation de la solution et à l'évaluation du système final.

Enfin, on clôture ce rapport par une conclusion générale qui récapitule le travail achevé tout en ouvrant de nouvelles perspectives.

I Chapitre 1 : Cadre général du projet

I.1 Introduction

Dans ce chapitre, on commence par décrire le cadre général du projet. Tout d'abord, on présente l'entreprise d'accueil. Ensuite, on mentionne la problématique et quelques aspects théoriques indispensables pour la compréhension des différentes étapes du projet.

I.2 Organisme d'accueil

Ce stage de fin d'études est proposé par TELNET HOLDING qui est un groupe spécialisé dans : L'ingénierie logicielle matérielle, l'intégration réseaux et télécom et les études mécaniques. Cette société Tunisienne a été crée en 1994 sous la dénomination TELECOM NETWORKS ENGINEERING en abrégé "TELNET".

Avec plus de 20 ans d'existence et plus que 600 collaborateurs, le groupe TELNET a su se forger une expertise solide dans l'ingénierie produit et ce dans plusieurs secteurs d'activités, notamment les secteurs de Télécoms et Multimedia, de l'industrie et l'energie, de l'automobile, de l'avionique, de la sécurité et système, et de la monétique.

Dans les domaines de ses compétences, TELNET s'est fixé pour objectifs :

- Maîtriser les technologies transverses dans les différents métiers de l'entreprise(soft embarqué, électronique, micro électronique et mécanique)

- Être au diapason des innovations technologiques dans chaque métier et viser de nouveaux secteurs porteurs

- Lancer des collaborations en Recherche et Innovation, particulièrement dans le domaine des technologies de l'Information et de la Communication (TIC), à travers des programmes et des partenariats de recherche développement et Innovation

- Structurer et développer l'approche produit en parallèle avec les activités d'ingénierie et de prototypage

TELNET dispose de plusieurs entités travaillant dans divers secteurs. Ce projet est accueilli par l'entité Recherche et innovation située au parc technologique El Ghazela.

Vu l'importance de la recherche et l'innovation au sein de TELNET, le groupe a créé en 2013 une entité dédiée ; la société TELNET Innovation Labs (TIL), qui assure essentiellement des activités de RDI, en s'appuyant sur une équipe multidisciplinaire d'une dizaine de thésards, chercheurs, experts scientifiques et ingénieurs de prototypage.

Les activités de RDI se basent sur un processus continu d'échanges interactifs avec les composantes «produit» et «Marketing» et une collaboration forte avec le Commissariat Français à l'énergie atomique et aux énergies alternatives « CEA-Tech », les laboratoires de recherches et les écoles d'ingénieurs, en Tunisie, en France et en Allemagne.

La vocation de TELNET Innovation Labs (TIL) est de valoriser les activités du groupe TELNET, par la conception et le développement de solutions innovantes dans le domaine des nouvelles Technologies de l'Information et de la Communication (NTIC). Ainsi, TIL a adopté une approche transversale sur l'ingénierie de produits (logiciels embarqués, électronique, micro-électronique et mécanique) dans les différents métiers du groupe TELNET.

Les projets de recherche amont de TIL se basent sur un volet stratégique de veille technologique dans le domaine des NTIC, et qui s'appuie d'une part sur les partenariats académiques de TIL et d'autre part sur les expressions fonctionnelles et techniques des différentes activités et sociétés du groupe. Le cadre des sujets de prototypage de TIL est le « SMART Country Concept » et ce pour des applications notamment dans les domaines de la santé, la mobilité, le transport, les réseaux intelligents et la ville intelligente.

Telnet innovation Labs accueille plusieurs projet dans des différents domaines : Réalité Augmentée 2D-3D, Biométrie, Signature Numérique, Multimédia, Traitement d'images, Télésurveillance, Robotique, Guidage radar...

Ce projet est exécutée en collaboration avec les membres de Telnet Innovation Labs.

Figure 1 – TELNET HOLDING

I.3 Problématique

L'authentification humaine est devenue importante à cette époque menaçante et suite à plusieurs années de recherche et d'analyse, il s'est avéré que la biométrie est l'un des moyens les plus efficaces pour l'authentification et elle est devenue largement utilisée dans les zones nécessitant une sécurisation d accès malgré les quelques inconvénients qui menacent la fiabilité de la plupart des caractéristiques biométriques [13].

La biométrie n'est pas restreinte à la reconnaissance vocale, faciale, des empreintes digitales ou de la lecture de l'iris car notre visage, notre voix et nos yeux ne sont pas les seuls caractéristiques propres à chaque individu avec lesquels on pourrait s'authentifier. En effet, les recherches ont montré que notre cœur, plus précisément sa forme et ses dimensions, peut être utilisé comme outil biométrique pour nous authentifier sur un smartphone ou un ordinateur [2].

Le battement du coeur d'une personne ne peut pas être modifié pour masquer l'identité. Le signal électrocardiogramme varie d'une personne à une autre suivant le changement de la taille, la position et l'anatomie du coeur, la configuration de la poitrine et divers autres facteurs.

Plusieurs recherches, ont prouvé l'unicité du signal ECG et ont confirmé son utilisation pour l'authentification. Les défis actuels sont l'extraction de descripteurs significatifs à partir de ce signal et la conception des modèles de classification robustes dans le but de prouver la stabilité à long terme de cette biomètrie.

Le signal électrocardiogramme peut également être utilisé dans des applications médicales. Une analyse, pas trop complexe de ce signal, permet de détecter si le rythme cardiaque est anormal, ou si certaines formes d'ondes présentent des anomalies[14]. Ceci est considéré comme une évolution remarquable dans le domaine médical, grâce aux travaux combinant la recherche et l'ingénierie.

Pour conclure, le signal électrocardiogramme et l'une des caractéristiques physiologiques authentiques et fiables de chaque individu. Par conséquent, deux axes d'application se présentent : le domaine de la sécurité informatique et le domaine médical. Ceci nous pousse à travailler sur ces deux aspects, dans le but de faciliter la vie des individus en offrant plus de protection digitale ainsi que pour prévenir des anomalies cardiaques.

Donc et en se basant sur ce qui précède, le signal électrocardiogramme sera utilisé dans notre projet comme moyen d'authentification. Notre objectif principal,plus précisément, est de concevoir et implémenter une solution pour l'authentification des individus par signal ECG.

Ce projet traite l'utilisation de la biomètrie basée sur l'activité électrique du coeur humain, d'où le traitement du signal électrocardiogramme.

Suite à une phase de documentation approfondie, il s'est avéré que le projet doit être décomposé en plusieurs parties consistantes.

Pour commencer, une phase de traitement de signal s'est présentée. De plus, une deuxième phase de Machine learning est indispensable. La phase finale est une phase de déploiement et de développement.

Ce projet a été proposé par TELNET INNOVATION LABS en 2017. La solution développée était basée sur une approche différente de celle traitée dans ce rapport. Les résultats ne sont pas satisfaisants et le système n'est pas fiable, d'après l'équipe de TELNET INNOVATION LABS. La solution proposée était basée sur un matching direct entre un certain nombre de points caractéristiques du signal en entrée et un signal de la même personne enregistrée dans une base donnée. Ceci a négligé le fait qu'un signal électrocardiogramme d'une même personne peut subir des légères variations en le mesurant dans des intervalles de temps différents. La partie du Machine Learning n'été pas traité, malgré son importance dans la classification et la prédiction.

La figure 2 présente la variation du signal ECG d'un individu dans une durée de 35 minutes.

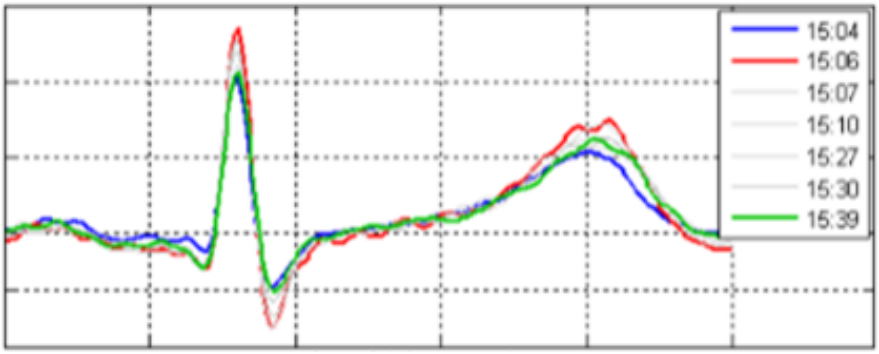


Figure 2 – Signal ECG d'un patient pendant 1 heure

La figure 3 présente les variations subies par le signal ECG d'un individu en le mesurant dans une durée de 5 mois.

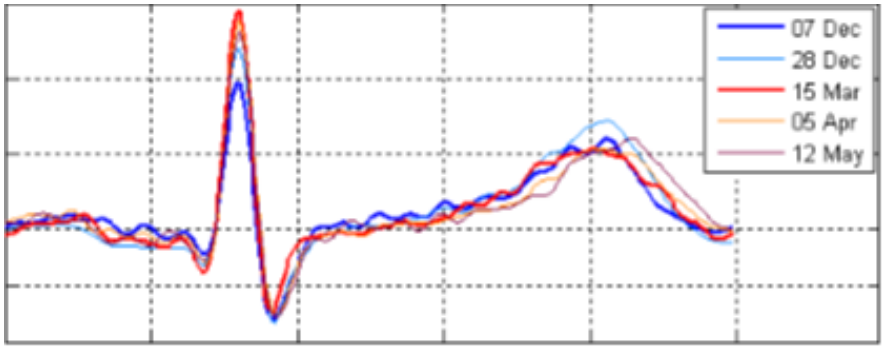


Figure 3 – Signal ECG d'un patient pendant 6 mois

I.4 Aspect théorique

a. La biométrie

Définition de la biométrie

La biomètrie représente la mesure biologique ou les caractéristiques physiques qui peuvent être utilisées pour identifier les individus. Il existent deux catégories de ces caractéristiques : Physiologiques et comportementales[15]. Les caractéristiques physiologiques sont liées à la physiologie de l'être humain : empreintes, ADN, Iris... Les caractéristiques comportementales sont liées directement au comportement comme la voix. Les chercheurs affirment que la forme de l'oreille, la manière de s'assoir et de marcher, et les odeurs corporelles sont des identificateurs uniques[13].

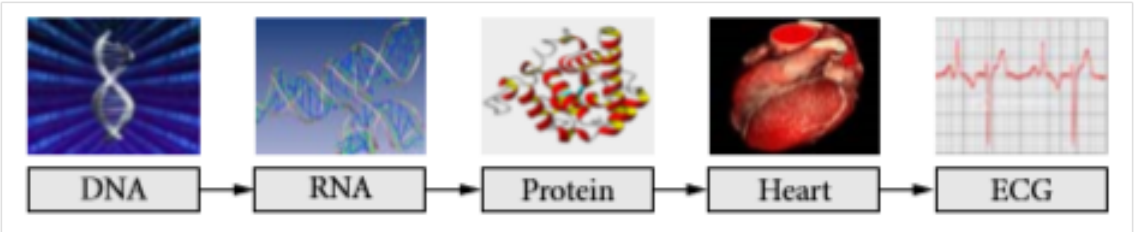


Figure 4 – Traits biométriques[1]

Puisque ces paramètres sont relativement personnalisées, ils sont utilisées pour remplacer ou au moins accroître la sécurité des système informatiques : déverrouillage smartphone, accées restreint aux immeubles...

La plupart des systèmes de sécurité nécessitent l'existence d'une clé, que l'utilisateur doit toujours s'en souvenir(mot de passe, code pin), ou de questions de sécurité.

La biométrie avancée est également utilisée pour protéger les documents sensibles. Citibank utilise déjà la reconnaissance vocale et la banque britannique Halifax teste des appareils qui contrôlent les battements du cœur pour vérifier l'identité des clients. Ford envisage même de placer des capteurs biométriques dans les voitures.

L'**authentification biométrique** est le processus consistant à comparer les données des caractéristiques de la personne au «modèle» biométrique de cette personne afin de déterminer la ressemblance. Le modèle de référence est le premier enregistré dans une base de données ou un élément portable sécurisé comme une carte à puce. Les données stockées sont ensuite comparées aux données biométriques de la personne à authentifier.

L'**identification biométrique** consiste à déterminer l'identité d'une personne. L'objectif est de capturer un élément de données biométriques de cette personne. Cela peut être une photo de leur visage, un enregistrement de leur voix ou une image de leur empreinte digitale. Ces données sont ensuite comparées aux données biométriques de plusieurs autres personnes conservées dans une base de données.

Pour que la biométrie soit applicable pour le contrôle d'accès et la sécurité, il faut qu'elle vérifie plusieurs conditions. Les traits biométriques sont[15] :

- **universels :** La biomètrie est présente et peut être mesurée pour chaque personne
- **Uniques :** La biomètrie est suffisamment différente d'une personne à une autre
- **Stable :** Les paramètres biomètriques restent invariants tout au long la vie des individus.
- **Performants :** La biométrie est robuste, fiable, et facile à analyser.
- **Acceptables :** La collecte et l'utilisation des traits biomètriques sont admis du point de vue de la société.

De plus, les traits biomètriques doivent être facilement mesurés.

Bien que la biométrie offre de nombreux avantages par rapport aux moyens de sécurité traditionnels, il y'a encore de sérieuses inquiétudes sur la fiabilité de la biométrie stockée. La biométrie ne peut pas être facilement changée, car elle dépend de caractéristiques physiologiques ou comportementales de l'individu. Même si les traits biométriques ne peuvent pas être facilement extraits pour fausser l'identité, il existent des moyens pour collecter la biométrie requise pour lancer une attaque. Par exemple, les descripteurs faciales peuvent être facilement collectés à partir des photos disponibles sur les réseaux sociaux, et la voix peut être obtenue par l'enregistrement des appels téléphoniques[13].

Signal ECG en tant que caractère biométrique

La physiologie du coeur et la genése du signal électrocardiogramme sont détaillés dans les annexes. On peut maintenant considérer l'ECG comme une biométrie fiable. Certaines caractéristiques doivent être vérifiées pour que le signal ECG soit applicable à la sécurité et au controle d'accés.

De toute évidence, l'ECG est universel, car il est produit suite à l'activité électrique du cœur, qui se produit dans chaque être vivant. La plupart des travaux existants sont basées sur l'établissement de l'unicité et de la stabilité du signal ECG[16].

Stabilité : La stabilité exige que les données soient extraits à partir d'une seule personne sur une période de temps suffisamment longue. La création de grandes bases de données est coûteuse et implique un investissement temporel significatif, ce qui explique que le nombre d'études qui examinent la stabilité de l'ECG sont limités.

Collectabilité : Les machines ECG à 12 dérivations traditionnelles exigent que 10 électrodes soient placées sur le thorax des membres du sujet. L'enregistrement de l'ECG à l'aide de moniteurs de qualité médicale est également invasif, ce qui oblige à exposer leur poitrine et leurs membres. Avec l'essor du domaine médical, les moniteurs ECG mono-lead sont de plus en plus répandus. Ces moniteurs sont portatifs et peuvent être utilisés pour enregistrer une trace d'ECG à un seul fil à l'aide d'électrodes qui font le contact avec les poignets (par exemple, les bandes de SmartWatch) ou les doigts (par exemple, les capteurs spécifiques). Bien que ces appreils fournissent moins de données que les machines médicales à 12 dérivations, ils peuvent être utilisés pour enregistrer le signal ECG d'une façon fiable et utile.

Performance : La performance du système biométrique dépend principalement de la qualité du signal utilisé et des descripteurs extraits. Certaines techniques de prétraitement et d'extraction de paramètres sont considérées fibales et performantes.

Acceptabilité : Avec l'apparition des capteurs ECG fiables, il y a eu plus de possibilités de créer des systèmes biométriques à base d'ECG.

Pour conclure, le signal ECG est un candidat fort à considérer dans le domaine de l'authentification des personnes. D'une part, l'introduction de capteurs ECG à faible coût fournit une opportunité pour pouvoir instaurer des solutions se basant sur l'ECG. D'autre part, plusieurs études doivent traiter les méthodes d'extraction de descripteurs afin d'améliorer la performance.

Composition du signal ECG :

Le signal ECG, pour un seul cycle cardiaque, est composé de plusieurs parties(ondes)[17] :

L'onde P : C'est la première onde détectable. Elle apparaît quand l'impulsion électrique se propage à partir du nœud sinusal pour dépolariser les oreillettes

Le complexe QRS : C'est un ensemble de déflexions positives et négatives qui correspondent à la contraction des ventricules. Ce complexe est constitué de trois ondes :

- **Onde Q :** première déflexion négative
- **Onde R :** première déflexion positive
- **Onde S :** déflexion négative qui suit l'onde R. Sa forme est variable selon les dérivations utilisées (emplacement des électrodes) ou une arythmie donnée.

L'onde T : Elle correspond à la repolarisation ventriculaire. Elle est normalement de faible amplitude et ne témoigne d'aucun événement mécanique. Cette onde succède le complexe QRS.
L'onde U : Dans certaines occasions, une onde, dite onde U, peut être observée après l'onde T. C'est une onde de faible amplitude et elle est visible dans certaines dérivations notamment chez les athlètes.

En plus des différents ondes qui sont les paramètres de base d'un signal ECG, il existent plusieurs intervalles et segments qui présentent des informations utiles [18].

Ces intervalles sont :

Intervalle RR : L'intervalle RR correspond au délai entre deux dépolarisations des ventricules. C'est cet intervalle qui permet de calculer la fréquence cardiaque.
Segment PR : Le segment PR correspond au délai entre la fin de la dépolarisation des oreillettes et le début de celle des ventricules.
Intervalle PR : L'intervalle PR correspond à la durée de propagation de l'onde de dépolarisation du nœud sinusal jusqu'aux cellules myocardiques ventriculaires.
Intervalle QT : Cet intervalle correspond au temps de systole ventriculaire, qui va du début de l'excitation des ventricules jusqu'à la fin de leur relaxation.
Segment ST : Le segment ST correspond à la phase pendant laquelle les cellules ventriculaires sont toutes dépolarisées.

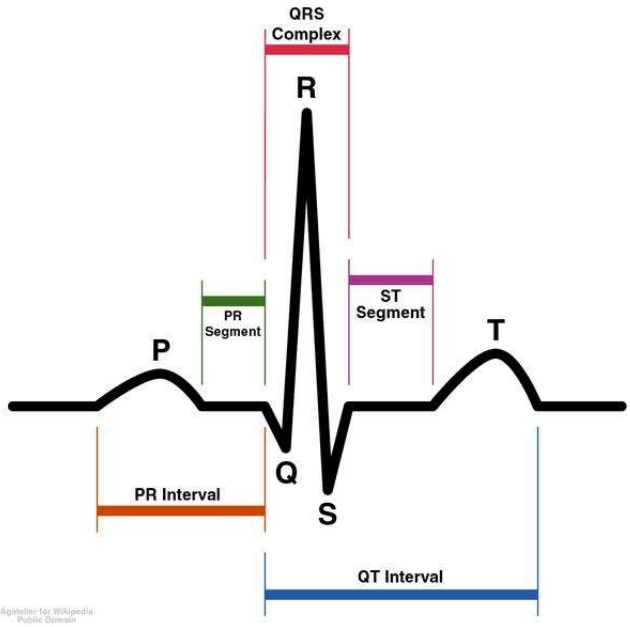


Figure 5 – Les différents intervalles du signal ECG[2]

b. Machine Learning

Machine Learning est une application de l'intelligence artificielle qui fournit aux systèmes la capacité d'apprendre et d'améliorer automatiquement de l'expérience sans être explicitement programmé.

Le processus du Machine Learning commence par des observations ou des données, comme des exemples, une expérience directe ou une instruction, afin de chercher des modèles de données et de prendre de meilleures décisions à l'avenir en fonction des exemples fournis. L'objectif principal est de permettre aux ordinateurs d'apprendre automatiquement sans intervention ou assistance humaine[19].

Ce concept est né sur la base de la reconnaissance et de la théorie que les ordinateurs peuvent apprendre sans être programmé pour effectuer des tâches spécifiques. Les chercheurs, intéressés par l'intelligence artificielle, voulaient voir si les ordinateurs pouvaient apprendre à partir des données. L'aspect itératif du Machine Learning est important car, comme les modèles sont exposés à de nouvelles données, ils sont capables de s'adapter indépendamment. Ils apprennent des calculs précédents pour produire des décisions et des résultats fiables.

Machine Learning est utilisé dans un large éventail d'applications aujourd'hui. L'un des exemples les plus connus est le fil d'actualité de Facebook.

En outre, les systèmes de gestion de la relation client utilisent des modèles d'apprentissage pour analyser les e-mails et inciter les membres de l'équipe commerciale à répondre aux messages les plus importants en premier.

Des systèmes plus avancés peuvent même recommander des réponses potentiellement efficaces.

Les systèmes de ressources humaines (RH) utilisent des modèles de Machine Learning pour identifier les caractéristiques des employés efficaces et s'appuient sur ces connaissances pour trouver les meilleurs candidats aux postes vacants.

Ils existent quatre catégories des algorithmes de Machine Learning[20].

— **Apprentissage supervisé :** Ces algorithmes peuvent appliquer ce qui a été appris par le passé à de nouvelles données en utilisant des exemples étiquetés pour prédire les événements futurs. À partir de l'analyse d'un jeu de données d'apprentissage connu, l'algorithme produit une fonction pour faire des prédictions sur les valeurs de sortie. Le système est en mesure de fournir des sorties pour toute nouvelle entrée après une formation suffisante. Chaque élément de données possède un étiquette indiquant sa classe. Le modèle apprend ainsi à reconnaître la classe en comparant la sortie avec l'étiquette correcte.

— **Apprentissage non supervisé :** Ces algorithmes sont utilisés quand les données d'apprentissage fournis, ne sont ni classés ni étiquetés. Ces algorithmes essaient de déduire des fonctions décrivant des structures cachées à partir de données non labélisées.

— **Apprentissage semi supervisé :** Ces algorithmes combinent l'apprentissage supervisé et non supervisé. La technique repose sur l'utilisation d'une quantité de données étiquetée et d'une quantité de données non étiquetée pour générer le modèle.

— **Apprentissage par renforcement :** C'est une méthode d'apprentissage qui interagit avec son environnement en produisant des actions et en découvrant des erreurs ou des récompenses. Une façon de comprendre l'apprentissage par renforcement, consiste à réfléchir à la manière dont une personne peut apprendre à jouer pour la première fois à un jeu, alors qu'elle n'est pas familiarisée avec les règles du jeu.

Pour chaque type de Machine Learning il existent plusieurs algorithmes ayant des caractéristiques différentes. Ces algorithmes sont détaillés dans les annexes.
Le choix de l'algorithme de Machine Learning à utiliser dépend de plusieurs paramètres notamment le problème envisagé et la taille de données disponibles.

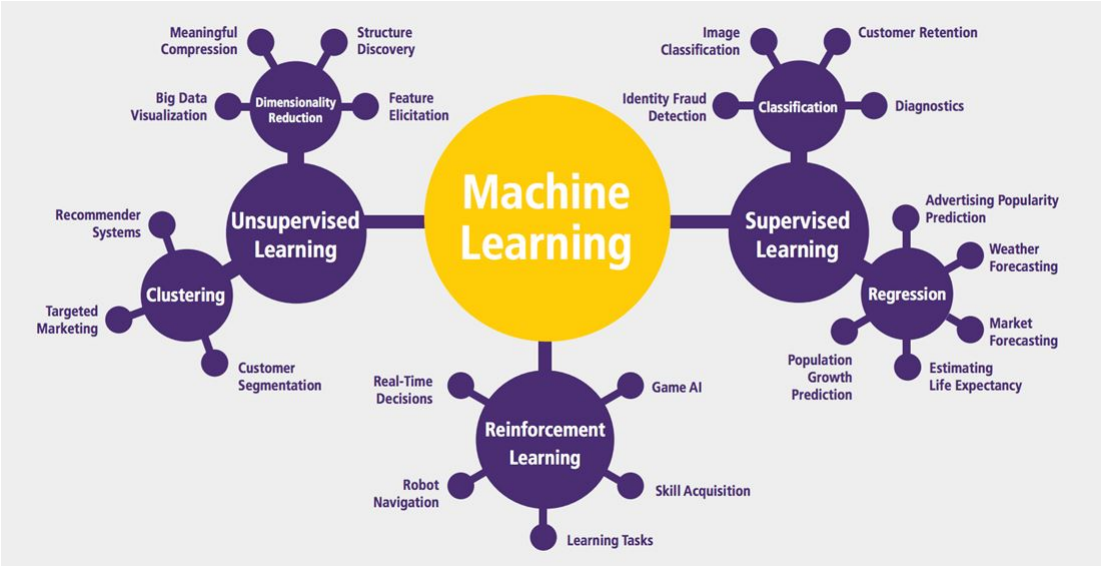


Figure 6 – Algorithmes de Machine Learning[3]

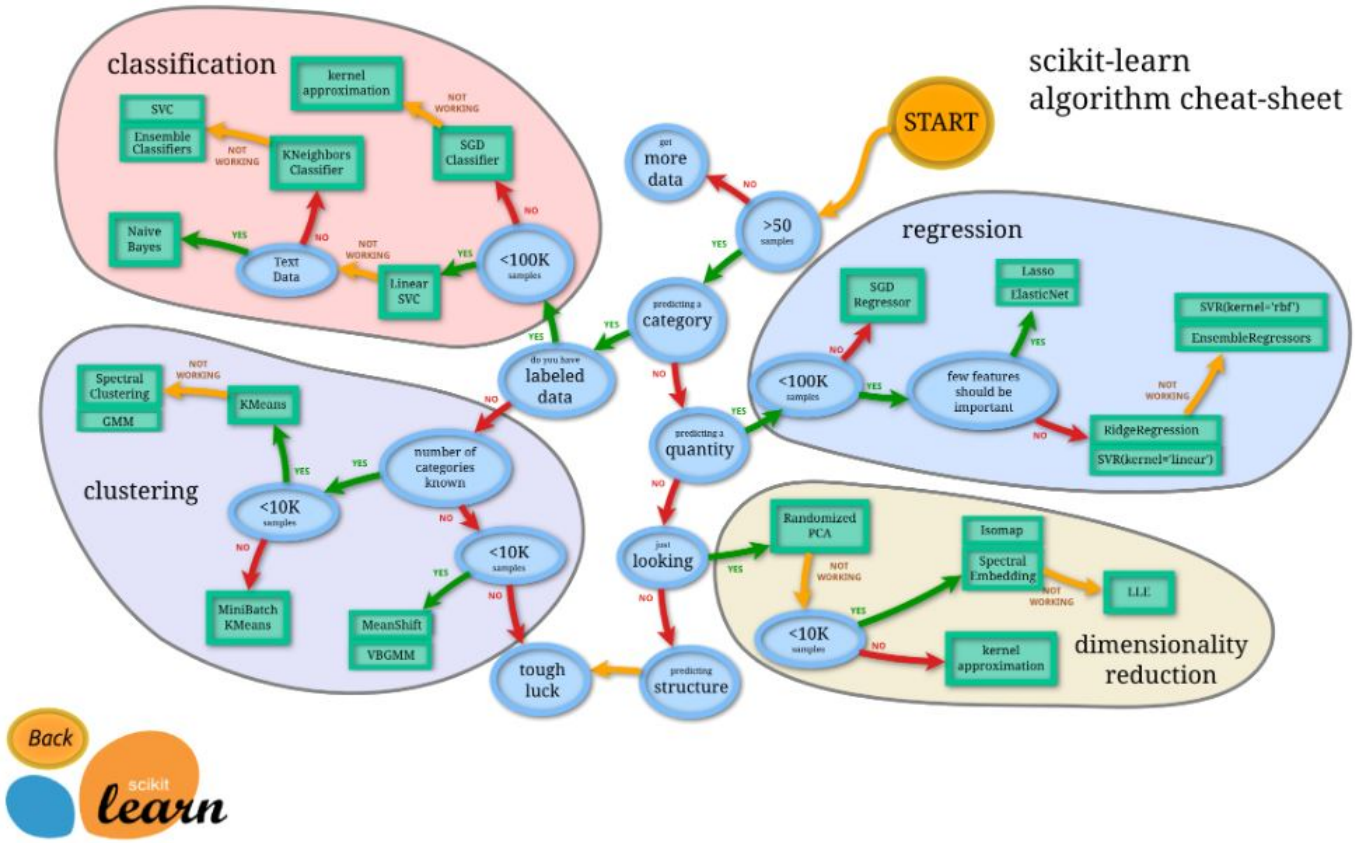


Figure 7 – Choix de l'algorithme de Machine Learning à utiliser[4]

La construction d'un modèle Machine Learning nécessite le passage par plusieurs étapes indispensables[5].

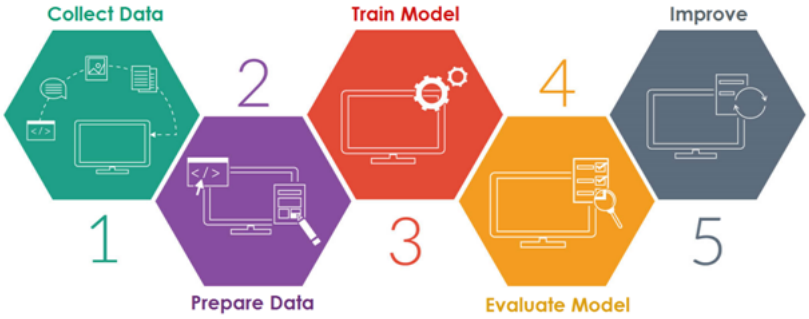


Figure 8 – Etapes de construction d'un modèle Machine Learning[5]

Rassemblage des données : Cette étape est très importante car la qualité et la quantité de données collectées vont affecter directement la qualité du modèle de classification.

Préparation des données : Dans cette étape, on charge les données dans un format spécifique, pour être utilisé dans le phases suivantes. . Nous allons d'abord rassembler toutes nos données, puis randomiser les commandes. L'ordre des données ne doit pas affecter la performance du modèle de classification, car cela ne fait pas partie de la prédiction.

Les données rassemblées ont besoin d'autres formes d'adaptation et de manipulation comme la duplication ou la normalisation. Tout cela se passerait à l'étape de la préparation des données.

Choix du modèle : Il existe plusieurs modèles utilisés par les chercheurs et les scientifiques. C'est un problème de classification. Dans ce sens, quelques algorithmes sont efficaces.

Plusieurs travaux de recherche ont montré une excellente performance de l'algorithme de classification SVM[21] : Support vector Machine. Cet algorithme utilise une technique de représentation de données sur un vecteur de dimension N(dimension de l'espace des descripteurs).La classification est obtenue au moyen d'hypers-plans assurant une séparation optimale, qui visent à obtenir la plus grande marge entre les points de données les plus proches appartenant à des classes différentes.

Les SVM dépendent des fonctions du noyau dans le processus de classification. Le choix du noyau influe les performances de classification, ce qui est considéré comme une tâche importante. La marge de séparation des données la plus grande est représentée par un hyper-plan de séparation qui sépare les classes différentes dans un problème à deux classes.

Les performances de SVM peuvent être modifiées à l'aide du terme C et des paramètres du noyau. Ces paramètres ont une grande influence sur la précision du modèle SVM et la marge de séparation maximisée. ‘

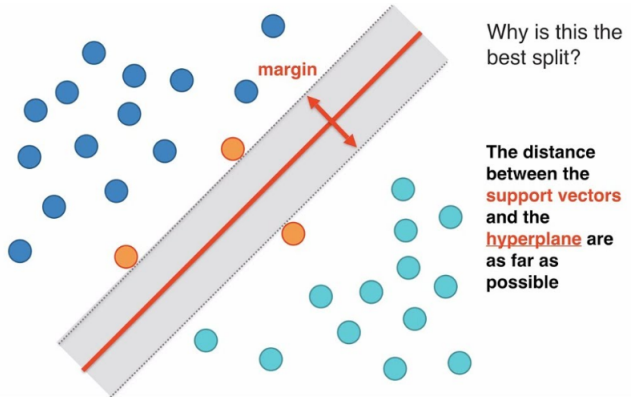


Figure 9 – Support Vector Machine

Les paramètres C et γ (gamma) affectent considérablement la précision de la classification du modèle. Cependant, on ne sait pas à l'avance quelles valeurs des paramètres conviennent mieux pour notre fichier de données.

Phase d'apprentissage : Dans cette étape on manipule les données dans le but de maximiser le pouvoir de détection du modèle.

Le processus d'apprentissage d'un modèle Machine Learning implique la fourniture d'un algorithme Machine Learning avec des données d'apprentissage à partir desquelles, il va extraire des similarités dans les données d'éntrée pour pouvoir "apprendre".

Dans l'apprentissage supervisé, les algorithmes tirent des enseignements de données étiquetées. Après avoir compris les données, l'algorithme détermine l'étiquette à attribuer aux nouvelles données en fonction du modèle et en associant les modèles aux nouvelles données non étiquetées.

Les données d'apprentissage doivent contenir la réponse correcte, appelée cible. L'algorithme d'apprentissage trouve dans les données d'apprentissage des modèles qui mappent les attributs de données d'entrée à la cible et génère un modèle ML qui capture ces modèles.

Évaluation : Une fois le modèle est construit, il est temps de l'évaluer en terme de performance. Dans cette phase, l'ensemble de données mises à coté et non utilisées dans la phase d'apprentissage entre en jeu. L'évaluation nous permet de tester notre modèle par rapport à de nouvelles données, non utilisé pour le training. Cette métrique permet de savoir le comportement du classifieur à l'égard de données qu'il n'a pas encore vues. Ceci représente la façon avec laquelle le modèle pourrait fonctionner dans le monde réel.

Variation des paramètres : Suite à la phase d'évaluation, il est nécessaire de savoir si on peut améliorer le modèle construit. Ceci peut être achevé à l'aide de l'ajustement des paramètres. Ils existent quelques paramètres implicitement fixés, lors du training, et c'est le

moment de tester d'autres hypothèses et d'essayer d'autres valeurs.

On peut modifier la taille des données utilisés pour le training ou changer les valeurs des différents paramètres propres aux modèles.

Prédiction : C'est la dernière étape. On peut finalement utiliser notre modèle pour prédire le résultat en se basant sur des nouvelles données d'entrée.

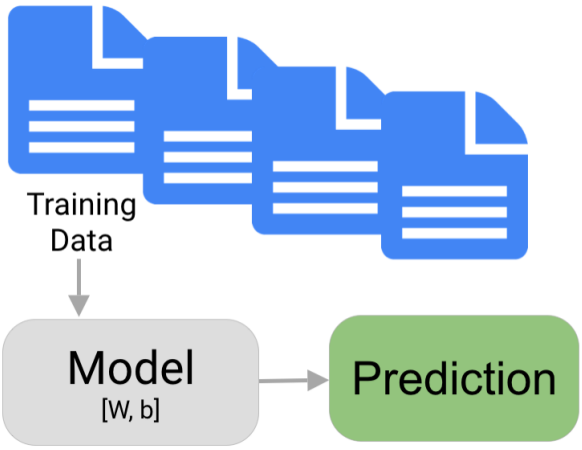


Figure 10 – Phase de test d'un mdoèle Machine Learning[5]

I.5 Conclusion

Ce premier chapitre est une initiation sur le cadre général du projet ainsi que sur les conditions d'exécution du projet. Nous passerons dans le chapitre suivant à une description plus détaillée des éléments techniques constituant notre projet.

II Chapitre 2 : Conception de la solution

II.1 Introduction

L'élaboration d'un bon plan de développement est une étape cruciale pour la création d'une application et c'est à ce moment que l'architecture vient jouer un rôle important. Pour cela on va détailler, dans ce chapitre, l'architecture de la solution proposée et les différentes étapes du déroulement du projet. Ensuite, on va détailler le choix algorithmique qu'on essayera de valider à travers de différents outils .

II.2 Architecture de la solution

L'architecture globale de la solution peut se résumer dans la figure ci-dessous. D'un point de vue général, deux phases sont indispensables. Une phase d'apprentissage dans laquelle les signaux sont extraits, prétraités. Ensuite, un modèle de classification Machine Learning est construit. Une deuxième phase d'authentification qui est la phase finale dans laquelle l'utilisateur peut s'authentifier.

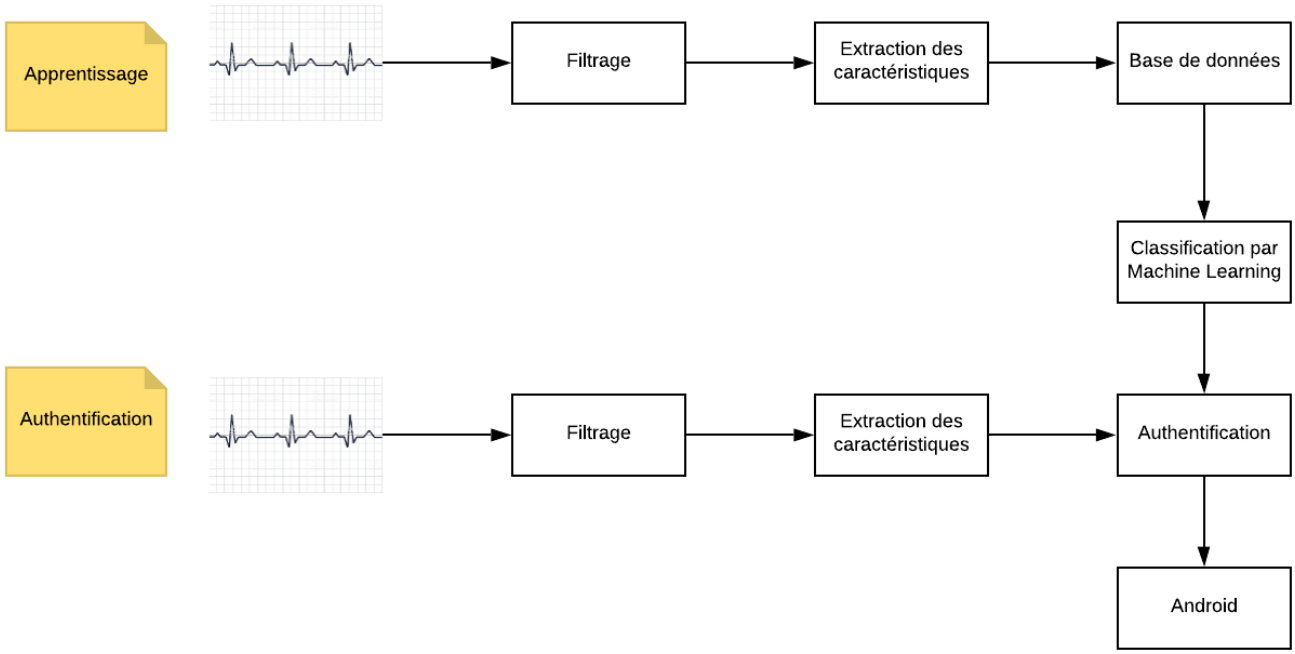


Figure 11 – Architecture globale de la solution

Le produit final est une application android permettant l'authentification de l'utilisateur. Ce dernier, doit positionner ces doigts sur un capteur d'acquisition pour que son signal cardiaque instantané soit enregistré. La communication entre l'application Android et le capteur est établie via le protocole Bluetooth.

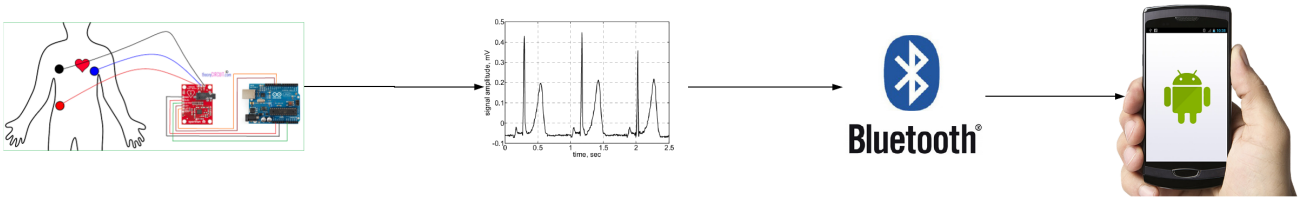


Figure 12 – Envoi du signal ECG du capteur vers l'application Android

Le signal électrocardiogramme de l'utilisateur déjà enregistré, est directement transmis à une base donnée temps réel. L'ECG subit un traitement spécifique à fin de pouvoir le filtrer, et d'en extraire ses descripteurs. Ensuite à l'aide du modèle de classification l'utilisateur l'utilisateur peut s'authentifer.

Avant de pouvoir mettre en place l'application finale, on a collecté les signaux ECG, on les a prétraité et on a construit le modèle de classification. Ceci est détaillé dans le chapitre suivant.

II.3 Étapes du déroulement du processus

Pour récapituler, le déroulement du processus aboutissant a l'application de notre solution et comme l'indique le schéma ci-dessous.

Étape 1 : Un capteur d'acquisition permet d'enregistrer le signal instantané de l'utilisateur.

Étape 2 : Une carte Arduino connectée au capteur assure l'envoi du signal à l'application android en utilisant le protocole de communication Bluetooth.

Étape 3 : L'application android permet un envoi instantané des valeurs reçues par le capteur à la base de données temps réel de Firebase. Par conséquent, le signal est totalement enregistré sur cette base.

Étape 4 : Suite à l'envoi du signal, l'utilisateur est redirigé vers une deuxième page permettant le lancement du processus de prédiction.

Étape 5,6,7,8 : L'utilisateur, par le biais du bouton "predict", provoque le déclenchement

des processus de traitement de signal. En effet, le signal est extrait de la base de données. En suite, l'étape de filtrage est exécutée.

Étape 9,10 : Les descripeurs sont extraits à partir du signal filtré en entrée.

Étape 11,12 : La dernière phase est la phase de Machine Learning. Le modèle de classification a pour entrée l'ensemble de features extraits. L'output du modèle est une classe bien déterminée..

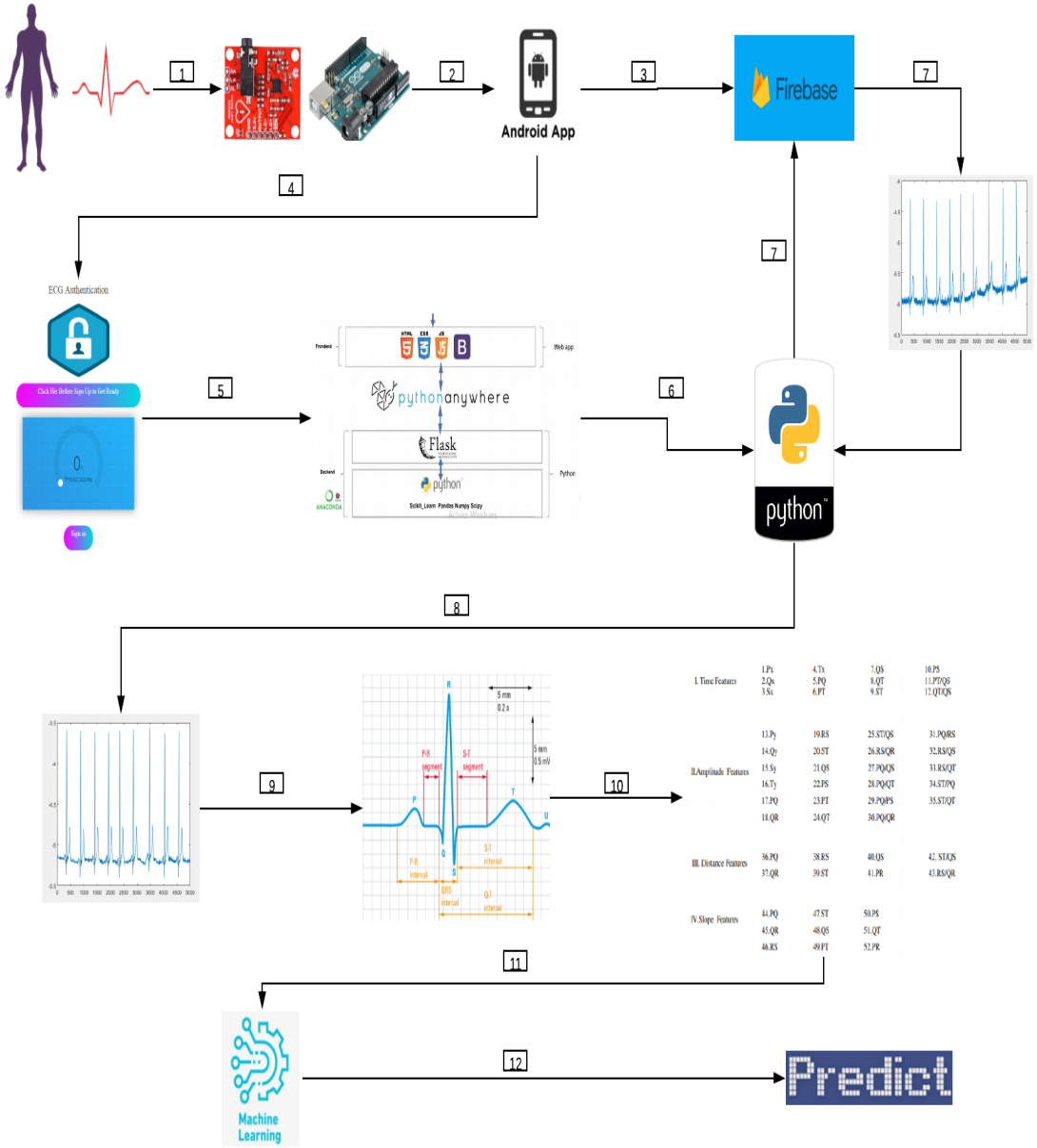


Figure 13 – Architecture de la solution

II.4 Choix Algorithmiques

a. Prétraitement du signal

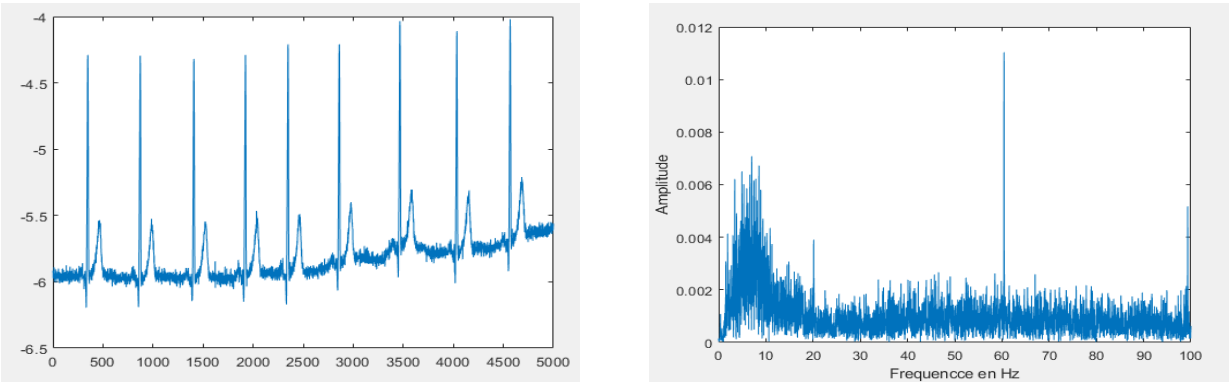
Dans tous les appareils électroniques, l'immunité au bruit est une essentielle caractéristique. Les signaux ECG du monde réel sont souvent perturbés par le bruit et dépendent fortement du patient, il est obligatoire de procéder à un débruitage pour obtenir un modèle de reconnaissance robuste. Le débruitage consiste à éliminer toutes sortes de bruit.

Les principales causes du bruit sont :

— **Power-line interference :** La ligne de puissance provoque des champs électromagnétiques qui sont supposés être une source de bruit commune pour un signal ECG. Ceux-ci sont caractérisés par l'interférence sinusoïdale de 50-60 Hz accompagnée d'un certain nombre d'harmoniques. La bande étroite du signal ECG rend difficile son analyse et son interprétation. Diverses mesures peuvent être prises pour réduire l'effet d'interférences de lignes électriques.
Dans ce projet, on a utilisé un filtrage linéaire pour gérer ces interférences.

— **Baseline Wander :** La dérive peut être éliminée par la conception et l'implémentation d'un filtre passe-haut linéaire. Les paramètres critiques à prendre en compte lors de la conception du filtre sont la fréquence de coupure et la réponse en phase. La fréquence de coupure doit être choisie judicieusement car les informations médicales doivent rester inchangées.

Power-line interference et Baseline Wander sont les deux principaux facteurs du bruit présent dans le signal électrocardiogramme.



(a) Signal ECG bruité (b) Spectre du signal ECG bruité

Figure 14 – Analyse du signal ECG bruité

Pour résoudre ce problème on a utilisé, un filtre passe bande, de type Butterworth, et d'ordre 2. Les fréquences de coupures sont : 2 Hz et 40 Hz[22]. Suite à son traitement, le

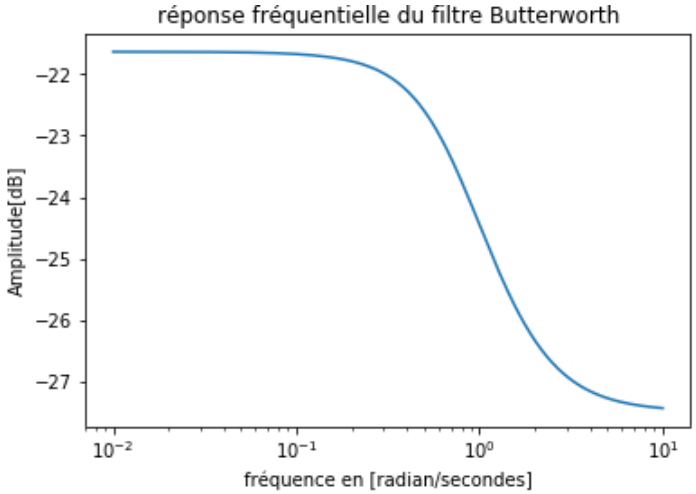
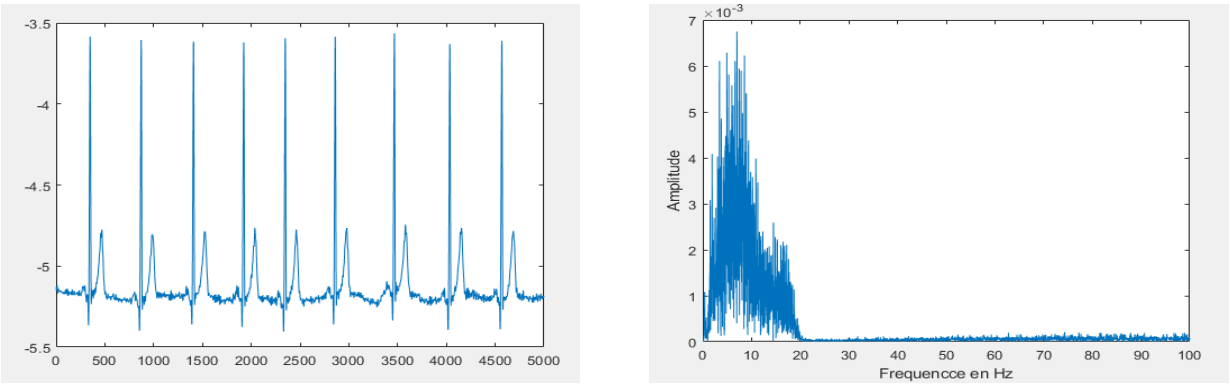


Figure 15 – Réponse fréquentielle du filtre pass-bande

signal présente une amélioration remarquable dans sa forme ainsi que dans son spectre.



(a) Signal ECG filtré (b) Spectre du signal ECG filtré

Figure 16 – Analyse du signal ECG filtré

b. Extraction des descripteurs

Il est vrai que le signal ECG est considéré stable et unique, mais cette unicité n'est pas évidemment identifiable. Le signal d'une seule personne peut présenter des légères variations au cours du temps, même s'il est mesuré dans les mêmes conditions physiologiques et climatiques. La problématique revient donc à pouvoir caractériser la signature unique pour chaque individu.

Pour ce faire, des paramètres décrivant la signature unique du signal doivent être extraits. Ces paramètres sont des paramètres temporels, en terme d'amplitude et en terme de distance euclidienne.

Le complexe QRS est la forme d'onde la plus pertinente au sein de l'ECG, car il reflète l'activité électrique du cœur pendant la contraction ventriculaire. La détection de QRS est difficile, non seulement en raison de la variabilité physiologique des complexes QRS, mais aussi en raison des différents types de bruit qui peuvent être présents dans le signal ECG.

Plusieurs algorithmes sont utilisés pour la détection du complexe QRS dans le signal électro-cardiogramme et plusieurs travaux de recherche ont pour but de comparer ces algorithmes en terme de performance.

Parmi ces algorithmes on a choisi celui de Pan-Tompkins pour la réalisation de ce projet. La performance de cet algorithme est confirmée par plusieurs travaux de recherche, ainsi que par les réalisations pratiques[6]. Le logiciel Matlab fournit une implémentation complète de cet algorithme, ce qui confirme son efficacité[23].

L'algorithme de détection des complexes QRS se base sur plusieurs étapes :
Premièrement et afin d'atténuer le bruit, le signal passe par un filtre passe bande ayant pour fréquences de coupure 5Hz et 15Hz.
Le processus suivant est la différenciation qui est suivie par une quadrature puis une intégration fenêtrée. Des informations sur la pente du complexe QRS peuvent être obtenues à la phase du dérivé. Le processus de quadrature intensifie la pente de la réponse en fréquence de la dérivée.

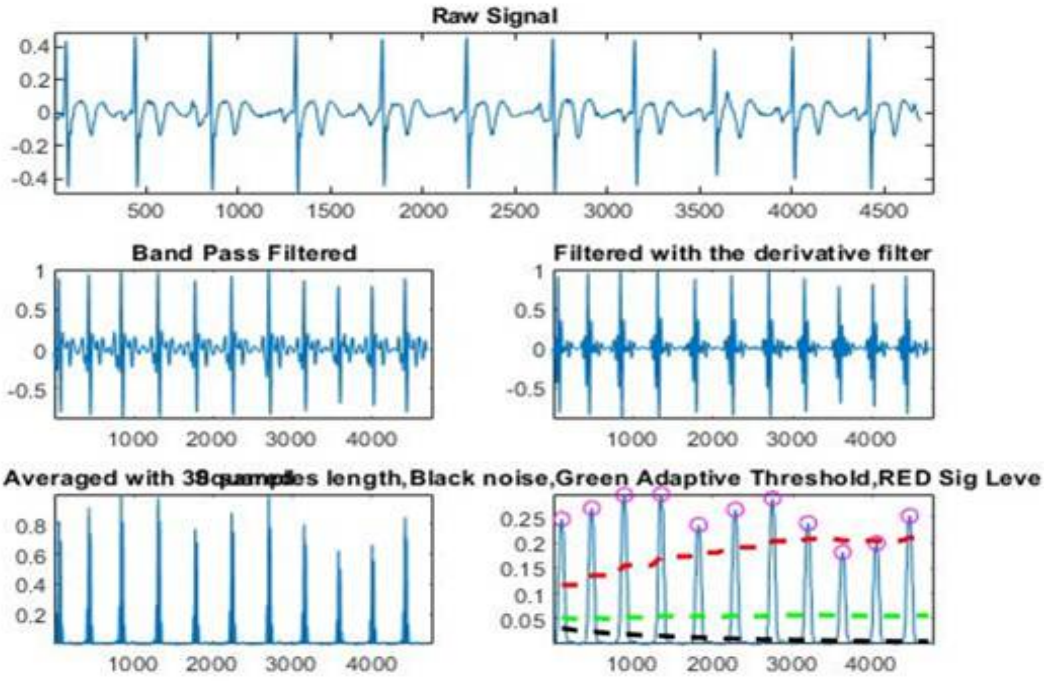


Figure 17 – Principe de l'algorithme Pan-Tompkins

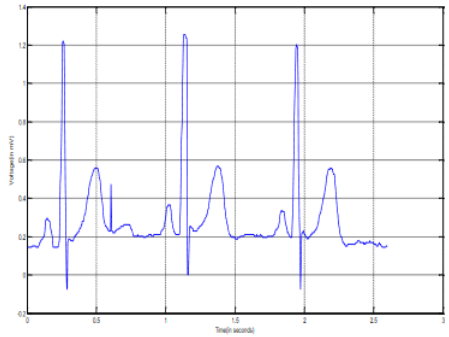


Figure 18 – Signal ECG à l'entrée de l'algorithme Pan-Tompkins[6]

Au début, le signal ECG est filtré à l'aide d'un filtre passebande numérique afin d'éliminer toute sorte de bruit en conservant les caractéristiques principales du signal : fréquentielles et temporelles.

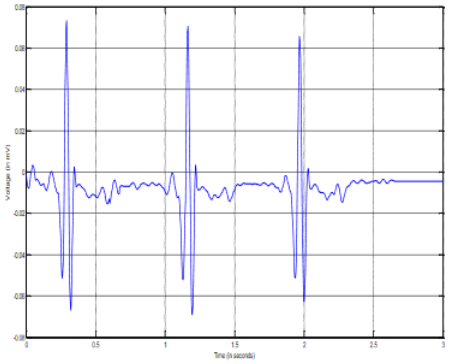


Figure 19 – Signal filtré entre [5Hz,15Hz][6]

Le signal est différencié pour en extraire les informations des pentes des complexes QRS.

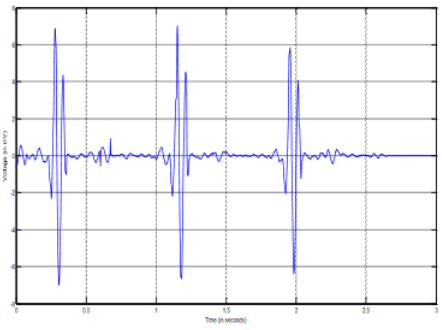


Figure 20 – Signal dérivé[6]

Le carré du signal dérivé est calculé.

Le but de l'intégration de fenêtre de déplacement est d'obtenir des caractéristiques de la forme d'onde en plus de la pente de l'onde R.

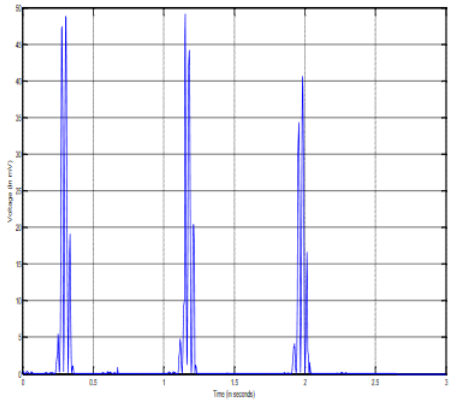


Figure 21 – Carré du signal dérivé[6]

Finalement, les seuils appropriés sont ajustés pour mieux détecter les R-pics.

Une fois les pics R sont détectés, les formes d'ondes restantes caractérisant le signal ECG doivent être extraites : P,Q,S,T. Dans ce projet on a utilisé un algorithme développé par un travail de recherche, afin de pouvoir extraire les différentes ondes du signal en se basant sur les positions des pics R déjà détectés[2]. Cette méthode se base sur le domaine temporelle. L'intervalle R-R est calculé avec : $Rloc(n+1) - Rloc(n)/fs(sec)$

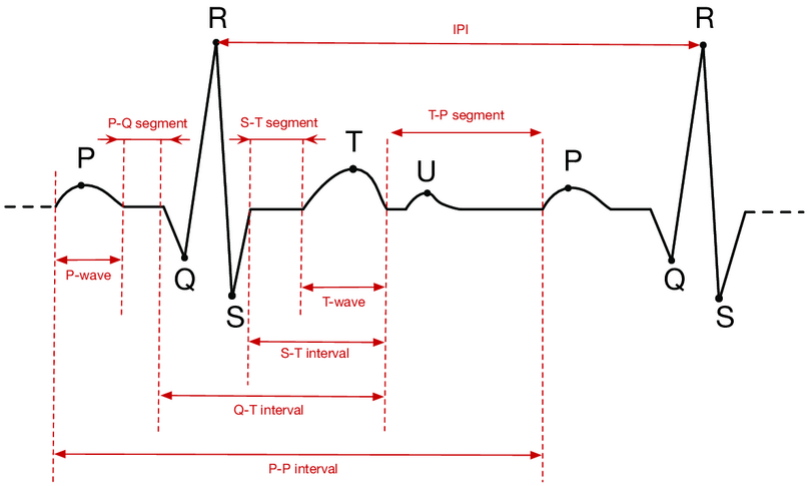


Figure 22 – Caractéristiques principales du signal ECG[7]

fs : fréquence d'échantillonage
Rloc : position temporelle du pic R
Pour identifier l'onde P, une fenêtre temporelle est créée avec des bornes allant de 65% à 95% de l'intervalle R-R. La valeur maximale du signal dans cette fenêtre représente l'onde P.
L'onde P est identifiée en choisissant la valeur minimale du signal dans un intervalle temporel commençant 20ms avant le pic R correspondant.

L'onde S est la valeur minimale du signal dans un intervalle commençant après le pic R correspondant.
Pour identifier le pic T, une fenêtre temporelle est créée avec des bornes allant de 15% à 55% de l'intervalle R-R. la valeur maximale du signal, représente l'onde T.
Un ensemble de 51 descripteurs sont extrais en se basant sur les différents ondes du signal. Tous les descripteurs sont calculés en considérant que le pic R correspondant est une origine.

Les descripteurs extraits du signal ECG sont liés avec l'amplitude(y), les intervalles temporels(x), différence entre les amplitudes, différence entre les indices temporelles, les pentes, et des ratios entre certains descripteurs.

I. Time Features	1.Px	4.Tx	7.QS	10.PS
	2.Qx	5.PQ	8.QT	11.PT/QS
	3.Sx	6.PT	9.ST	12.QT/QS
II.Amplitude Features	13.Py	19.RS	25.ST/QS	31.PQ/RS
	14.Qy	20.ST	26.RS/QR	32.RS/QS
	15.Sy	21.QS	27.PQ/QS	33.RS/QT
	16.Ty	22.PS	28.PQ/QT	34.ST/PQ
	17.PQ	23.PT	29.PQ/PS	35.ST/QT
	18.QR	24.QT	30.PQ/QR	
III. Distance Features	36.PQ	38.RS	40.QS	42. ST/QS
	37.QR	39.ST	41.PR	43.RS/QR
IV.Slope Features	44.PQ	47.ST	50.PS	
	45.QR	48.QS	51.QT	
	46.RS	49.PT	52.PR	

Figure 23 – L'ensemble des descripteurs

II.5 Validation du choix Algorithmique

Comme nous l'avons déjà mentionné, l'algorithme d'extraction des pics R sur lequel on s'est basé est celui proposée par Pan et Tompkins, proposé en 1985. C'est un algorithme d'extraction des complexes QRS en temps réel et il a été cible de plusieurs travaux de recherches.

Vu que MathWorks fournit une implémentation complète de de cet algorithme, la validation de la solution dans ce projet est possible et sera réalisé grâce â Matlab.

Les étapes de l'algorithme ont été présentées en détails dans la partie précédente et dans cette partie, on se concentrera sur des points significatifs dans l'exécution de l'algorithme[8].

Un seuil adaptatif ainsi que plusieurs autres règles de décision doivent être prises en consi-

dération pour une détection précise et fiable.

Seuil adaptatif : Lors de l'analyse de l'amplitude du signal, l'algorithme utilise deux seuils(THR(SIG) et THR(NOISE)) qui sont adaptés d'une façon continue selon la qualité du signal ECG.

Recherche des complexes QRS non détectés : D'après l'étape précédente si l'amplitude du pic R détecté est inférieure au seuil THR(SIG), le pic ne peut pas être considéré comme un complexe QRS. Par contre, si une longue période s'est terminée sans pic réel, l'algorithme assume qu'un pic QRS est négligé ce qui déclenche un "searchback".

Élimination de pics non significatifs : Il est impossible que deux pics QRS se présentent dans une durée inférieure à 200ms. C'est une contrainte physiologique, qui correspond à la dépolarisation des ventricules.

La figure ci-dessous présente une évaluation de l'implémentation de l'algorithme Pan-Tompkins fournie par Matlab sur des signaux ECG à partir de la base de données de MIT.

Record (No.)	Total (No. Beats)	FP (Beats)	FN (Beats)	Failed (Beats)	Failed (%)
100	2274	0	0	0	0
101	1874	0	6	6	0.32
102	2187	0	0	0	0
104	2230	1	2	3	0.13
105	2572	48	32	80	3.11
108	1824	61	71	132	7.24
200	2601	1	3	4	0.15
202	2146	0	6	6	0.279
219	2312	0	0	0	0
222	2634	131	2	133	5.04

Figure 24 – Évaluation de l'algorithme Pan-Tompkins sur la base données de MIT[8]

II.6 Conclusion

Une fois qu'on a construit le squelette de notre système en comprenant l'architecture détaillée de l'application et en montrant la validité de notre choix algorithmique comme a été fait tout au long de ce chapitre, nous passerons dans le chapitre suivant à la mise en oeuvre de notre application.

III Chapitre 3 : Réalisation de la solution

III.1 Introduction

Après avoir déterminé l'architecture adéquate pour la réalisation de notre solution ainsi que les étapes du déroulement de notre processus tout en se basant sur les résultats des algorithmes sélectionnés, nous passons a l'implémentation, étape qui nécessite la présence d'un environnement facilitant le développement qui s'achèvera par une évaluation de notre système. Ceci sera traité au cours de ce chapitre.

III.2 Environnement du développement

a. Environnement Logiciel

Afin de réussir nos taches, on a eu recours à l'utilisation de plusieurs logiciels et outils de développement.

Figure 25 – Python[9]

La solution globale est développée en **python**. Python est un langage de programmation interprété. Il favorise la programmation impérative structurée, fonctionnelle et orientée objet. Le langage Python est placé sous une licence libre proche de la licence BSD8 et fonctionne sur la plupart des plates-formes informatiques, des smartphones aux ordinateurs centraux, de Windows à Unix avec notamment GNU/Linux en passant par macOS, ou encore Android, iOS, et peut aussi être traduit en Java ou .NET. Il est conçu pour optimiser la productivité des programmeurs en offrant des outils de haut niveau et une syntaxe simple à utiliser.

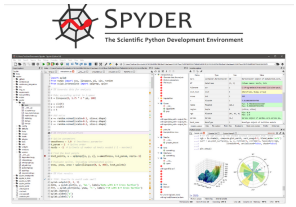


Figure 26 – Spyder[10]

Pour ce faire, l'environnement de développement utilisé est **SPYDER**. Spyder est un environnement scientifique puissant développé en Python, pour Python, et conçu par et pour les scientifiques, les ingénieurs et les analystes de données. Il offre une combinaison unique de la fonctionnalité avancée d'édition, d'analyse, de débogage et de profilage d'un outil de développement complet avec l'exploration de données, l'exécution interactive, l'inspection approfondie et de capacités de visualisation de paquets scientifiques.

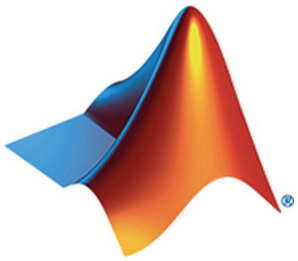


Figure 27 – Matlab

Au début, on a essayé d'utiliser le logiciel **Matlab**, qui est trop puissant dans les calculs numériques et le traitement du signal, pour la partie de prétraitement et extraction des paramètres. Ceci présentait plusieurs problèmes.

Figure 28 – Arduino

La phase d'acquisition de signaux réels est exécutée à l'aide du capteur AD8232, ainsi qu'une carte Arduino Mega permettant l'envoie du signal enregistrée à une base de données. La manipulation de cette carte électronique est possible à l'aide de ARDUINO IDE, qui est l'environnement de développement intégré adéquat. ARDUINO IDE est une application multiplateforme qui est développée en langage de programmation Java. Il est utilisé pour programmer des cartes compatibles.

Figure 29 – Android Studio

Le produit final est une application android. Pour ce faire, on a utilisé **Android Studio** qui est l'environnement de développement intégré officiel pour le système d'exploitation Android de Google, construit sur le logiciel IntelliJ IDEA de JetBrains et conçu spécifiquement pour le développement Android. Il est disponible en téléchargement sur les systèmes d'exploitation Windows, macOS et Linux.

b. Environnement Matériel

Le projet est exécuté sur un Laptop ayant les caractéristiques suivantes :

- Microprocesseur : Intel(R) Core(TM) i7-5500U CPU 2.4GHz
- RAM : 8Go
- Disque Dur : 1To

La phase d'acquisition était réalisé à l'aide d'une carte Arduino Uno/Mega/Nano, un capteur ECG(AD8232), 3 électrodes et des fils de connexion.

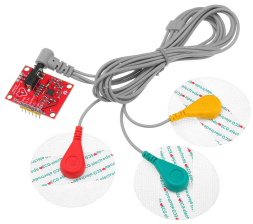


Figure 30 – Capteur ECG comportant 3 électrodes

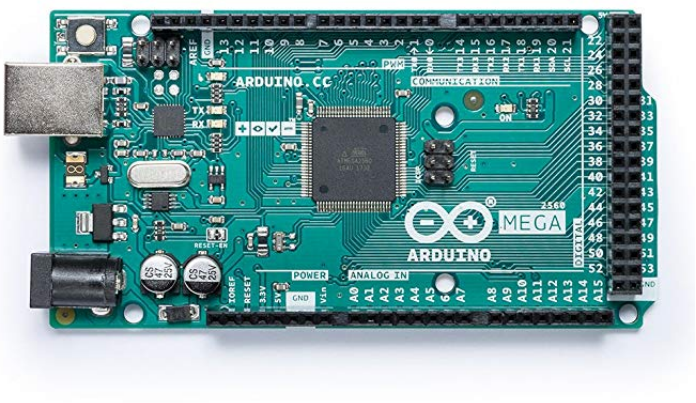


Figure 31 – Carte Arduino Mega

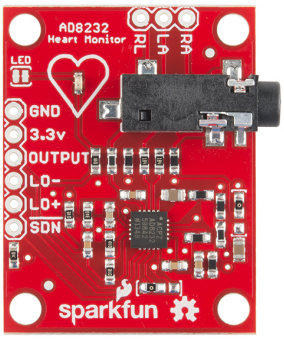


Figure 32 – Capteur ECG AD8232

L'AD8232 est un bloc de conditionnement de signal intégré pour le signal ECG et d'autres applications de mesure de biopotentiel. Il permet d'extraire, amplifier et filtrer de petits signaux biopotentiels dans le présence de conditions bruyantes, telles que celles créées par le mouvement ou le placement à distance des électrodes. Le capteur dispose d'un ultra-faible convertisseur analogique-numérique (ADC) pour acquérir facilement le signal de sortie.

Spécifications
Voltage : 2.0V à 3.5V
Courant : 170uA
Rejection en mode commun : 80dB(60Hz)
Nombre d'électrodes : 3

L'envoi du signal ECG du capteur vers l'application android via Bluetooth nécessie la présence d'un module Bluetooth connecté à la carte Arduino Mega.

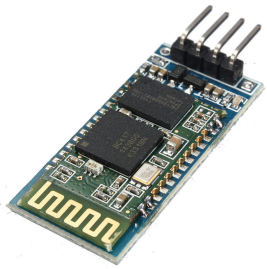


Figure 33 – Module Bluetooth HC-06

III.3 Implémentation et Résultats

Dans cette partie, on présente l'implémentation de la solution conçue. La solution est d'abord implémentée sur des signaux extraits à partir d'une base de données publiques. Ensuite, l'implémentation est achevée sur des signaux acquis réellement durant la période du stage.

a. Choix de la base de données théorique

PhysioNet
the research resource for
complex physiologic signals

Figure 34 – Physionet[11]

Pour valider le travail théoriquement, on a eu recours à l'utilisation des signaux de la base de donnée de PHYSIONET. Cette base a été choisie parce qu'elle contient plusieurs signaux pour chaque sujet.

La référence exacte de la base utilisée est : The ECG-ID Database[24]. Cette base a été crée en 2005, et contient 310 enregistrements ECG prélevés de 90 personnes. Chaque signal dure environ 10 secondes et présente 10 pics R(10 battements du coeur).

Les enregistrements ont été obtenus auprès de volontaires (44 hommes et 46 femmes) âgés de 13 à 75 ans. Le nombre d'enregistrements pour chaque personne varie de 2 (collectés au cours d'une journée) à 20 (collectés périodiquement sur 6 mois).

Cette base de données a été cible de plusieurs travaux de recherche grâce à sa diversité. Les signaux peuvent être telechargés sous différents formats.

b. Acquisition des signaux réels

Un capteur ECG avec 3 électrodes doit être fixer sur 3 points cibles du corps humain pour détecter le signal cardiaque. Les électrodes du capteur ECG convertissent le battement du coeur en signal électrique. Le capteur utilisé est le capteur AD8232 de Analaog devices. Ce capteur léger et fin permet de mesurer le rythme cardiaque continu et fournit des informations sur les ondes composant le signal ECG[12].

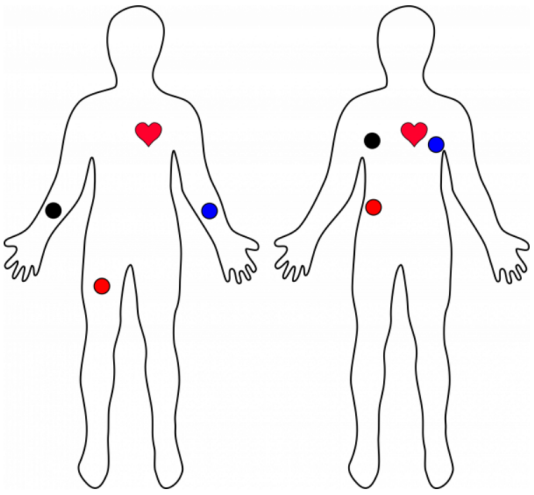


Figure 35 – Placement typique des électrodes[12]

Plus les électrodes sont proches du coeur, meilleure est la mesure. Les câbles sont codés par couleur pour aider à identifier le placement approprié. Les capteurs peuvent être placés sur les avant-bras et la jambe. Ils peuvent aussi être placés sur la poitrine, près des bras et au-dessus du bas de l'abdomen droit (c'est-à-dire juste au-dessus de la hanche droite).

A l'aide du capteur mentionné, on a pu acquérir des signaux réels des stagiaires présents dans la plateforme. Les signaux de 3 personnes sont enregistrés sur plusieurs reprises. On a configuré un fichier de données comportant les descripteurs de chaque signal réel. De chaque signal réel en entrée, quatre complexes QRS sont détectés. On a prétraité les signaux et on a extrait les descripteurs.

c. Construction du modèle de classification Machine Learning

La signature unique du signal ECG ne peut pas être identifiée à l'oeil nu. L'efficacité du machine Learning consiste à pouvoir extraire cette unicité à l'aide des différents descripteurs fournis.

Dans une première étape les données utilisés pour l'implémentation sont les signaux ECG de 15 personnes. Ensuite, pour l'implementation réelle les signaux de 3 personnes sont utilisés. On a extrait les différents descripteurs de chaque signal de chaque sujet.

Une normalisation des données est indispensable pour avoir des valeurs sur une même échelle.

L'algorithme choisie est Support Vector Machine.

Dans ce projet, les données collectées sont non linéaires. Dans ce cas, le SVM mappe les points vers un espace de fonctions de grande dimension basé sur la fonction du noyau, ce qui rend la séparabilité des données plus claire. Le noyau utilisé est la distribution gaussienne de données : RBF.

Le Kernel RBF crée une combinaison non-linéaire de descripteurs par transformer les points de données à un espace de dimension supérieure dans lequel on utilise des limites de décision linéaires afin de séparer les classes.

Des méthodes comme grid search vont permettre un changement itératif des paramètres mentionnés à fin de maximiser les performances.

Pour qu'on puisse choisir le meilleur modèle, on a testé plusieurs algorithmes de classification comme K-nearest neighbor et Decision Trees. La meilleure performance correspond au modèle SVM.

Dans cette étape on a utilisé les données dans le but de maximiser le pouvoir de détection du modèle.

Le modèle de classification construit est un modèle de classification Machine Learning SVM multiclasse, à kernel gaussian. Les données d'entrée sont partagées en 2 parties : apprentissage et évaluation. Le partage est assuré par la fonction train test split de la bibliothèque Sklearn du langage python.

La taille des données test est 26%. Ensuite, la normalisation des données est assurée par StandardScaler de la bibliothèque Sklearn.processing. Le modèle fonctionne avec une accuracy = 0.9375.

Chaque personne représente une classe dans le modèle. Donc le modèle dispose de 15 classes. La prédiction d'une personne non introduite pour le training du modèle est un problème. Pour être clair, le modèle est censé avoir un output même si le signal introduit ne correspond à aucune classe. Par suite, l'output est une classe que le modèle a trouvé similaire à la classe de features introduits. Pour surmonter ce problème, on a utilisé le pourcentage des pics correctement classées avec un seuil pour prédire si la classe de sortie appartient au

modèle. Si le pourcentage de prédiction est inférieure au seuil, la personne test est anonyme.

L'algorithme SVM est utilisée dans l'implémentation de la solution à l'aide des signaux réels. Le modèle fonctionne avec une accuracy = 91%.

d. Déploiement du modèle dans une application android

Le déploiement du la solution complète est nécessaire pour l'atteinte des objectifs. Ceci présentait un problème dans une première phase. En effet, le logiciel Matlab été utilisé pour le traitement de signal et l'extraction des paramètres. On se trouvait face à un problème réel puisque le déploiement de l'application en utilisant Matlab peut ne pas être réalisable. En effet, Matlab n'est pas simple à communiquer avec les bases de données. D'une part, ce logiciel ne peut pas être utilisé sur le cloud d'une facon automatisé. D'autre part, c'est compliqué de convertir des fonctions développées avec Matlab pour être utilisé dans le système android. Ces facteurs nous ont poussé à implémenter la solution en langage Python. Plusieurs bibliothèques mathématiques et de traitement de signal ont été utiles.

Il existe plusieurs solutions permettant le déploiement d'un code python en une application web. Notamment, Pythonanywhere qui nous offre un environnement de développement propres à nous. Au début, on doit configurer l'environnement avec le langage et les bibliothèques souhaités. Flask ,qui est un micro framework développé en python, assure le developpement de l'application web à l'aide du code python. Ensuite, il suffit d'inclure les fichiers de codes de l'application web dans le répertoire spécifique et donc ces fichiers seront hébergés sur pythonanywhere.
On a développé une page web simple, comportant deux boutons l'une pour déclencher le processus total de prédiction et l'autre pour pouvoir afficher le profil de la personne détectée. L'affichage du profil de la personne est une fonction dans l'application android. Une sélection basée sur une clé(la référence de la personne détectée) dans une base de données de type SQLite est exécutée dans ce sens. La base de données contient des informations des personnes avec lesquelles le modèle est construit : nom , prénom, photo de profil et date de naissance.

Cette page web est utilisée dans l'application android finale grâce aux objets WebView. Ces objets permettent d'afficher le contenu d'une page web dans une activité, mais ne possèdent pas toutes les fonctionnalités du navigateur. Une WebView est utile lorsque vous avez besoin d'un contrôle accru sur l'interface utilisateur et d'options de configuration avancées vous permettant d'intégrer des pages Web dans un environnement spécialement conçu pour votre application.

Cette activité android principale, assure le déclenchement des processus mentionnés. Le résultat de prédiction est une classe, sous forme d'entier. Cette classe sera affichée avec un pourcentage de détection. L'entier, représentant la classe de la personne détectée, est utilisé comme référence ou clé pour la sélection de cet individu de la base de données SQLite liée à l'application. Une fois le bouton "Sign in" est utilisé, le profil de la personne identifiée s'affiche.

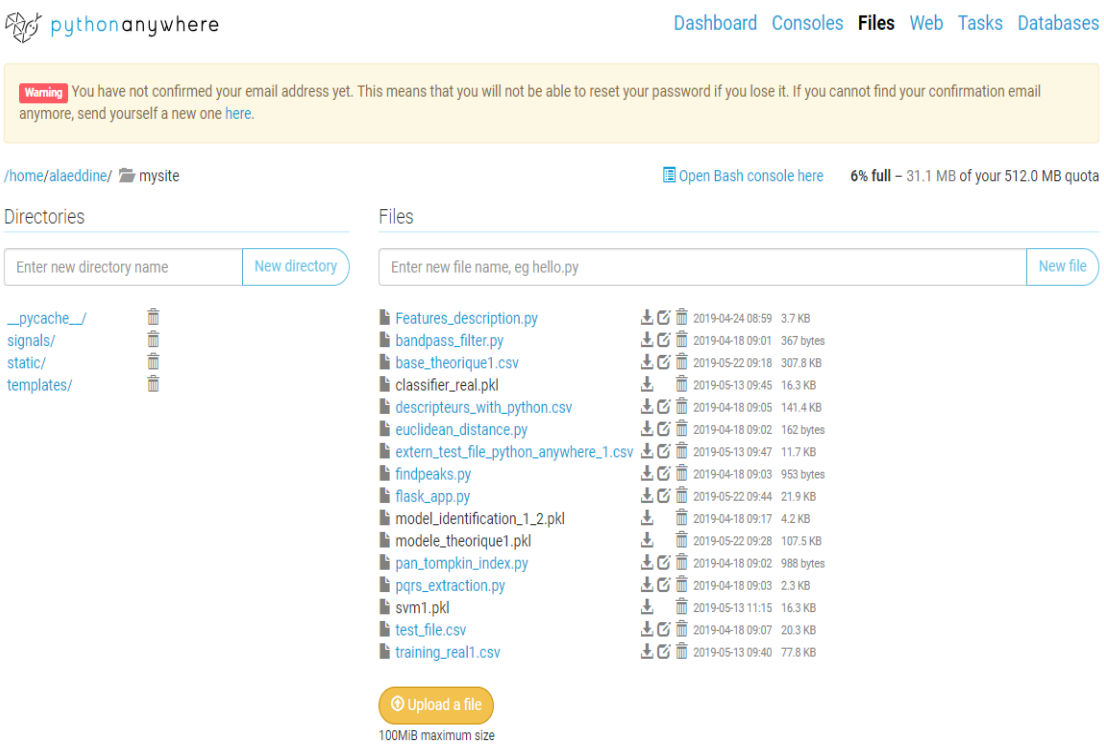


Figure 36 – Hébergements des fichiers de l'application

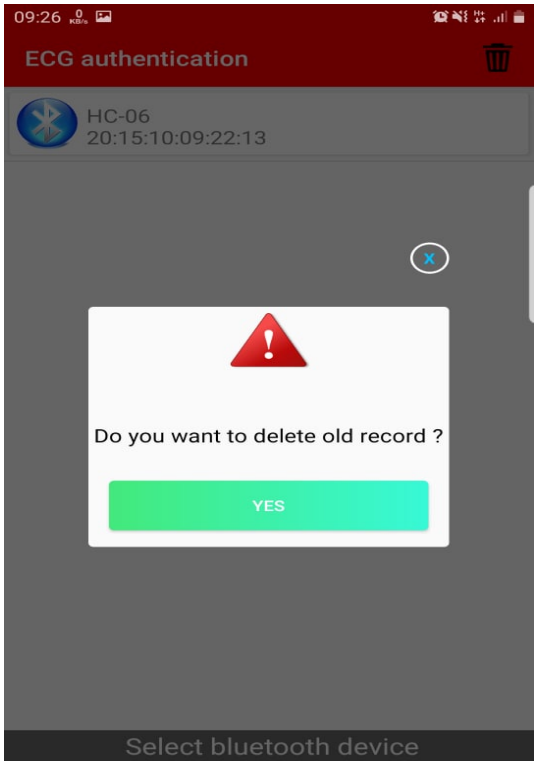


Figure 37 – Vider la base de données

La base de données Real-time de Firebase peut contenir un signal préalablement enregistré. Il faut le supprimer.

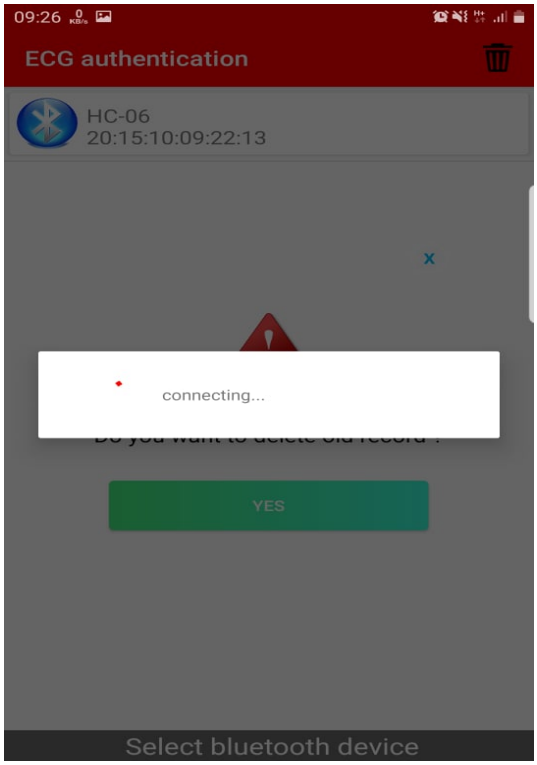


Figure 38 – Connexion au module HC-06 via Bluetooth

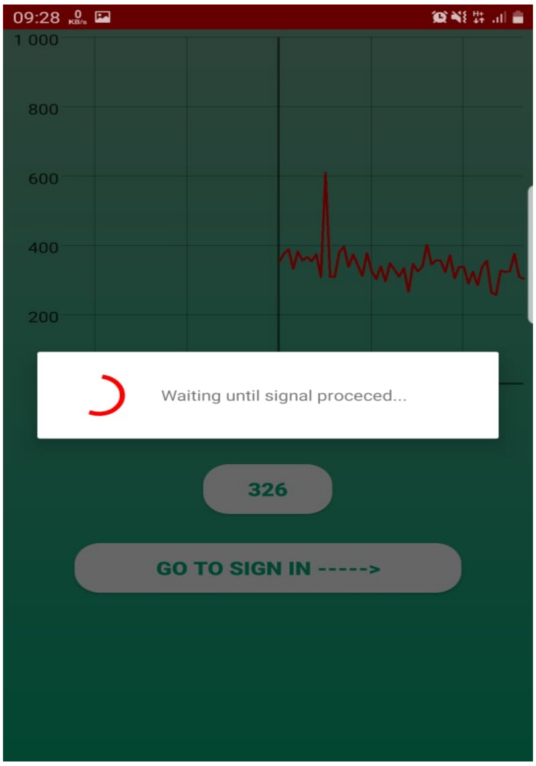


Figure 39 – Réception des valeurs du signal ECG

Figure 40 – visualisation du signal ECG

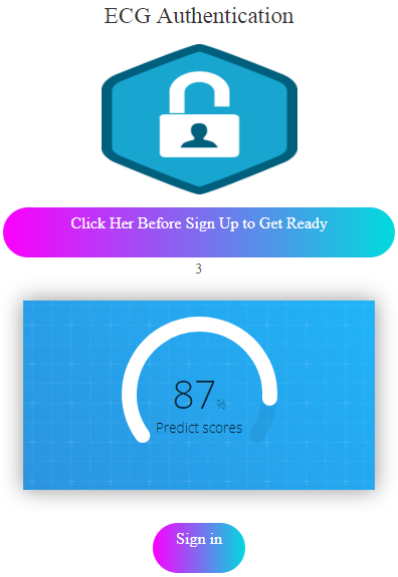


Figure 41 – Interface de l'authentification

Suite à l'authentification, l'application android affiche le profil de l'utilisateur.

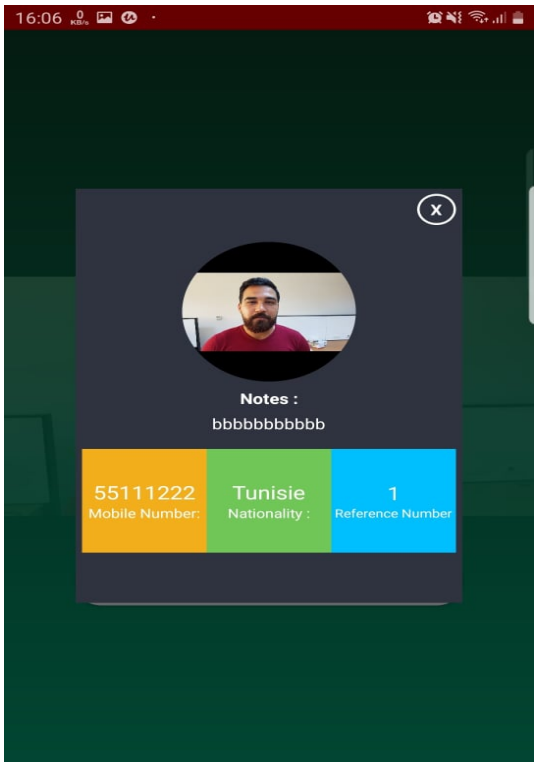


Figure 42 – Profil de la personne authentifiée

Dans le cas ou la personne test(l'utilisateur de l'application) n'est pas identifiée(elle ne figure pas dans les données avec lesquelles le modèle de classification est construit) l'application la considère comme anonyme.

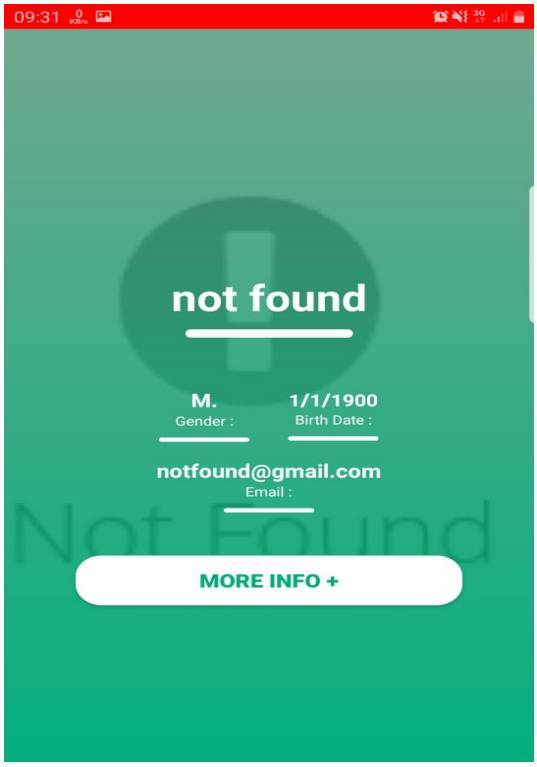


Figure 43 – Classe anonyme

III.4 Évaluation système

L'évaluation est une partie nécessaire et importante dans la conception d'un système viable. Cette partie du rapport traite les différentes métriques utilisés afin de confirmer la performance et la fiabilité d'un système d'authentification biométrique. Tout au long de cette partie, on assume que les données d'entrée du système sont partagées en données d'apprentissage(Training) et de test. L'ensemble des données d'apprentissage est utilisé pour optimiser les paramètres du système biomètrique et les données de test sont utilisées uniquement pour obtenir la mesure de performance du système final. Dans le modèle de classification final, 74% de données initiales sont utilisées pour le training(la construction du modèle de classification) et 26% sont utilisées pour la phase de test.

a. Précision

L'Accuracy est la précision de classification, C'est le rapport entre le nombre de prédictions correctes et le nombre total d'échantillons d'entrée. L'accuracy peut nous indiquer immédiatement si un modèle est correctement construit et comment il peut fonctionner d'une

façon générale. Toutefois, elle ne donne pas d'informations détaillées concernant le modèle et son efficacité aux nouveaux points de données.

Il est vrai que la précision nous informe globalement sur le modèle de classification, mais ceci ne peut pas être un bon indicateur lorsque les données de différentes classes ne sont pas équilibrées.

$$prcision = (Vraispositifs + Fauxngatifs)/(Nombretotald'chantillonsdedonnes)$$

b. Matrice de confusion

La matrice de confusion comme son nom l'indique nous donne comme sortie une matrice décrivant la performance totale du modèle.

Dans le cas d'un problème binaire, les échantillons de données appartiennent à deux classes.

n=165	**Predicted: NO**	**Predicted: YES**
Actual: NO	50	10
Actual: YES	5	100

Figure 44 – Matrice de confusion d'un modèle Machine Learning multiclasse

Les valeurs présentes dans cette matrice, selon leurs positions, indiquent quatre types d'éléments.

-Vrai négatif(True Negatif) : le résultat prédit et le résultat correct sont : négatif
-Vrai positif(True positif) : le résultat prédit et le résultat correct sont : positif
-faux négatif(False Negatif) : le résultat prédit est positif et le résultat correct est négatif
-Vrai négatif(False Positif) : le résultat prédit est positif et le résultat correct est négatif

La figure ci dessous représente la Matrice de confusion du système implémentée de type SVM.

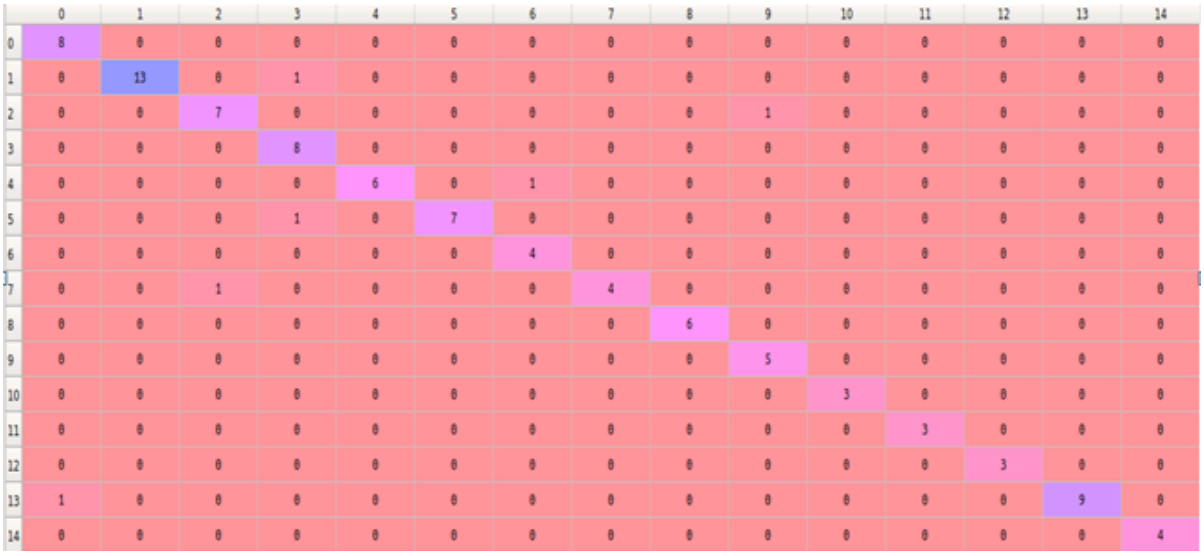


Figure 45 – Confusion Matrix du système d'authentification ECG

c. Rapport de classification

Le rapport de classification contient plusieurs métriques d'évaluation : precision, recall et f1-score. Les métriques sont définies en termes de vrais et faux positifs, et vrais et faux négatifs. Positif et négatif dans ce cas sont des noms génériques pour les classes d'un problème de classification.

	precision	recall	f1-score	support
1.0	0.89	1.00	0.94	8
2.0	1.00	0.93	0.96	14
3.0	0.88	0.88	0.88	8
9.0	0.80	1.00	0.89	8
10.0	1.00	0.86	0.92	7
14.0	1.00	0.88	0.93	8
16.0	0.80	1.00	0.89	4
24.0	1.00	0.80	0.89	5
25.0	1.00	1.00	1.00	6
26.0	0.83	1.00	0.91	5
28.0	1.00	1.00	1.00	3
30.0	1.00	1.00	1.00	3
46.0	1.00	1.00	1.00	3
52.0	1.00	0.90	0.95	10
53.0	1.00	1.00	1.00	4
micro avg	0.94	0.94	0.94	96
macro avg	0.95	0.95	0.94	96
weighted avg	0.95	0.94	0.94	96

Figure 46 – Classification report

Precision La précision est la capacité d'un modèle de classification à ne pas attribuer une instance comme positive alors qu'elle est réellement négative. Pour chaque classe, il est défini comme le ratio de vrais positifs à la somme des vrais et faux positifs.

Recall C'est la fraction des positifs qui sont correctement identifiés.

$$Recall = TruePositif/(TruePositif + FalseNegatif)$$

F1-score Le score F1 est une moyenne harmonique pondérée de Precision et de Recall de telle sorte que le meilleur score est 1,0 et le pire est 0,0. En général, les scores F1 sont inférieurs aux mesures de précision car ils intègrent la Precision et le Recall dans leur calcul.

$$F1Score = 2 * (Recall * Precision)/(Recall + Precision)$$

En plus de ces différentes métriques, on a utilisé une métrique d'évaluation du modèle qui est : Cross Validation.

Cross validation : c'est une technique de validation des modèles, qui permet d'évaluer la généralisation des résultats d'une analyse statistique à un ensemble de données indépendant.

Le but principal du Cross validation est de définir la taille de données utilisées pour l'apprentissage afin de définir plusieurs problèmes de construction comme : Overfitting et underfitting.

Cela nous aide à évaluer la qualité du modèle, à sélectionner le meilleur modèle et à se mettre à l'abri de l'underfitting(le modèle ne peut pas déterminer suffisamment des descripteurs à partir de données en entrée) et de l'overfitting(le modèle se base sur des échantillons bruités).

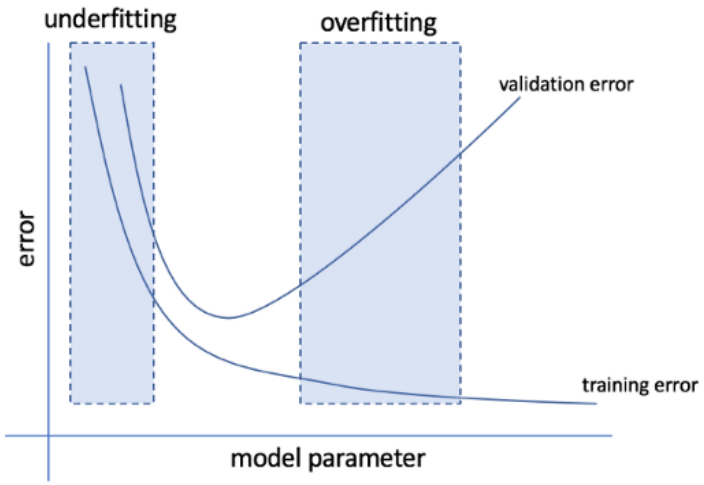


Figure 47 – Variation de l'erreur du modèle en fonction de ses paramètres

Cross validation est un indicateur de performance plus précis que la méthode de partage des données traditionnelles, nommée Train/test Split, dans laquelle l'ensemble de données est partagée en Train et test de façon indépendante. Cross validation peut être exécutée de plusieurs manières. Dans ce projet, on a utilisée K-fold Cross Validation.

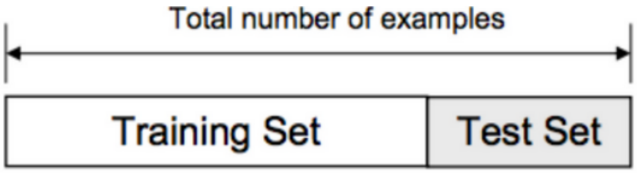


Figure 48 – Train/test Split

L'idée est que chaque point de donnée doit figurer une seule fois dans les données de test, et (K - 1) fois dans les données d'apprentissage du modèle.

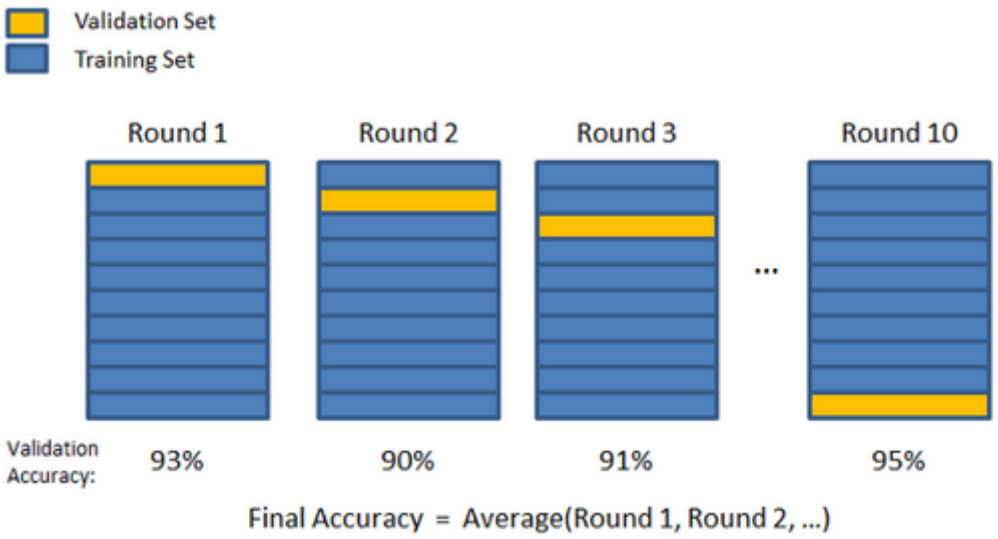


Figure 49 – K-fold Cross Validation

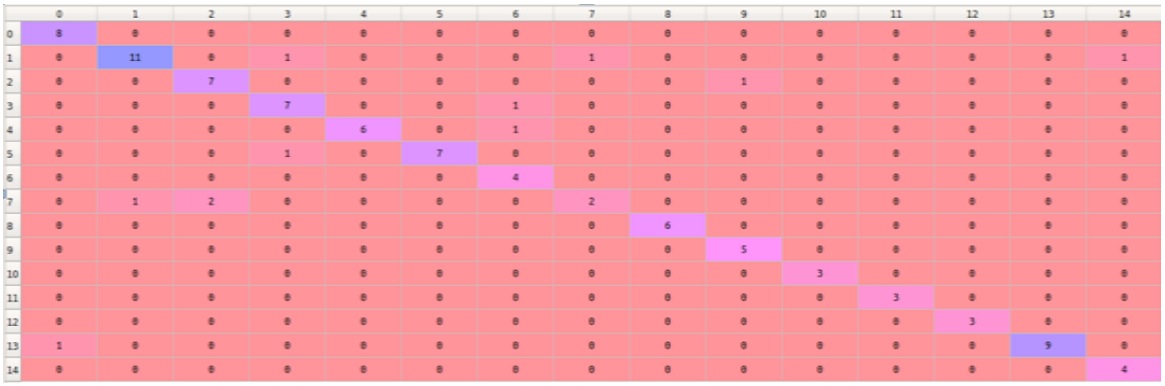
Ces techniques sont utilisées pour évaluer la performance du modèle construit sur des signaux extraits à partir de la base de données de Physionet. Ce modèle, ayant pour but de classifier le signal ECG en entrée, et par suite d'authentifier la personne. L'accuracy du modèle final est 0.9375. Cette valeur n'est obtenu qu'après plusieurs tests. Plusieurs algorithmes de classification sont utilisés comme SVM, KNN et Decision Tree.

La figure ci-dessous présente une comparaison des différents algorithmes appliqués au problème.

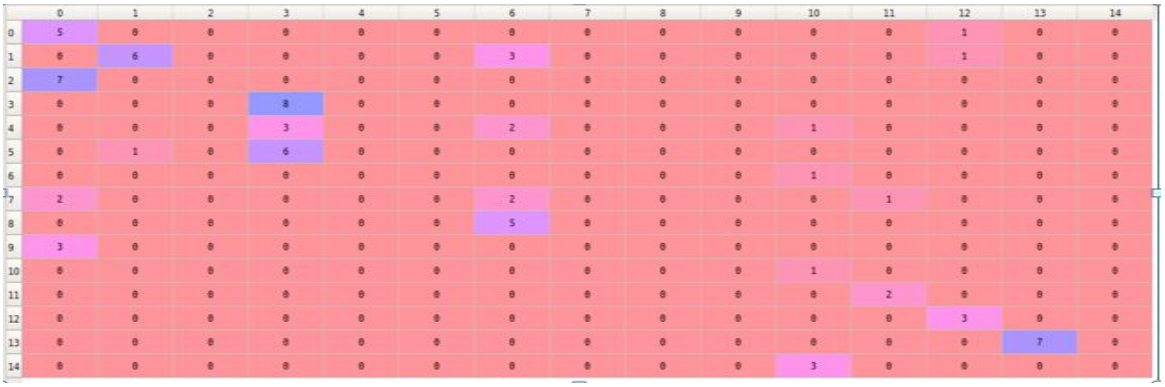
	SVM	KNN	Decision Trees
Temps de construction de modèle(secondes)	0.01	1.03 secondes	72.03 secondes
Taille du modèle(ko)	110.1	128.6	128.6

Figure 50 – Comparaison des algorithmes

L'algorithme SVM est choisi puisqu'il présentait une performance supérieure aux autres algorithmes testés.

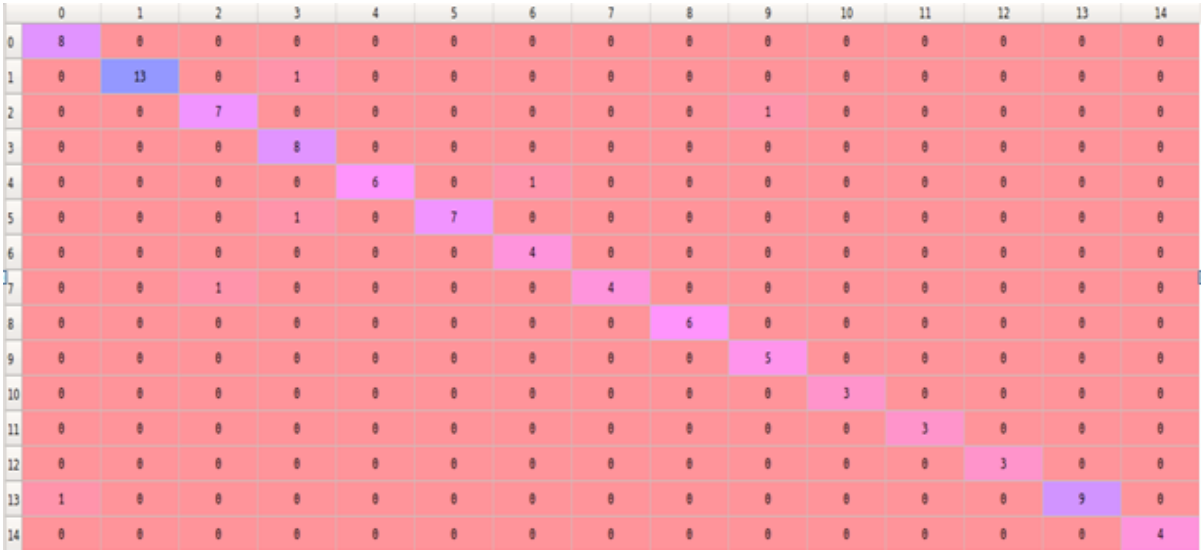


(a) Matrice de confusion de l'algorithme KNN



(b) Matrice de confusion de l'algorithme Decision Trees

Figure 51 – Matrices de confusion des algorithmes utilisés



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	13	0	1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	7	0	0	0	0	0	0	1	0	0	0	0	0
3	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	6	0	1	0	0	0	0	0	0	0	0
5	0	0	0	1	0	7	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
7	0	0	1	0	0	0	0	4	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
13	1	0	0	0	0	0	0	0	0	0	0	0	0	9	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4

Figure 52 – Matrice de confusion du modèle SVM

	precision	recall	f1-score	support
1.0	0.89	1.00	0.94	8
2.0	1.00	0.93	0.96	14
3.0	0.88	0.88	0.88	8
9.0	0.80	1.00	0.89	8
10.0	1.00	0.86	0.92	7
14.0	1.00	0.88	0.93	8
16.0	0.80	1.00	0.89	4
24.0	1.00	0.80	0.89	5
25.0	1.00	1.00	1.00	6
26.0	0.83	1.00	0.91	5
28.0	1.00	1.00	1.00	3
30.0	1.00	1.00	1.00	3
46.0	1.00	1.00	1.00	3
52.0	1.00	0.90	0.95	10
53.0	1.00	1.00	1.00	4
micro avg	0.94	0.94	0.94	96
macro avg	0.95	0.95	0.94	96
weighted avg	0.95	0.94	0.94	96

Figure 53 – Rapport de classification de l'algorithme SVM

	precision	recall	f1-score	support
1.0	0.89	1.00	0.94	8
2.0	0.92	0.79	0.85	14
3.0	0.78	0.88	0.82	8
9.0	0.78	0.88	0.82	8
10.0	1.00	0.86	0.92	7
14.0	1.00	0.88	0.93	8
16.0	0.67	1.00	0.80	4
24.0	0.67	0.40	0.50	5
25.0	1.00	1.00	1.00	6
26.0	0.83	1.00	0.91	5
28.0	1.00	1.00	1.00	3
30.0	1.00	1.00	1.00	3
46.0	1.00	1.00	1.00	3
52.0	1.00	0.90	0.95	10
53.0	0.80	1.00	0.89	4
micro avg	0.89	0.89	0.89	96
macro avg	0.89	0.90	0.89	96
weighted avg	0.89	0.89	0.88	96

(a) Rapport de classification de l'algorithme KNN

	precision	recall	f1-score	support
1.0	0.29	0.83	0.43	6
2.0	0.86	0.60	0.71	10
3.0	0.00	0.00	0.00	7
9.0	0.47	1.00	0.64	8
10.0	0.00	0.00	0.00	6
14.0	0.00	0.00	0.00	7
16.0	0.00	0.00	0.00	1
24.0	0.00	0.00	0.00	5
25.0	0.00	0.00	0.00	5
26.0	0.00	0.00	0.00	3
28.0	0.17	1.00	0.29	1
30.0	0.67	1.00	0.80	2
46.0	0.60	1.00	0.75	3
52.0	1.00	1.00	1.00	7
53.0	0.00	0.00	0.00	3
micro avg	0.43	0.43	0.43	74
macro avg	0.27	0.43	0.31	74
weighted avg	0.33	0.43	0.35	74

(b) Rapport de classification de l'algorithme Decision Trees

Figure 54 – Rapports de classification des algorithmes utilisés

La technique Cross validation est utilisée pour s'assurer de la performance du modèle SVM avec le paramètre k = 10. Ensuite, l'accuracy moyenne est calculée est égale à 91%.

III.5 Conclusion

Ce chapitre décrit de manière détaillée la progression de notre travail en partant de la préparation du bon environnement de développement jusqu'à l'étape d'évaluation du système après son implementation, l'exécution du code développé et la comparaison des résultats pratiques avec la théorie.

Conclusion générale

De nos jours, les entreprises et vu la confidentialité des données qu'elles peuvent en disposer, cherchent à atteindre un niveaux élevé de sécurisation des accès. Dans ce contexte, les recherches ont prouvé que la biométrie est un moyen efficace pour la sécurité informatique.

Dans ce projet ayant lieu à Telnet Holding, on s'est basé sur cette idée mais tout en sortant du cadre classique de la biométeie comportant l'identification des personnes par leurs visages, leurs empreintes ou leurs iris pour s'étendre à l'utilisation des battements du coeur comme caractéristique unique de chaque individu.

Les résultats obtenus à la fin du projet sont comparables aux études récentes et prouvent la fiabilité de l'utilisation des capteurs ECG(non médicaux) pour l'authentification des personnes à court terme. L'amélioration de la performance du système biométrique sur une longue durée peut se faire par la synchronisation de la biométrie stockée avec les nouveaux signaux, suite à chaque authentification correcte.

Le système actuel nécessite que l'utilisateur soit au repos. Ceci ouvre plusieurs perspectives. En effet, la solution peut être extensible en ajoutant la possibilité de détecter si le signal électrocardiogramme enregistré correspond à une personne en activité.

De plus on peut s'approfondir dans l'étude de l'axe médical, en se basant sur le fait que l'analyse et le traitement du signal électrocardiogramme permettent d'avoir une idée sur le rythme cardiaque de l'individu et donc de détecter la présence de quelques anomalies.

Ce rapport était un support écrit décrivant dans ses trois grandes parties l'évolution de notre projet de fin d'études. Dans la première partie, on a introduit globalement le sujet puis dans le deuxième chapitre on a détaillé le premier niveau de construction de notre application qui est la conception et enfin dans la troisieme partie on a détaillé l'implémentation de la solution.

Ce stage était une bonne occasion pour découvrir avec un grand plaisir le champs d'étude de Machine Learning. C'était également une occasion pour appliquer ce qui a été acquis durant le parcours à SUP'COM. Enfin, on a pu découvrir de près la vie professionnelle et apprécier le travail en équipe dans une entreprise qui instaure le respect et l'harmonie.

IV Annexe

IV.1 Physiologie du coeur

Le cœur est le muscle qui pompe le sang rempli d'oxygène et de nutriments par les vaisseaux sanguins vers les tissus corporels. Le cœur contient quatre chambres : les chambres supérieures (Atria gauche et droite) sont des points d'entrée dans le cœur, tandis que les deux chambres inférieures (ventricules gauche et droite) sont des chambres de contraction envoyant du sang dans la circulation.

Le cycle cardiaque se réfère à un battement de coeur complet de sa génération au début du prochain battement, comprenant plusieurs étapes de remplissage et de vidange des chambres. La fréquence du cycle cardiaque est connue sous le nom de «fréquence cardiaque» (nombre de battements par minute, BPM).

Afin de pomper le sang, le muscle cardiaque doit contracter, ce qui nécessite une impulsion électrique. Cette impulsion vient du nœud sinus (situé dans l'atrium droit), qui est transmis par des voies spécifiques à travers le cœur, permettant un contraction et une relaxation régulière.

L'impulsion électrique générée par le cœur peut être détectée sur la surface du corps à l'aide d'électrodes placées sur la peau, ce qui se fait lors d'un test d' électrocardiogramme (ECG ou EKG).

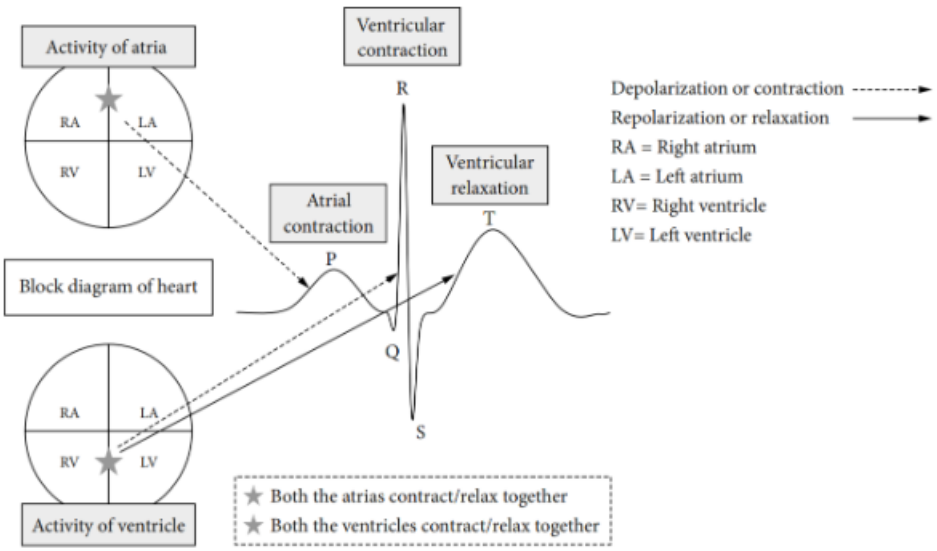


Figure 55 – Génération du signal ECG à partir de l'activité électrique du coeur

IV.2 Algorithmes de Machine Learning

Les algorithmes de Machine sont partagés selon le problème envisagé.

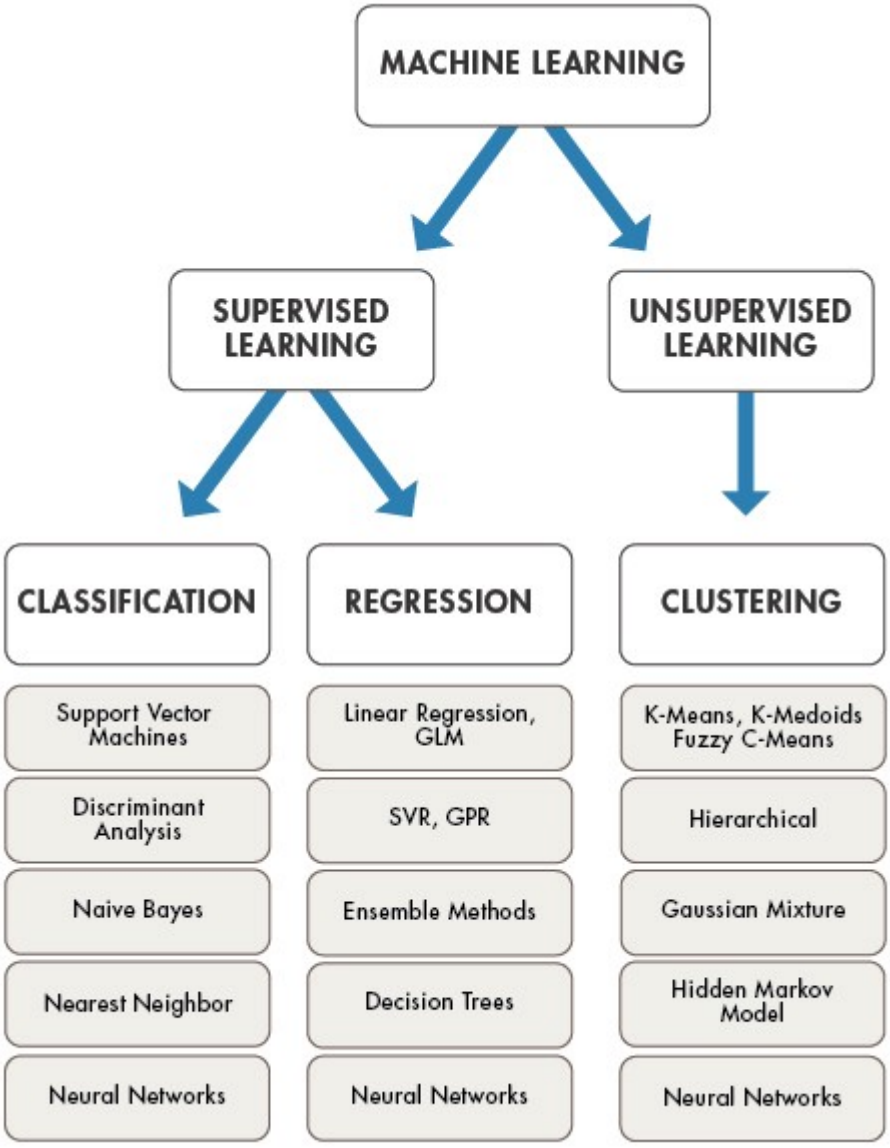


Figure 56 – Algorithmes de Machine Learning

Apprentissage supervisé :

Régression

-Linear Regression : La régression modélise une valeur de prédiction cible basée sur des variables indépendantes. Il est principalement utilisé pour découvrir la relation entre les variables et les prévisions. La régression linéaire effectue la tâche de prédiction d'une valeur de variable dépendante (y) basée sur une variable indépendante donnée (x). Ainsi, cette technique de régression découvre une relation linéaire entre x (Input) et y (output).

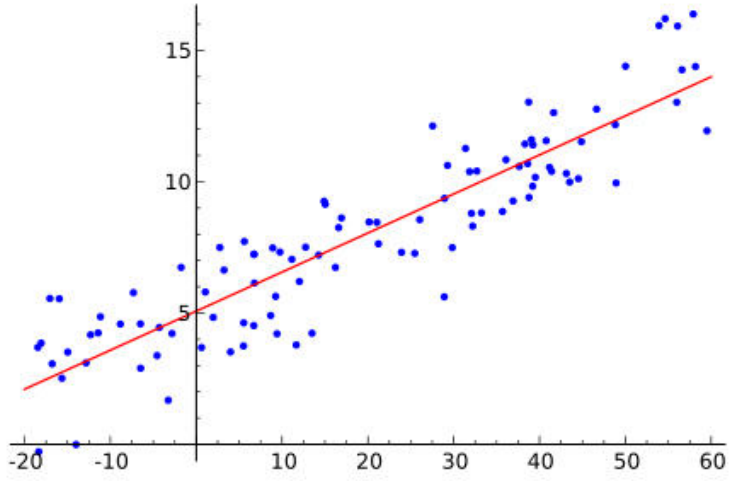


Figure 57 – Linear Regression

Classification

-Decision Trees : C'est un outil d'aide à la décision qui utilise un graphique ou un modèle de décisions arborescentes et leurs conséquences possibles. Ceci peut être représenté par une série de nœuds, un graphe qui démarre à la base avec un nœud unique et s'étend aux nombreux nœuds de feuille qui représentent les catégories que l'arborescence peut classer.

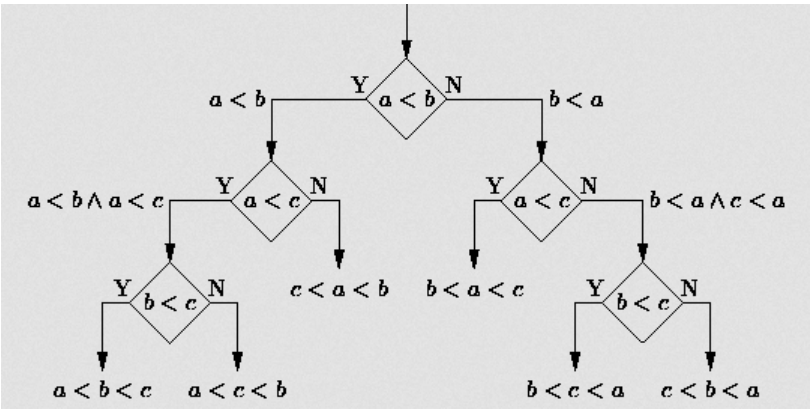


Figure 58 – Decision Tree

Cet algorithme permet d'aborder le problème d'une manière structurée et systématique pour arriver à une conclusion logique.

-Naive Bayes : C'est une famille de modèle de classifications probabilistes simples basés sur l'application du théorème de Bayes avec des hypothèses d'indépendance fortes entre les caractéristiques. Cet algorithme est simple, efficace et couramment utilisé. Ce modèle est facile à construire et particulièrement utile pour les ensembles de données très volumineux.

$$P(c|x) = (P(x|c) * P(c))/P(x)$$

P (c | x) : probabilité de la classe c sachant l'attribut x
P (c) : probabilité de la classe c
P (x | c) : probabilité de l'attribut x sachant la classe c
P (x) : probabilité de l'attribut x

- Support vector Machines : SVM est un algorithme de classification binaire. Étant donné un ensemble de points de 2 types en N dimensions, SVM génère un hyperplan de dimension (N-1) pour séparer ces points en 2 groupes. Cet algorithme trace chaque élément de données sous forme de point dans un espace à N dimensions(N est le nombre d'attributs). Ensuite, le modèle cherche à séparer les deux classes de la meilleure façon.

Dans plusieurs cas, les données fournies au modèle ne sont pas linéairement séparables. Par suite, le modèle SVM utilise le "Kernel trick" pour le mapping de données vers un espace de dimensions supérieures. Par conséquent, il sera plus facile de séparer les données. Pour être clair, pour résoudre le problème de séparation non linéaire, SVM effectue des transformations de données extrêmement complexes, puis découvre le processus permettant la séparation des données en fonction des étiquettes ou de sorties définies.

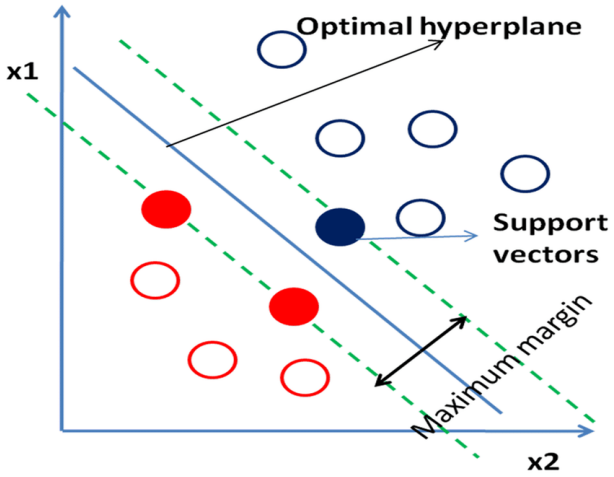


Figure 59 – Support Vector Machine

-Random Forest : C'est un algorithme de Machine Learning flexible et facile à utiliser qui est performant dans plusieurs situations. Il est également l'un des algorithmes les plus utilisés, grâce à sa simplicité et au fait qu'il peut servir aux tâches de classification et de régression. D'abord, l'algorithme choisit K points de données à partir des données d'apprentissage d'une façon aléatoire. Ensuite, il construit l'arbre de décision correspondante à ces

k points de données. Puis, il faut choisir le nombre Ntree d'arbres à construire et on exécute les deux étapes précédentes. Pour un nouvel échantillon de données, chaque arbre de décision doit prédire la catégorie à laquelle correspond ce point, et l'algorithme doit assigner le nouveau point à la classe ayant la majorité des votes. Random Forest a presque les mêmes hyperparamètres que "Decision Trees".

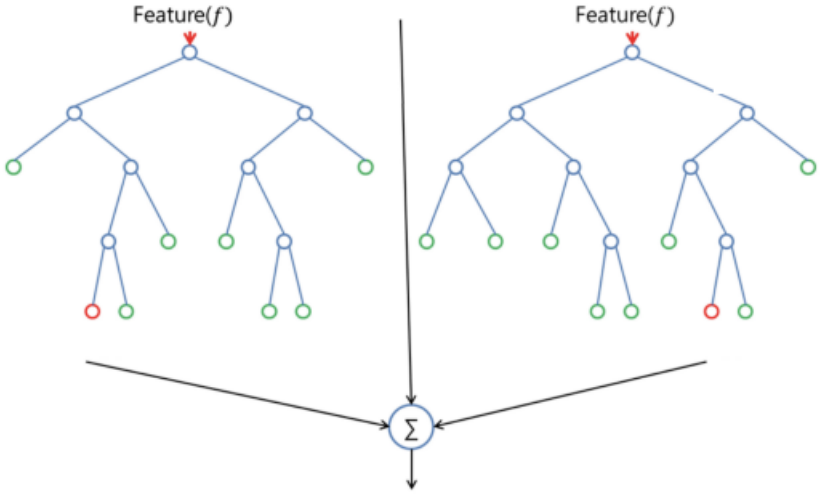


Figure 60 – Random Forest

KNN : K-nearest neighbor. KNN peut être utilisé pour les problèmes de classification et de régression. Cependant, il est largement utilisé dans les problèmes de classification.

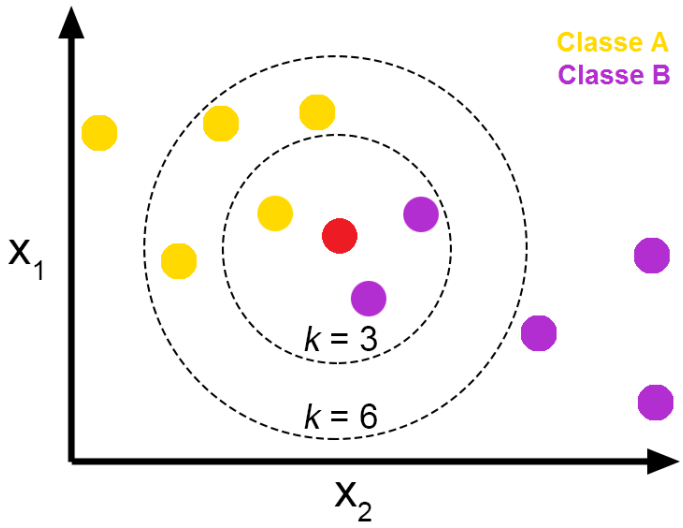


Figure 61 – K nearest neghibors

KNN utilise une base de données dans laquelle les données appartiennent aux différentes classes mais ne sont pas etiquetés. Cet algorithme est utilisé quand on n'a que peux de savoir sur l'ensemble données. KNN ne forme aucune hypothèse sur le processus de classification, et toutes les données utilisées pour la phase d'apprentissage sont nécessaires lors de la phase de test.

Tout d'abord il faut choisir le nombre k de voisins et la métrique à utiliser : distance Euclidienne, Chebychev ... Puis, il faut extraire les K plus proches voisins du nouvel échantillon (point de donnée) à l'aide de la métrique choisie. Finalement, l'échantillon de donnée est assigné à la classe comportant plus de voisins.

Apprentissage non supervisé : Ces algorithmes sont généralement les algorithmes de "Clustering" ou de regroupement. Le "Clustering" est la tâche consistant à regrouper un ensemble d'objets de sorte que les objets du même groupe (cluster) soient plus semblables les uns aux autres que ceux des autres groupes. On peut utiliser le "Clustering" pour obtenir des informations significatives à partir des données, en les analysant et en déterminant les groupes auxquels ils appartiennent.

Un exemple d'algorithme de "Clustering" est **K-means Clustering**. Les algorithmes K-means visent à regrouper dans des classes les points de données qui sont similaires. D'abord, on choisit le nombre K de "clusters" à considérer. Ensuite, l'algorithme choisit aléatoirement les centroïde de chaque "cluster". Les points de données sont assignées au "cluster" dont le centroïde est le plus proche. Puis, les nouveaux centroïdes sont calculés à partir de la moyenne des points de données appartenant au "cluster". La troisième étape est réexecutée.

Références

[1] F. Sufi, I. Khalil, and J. Hu, “Ecg-based authentication,” in *Handbook of information and communication security*. Springer, 2010, pp. 309–331.

[2] K. K. Patro and P. R. Kumar, “Effective feature extraction of ecg for biometric application,” *Procedia computer science*, vol. 115, pp. 296–306, 2017.

[3] Machine Learning Tutorial : https://www.guru99.com/machine-learning-tutorial.html.

[4] Choosing the right estimator : https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html.

[5] The 7 steps of Machine Learning : https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e.

[6] P. Trivedi and S. Ayub, “Detection of r peak in electrocardiogram,” *International Journal of Computer Applications*, vol. 20, pp. 0975–8887, 2014.

[7] L. Ortiz-Martin, P. Picazo-Sanchez, P. Peris-Lopez, and J. Tapiador, “Heartbeats do not make good pseudo-random number generators : an analysis of the randomness of interpulse intervals,” *Entropy*, vol. 20, no. 2, p. 94, 2018.

[8] H. Sedghamiz, “Matlab implementation of pan tompkins ecg qrs detector,” *Code available at the File Exchange site of MathWorks. URL https ://fr. mathworks. com/matlabcentral/fileexchange/45840-complete-pan-tompkins-implementationecg-qrs-detector*, 2014.

[9] Best Python Training : http://site-1568605-6817-6085.strikingly.com/blog/best-python-training-institute-in-noida-92c97d6d-9da0-48a4-b98c-0e6a479e4789.

[10] Best Python IDE : https://kiloretad.pw/Top-10-Best-Python-IDE-For-Windows-And-Linux-t.html.

[11] https://physionet.org/.

[12] AD8232 Heart Rate Monitor : https://learn.sparkfun.com/tutorials/ad8232-heart-rate-monitor-hookup-guide/all.

[13] N. Samarin, “A key to your heart : Biometric authentication based on ecg signals,” *Project Report, Computer Science School of Informatics University of Edinburgh*, 2018.

[14] Y. Özbay, R. Ceylan, and B. Karlik, “A fuzzy clustering neural network architecture for classification of ecg arrhythmias,” *Computers in Biology and Medicine*, vol. 36, no. 4, pp. 376–388, 2006.

[15] P. Stavroulakis and M. Stamp, *Handbook of information and communication security*. Springer Science & Business Media, 2010.

[16] What Is ECG and How Does It Work ?: https://imotions.com/blog/what-is-ecg/.

[17] N. V. Thakor and Y.-S. Zhu, “Applications of adaptive filtering to ecg analysis : noise cancellation and arrhythmia detection,” *IEEE transactions on biomedical engineering*, vol. 38, no. 8, pp. 785–794, 1991.

[18] E. A. Ashley and J. Niebauer, *Cardiology explained*. Remedica, 2004.

[19] Machine Learning : https://searchenterpriseai.techtarget.com/definition/machine-learning-ML/.

[20] Commonly used Machine Learning Algorithms : https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/.

[21] M. Sansone, R. Fusco, A. Pepino, and C. Sansone, “Electrocardiogram pattern recognition and analysis based on artificial neural networks and support vector machines : a review,” *Journal of healthcare engineering*, vol. 4, no. 4, pp. 465–504, 2013.

[22] P. K. Gakare, A. M. Patel, J. R. Vaghela, and R. Awale, “Real time feature extraction of ecg signal on android platform ieee conference on communication,” *Information & Computing Technology, Mumbai, India*, 2012.

[23] Complete Pan Tompkins Implementation ECG QRS Detector : https://www.mathworks.com/matlabcentral/fileexchange/45840-complete-pan-tompkins-implementation-ecg-qrs-detector.

[24] https://physionet.org/cgi-bin/atm/ATM.