الشعبية الديمقراطية الجزائرية الجمهورية

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

Université de 8 Mai 1945 – Guelma -

Faculté des Mathématiques, de l'Informatique et des Sciences de la matière

Département d'Informatique

Mémoire de Fin d'études Master

Filière : Informatique

Option : Systèmes Informatiques.

Thème :

Exploitation des séries chronologiques pour l'étude des données évolutives

Encadré Par :

Dr. Aicha AGGOUNE

Présenté par :

Mme. Benretem Zineb

Septembre 2021

Remerciement

Je remercie d'abord DIEU de m'avoir permis de terminer ce travail dans les meilleures conditions et qui a éclairé mon chemin et m'a doué de la connaissance.

Je remercie mon encadrante Dr. Aicha AGGOUNE qui m'a toujours soutenu par son aide et ses précieux conseils.

J'exprime mon remerciement à tous les enseignants de département d'informatique qui m'ont suivi durant mon parcours académique et qui ont su me transmettre leurs savoirs faire.

Enfin, je remercie tous ceux qui ont contribué à l'élaboration de ce travail de près ou de loin et qui méritent d'y trouver leurs noms.

Dédicaces

Je dédie ce travail,

À mes très chers parents et mes frères, qui

m'ont permis de

devenir ce que je suis aujourd'hui.

À mon marie qui m'encourage de terminer mes études

À mes enfants,

Maram, Chahd, Ahmed

À mes collègues de travail et amis,

Nassima , lamia et Yacin

À Mr le directeur de l'ESPS de oued zenati

Résumé

L’intérêt d’étudier les données évolutives au fil du temps a considérablement augmenté à l'ère de technologie de l’information et de communication. Les données ECG de l´électrocardiogramme est un exemple de données évolutives en pleine utilisation afin d’établir un diagnostic des maladies cardiovasculaires. Ces données peuvent être considérées comme une série de valeurs liées au temps. L’objectif de ce travail est double : la modélisation de données évolutives de l’ECG par les séries chronologiques (ou temporelles) avec le stockage de données dans un système de gestion de données temporelles à large échelle appelé InfluxDB et l’analyse descriptive de ces données en se basant sur la visualisation graphique de données en vue d'une classification supervisée de données. En effet, le système développé est basé sur l’utilisation de framework Chronograf et Grafana qui offrent non seulement des dashboards de haute qualité mais aussi des opérations de détection des pics par des seuils d'alertes et même d’interrogation de données.

Mots clés : Données évolutives, Séries chronologiques, Données ECG, Visualisation de données, Stockage de données.

Abstract

In the age of information and communication technologies, there has been a surge in interest in the study of changing data over time. ECG data from an electrocardiogram is an example of evolving data that is fully used for the diagnosis of cardiovascular disease. These data can be viewed as a series of time-related values. The aim of this work is twofold: ECG evolving data modeling using time series with the data storage within temporal large scale data management system, called InfluxDB and descriptive analysis of these data based on data graphical visualization for data supervised classification. Indeed, the developed system is based on the use of Chronograf and Grafana frameworks which offer not only high-quality dashboards but also even operations such as the peak detection through alert thresholds and data querying.

Keywords: Evolving Data, Time Series, ECG Data, Deep Learning, Data visualization, Data storage.

Table des Matières

Chapitre 03. ***Etat de l'art sur l'étude des données ECG modélisées par les séries chronologiques***

Chapitre 04. *Conception et implémentation*

Liste des figures

Chapitre 01

Chapitre 02

Chapitre 03

Chapitre 04

Liste des tables

Introduction Générale

Avec le développement rapide du matériel informatique et des appareils intelligents, les données sont de plus en plus évolutives au cours du temps. Cette évolution peut être liée à la nature de données elles même comme les données temporelles (date, heure, instant, période, …etc.) ou à la mise à jour continuelle de données, par exemple le contenu de fichiers log, les données de la démographie, ou encore les données issues de l'internet des objets (IoT Internet of Things). Les données qui sont influencées par des facteurs temporels dites données évolutives. Parmi les domaines d'applications extrêmement utilisant les données évolutives, nous citons : la médecine, la météorologie, la finance et la bourse, etc.

Dans le domaine médical, les données issues des dispositifs médicaux intelligents sont aujourd'hui les plus répondus. L'électrocardiographe est l'un des appareils médicaux qui produit des données évolutives, appelées électrocardiogrammes (ECG) utilisés par les cardiologues. L'ECG détermine la fréquence cardiaque, c'est-à-dire le nombre de cycle cardiaque par unité de temps. Le médecin interprétera le tracé obtenu pour y déceler d'éventuelles anomalies (arythmie, insuffisance coronarienne, etc.). Or, l'analyse superficielle de l'ECG par le médecin peut conduire à des fausses interprétations et donc mauvaises orientations. L'analyse automatique de telles données s'avère en effet nécessaire pour mieux aider le médecin à diagnostiquer profondément le patient, prévoir son évolution future et planifier les thérapies à adopter.

Pour une analyse automatique de données ECG, plusieurs travaux ont été proposés dans la littérature et qui sont basés sur des techniques d'apprentissage automatique (Machine Learning), et des systèmes experts qui permettent l'exploitation d'un certain nombre de connaissances fournies explicitement par des experts du domaine [67].

Actuellement, les données ECG sont devenues très évolutives en termes de volume et de fréquence utilisation par non seulement des médecins spécialistes mais aussi par les médecins généralistes ou encore par les chercheurs des instituts de cardiologies[1], etc. L'analyse de ces données est devenue une tâche fastidieuse via les techniques traditionnelles comme système

[1] https://www.ottawaheart.ca/fr/chercheurs

expert et des techniques de machine learning qui reposent sur l'ingénierie et la sélection des caractéristiques de domaine étudié à l'aide des experts de domaine. L'apprentissage profond (Deep Learning) qui fait partie de machine learning permet d'exploiter des données volumineuses avec une sélection des caractéristiques d'une manière implicite sans intervention d'un expert de domaine. Cette analyse nécessite une étude descriptive appelée aussi analyse descriptive afin de mieux comprendre les données (Data Understanding).

Par ailleurs, la gestion efficace de données évolutives nécessite la mise en place d'un modèle de données à base de temps. Les séries chronologiques (Time series) permettent de représenter les données comme des suites d'observations indexées par le temps.

L'objectif principal de ce projet de fin d'étude est de réaliser une étude de données évolutives modélisées par les séries chronologiques. L'étude choisie dans ce travail concerne l'analyse descriptive et visualisation graphique de données (Data visualization). La méthode proposée est fondée sur l'utilisation à la fois de SGBD temporelles à large échelle InfluxDB plutôt que d'utiliser des simples fichiers CSV et de framework Chronograf et Grafana pour la visualisation et la manipulation de données évolutives. Cette étude est une étape importante de la classification automatique de données pour en tirer des informations utiles.

Hormis la conclusion générale, le reste de ce mémoire est structuré en 04 chapitres :

Chapitre 01 : Les données évolutives et les séries chronologiques : Ce chapitre représente un état de l'art sur les données évolutives et les séries chronologiques ainsi que les relations entre elles.

Chapitre 02 : Les signaux ECG : Dans ce chapitre, nous présentons un exemple de données évolutives de domaine médical utilisées pour réaliser notre étude travail. Il s'agit de données issues de l'appareil d'Electrocardiogramme qui génère des signaux ECG présentant l'activité électrique accompagnant les contractions du cœur. La compréhension de données est une étape très importante dans l'analyse de données et le data mining.

Chapitre 03 : Etat de l'art sur l'étude de données ECG modélisées par les séries chronologiques : Ce chapitre présente d'une part, les différentes techniques de deep learning appliquées aux séries chronologiques, l'analyse descriptive et une synthèse des travaux récents d'apprentissage supervisé de données ECG d'autre part.

Chapitre 04 : Conception et implémentation : Après la réalisation d'une recherche scientifique sur les notions de base liées à notre thématique, le chapitre 04 présente une deuxième

contribution qui sert dans un premier temps à concevoir un modèle conceptuel à base de séries chronologiques de données ECG. Dans un second temps, nous réalisons une étude de ces données en passant par une visualisation et un traitement à des fins de classification automatique de données.

Chapitre 01 :

Les données évolutives et les séries chronologiques

1. Introduction

De nos jours, les progrès technologiques et scientifiques ont conduit à l'étude de données qui deviennent de plus en plus évolutives. Les méthodes de collecte, de stockage, d'analyse et de traitement de données prennent un grand intérêt pour une bonne gestion de données. Les données évolutives se réfèrent aux données qui changent de valeurs au cours du temps. Elles peuvent être considérées comme une suite d'observations indexées par le temps appelée série chronologique ou temporelle.

Ce chapitre vise à présenter dans un premier temps les différentes notions liées aux données évolutives, notamment Big data, IoT puis dans un second temps, les séries chronologiques, leurs composants, leurs caractéristiques et leurs types. Nous allons présenter par la suite les différentes techniques de stockage et de visualisation des séries chronologiques. Nous terminerons par quelques domaines d'utilisation des séries chronologiques.

2. Données évolutives

L'émergence des applications informatique avancées, notamment IoT, DevOps, Cloud a conduit à une évolution importante de données. Les données évolutives (en anglais Evolving data) sont des données qui changent leurs valeurs dans des intervalles de temps réguliers ou irréguliers [1]. En d'autres termes, les données évolutives sont considérées comme de données variables dans le temps (Time-varying data) ou encore de données temporelles.

Les données évolutives existent dans notre vie habituelle, elles représentent des observations relatives au temps passé, présent et futur [15]. Ces observations peuvent être liées à un évènement donné comme par exemple la navigation sur le Web, la gestion de transactions bancaires, ou être liées au résultat d'une mesure donnée à titre d'exemple ,un compteur électrique intelligent qui enregistre les données de consommation d'électricité par heure et génère des données de facturation en temps réel, un moulin à vent sur une haute montagne obtient les données en temps réel de la vitesse de rotation et génère les données de la vitesse du vent et de la production d'électricité [15] [16]. En outre, les données évolutives sont volumineuses et fortement liées aux notions de Big data et IoT.

2.1. Données évolutives et le BIG DATA

L'évolution continue de données comme les données temporelles, les vidéos, l'audio, données médicales a introduit le problème de BIG data. La notion de BIG DATA (grosses données) fait référence à une collection de données massives qui sont généralement volumineuses, complexes, et hétérogènes [2] [17]. Ces grosses données évolutives ne peuvent pas être gérées par des systèmes de gestion de bases de données relationnelles.

Par conséquent, ce type de données requièrent des systèmes massivement parallèles qui s'exécutent sur de nombreux serveurs assurant efficacement une collecte, un stockage, une recherche, un partage, un transfert, une analyse et une visualisation des données [2].

Cette dernière décennie, l'un des grands défis de manipulation de données est le traitement des données évolutives au cours du temps. En raison du délai considérable lorsque le volume des données s'augmente, de la présence de redondance et du manque inné de structures de données évolutives, les modèle de données structurelles comme le relationnel ne semblent pas être suffisamment en mesure d'analyser ce type de données. De plus, généralement le modèle relationnel ne prend pas en charge des opérateurs propres aux données temporelles.

2.2. Données évolutives et IoT

La plupart de données évolutives ont été collectées à partir des appareils intelligents et des capteurs connectés entre eux et formants ce qu'on appelle un réseau des objets [18]. Le réseau des objets connectés via une connexion Internet est connu sous le nom de l'Internet des objets (en anglais Internet of Things IoT, terme le plus utiliser) [19].

« *L'Internet des objets est une vision où chaque objet dans le monde a le potentiel de se connecter à Internet et de fournir ses données afin d'en tirer des informations exploitables seul ou à travers d'autres objets connectés* » [3]

La majorité des surfaces disponibles dans le monde physique sont instrumentées avec des capteurs: rues, voitures, usines, réseaux électriques, calottes glaciaires, satellites, vêtements, téléphones, micro-ondes, récipients à lait, planètes, voire corps humains.

Par l'équipe de base de données Alibaba Cloud, les données évolutives représentent un type de données qui indiquent les changements au fil du temps dans un périphérique physique, un système,

un processus d'application ou un comportement. Il est largement utilisé dans des scénarios tels que l'Internet des objets (IoT) [4].

3. Séries chronologiques : un modèle de données évolutives

Les données évolutives sont des données temporelles dans le sens où chaque changement de données est lié au temps. Les séries chronologiques ou temporelles représentent un modèle de données de séquence liées au temps [20]. Plus généralement, on désigne par données de séquences toute collection de données ordonnées selon un critère qui peut être sémantique, biologique, temporel ou autre [21]. Par exemple, des séquences de mots dans un texte ; on parle alors d'ordre syntaxique, de séquences d'acides aminés composant une chaine d'ADN de peptides constituants une protéine, une séquence d'image et audio formant une vidéo, etc. Le changement de l'ordre de ces valeurs pourrait changer la signification des données.

Les données temporelles sont omniprésentes, elles constituent la catégorie de données de séquences la plus fréquemment rencontrée dans les applications. Elles apparaissent dans des applications émergentes visant à analyser, par exemple, le comportement des utilisateurs du web, l'évolution structurelle ou informationnelle au sein de réseaux sociaux, les courbes de charge de consommation d'énergie ou l'évolution des cours boursiers etc. Dans le domaine médical, les données temporelles sont également impliquées dans différentes applications telles que l'analyse de signaux décrivant la progression de paramètres physiologiques comme IRM, EEG, ECG, les profils d'expression de gènes, etc. [5].

3.1. Définition

Une série chronologique est une séquence ordonnée de valeurs d'une variable (par exemple, la température) à des intervalles de temps également espacés (par exemple, toutes les heures). Il s'agit une séquence de données à temps discret. Par exemple, les fichiers journaux (fichier log) et les résultats de mesures d'appareils électroniques et de capteurs IoT, peuvent être considérées comme des séries chronologiques, exprimant l'évolution de données [6].

De plus, une série chronologique décrit une séquence de données variées dans des périodes. Il s'agit d'une séquence de données à temps continu, par exemple la conservation de trace de l'ancien poste et de l'ancien salaire d'un employé.

3.2. Les composants d'une série chronologique

Une série chronologique est vue comme un tableau de valeurs observées à un instant t données.

La représentation formelle d'une série chronologique S est la suivante [22] :

$$S(t_n)_{n\in\mathbb{N}}$$

Avec :

- s_i un élément de série chronologique S à un instant t_i. Il est défini comme suit : $s_i = S(t_i)$.

- n est le nombre d'éléments de S.

Une série chronologique peut contenir tout ou une partie de composants suivants [9]:

Effet saisonnier (S), Changement cyclique (C), Tendance (T) et Variation irrégulière (I).

- **Effet saisonnier** : La saisonnalité se produit lorsque la série chronologique présente des fluctuations régulières dans des périodes moins d'une année (un mois, un trimestre, un semestre, etc.). Par exemple, les ventes au détail culminent au cours du mois de décembre.

- **Changement cyclique :** Tout modèle montrant changement qui se répète périodiquement est appelé un changement cyclique. La variation cyclique est une composante non saisonnière qui varie dans un cycle reconnaissable, par exemple les battements de cœur.
- **Tendance** : La tendance est le modèle à long terme d'une série chronologique. Une tendance peut être positive ou négative selon que la série chronologique présente une tendance à long terme croissante ou une tendance à long terme décroissante. Par exemple l'augmentation de prix d'achat, la croissance de nombre de morts dus à la pandémie de coronavirus. Si une série chronologique ne montre pas de modèle croissant ou décroissant, la série est stationnaire dans la moyenne.
- **Variation irrégulière** : Ce composant est imprévisible. Chaque série chronologique a une composante imprévisible qui en fait une variable aléatoire. Une série chronologique sans tendance et sans variations cycliques, elle contient des variations aléatoires ou irrégulières.

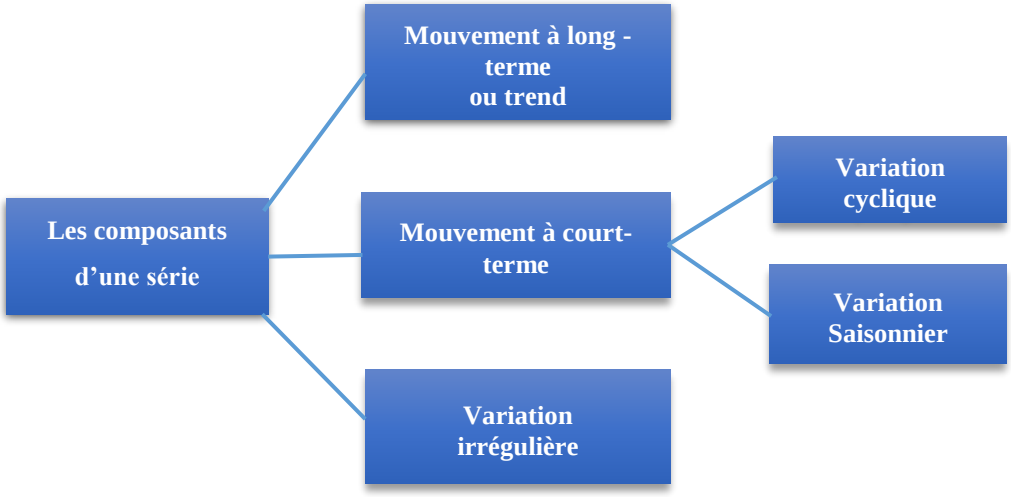


Figure I.1 : Les Composantes d'une série chronologique

Ces composants peuvent être combinés de différentes manières. On suppose généralement qu'elles sont multipliées ou ajoutées [9]:

$$Y = T \times C \times S \times I$$

$$Y = T + C + S + I$$

Pour corriger la tendance, la première expression est divisée par la tendance (T) tandis que on soustrait la tendance de la deuxième expression.

3.3. Propriétés de données d'une série chronologique

Les principales propriétés qui différencient les données de séries chronologiques aux données régulières sont la synthèse, la gestion du cycle de vie de données et l'analyse à grande échelle de nombreux enregistrements. Les données de série chronologique (DSC) se caractérisent par les propriétés suivantes [7]:

• **Emplacement des données :** Si les données associées ne sont pas localisées ensemble dans le stockage physique, les requêtes de données peuvent être très lentes et même entraîner des délais d'attente en raison des E/S. Une DSC co-localise des partitions de données dans la même

plage de temps sur la même partie physique du cluster de base de données et permet donc un accès rapide pour une analyse plus rapide et plus précise.

• **Requêtes de plage rapides et faciles** : une requête de plage permet de retourner tous les enregistrements dont une valeur est comprise entre une limite supérieure et inférieure [6]. Les DSC garantisse que les requêtes de plage sont rapides.

En outre, il convient de prendre en considération le fait que le langage de requête utilisé devrait permettre aux utilisateurs d'écrire plus facilement de telles requêtes.

• **Performances d'écriture élevées** : de nombreuses bases de données ne sont pas en mesure de répondre aux requêtes de manière rapide pendant les pics. Les données de séries chronologiques sont généralement enregistrées toutes les secondes ou même moins que cela, les opérations d'écriture doivent donc être rapides.

• **Compression des données** : comme les données de séries chronologiques sont principalement enregistrées par seconde ou même avec moins de granularité, elles nécessitent généralement une meilleure technique de compression des données. Les DSC doivent donc fournir des fonctionnalités pour effectuer des cumuls dans de tels scénarios pour le compactage des données.

• **Évolutivité** : les données de séries chronologiques augmentent très rapidement et les bases de données classiques ne sont pas conçues pour gérer cette évolutivité. D'autre part, les bases de données de séries chronologiques sont conçues pour prendre en charge l'échelle en introduisant des fonctionnalités qui ne sont possibles que lorsqu'on traite le temps. Cela peut entraîner des améliorations des performances, notamment : des taux d'insertion plus élevés, des requêtes plus rapides à grande échelle et une meilleure compression des données.

• **Convivialité**: les DSC incluent généralement des fonctions et des opérations pour l'analyse de données de séries chronologiques. Par exemple, ils utilisent des politiques de conservation des données, des requêtes continues, des agrégations de temps flexibles, des requêtes de plage, etc. Cela augmente donc la convivialité en améliorant l'expérience utilisateur en cas d'analyse liée au temps.

4. Types des séries chronologiques

Les séries chronologiques sont classées selon les critères suivants [29]: La nature de données, la nature de temps d'observation, la classification de données, les variations de données et l'intervalle du temps.

4.1. Séries chronologiques liées à la nature de données

La nature des valeurs des séries en termes de valeurs connectées ou déconnectées, et ce critère conduit aux deux types suivants :

- Série chronologique continue : Il s'agit de la série chronologique dans laquelle nous mesurons l'évolution des valeurs apparentes dans une période de temps, telle que l'heure, le jour, la semaine, le mois, le trimestre, etc. Par exemple courant électrique résultat d'un électrocardiogramme, pourcentage de naissances au cours de l'année, etc.
- Série chronologique discontinue : Il s'agit de la série chronologique dans laquelle nous mesurons l'évolution des valeurs apparentes à un moment donné, et un exemple de ces séries est le nombre d'habitants d'une ville en premier jour de chaque année.

4.2. Séries chronologiques liées à la nature de temps

La nature du temps dans lequel les valeurs de la série chronologique se produisent. Le temps peut être prédéterminé ou non spécifié, et cette échelle conduit aux deux types suivants :

- Séries chronologiques ponctuelles : elles sont mesurées à des moments prédéterminés. Par exemple : les revenus annuels moyens des particuliers, mesurées chaque fin d'année.
- Séries chronologiques non ponctuelles : dont la valeur est mesurée à des moments inattendus, tels que les chaînes de catastrophe, les accidents d'avion, les accidents de train, les accidents de voiture et une série de tremblements de terre.

4.3. Séries chronologiques liées à la classification de données

Le nombre de valeurs indique la classification de données en données binaires et données non binaires. Cette classification conduit aux deux types de séries chronologiques :

- Séries chronologiques binaires : ce sont les chaînes qui prennent l'une des deux valeurs, zéro ou un (échec ou succès). Ces chaînes apparaissent dans l'électrotechnique et la théorie de la communication.
- Les séries chronologiques non binaires : qui prennent plus de deux valeurs comme par exemple le nombre de population.

4.4. Séries chronologiques liées à la variation de données

Les changements des valeurs des séries chronologiques d'un façon croissante ou décroissante:

- Séries à tendance croissante : ce sont des chaînes dont les points peuvent même être médiatisés par une ligne droite croissante (avec une pente positive). Des exemples de ces chaînes sont celles qui représentent la population, les chaînes nationales de revenus et les chaînes d'accidents de voiture.
- Séries à tendance décroissante : ce sont les chaînes dont les points peuvent même être médiatisés par une ligne droite décroissante (avec une pente négative). Un exemple en est les chaînes de superficies agricoles dans une certaine zone, qui sont en décroissance continue en raison de la propagation de bâtiments.
- Séries à direction fixe : ce sont des chaînes dont les points peuvent même être médiés par une ligne droite fixe (avec une inclinaison nulle), et un exemple est la série d'énergie électrique consommée dans l'éclairage des feux de signalisation et les rues principales d'une ville.
- Séries avec des changements fréquents à de longs intervalles : ce sont les chaînes dont les points peuvent même être médiés par une ligne qui ressemble à la courbe du couplage sinus (ou cosinus, par exemple la Série de vente de vêtements en laine qui a lieu tous les jours de l'année, mais elle augmente en hiver et diminue en été.

4.5. Les séries chronologiques liées à l'intervalle du temps

Les intervalles du temps où les résultats de mesures ont été collectées :

- Série chronologique avec des mesures recueillies à intervalles de temps réguliers. Elle correspondant aux données métriques comme par exemple la température.
- Série chronologique avec des mesures recueillies à des intervalles de temps irréguliers. Elle correspondant aux données des événements comme par exemple le contenu de fichier log.

5. Stockage des séries chronologiques

D'après Schmidt et al. [23], les bases de données temporelles sont différentes aux bases de données de séries chronologiques :

- Une base de données temporelles est une base de données qui conservent l'historique des données dans des intervalles temporels constants indiquant leur période de validité. Par exemple, le changement de statut d'un employé.

- Une base de données de séries chronologiques contient des données recueillies à des périodes adjacentes et qui sont ordonnées chronologiquement. Par exemple, surveillance de la fréquence cardiaque.

Dans notre travail, nous nous intéressons aux données de séries chronologiques pour la représentation et l'exploration de données évolutives de l'électrocardiogramme de maladies cardiovasculaire.

Généralement, il n'existe pas une méthode standard bien définie pour la modélisation et le stockage de données de séries chronologiques [24]. La plupart des data sets des séries chronologiques sont stockées dans des fichiers aux formats CSV, JSON, XML.

Des bases de données de séries temporelles (Time series database TSDB) sont particulièrement conçues pour gérer des métriques, des événements ou des mesures horodatées, et optimisé pour mesurer le changement dans le temps. La majorité des entreprises génèrent un flux extrêmement important de mesures et d'événements (des séries chronologiques) et, par conséquent, le besoin d'un TSDB est inévitable [7].

De nos jours, de nombreuses bases de données de différents types existent et répondent chacune à des besoins bien précis. Nous pouvons citer[52] :

- Les bases de données relationnelles propres pour le traitement de séries chronologiques et offrent des techniques de stockage et de manipulation de temps comme par exemple le système TokuDB fondé sur le SGBD MySQL [25].
- Les entrepôts de données (Data warehouse) : sont conçues pour stocker des données liées à une période temporelle (des historiques), et pour gérer les gros volumes issus de multiples sources de données. Exemple la plate-forme SHAPE basée sur un entrepôt de données stockant la consommation quotidienne d'électricité.
- Les bases de données NoSQL : Hadoop TS permet de distribuer le stockage et la charge de calcul sur plusieurs machines, OpenTSDB basé sur le système HBASE et le paradigme Mapreduce
- Les bases de données tourniquet (Round-Robin database RRD) : un exemple de l'outil Round-Robin database Tool (RRDtool) pour la sauvegarde de données cycliques et le tracé de graphiques, de données chronologiques.

- Les bases de données à grande échelle : TimescaleDB qui rend SQL évolutif pour les données de séries chronologiques. Il est basé sur le SGBD relationnel PostgreSQL qui est très puissant pour la manipulation des bases de données temporelles. Alibaba Cloud High-Performance Time Series Database (HiTSDB) prend en charge l'écriture fiable de données de séries chronologiques à grande échelle. InfluxDB qui représente le système le plus performant pour la gestion de séries temporelles d'après le site officiel de classement des systèmes de gestion de bases de données de séries temporelles SGBDST [26].

☐ include secondary database models — 35 systems in ranking, April 2021

Rank Apr 2021	Rank Mar 2021	Rank Apr 2020	DBMS	Database Model	Score Apr 2021	Score Mar 2021	Score Apr 2020
1.	1.	1.	InfluxDB	Time Series, Multi-model	26.55	-0.31	+4.93
2.	2.	2.	Kdb+	Time Series, Multi-model	7.85	+0.09	+2.58
3.	3.	3.	Prometheus	Time Series	5.73	-0.17	+1.48
4.	4.	4.	Graphite	Time Series	4.53	-0.14	+1.10
5.	↑ 6.	↑ 8.	TimescaleDB	Time Series, Multi-model	2.77	-0.06	+0.90
6.	↓ 5.	↓ 5.	RRDtool	Time Series	2.71	-0.20	+0.10
7.	7.	7.	Apache Druid	Multi-model	2.63	-0.06	+0.71
8.	↑ 9.	↓ 6.	OpenTSDB	Time Series	1.76	-0.02	-0.25
9.	↓ 8.	9.	Fauna	Multi-model	1.52	-0.31	+0.65
10.	10.	↑ 11.	GridDB	Time Series, Multi-model	0.98	+0.01	+0.54
11.	11.	↑ 15.	DolphinDB	Time Series	0.84	+0.00	+0.53
12.	↑ 13.	↑ 13.	eXtremeDB	Multi-model	0.76	+0.04	+0.38
13.	↓ 12.	↓ 10.	KairosDB	Time Series	0.73	+0.00	+0.18
14.	14.	14.	Amazon Timestream	Time Series	0.62	+0.01	+0.28
15.	↑ 16.	↑ 20.	QuestDB	Time Series, Multi-model	0.40	+0.01	+0.23
16.	↓ 15.	↓ 12.	Alibaba Cloud TSDB	Time Series	0.39	-0.06	-0.01
17.	↑ 19.	17.	Riak TS	Time Series	0.37	+0.02	+0.10
18.	↓ 17.	↓ 16.	IBM Db2 Event Store	Multi-model	0.36	-0.04	+0.06

Figure I.2: Classement de SGBDST en mois d'Avril 2021TSDBs[26]

Dans notre travail, nous utilisons le SGBD le plus populaire InfluxDB pour le stockage de données évolutives modélisées par les séries chronologiques.

6. Visualisation des séries chronologiques

La visualisation de données est une fonction très importante pour l'analyse et l'exploration des séries chronologiques. Elle est effectuée à travers des graphiques qui tracent les valeurs observées sur l'axe des y par rapport à un incrément de temps sur l'axe des x (voir la figure I.3). Ces graphiques mettent en évidence visuellement le comportement et les modèles des données.

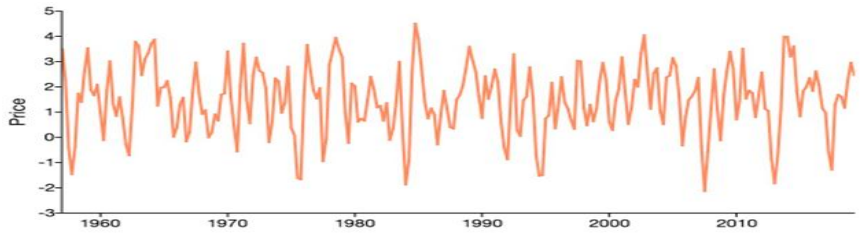


Figure I.3. Exemple d'une représentation graphique d'une série chronologique [28]

Un graphique de série chronologique fournit un outil d'inspection visuelle pour savoir le comportement de données : données convergentes, une tendance temporelle, saisonnalité, des données aberrantes, etc.

La visualisation des séries chronologiques peut être effectuée par un programme en langage python ou R en exploitant leurs librairies comme par exemple, Seaborn, Matplotlib de python et ggplot2 et Lattice de langage R [30]. Il existe aussi des outils puissants pour la visualisation et l'analyse de ce type de données comme Chronograf de système InfluxDB et Grafana.

Chronograf : Chronograf est un outil de InfluxDB qui permet de visualiser rapidement les données stockées dans InfluxDB. Il offre des modèles et des bibliothèques pour créer rapidement des tableaux de bord avec des visualisations en temps réel.

Figure I.4. Exemple de visualisation de données par Chronograf[1].

[1] *https://www.influxdata.com*

Grafana : est une plateforme open source pour la surveillance, l'analyse et la visualisation des métriques. Il est conçu pour générer des dashboards sur la base de métriques et données temporelles de plusieurs sources de données comme InfluxDB, OpenTSDB, Graphite.

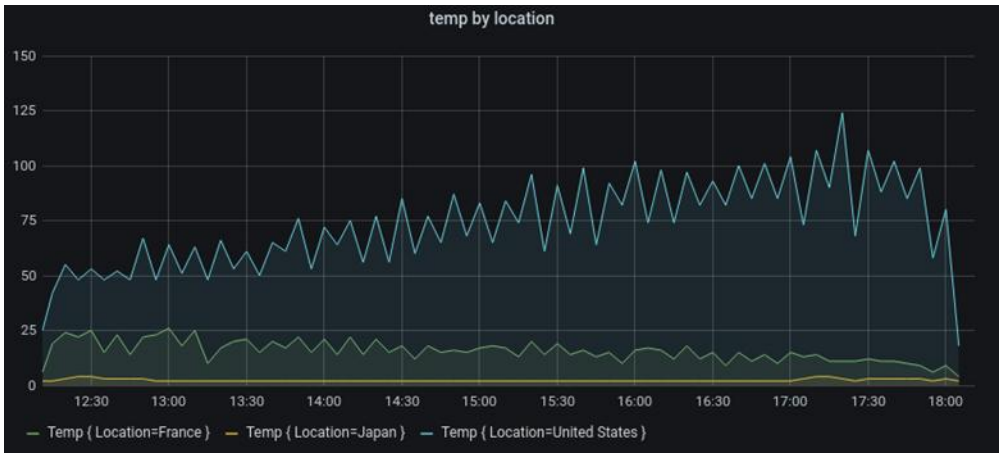


Figure I.5. Exemple De visualisation de données par Grafana[1]

Dans notre travail, nous utilisons les deux plateformes Chronograf (beaucoup plus) et Grafana pour la visualisation de données stockées dans InfluxDB (plus de détail dans le chapitre 04).

7. Types d'analyse des séries chronologiques

Les analyses ou études de séries chronologiques sont souvent définies comme un ensemble de méthodes permettant d'explorer, de décrire, de classifier, de prédire des données de base de données ou encore de l'ensemble de données « Dataset ». Il existe différents types et modèles d'analyse de données les plus souvent sont :[14].

- **Classification :** ou regroupement ou encore partitionnement (Clustering) qui permet d'identifier et d'attribuer des catégories aux données.
- **Ajustement de courbe :** trace les données le long d'une courbe pour étudier les relations des variables au sein des données.
- **Analyse descriptive :** identifie les modèles dans les données de séries chronologiques, comme les tendances, les cycles ou les variations saisonnières.

[1] *https://grafana.com*

- **Analyse explicative :** tente de comprendre les données et les relations qui les composent, ainsi que la cause et l'effet.
- **Analyse exploratoire** : met en évidence les principales caractéristiques des données de séries chronologiques, généralement dans un format visuel.
- **Prévision :** prédit les données futures. Ce type d'analyse est basé sur des tendances historiques. Il utilise les données historiques comme modèle pour les données futures, prédisant les scénarios qui pourraient se produire le long des futurs points de tracé.
- **Analyse d'intervention :** étudie comment un événement peut modifier les données.
- **Segmentation :** divise les données en segments pour afficher les propriétés sous-jacentes des informations source.

Dans notre travail, nous nous focalisons beaucoup plus sur l'analyse descriptive avec une visualisation de données.

8. Les domaines d'applications des séries chronologiques

Les séries chronologiques ou temporelles, se rencontrent de nombreux domaines d'application [14] [29] :

- Finance et l'économétrie : utilisent fréquemment l'analyse de séries chronologiques (évolutions des indices boursiers, changement des ventes). L'analyse boursière est un excellent exemple d'analyse de séries chronologiques en action, en particulier avec des algorithmes de trading automatisés.
- Météorologie : l'analyse des séries chronologiques est idéale pour prévoir les changements météorologiques, aider les météorologues à tout prévoir, du bulletin météo de demain aux années futures de changement climatique, mesure de précipitations, etc.
- Médecine :
  - Lectures de température
  - Surveillance de la fréquence cardiaque (ECG), Surveillance cérébrale (EEG), prédiction et prévention des maladies, etc.
- Astronomie, Biologie, etc.

9. Conclusion

Comme on a vue l'analyse des séries chronologiques comprend de nombreuses catégories ou variations de données, les analystes doivent parfois créer des modèles complexes. Cependant, les analystes ne peuvent pas tenir compte de toutes les variances. Dans ce travail, nous étudierons les données évolutives de l'ECG comme domaine d'étude pour la détection et le diagnostic des maladies, cardiovasculaires. La modélisation et l'étude de ce type de données sont basées sur les séries chronologiques. Nous présenterons dans le prochain chapitre la physiologie du cœur et la description de données ECG.

Chapitre 02 :
Les signaux ECG

1. Introduction

L'un des défis majeurs dans le domaine médical aujourd'hui est de savoir comment exploiter l'énorme quantité de données santé. L'exploitation de signaux de données médicales comme l'électrocardiogramme ECG, l'électroencéphalogramme EEG et l'électromyogramme EMG, est une nécessité fondamentale dans plusieurs disciplines de la médecine. Particulièrement, la détection et la prédiction des anomalies, la classification des événements dans un signal sont des tâches qui suscitent un intérêt depuis plusieurs années dans le domaine de l'ingénierie biomédicale.

L'objectif de ce chapitre est de décrire les signaux ECG qui jouent un rôle très important pour diagnostiquer un problème cardiaque. Pour cela, en passant par la présentation de la morphologie du cœur humain, les principaux composants d'ECG, et les diverses formes du tracé ECG (dérivation). Ensuite, nous présenterons les pathologies cardiaques les plus fréquents et nous terminerons par la description de l'électrocardiographe ambulatoire « Holter ».

2. Morphologie du cœur humain

Le cœur humain est un organe musculeux creux assurant la circulation sanguine dans le corps grâce à ses contractions régulières. Il est l'élément principal pour le fonctionnement du système cardiovasculaire. Il permet de pomper le sang en le faisant circuler dans tous les tissus du corps [32]. Les parois du cœur sont constituées par le muscle cardiaque, appelé myocarde, est composé de deux parties : côté droit et côté gauche qui sont séparés par un muscle, c'est la cloison. Chaque partie est composée d'une cavité supérieure "**oreillette**" et d'une cavité inferieure "**ventricule**" qui se communiquent par des valves d'admission qui laissent passer le sang de l'oreillette vers le ventricule [33]. La partie gauche du myocarde (de l'oreillette vers le ventricule) effectue le travail le plus important, qui consiste à envoyer le sang oxygéné vers le reste du corps, où la pression est considérablement plus grande que celle de la circulation pulmonaire pompée par la partie droite (le ventricule droit envoie le sang qui transporte le CO2 vers les poumons).

La contraction musculaire du myocarde est provoquée par la propagation des impulsions électriques qui prennent naissance dans le nœud sino-auriculaire (SA), entraînant la dépolarisation des cellules musculaires du myocarde. La fréquence de production des impulsions électriques détermine celle du rythme normal (nombre de battements de cœur) dit sinusal. Ces impulsions peuvent changer en fonction des exigences physiques, le stress, ou les anomalies cardiaques [63].

Le système responsable à l'excitation et la conduction électrique comprend : le nœud sinusal, les voies spécialisées internodales, le nœud auriculo-ventriculaire (NAV), le faisceau de His, les branches droite et gauche et les fibres de Purkinje (voir la figure II.1).

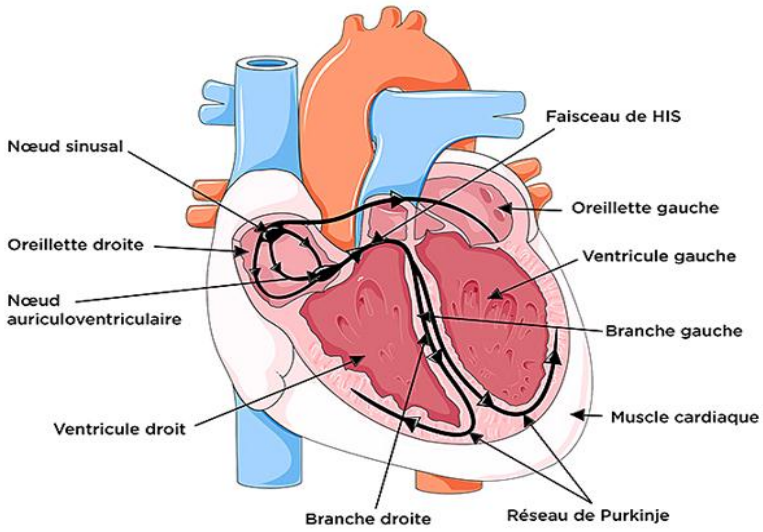


Figure II.1. Morphologie de cœur[2]

L'activité électrique normale du cœur suit la séquence d'activation suivante :

- **Le nœud sinusal (NS):**

Dans la partie supérieure de l'oreillette droite se trouve un petit morceau de tissu cardiaque particulier appelé le nœud sinusal qui se présente comme une petite traînée blanche située entre l'abouchement des deux veines caves, à proximité de la veine cave supérieure. C'est à cette proximité des gros troncs veineux qu'il doit son nom de nœud sinusal, par analogie avec le cœur de batracien où le sang veineux se déverse dans un sinus. Il donne l'impulsion électrique à tout le cœur en battant spontanément à une fréquence élevée de 90 battements par minute, régulée à 70 battements par minute. Il distribue le courant électrique à travers des faisceaux atrioventriculaires qui cheminent dans la paroi des atriums droit et gauche, pour rejoindre le nœud atrioventriculaire.

- **Le nœud auriculo-ventriculaire (NAV) :**

NAV est situé en bas de l'oreillette droite et est constitué de cellules qui présentent une conduction électrique lente. L'activation électrique qui arrive au NAV est physiologiquement ralentie

[2] https://www.chumontreal.qc.ca/patients/centre-cardiovasculaire/

(approximativement 100 ms) avant d'arriver au faisceau de His (faisceau de cellules musculaires cardiaques spécialisées dans la conduction électrique). Cette propriété physiologique du NAV permet de protéger les ventricules d'un nombre excessif d'activations du NAV et d'activations auriculaires et donne aux oreillettes un temps de vidange plus grand, ce qui optimise la contraction ventriculaire.

- **Le faisceau de His :**

Il est situé dans la partie haute du septum interventriculaire et ses fibres traversent le tissu connectif (non excitable) qui sépare électriquement les oreillettes des ventricules. Dans les cas normaux, le NAV et le faisceau de His constituent la seule voie de propagation de l'activité électrique cardiaque entre les oreillettes et les ventricules. L'ensemble de ces deux structures est souvent appelé la jonction auriculo-ventriculaire. Le faisceau de His comprend un tronc initial qui se divise en deux parties, droite pour le ventricule droit et gauche pour le ventricule gauche.

- **Les fibres de Purkinje :**

Les branches du faisceau de His finissent dans un réseau de fibres qui arrivent dans les parois ventriculaires. Ils terminent en anastomoses avec les fibres myocardiques musculaires, facilitant leur excitation

- **Réseau de Purkinje**

Réseau spécialisé dans la conduction intraventriculaire (tissus nodal). Il associe le faisceau de His et les fibres de Purkinje qui le prolongent et se distribuent par arborescence dans le myocarde ventriculaire.

3. Description du signal d'électrocardiogramme (ECG)

Le signal d'électrocardiogramme (ECG) est une représentation graphique (un tracé) de séquence de données qui représente les impulsions électriques du myocarde [53] Un signal d'ECG est enregistré à partir de nombreuses électrodes(capteurs) fixées sur la peau. La plupart des médecins préfèrent utiliser l'ECG comme outil non invasif pour détecter et diagnostiquer les maladies cardiaques à l'aide de l'appareil appelé électrocardiographe.

Les deux caractéristiques importantes d'un signal ECG sont : Les multiples enregistrements de signaux provenant de diverses positions du myocarde et la forme d'onde périodique synchronisée avec un cycle cardiaque.

La morphologie ECG du tracé est simplement la forme d'onde ou la perspective de l'activité électrique du muscle cardiaque, de la dépolarisation et de la repolarisation, dans un cycle cardiaque.

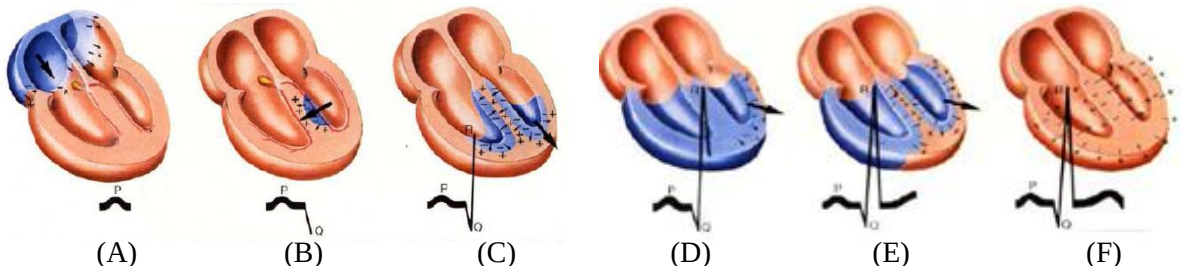
- **La dépolarisation** des cellules qui provoque la systole, la phase de contraction.

  - L'onde P : Dépolarisation des oreillettes (systole auriculaire = contraction des oreillettes).
  - L'espace PR ou espace PQ : Temps de conduction auriculo-ventriculaire.
  - Le complexe QRS : Dépolarisation des ventricules (systole ventriculaire = contraction des ventricules).

- **La repolarisation** des cellules qui entraîne la diastole, la phase de relâchement qui permet le remplissage sanguin des cavités auriculaires et ventriculaires.

  - La repolarisation des oreillettes (diastole auriculaire = relâchement des oreillettes) : se produit pendant la dépolarisation ventriculaire.
  - Le segment ST : Temps de repolarisation complète des ventricules
  - L'onde T : Repolarisation des ventricules (diastole ventriculaire = relâchement des ventricules).

  La repolarisation est plus longue en durée que la dépolarisation car la vitesse de conduction de l'onde de repolarisation est plus faible que celle de l'onde de dépolarisation.



(A) (B) (C) (D) (E) (F)

Figure II.2. Morphologie de l'ECG VS les impulsions électriques dans myocardes [64]

En se référant à la figure II.2, l'impulsion électrique se propage dans le muscle cardiaque et induit sa contraction. Elle prend naissance dans le sinus, puis se propage dans les oreillettes, entraînant leurs contractions (Partie (A) Figure II.2). L'impulsion arrive alors au nœud auriculo-ventriculaire (Partie (B) Figure II.2) une courte pause est alors introduite (permettant au sang de pénétrer dans les ventricules) juste avant la propagation dans les fibres constituant le faisceau de His (Partie (C) Figure II.2). Au passage de l'impulsion électrique les ventricules se contractent à leur tour (Partie (D) Figure II.2). Après la diastole les cellules se repolarisent (Partie (E) Figure II.2). Le faisceau de His est complétées par les fibres de Purkinje, qui grâce à leur conduction rapide, propagent l'impulsion électrique en plusieurs points des ventricules, et permettent ainsi une dépolarisation quasi instantanée de l'ensemble du muscle ventriculaire, ce qui assure une efficacité optimale dans la propulsion du sang. Cette contraction constitue la phase de systole ventriculaire, puis suit la diastole ventriculaire (Partie (F) Figure II.2) et les fibres musculaires se répolarisent et reviennent ainsi dans leur état initial.

4. Composants du signal d'ECG

Un ECG normal se compose des segments morphologiques et des intervalles, formant l'onde **PQRST** qui correspondent à la conductivité électrique tout au long du cycle cardiaque.

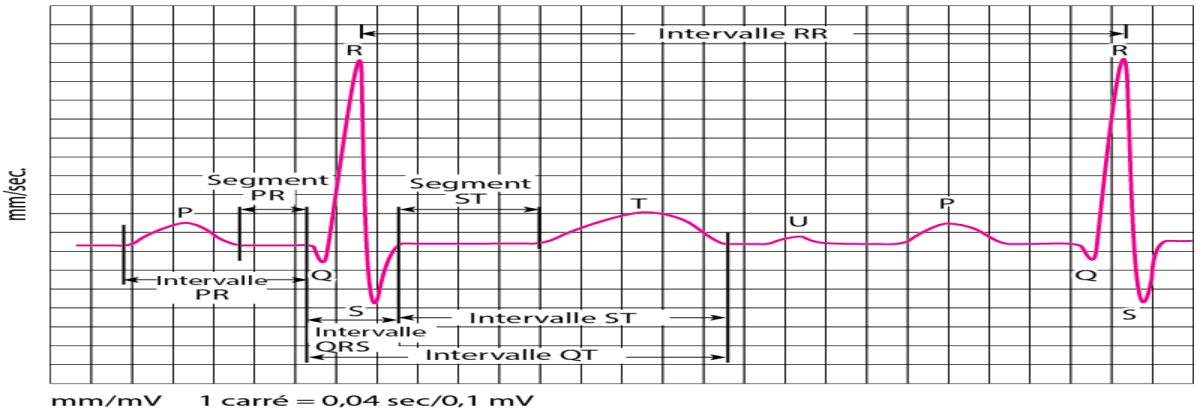


Figure II.3 : Les ondes et les intervalles dans un ECG [53].

(1) **L'onde P** se réfère à l'activation électrique de la dépolarisation auriculaire qui provoque la conduction de l'impulsion électrique à travers les oreillettes.

(2) **Le complexe QRS** montre une dépolarisation ventriculaire qui provoque une contraction des ventricules.

(3) **L'intervalle PR** commence à partir du début de la P onde à un début du complexe QRS.

(4) **L'onde T** affiche la repolarisation des ventricules pendant le temps où les ventricules retournent à leur état électrique de repos.

(5) **L'intervalle QT** commence à partir d'un début du complexe QRS jusqu'à la fin de l'onde T. L'intervalle QT présente une dépolarisation et une repolarisation ventriculaires.

(6) **Le segment TP** commence à partir de la fin de l'onde T du cycle d'ECG (ou battement de cœur) précédent jusqu'au début de l'onde P du cycle ECG suivant. Le segment TP représente le moment où les cellules du muscle cardiaque sont électriquement silencieuses. Ainsi, il est toujours illustré par un intervalle isoélectrique qui représente une ligne zéro, une ligne de base ou une ligne électrique.

(7) L'intervalle RR il est délimité par deux pic R successives et d'où est évaluée la fréquence cardiaque instantanée. Cet intervalle est utilisé pour la détection des arythmies ainsi que pour l'étude de la variabilité de la fréquence cardiaque.

(8) Le segment ST il représente l'intervalle durant lequel les ventricules restent dans un état de dépolarisation actif. Il est aussi défini comme la durée entre la fin de l'onde S et le début de l'onde T.

(9) L'onde S elle représente la durée de dépolarisation ventriculaire (Les chambres en bas à droite et à gauche) précèdent l'effet mécanique de contraction. Sa durée normale est comprise entre 85ms et 95ms.

(10) L'onde U Signal électrique de base amplitude et de basse fréquence (« phénomène mécano-électrique ») qui survient après l'onde T ou fusionne avec elle. Sa signification n'est pas bien connue, elle pourrait correspondre à l'onde T de repolarisation des cellules de Purkinje.

5. Dérivations électrocardiographiques

L'enregistrement de l'activité électrique du cœur au fur et à mesure des battements cardiaques se fait par l'appareil Electrocardiographe relié à des électrodes de détection (des capteurs d'acquisition de données), permettant de mesurer des différences de potentiels électriques générées par la contraction musculaire [32].

Les signaux captés étant particulièrement faibles, des applications de hautes performances (amplification, quantification, filtrage...) sont souvent nécessaires. Divers groupements de ces électrodes, correspondant à différents circuits d'enregistrement, sont reliés à un stylet qui donne une trace correspondant à une dérivation (reflet localise de l'activité électrique du cœur). Selon les différentes dérivations, l'électrocardiogramme se présente différemment (typiquement 3, 5, 12 dérivations). Généralement, l'ECG de douze (12) dérivations produit un tracé plus précis [32]. L'enregistrement graphique s'effectue habituellement à une vitesse de déroulement du papier de 25mm/s et la détection d'une tension de 1mV provoque une déflexion verticale de 1cm.

Figure II. 4. Electrocardiographe [54]

Les électrodes : Les électrodes doivent être positionnées à des endroits bien définis du corps et directement sur la peau. Quatre électrodes sont placées sur les poignets et les chevilles et six autres ont des points détermines de la surface du thorax [34]. L'enregistrement de l'ECG indique une tension positive lorsque l'onde de dépolarisation se déplace vers l'électrode et une tension négative lorsqu'elle s'éloigne de l'électrode.

Les dérivations : : Une dérivation est un circuit électrique détermine par deux points l'observation (deux électrodes) de l'activité électrique du cœur à partir des quels on mesure une différence de potentiel électrique [34].

L'emplacement de ces électrodes est choisi de manière à explorer la quasi-totalité du champ électrique cardiaque résultant de la contraction du myocarde (voir la figure II.5).

Il existe deux catégories de dérivations : dérivations périphériques et dérivations précordiales.

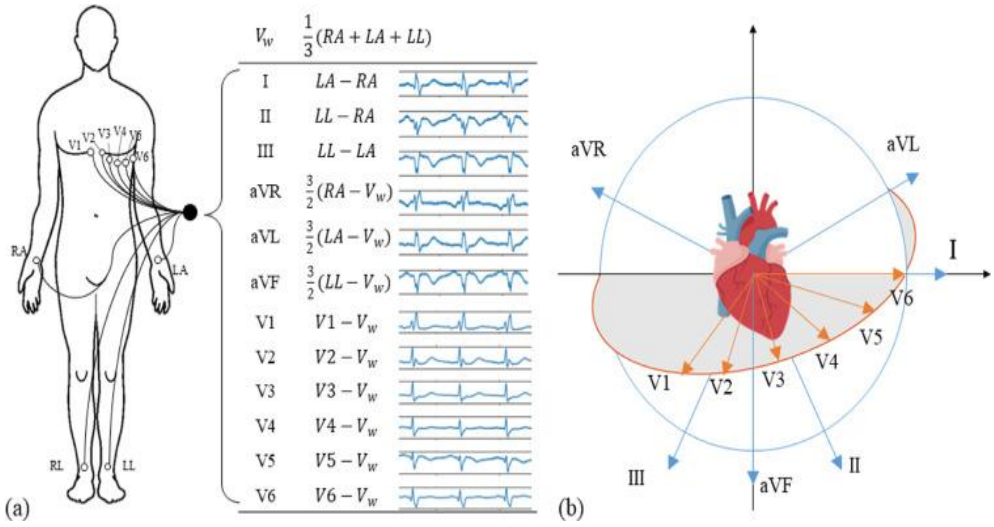


Figure II.5 : Position de 12 dérivations d'un ECG [50]

5.1. Dérivations périphériques

Les dérivations périphériques (ou dérivations des membres) permettent d'apprendre l'activité électrique du cœur sur le plan frontal. Elles sont obtenues au moyen de 4 électrodes appliquées au bras droit, au bras gauche, à la jambe gauche, et l'électrode de la jambe droite étant une électrode neutre destinée à éliminer les parasites électriques [34].

- **Les dérivations périphériques bipolaires (DI, DII, DIII):** Ce sont les dérivations cardiaques classiques de l'électrocardiogramme décrites par Einthoven [34]. Elles enregistrent la différence de potentiel entre deux électrodes placées à des extrémités différentes. Ces dérivations utilisent trois électrodes placées sur les bras droit et gauche et sur la jambe gauche pour former un triangle (triangle d'Einthoven). Ces dérivations sont dites bipolaires parce qu'elles mesurent une différence de potentiel entre deux électrodes. Chaque côté du triangle formé par les trois électrodes représente une dérivation en utilisant une paire d'électrodes différente pour chacune des dérivations.
- **Les dérivations unipolaires (aVR, aVL, aVF):** des extrémités enregistrent la différence de potentiel entre un point théorique au centre du Triangle de Einthoven, ayant une valeur de 0 et l'électrode de chaque extrémité, permettant ainsi de connaître le potentiel absolu dans ladite électrode [34]. Ces dérivations sont anciennement nommées VR, VL et VF où V signifie vecteur et R, L, F : (Right) droite, (Left) gauche et (Foot) pied. Elles sont par la suite amplifiées (augmenter leur puissance), présentant aVR, aVL et aVF (le 'a' minuscule

signifie amplifiée) qui indiquent les dérivations unipolaires amplifiées par rapport aux premières.

5.2. Dérivations précordiales

Les dérivations précordiales sont les dérivations unipolaires les mieux adaptées pour dépister les anomalies du ventricule gauche, en particulier celles des parois antérieure et postérieure. Elles n'utilisent qu'une seule électrode exploratrice sur la paroi thoracique antérieure. Il existe six pistes précordiales désignés par un V majuscule et un nombre compris entre 1 et 6. Pour un ECG normal, le complexe QRS est principalement négatif dans les dérivations V1 et V2 et majoritairement positif dans V4 à V6 (profil Rs). Ces dérivations sont positionnées comme suite (Figure II.5) :

- V1 : 4ème espace intercostal, bord droit du sternum (ligne parasternale).
- V2 : 4ème espace intercostal, bord gauche du sternum (ligne parasternale).
- V3 : à mi-distance entre V2 et V4.
- V4 : 5ème espace intercostal, ligne médio-claviculaire gauche.
- V5 : à mi-distance entre V4 et V6, sur la ligne axillaire antérieure.
- V6 : même niveau horizontal que V4 et V5, ligne axillaire moyenne.

6. Interprétation de l'ECG et diagnostic de pathologies cardiaques

L'électrocardiogramme met en évidence énormément de problèmes cardiaques tels que les blocs auriculo-ventriculaires (mauvaise conduction de l'influx électrique) avec l'étude de l'onde P, des bradycardies et tachycardies avec le ralentissement ou l'accélération de ces complexes sur le dessin mais aussi les fibrillations auriculaires et ventriculaires symbolisées par une contraction anarchique des ventricules [31].

La morphologie de l'ECG reflète l'état cardiaque. En général, deux principaux paramètres pour diagnostiquer si le tracé ECG est normal ou anormal :

1. La durée des ondes : un cardiologue peut déterminer combien de temps l'onde électrique prend pour traverser le système de conduction électrique du cœur. Cette information permet de savoir si l'activité électrique est régulière (l'intervalle R-R est quasi-constant sur tout le tracé) ou irrégulière, rapide ou lente.
2. La distance entre les ondes : force de l'activité électrique : un cardiologue est en mesure de savoir si les parties du cœur sont trop grosses ou sont surmenées.

L'irrégularité des battements cardiaque, communément appelé Arythmie, est une caractéristique importante qui est souvent associé à un trouble de la production ou de la conduction de l'impulsion électrique [53]. Dans le reste de cette section, nous allons présenter brièvement les pathologies les plus courants d'arythmie en se basant sur les références [36] [53].

6.1.Fibrillation auriculaire (FA)

La FA se produit lorsque les potentiels d'action se déclenchent très rapidement dans l'oreillette, entraînant une fréquence auriculaire rapide (environ 400 à 600 battements/minute). Par conséquent, les ondes P ne seront pas vues depuis le taux avec un faible niveau d'amplitude.

6.2.Bloc de dérivation de faisceau droit (RBBB) et gauche (LBBB)

Bundle Branch Block (BBB) est une interruption du système de conduction régulier qui conduit à une morphologie QRS anormale. Typiquement, le faisceau droit dépolarise le ventricule droit (VR). Dans un RBBB, le paquet droit ne s'activé pas. Le ventricule droit est plutôt dépolarisé en diffusant l'impulsion du faisceau gauche à travers le ventricule gauche (LV) puis vers le ventricule droit (VR). Ce modèle de propagation électrique crée une morphologie QRS aberrante.

Dans un LBBB, le faisceau de gauche n'est pas activé. Le LVG est plutôt dépolarisé par la propagation de l'impulsion du faisceau droit au RV puis au LV. Cette le motif de propagation électrique crée une morphologie QRS aberrante.

6.3.Contraction auriculaire prématurée (PAC) et ventriculaire prématurée (PVC)

PAC et PVC se produisent lorsque le rythme régulier du cœur est interrompu par un battement prématuré ou précoce au niveau des oreillettes et des ventricules, respectivement. QRS prématuré, non précédé d'onde P, de morphologie différente des complexes de base, élargi (> ou égal à 0,12 seconde).

6.4.Battements ectopiques

Les rythmes auriculaires ectopiques se produisent lorsqu'un site en dehors du sinus nœud dans les oreillettes crée des potentiels d'action plus rapidement que le nœud sinusal (avec une fréquence auriculaire inférieure à 100 battements/minute). Depuis cette activité électrique ne provient pas du nœud sinusal, l'onde P n'aurait pas son aspect sinusal normal. Les battements ectopiques sont également fréquents pendant les périodes de stress ou d'exercice, et ils peuvent se produire par la consommation de certains aliments tels que l'alcool.

6.5.Infarctus du myocarde (MI)

L'MI (c'est-à-dire la crise cardiaque) survient lorsque le flux sanguin diminue ou s'arrête dans une partie du cœur, causant des dommages permanents au muscle cardiaque ou des artères.

6.6.Battement de fusion

Un battement de fusion se produit lorsque des impulsions électriques provenant de différentes sources agissent simultanément sur la même région du cœur. Il est appelé battements de fusion ventriculaire (VFB) s'il agit sur les chambres ventriculaires, tandis que les courants de collision dans les chambres auriculaires produisent Atrial Fusion Beats (AFB).

6.7.Bradycardie sinusale

La bradycardie sinusale est un rythme sinusal avec une fréquence inférieure à la normale (moins de 60 battements par minute). Le cœur diminué entraîne une diminution du débit cardiaque entraînant des symptômes tels que comme étourdissements, hypotension, vertiges et syncope.

6.8.Tachycardie

La tachycardie survient lorsque la fréquence cardiaque dépasse la norme de taux de repos (appelée tachyarythmie). Généralement, un cœur au repos a fréquence supérieure à 100 battements par minute chez l'adulte est considérée comme une tachycardie. Les types de tachycardies comprennent :

1. Tachycardie auriculaire ou supraventriculaire (TSV) : est une fréquence cardiaque rapide regardant dans les cavités cardiaques supérieures.

2. Tachycardie sinusale : survient lorsque le cœur envoie des signaux électriques plus rapidement que d'habitude.

3. Tachycardie ventriculaire (TV) : est une série de plus de trois rythmes cardiaques complexes QRS consécutifs abnormaux d'une durée au-delà de 120 ms et le vecteur ST qui pointe en face du QRS déviation.

6.9.Flutter auriculaire (AFL)

L'AFL est un rythme cardiaque anormal répandu qui commence dans les chambres auriculaires du cœur. Lorsqu'elle survient, elle est généralement associée à une fréquence cardiaque rapide.

6.10. Flutter ventriculaire (FV)

Il s'agit d'une arythmie instable caractérisée par une tachycardie affectant les ventricules avec une fréquence de plus de 150 à 300 battements par minute. FV est une étape de transition possible entre la TV et la fibrillation qui peut provoquer une mort subite d'origine cardiaque. Une forme d'onde sinusoïdale le caractérise sans définition claire des ondes T et QRS.

6.11. Fibrillation ventriculaire (VFib)

VFib est une arythmie cardiaque dans laquelle le cœur tremble à la place de pompage dû à une activité électrique désorganisée dans les ventricules caractérisés par la présentation de complexes QRS irréguliers et non formés sans ondes P claires. Le VFib entraîne un arrêt cardiaque avec perte de connaissance suivi de la mort en l'absence de traitement.

6.12. Rythme idioventriculaire

Un rythme idioventriculaire est très similaire à la TV (Tachycardie ventriculaire) mais avec la fréquence ventriculaire inférieure à 60 battements par minute. Par conséquent, le rythme idioventriculaire est appelé tachycardie ventriculaire lente.

6.13. Bigéminisme

Le bigéminisme est un trouble du rythme cardiaque caractérisé par deux battements cardiaques très rapprochés, le second étant une extrasystole suivis d'une pause.
Selon la source de l'extrasystole, on distingue « bigéminisme auriculaire » et « bigéminisme ventriculaire ».

7. Dispositifs portables pour la surveillance cardiaque

La communauté des cardiologues commence à envisager des enregistrements Holter de très longues durées, ce qui correspond à plus de 2 millions de battements enregistrés. On comprend que l'analyse d'une telle quantité d'informations (donnes évolutives) n'est envisageable que parce qu'un traitement automatique des données enregistrées est aujourd'hui possible.

La surveillance ambulatoire des signes vitaux se limitait principalement à la surveillance ECG avec des dispositifs **Holter[65]**. Ces moniteurs sont utilisés depuis plus de 40 ans comme méthode non invasive de surveillance continue de la fréquence cardiaque et de l'ECG pendant des périodes définies. Les améliorations récentes de la technologie des batteries et de la transmission de données sans fil, parallèlement à l'avènement des smartphones, ont annoncé des avancées dans les

moniteurs portables (voir la figure II.6.). Au cours des 15 dernières années, des moniteurs portables ont été développés qui intègrent plusieurs capteurs, un traitement intelligent, des alarmes pour soutenir les décisions médicales et les interactions médicales :

- Patch adhésif de qualité médicale (Figure II.6 (A)) : de Sensium, Abingdon, Royaume-Uni.
- Vêtements avec capteurs intégrés (Figure II.6 (B)) : de Hexoskin, Montréal, Canada.
- Sangle de poitrine (Figure II.6 (C)) : de Medtronic, Maryland, États-Unis.
- Sangle de poitrine (Figure II.6 (D)) : de Polar Electro, Warwick, Royaume-Uni.
- Brassard supérieur (Figure II.6 (E)): de Current, Édimbourg, Royaume-Uni.
- Bracelet (Figure II.6 (F)) : de Fitbit, San Francisco, États-Unis.

Figure II.6 : Exemples de dispositifs de surveillance ambulatoire [65]

8. Conclusion

Dans ce chapitre, nous avons découvert les différentes composantes de l'ECG, le lien entre le fonctionnement physiologique et les caractéristiques du signal ECG, l'importance de l'électrocardiogramme dans la détection des anomalies cardiaques et l'établissement du diagnostic médical.

Après avoir décrire tout ce qui concerne les signaux ECG, il devra nécessaire de faire une synthèse sur les techniques les plus récentes pour l'étude descriptive de ces données pour la classification supervisée des signaux ECG. Cette partie fera l'objet du chapitre suivant.

Chapitre 03 :

Etat de l’art sur l’étude des données ECG modélisées par les séries chronologiques

1. Introduction

Rappelons que notre travail a pour but d'effectuer une analyse descriptive des données évolutives par l'exploitation des séries chronologies qui sont très adéquates pour modéliser ce type de données. L'exemple de données évolutives à étudier est celles de santé, plus précisément les données ECG (données de signal ECG). L'étude de données ECG s'avère très importante notamment que les maladies cardiovasculaires ou cardio-neurovasculaires sont la principale cause de mort subite.

Après avoir présenté les données évolutives et les séries chronologiques dans le chapitre 01 ainsi que les signaux ECG dans le chapitre 02, nous présenterons dans ce troisième chapitre un état de l'art sur l'étude de données évolutives. Nous allons présenter dans un premier temps les motivations d'analyser automatiquement les données ECG, ensuite nous présenterons les techniques de l'apprentissage profond avec les différentes architectures de réseau de neurones profonds. Le chapitre termine par la présentation de différents types d'analyse descriptive ainsi que les travaux récents sur la classification supervisée de données comme une analyse complète de données nécessitant à priori une analyse descriptive.

2. Motivation d'analyse automatique d'ECG

Le besoin de développer des applications informatiques fondées sur les techniques de l'intelligence artificielle pour l'analyse automatique de données ECG s'est appuyé sur les motivations suivantes:

- Tracé des ECG difficiles à interpréter à la main : d'après Jia Li, Chief Scientist Officer de la jeune pousse et le gagnant du concours mondial d'innovation en 2014[3], l'interprétation d'un examen n'est pas triviale, même pour un cardiologue. En effet, les applications de détection et de classification automatique permettent aux médecins généralistes ou praticiens de gagner du temps dans l'analyse de l'ECG dans le cas d'absence d'un cardiologue.
- Besoin d'optimiser les décisions médicales individuelles et donc réduire les coûts à long terme des soins de santé.
- Le besoin de traitement de données volumineuses ECG recueillies sur une longue durée par Holter.

[3] https://www.sciencesetavenir.fr/sante/coeur-et-cardio/cardiologs-un-algorithme-pour-detecter-les-pathologies-cardiaques-sur-ecg_111627

- Problème de consommation importante de papier thermique.
- La qualité d'impression de résultat d'ECG s'estompe avec le temps et selon les conditions de stockage.
- L'interprétation de plusieurs ECG en urgence est une tâche fastidieuse.
- La numérisation de données ECG est restreint à l'exportation en PDF.

3. Apprentissage automatique et Apprentissage profond

L'interprétation d'un signal ECG est vue comme une carte représentative de la façon dont une onde électrique est conduite dans les oreillettes et les ventricules du cœur [54]. Traditionnellement, l'analyse automatique d'ECG reposait sur le développement des systèmes expert basés sur l'élaboration des règles et cela se fait en deux étapes : la détermination des fonctionnalités utiles de données ECG brutes, appelées « fonctionnalités expertes », puis le déploiement des règles de décision.

Les modèles d'apprentissage automatique (Machine Learning ML) tels que les arbres de décision et les machines à vecteurs de support (SVM) consistent à apprendre une fonction à faire correspondre une entrée à une sortie en se basant sur des exemples connus (des paires entrée-sortie). Ces différents modèles de ML ne sont pas toujours utiles pour l'analyse de données ECG, notamment que lorsque les données contiennent plusieurs propriétés avec un volume très grand [66]. Le problème réside dans le fait que les modèles de ML exigent une phase critique appelée l'ingénierie des caractéristiques (Feature Ingeneering) où l'expert de domaine sélectionne les caractéristiques importantes des données. Ces modèles requièrent beaucoup temps non seulement pour le prétraitement de données mais aussi pour l'apprentissage, la résolution des problèmes de tendance et de saisonnalité, etc. [66].

Récemment, les méthodes d'apprentissage en profondeur (Deep leraning DL) ont obtenu des résultats prometteurs dans de nombreux domaines d'application tels que la vision par ordinateur, la reconnaissance vocale et le traitement du langage naturel, etc. avec le principal avantage est qu'elles ne nécessitent pas d'étape d'extraction de caractéristiques explicite à l'aide d'experts humains (voir la figure III.1) [39]. Au lieu de cela, l'extraction des caractéristiques est effectuée automatiquement et implicitement par des modèles d'apprentissage en profondeur basés sur leurs puissantes et capacités d'analyse et d'exploration de données.

Les modèles de deep learning sont basés sur des architectures de modèle de réseaux de neurones profonds qui disposaient de plus en plus de couches cachées et que les nombres élevés de couches devenait une source de problèmes [39].

Des modèles DL ont été appliqués aux données ECG pour résoudre des problèmes tels que la détection de maladie, l'annotation ou la localisation, la mise en scène du sommeil, l'identification biométrique humaine, débruitage, etc [40].

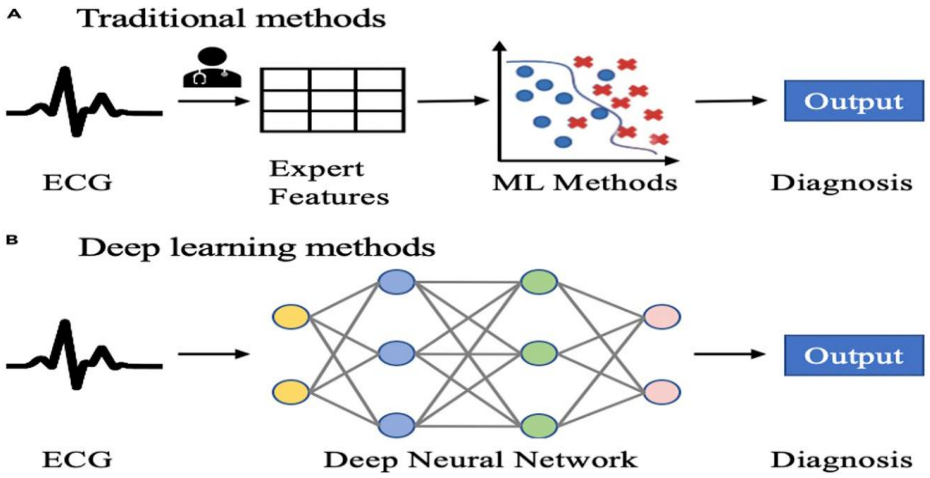


Figure III.1 : les techniques DL VS techniques ML [43]

Un réseau de neurones profonds est une composition de **N** fonctions paramétriques appelées couches où chaque couche est considérée comme une représentation du domaine d'entrée [45]. Pour n'importe quel type de ces réseaux ; il existe une Couche d'entrée (Input Layers), une ou plusieurs Couches cachées (Hidden Layers) et une Couche de sortie (Output Layer). Chaque paire de couches voisines est connectée par des connexions appelées synapses qui sont associés par des poids (**W**eights). La figure III.2 illustre une architecture standard d'un réseau de neurones profonds

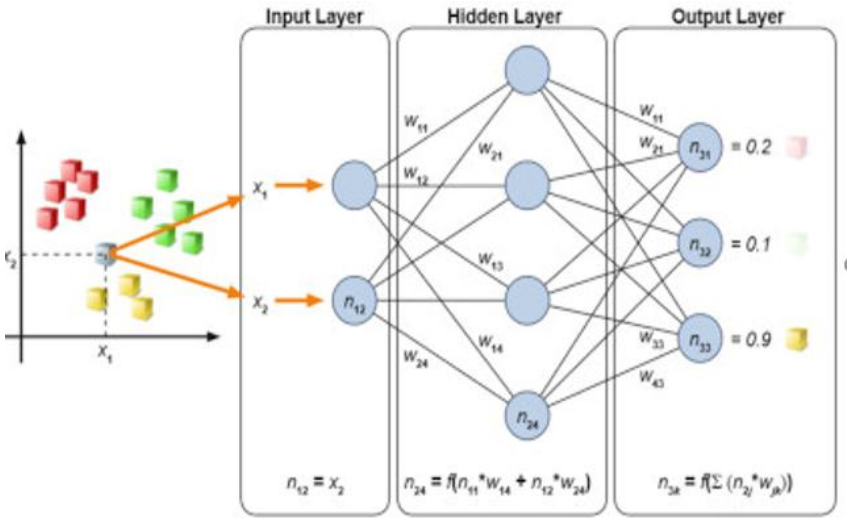


Figure III.2 : Architecture d'un DNN [44]

- Chaque couche L_i, i ∈ {1...N}, contient des neurones, prend comme entrée la sortie de la couche précédente L_{i-1} et applique une fonction (comme exemple sigmoïde) pour calculer la sortie de la couche.
- **Forward propagation :** Le comportement de ces transformations non linéaires sont contrôlées par un ensemble de paramètres appelés poids W qui relient l'entrée de la couche précédente à la sortie de la couche courante qui peut être écrit comme suit :

$$y = \delta\left(\sum_{i=1}^{n} w_i x_i + b\right) = \delta(W^T X + b)$$

Avec :

• W : est le vecteur des poids.

• X : est le vecteur des entrées.

• b : désigne le biais.

• δ : représente la fonction d'activation

- **Back propagation**, La perte de prédiction du modèle est calculée à l'aide d'une fonction de coût (Cost function), par exemple le « négative log likelihood ». Après, en utilisant descente de gradient (gradient descent), les poids sont mis à jour dans une passe en arrière pour propager l'erreur [46].

Les itérations de : « forward propagation et back propagation » indiquant une phase d'entraînement basée sur l'ajustement des poids de manière à minimiser la perte des données d'entraînement.

- Lors des tests, le classificateur probabiliste (le modèle) est testé sur des données différent. Cette phase est appelée phase d'évaluation, suivie d'une prédiction de classe. La prédiction correspond à la classe dont la probabilité est maximale et cela se fait par la mesure de performance de modèle obtenu.

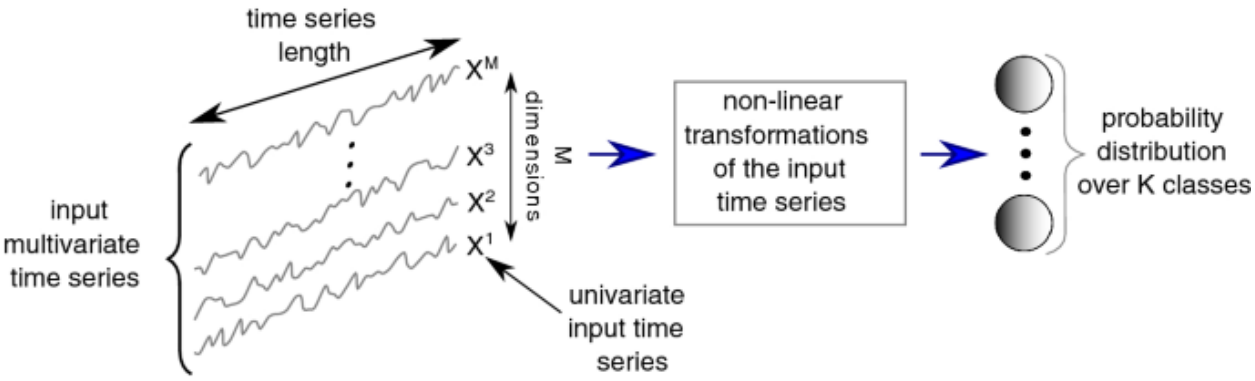


Figure III.3 : Un cadre de DL unifié pour la classification des séries chronologiques [44]

5. Architectures de DNN dédiées à la classification des séries chronologiques

Il existe de nombreuses architectures de réseaux de neurone profonds (Deep Neurale Network DNN). Dans ce chapitre, nous nous focalisons sur les principales architectures utilisées pour la classification des séries chronologiques (Time Series Classification TSC).

5.1.Réseaux de neurones convolutifs CNN

Un réseau de neurones convolutif CNN se compose de plusieurs couches connectées de manière anticipée [39]. Les couches principales comprennent la couche convolutive, couche de mise en commun et couche entièrement connectée. Les deux premières couches sont responsables de l'extraction des caractéristiques, tandis que les couches entièrement connectées sont chargées de la classification.

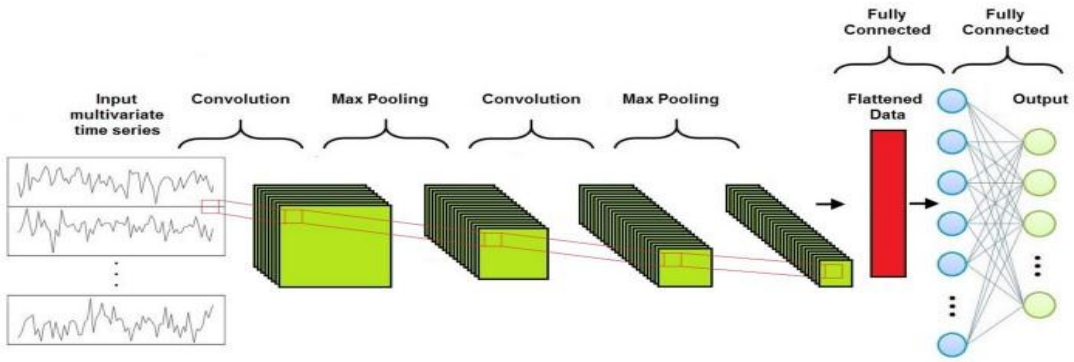


Figure III.4 : Architecture de CNN [56]

La figure III.4, illustre l'architecture du réseau de neurones convolutifs.

- **Couche convolutive (Convolutional Layer)**

La convolution est définie par un ensemble de filtres (des matrices de taille fixe) qui sont appliquées à une sous-matrice de la carte des caractéristiques (features map) en entrée pour donner en sortie la somme du produit de chaque élément du filtre avec l'élément dans la même position de la sous-matrice afin d'extraire les caractéristiques de haut niveau de données.

Les valeurs des filtres sont considérées comme des poids qui peuvent être entraînés et ensuite apprises pendant l'entraînement. Deux paramètres importants qui doivent être choisis pour la couche convolutive sont le stride et le Padding.

- **Stride :** il contrôle la façon dont le filtre s'articule autour d'une carte des caractéristiques en entrée. En particulier, la valeur de foulée indique combien d'unités doivent être décalées à la fois (voir la figure III.5).

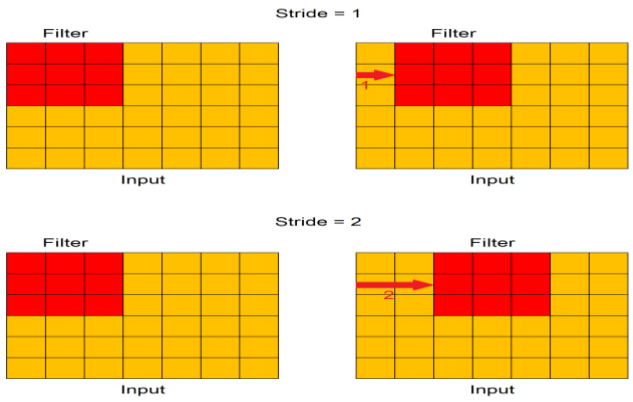


Figure III.5 : Mécanisme de stride [56]

- Padding indique le nombre de colonnes et de lignes supplémentaires à ajouter en entour d'une carte des caractéristiques en entrée, avant d'appliquer un filtre de convolution. Toutes les cellules des nouvelles colonnes et lignes ont une valeur fictive, généralement 0.

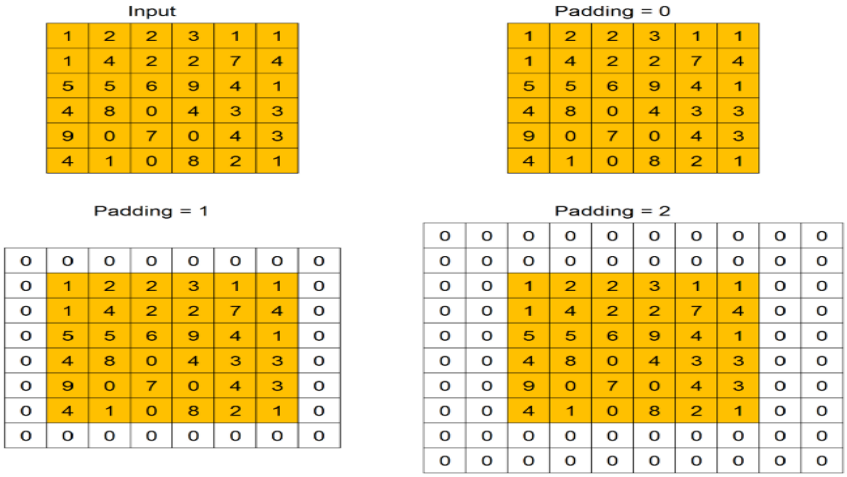


Figure III.6 : Mécanisme de Padding [56]

Le Padding est utilisé car lorsqu'un filtre de convolution est appliqué à une carte des caractéristiques en entrée de petite taille. Ensuite, après l'application de nombreux filtres, la taille peut devenir trop petite. En ajoutant des lignes et des colonnes supplémentaires, nous pouvons conserver la taille d'origine ou la ralentir.

- **Couche de mise en commun (Pooling Layer)**

Le but de l'opération de pooling est d'obtenir une réduction de dimension des cartes de caractéristiques, en préservant autant d'informations que possible. Il est également utile pour extraire les caractéristiques dominantes qui sont invariantes en rotation et en position. Son entrée est une série de cartes de caractéristiques et sa sortie est une série différente de cartes de caractéristiques, avec une dimension inférieure.

Le pooling est appliquée aux fenêtres coulissantes de taille fixe sur la largeur et la hauteur de chaque carte d'entités en entrée. Il existe deux types de pooling: Max Pooling and Average Pooling

Max Pooling fonctionne également comme un suppresseur de bruit, éliminant complètement les activations bruyantes. Par conséquent, il fonctionne généralement mieux que Average Pooling. Aussi pour la couche de pooling, le Stride et le Pading doivent être spécifiés.

- **Couche entièrement connectée (Fully-Connected Layer FCL)**

L'objectif de la FCL est d'apprendre des combinaisons non linéaires des caractéristiques de haut niveau représentées par la sortie de la couche convolutive et de la couche de mise en commun.

Après plusieurs opérations de convolution et de regroupement, la TS d'origine est représentée par une série de cartes de caractéristiques. Toutes ces cartes de caractéristiques sont aplaties dans un vecteur de colonne, c'est-à-dire la représentation finale de la série chronologique multivariée d'entrée d'origine. La colonne aplatie est connectée au Perceptron multicouche, dont la sortie a un nombre de neurones égal au nombre de classes possibles de séries temporelles.

La rétro propagation est appliquée à chaque itération de l'entrainement. Sur une série d'époques, le modèle est capable de distinguer les séries temporelles d'entrée grâce à leurs caractéristiques dominantes de haut niveau et de les classer.

5.2.Réseau Inception Time

L'architecture Inception Time (commencement) est une nouvelle architecture basée sur les réseaux de neurones convolutifs introduite par Google en 2014 [48]. Les composants de cette architecture sont (voir la figure III.7).

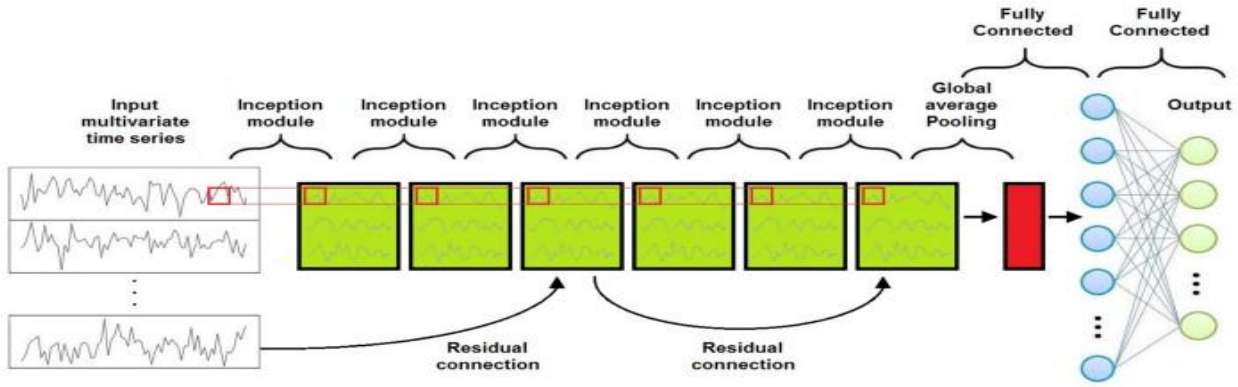


Figure III.7 : l'architecture de Inception time [56]

- La première couche est une couche de Bottleneck layer, qui réduit la dimensionnalité des entrées qui donc réduire le coût de calcul et le nombre de paramètres, accélérant la formation et améliorant la généralisation.
- Le deuxième composant majeur du module Inception est un ensemble de couches convolutives parallèles de différentes tailles agissant sur la même carte de caractéristiques d'entrée.
- La troisième couche est un MaxPooling, qui introduit la possibilité d'avoir un modèle invariant aux petites perturbations.
- La dernière couche est une couche de concaténation de profondeur, où la sortie de chaque convolution parallèle indépendante et du MaxPooling est concaténée pour former la série temporelle multivariée de sortie du module de création actuel.

En empilant plusieurs modules Inception et en entraînant les valeurs des filtres par rétro propagation, le réseau est capable d'extraire les caractéristiques hiérarchiques latentes de plusieurs résolutions grâce à l'utilisation de filtres de différentes tailles.

5.2. Réseaux de neurones récurrents RNN

Un réseau de neurones récurrents (Recurrent Neural Network RNN) est une autre architecture de DNN conçu pour modéliser des données séquentielles, telles que des séries temporelles, des séquences d'événements [57]. Dans un RNN, la sortie de pas précédent est utilisée comme entrée pour le pas en cours. Par itération mise à jour des états cachés et de la mémoire, un RNN est capable de mémoriser les informations dans l'ordre séquentiel à chaque neurone. Les neurones d'entrée et de sortie sont connectés uniquement aux couches cachées avec le même pas de temps assigné.

Le RNN prend en compte l'état sauvegardé précédent lors de la mise à jour du poids, les gradients lorsque l'entraînement devient de plus en plus petit et après quelques étapes, les erreurs n'ont pas pu être propagées à la fin du réseau. Il n'y aura pas de différence significative dans le résultat, donc il ne peut pas faire de mise à jour des poids. Ce problème appelé « Vanishing Gradients ». Pour surmonter ce problème, une architecture à mémoire longue et courte durée (LSTM) a été proposée et aussi des étapes supplémentaires appelées Gated Recurrent Units (GRU). Ces étapes ont été utilisées pour améliorer les performances et la précision des RNNs.

5.3. Réseaux d'état d'écho

Les réseaux d'état d'écho (en anglais *Echo* State Network (ESN)) sont un type de réseaux de neurones récurrents, conçus pour atténuer les problèmes des réseaux de neurones récurrents en éliminant le besoin de calculer le gradient pour les couches cachées, en réduisant le temps d'apprentissage et en évitant le problème du gradient de disparition. En fait, de nombreux résultats montrent que les réseaux d'état d'écho sont vraiment utiles pour gérer les séries chronologiques chaotiques [9].

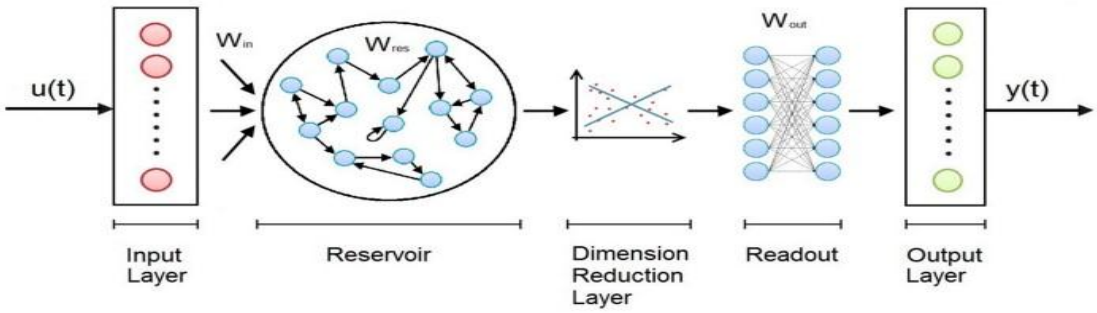


Figure III.8 : l'architecture de Réseaux d'état d'écho [56]

L'architecture d'un réseau d'état d'écho (voir la figure III.8) se compose d'une couche d'entrée, d'une couche cachée appelée réservoir, d'une couche de réduction de dimension, d'une couche entièrement connectée appelée lecture et d'une couche de sortie.

- Le réservoir est le bloc constitutif principal d'un réseau d'état d'écho. Il est vu comme un réseau neuronal récurrent aléatoire faiblement connecté.
- L'algorithme de réduction de dimension est généralement implémenté avec l'analyse en composantes principales.
- La lecture 'Readout' est généralement implémentée en tant que Perceptron multicouche.

Les poids entre la couche d'entrée et le réservoir et ceux dans le réservoir sont attribués au hasard et ne peuvent pas être entraînés. Les poids dans la lecture peuvent être entraînés, de sorte que le réseau puisse apprendre et reproduire des modèles spécifiques.

Le réservoir est connecté à la couche d'entrée et composé d'un ensemble de neurones internes faiblement connectés et ses propres neurones de sortie. Dans le réservoir, il existe 4 types de poids:

- Poids d'entrée entre la couche d'entrée et les neurones internes ;

- Poids internes, qui relient les neurones internes les uns aux autres ;
- Poids de sortie entre les neurones internes et la sortie ;
- Poids de rétropropagation, qui reconnectent la sortie aux neurones internes. Tous ces poids sont initialisés de manière aléatoire, égaux pour chaque pas de temps et qui ne sont pas entraînables.

Comme dans les RNN, la sortie du réservoir est calculée séparément pour chaque pas de temps puisque la sortie d'un pas de temps fait partie de l'entrée du pas de temps suivant. À chaque pas de temps, l'activation de chaque neurone interne et de sortie est calculée, et la sortie pour le pas de temps actuel est obtenue.

L'avantage des ESN est que le réservoir crée une intégration non linéaire récurrente de l'entrée dans une représentation de dimension supérieure, mais puisque seuls les poids dans la lecture peuvent être entraînés, le temps de calcul de l'entraînement reste faible.

En ce qui concerne la réduction de dimension, de nombreuses recherches montrent qu'il est possible de réduire le temps d'exécution sans diminuer la précision en fournissant une régularisation qui améliore la capacité de généralisation globale et la robustesse des modèles.

5.4.Architecture hybride

Actuellement, deux modèles d'apprentissage en profondeur les plus populaires et largement utilisés dans la classification des séries chronologiques : CNN et RNN. Ces modèles utilisent souvent une couche LSTM et une couche CNN empilée pour extraire des caractéristiques de la série chronologique, et une couche Softmax appliquée pour prédire les étiquettes [11].

6.Analyse Descriptive Des Données

L'analyse descriptive des données est une étude indispensable pour l'exploration et l'extraction de connaissances à partir de données, connue sous le nom de Fouille de données ou Data Mining en anglais. Quel que soit l'architecture de réseau de neurone profond utilisée, l'analyse descriptive joue un rôle important pour une classification automatique de qualité.

L'analyse de données permet de recueillir des données quantifiables qui peuvent être analysées à des fins statistiques portant sur une population cible. Elle vise essentiellement à décrire les caractéristiques d'un échantillon et à répondre aux questions de recherche [68].

L'analyse descriptive reposent sur l'utilisation des mesures de tendance centrale, de dispersion et de position ainsi que les analyses de fréquences [68] :

- **Mesures de tendance centrale :**
  - Le mode : la valeur de la variable statistique qui apparaît le plus souvent au sein d'un échantillon de données,
  - La médiane : réfère à la valeur de la variable telle qu'il y ait autant d'observations en dessous d'elle qu'au-dessus.
  - La moyenne : exprime la grandeur qu'aurait chacun des éléments d'un ensemble s'ils étaient tous identiques sans changer la dimension de l'ensemble.
- **Les mesures de dispersion et de position :**
  - L'écart-type : mesure de dispersion autour de la moyenne.
  - La variance : consiste en la somme des carrés des écarts par rapport à la moyenne, divisée par le nombre d'observations.
  - L'étendue : représente la différence entre les valeurs extrêmes d'une distribution/d'un ensemble.
  - Le minimum (la plus petite valeur) et le maximum (la plus grande valeur) d'une distribution.
- **L'analyse de la fréquence :** permet d'établir si les données suivent des distributions connues, comme la loi normale, ce qui influencera le choix du test statistique dans le cas d'une prédiction.

7.Travaux connexes pour la classification ECG basées sur DL

Après avoir présenté un aperçu sur différentes architectures d'apprentissage profond pour la TSC (Time Series Classification), nous allons présenter dans cette section une liste des travaux connexes pour la détection et le diagnostiques de données ECG modélisées par les séries chronologiques. Nous avons classé ces travaux en trois classes : approches basées sur CNN, approches basées sur RNN et les approches hybrides.

7.1.Les méthodes de classification ECG basées sur CNN

En général, l'utilisation des CNN s'inscrit dans l'approche d'apprentissage supervisé. Ils sont largement utilisés dans diverses applications telles que le filtrage du bruit, l'extraction des caractéristiques et la classification. Parmi les travaux récents basés sur l'architecture CNN pour le diagnostic de l'arythmie :

Atia et al. 2019 [36] ont proposé une approche à base de CNN pour identifier les dysfonctions ventriculaires gauche asymptomatique (ALVD). À l'aide de données appariées d'ECG à 12 dérivations, y compris la fraction d'éjection ventriculaire gauche (une mesure de la fonction contractile), de 44 959 patients à la clinique Mayo, ils ont formé un réseau neuronal convolutif pour identifier les patients présentant une dysfonction ventriculaire, définie comme la fraction d'éjection ≤ 35%, utilisant uniquement les données ECG. Lorsqu'il a été testé sur un ensemble indépendant de 52 870 patients, le modèle de réseau a donné des valeurs pour l'aire sous la courbe, la sensibilité, la spécificité et la précision de 93%, 86,3%, 85,7 %, respectivement.

X. Xu et H. Liu(2020) [59] proposent une méthode de classification des battements cardiaques ECG basée sur CNN. Sur la base de dataset d'arythmie MIT-BIH, l'approche proposée atteint une sensibilité de 99,2 % et une prédictive positive de 99,4 % dans la détection de VEB ; une sensibilité de 97,5% et une prédictive positive de 99,1% dans la détection de SVEB ; et une précision globale de 99,43 %. Le système développé peut être directement mis en œuvre sur des appareils portables pour surveiller les données ECG à long terme.

T. Mahmud, S. A. Fattah and M. Saquib (2020) [58] proposent une architecture efficace de réseau de neurones à convolution (CNN) basée sur la convolution temporelle en profondeur ainsi que d'un schéma de bout en bout robuste pour détecter et classer automatiquement l'arythmie à partir du signal d'électrocardiogramme (ECG) débruité, appelé comme `DeepArrNet'.Cette proposition a été validée sur un ensemble de données d'arythmie MIT BIH qui contient 48 demi-heures d'enregistrements ECG à deux canaux (MLII et V1) collectés auprès de 47 patients, avec environ 110 000 battements ECG. Plusieurs opérations de convolution temporelle sont effectuées en parallèle sur différentes tailles de noyau en même temps. Afin de limiter la complexité de calcul dans les opérations de convolution temporelle parallèles, une convolution temporelle échelonnée est effectuée, ce qui réduit également la longueur de la carte de caractéristiques de sortie.
Le modèle de réseau a donné des valeurs de performance pour la précision, la sensibilité, la valeur prédictive positive (VPP) et le F-score de 99.28 %, 99.13 %, 99.08 % et 99.11 %, respectivement.

7.2.Les méthodes de classification ECG basées sur RNN/LSTM/GRU

Les techniques RNN avec la plus grande précision sont expliquées ci-dessous.

Wang et al. (2019) [60] ont proposé un schéma de classification global et actualisable nommé Global Recurrent Neural Network (GRNN) qui a quatre couches au total. Dans la partie

morphologique, ils ont combiné la mémoire à long terme (LSTM) et un module de lancement pour la détection de CHF. Cinq bases de données open source ont été utilisées pour la formation et les tests, et trois types de longueur de segment RR (N = 500, 1000 et 2000) ont été utilisés pour la comparaison avec d'autres études. La méthode proposée a atteint une précision de 99,22 %, 98,85 % et 98,92 % sur le dataset BIDMC.

H. M. Lynn, S. B. Pan and P. Kim(2019) [61] ils proposent un réseau de neurones récurrents (RNN) profond basé sur l'unité récurrente fermée (GRU) de manière bidirectionnelle (BGRU) pour l'identification humaine à partir de la biométrie basée sur l'électrocardiogramme (ECG), une tâche de classification qui vise à identifier un sujet à partir de donnée séquentielle de série chronologique donnée. Bien qu'il y ait un problème majeur dans les réseaux RNN traditionnels dans lesquels ils apprennent des représentations à partir de séquences temporelles précédentes, le bidirectionnel est conçu pour apprendre les représentations à partir d'étapes de temps futures, ce qui permet une meilleure compréhension du contexte et élimine l'ambiguïté. Les résultats expérimentaux suggèrent que le modèle BGRU proposé, la combinaison de RNN avec une unité cellulaire GRU de manière bidirectionnelle, a atteint une précision de classification élevée de 98,55%.

Klosowski, G et al.(2020) [62] proposent une méthode de classification des signaux ECG basée sur l'extraction spectrale de caractéristiques à l'aide de la transformée logarithmique et l'utilisation du réseau de neurones RNN avec LSTM donne de bons résultats. Dans ce modèle, un seul signal ECG brut a été transformé en deux signaux de spectrogrammes, générés par diverses transformations temporelles, ce qui a augmenté l'efficacité de la prédiction. La précision moyenne du LSTM pour l'ensemble de tests était de 70,8 %.

7.3.Les méthodes de classification ECG basées sur les architectures hybrides

Q. Yao, R. Wang and X. Fan et al. (2020) [38] ont proposé un modèle de classification de L'ECG à 12 dérivations. Une génération d'une série temporelle prétraitée dans les couches entièrement convolutives. Cette série temporelle est ensuite introduite dans les cellules LSTM pour échanger des informations entre différents moments. Un module d'attention accepte la sortie des cellules LSTM, attribue des pondérations pour différents moments et génère un résultat final. Les données d'entraînement utilisées dans cette étude provenaient du 1er China Physiological Signal Challenge. Ils ont obtenu un F-score moyen de 81,2% dans la classification de 8 types d'arythmies et de rythme

sinusal, dépassant de 7,7% le modèle CNN de référence. Le mécanisme d'attention aide le modèle à localiser la partie informative des signaux et améliore l'interprétabilité.

Fawaz et al. (2020) [48]. Un modèle hybride composé de deux blocs de trois Modules de démarrage chacun, par opposition aux trois blocs de trois couches convolutionnelles traditionnelles dans le dataset ResNet. Ces blocs maintiennent des connexions résiduelles et sont suivis par la mise en commun globale moyenne et les couches Softmax. Ce modèle prend une série multivariée d'entrée de longueur m, dimensionnalité d, et utilise une couche de goulot d'étranglement avec une longueur et une foulée de 1 à réduire la dimensionnalité. Les sorties de ces convolutions sont combinées à une source supplémentaire de diversité, à Max Pooling suivi d'un goulot d'étranglement appliqué à la série chronologique d'origine, et tous empilés pour former les dimensions de la multivariée de sortie de séries temporelles à alimenter dans la couche suivante.

Zhang et al. (2020)[50]. Une nouvelle approche TapNet qui s'appuie sur les avantages des approches d'apprentissage traditionnelles et en profondeur. TapNet combine ces avantages pour produire une architecture de réseau qui peut être décomposée en trois modules distincts : Permutation de dimensions aléatoires, Séries temporelles multivariées, Encodage et apprentissage de prototypes attentionnels. L'expérimentation explorant l'effet de ce module a révélé que dans 22 des 33 ensembles de données dans les archives multivariées de dataset UEA, la précision a été augmentée parmis eux les signaux biologique comme ECG

8.Conclusion

Nous avons présenté les techniques de deep learning pour l'apprentissage automatique de données des séries chronologiques, comme les techniques de classification supervisées et analyse descriptive de données. La plupart des travaux existants sont basés sur l'utilisation de fichiers CSV pour effectuer leur analyse. Malgré la simplicité de ce type de format mais on ne peut pas assurer une liaison entre fichiers ni de présenter d'une façon structurée avec moins de redondance. L'utilisation de base de données des séries chronologiques s'avèrent nécessaire notamment pour les séries chronologies volumineuses. Le chapitre suivant, nous monterons comment représenter et manipuler les données ECG dans un SGBD propre pour les séries chronologiques, et faire une analyse descriptive après une préparation de données.

Chapitre 04 :

Conception et Implémentation

1. Introduction

L'objectif de ce chapitre est de décrire les contributions réalisées dans le contexte d'étude de données évolutives.

La première contribution concerne la modélisation des données ECG à l'aide des séries chronologiques. Le choix de type et composants des séries chronologiques sont la première tâche à faire après la compréhension de données (Data Understanding) et le prétraitement de données (Data Preprocessing) de Dataset sélectionné.

Le modèle temporel proposé va être implémenté en utilisant le SGBD des séries chronologiques à large échelle InfluxDB.

La deuxième contribution sert à effectuer une analyse descriptive basée sur la visualisation graphique de données pour l'interprétation rapide de l'évolution de données et aussi la détection des changements par des alertes. Pour cela, nous nous focalisons sur l'utilisation de Framework Chronograf et Grafana.

Les détails de ces contributions sont présentés dans le reste de ce chapitre.

2. Dataset utilisé

Pour réaliser notre étude, nous avons utilisé le Dataset pour les données ECG provient de Physionet (une banque de données médicale comprenant des datasets pour les signaux physiologiques complexes): PTB-XL.

Le PTB-XL[4] est un dataset ECG cliniques avec des modifications appliquées pour évaluer algorithmes d'apprentissage automatique. L'ensemble de données ECG PTB-XL contient 21 837 données cliniques à 12 dérivations ECG de 18 885 patients d'une durée de 10s, échantillonnés à 500 Hz et 100 Hz.

PTB-XL (Physikalisch-Technische Bundesanstalt en allemand (la fédération physio-technique)) composé de cinq clusters (**NORM** : ECG normal, **CD** : trouble de la conduction, **IM** : infarctus du myocarde, **HYP** :hypertrophie et **STTC** : changements ST/T) et 24 sous-classes est fourni (voir la figure IV.1)

[4] https://physionet.org/content/ptb-xl/1.0.1/

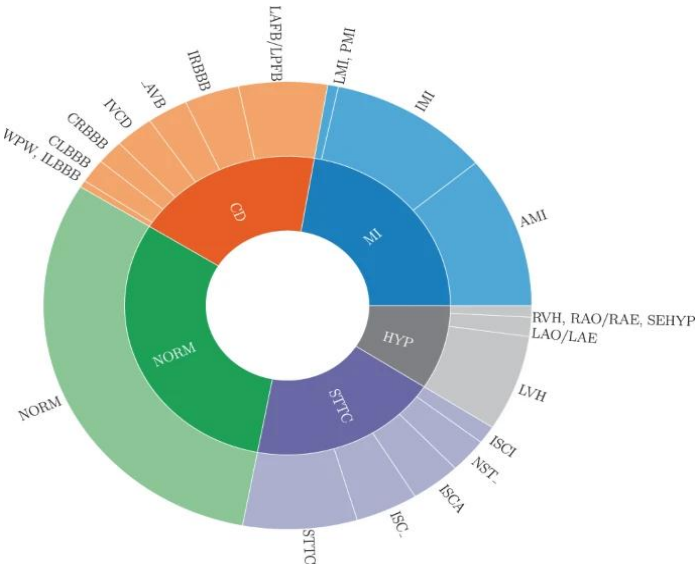


Figure IV.1 Dataset PTB-XL [4]

Tous les signaux ECG sont stockés dans le répertoire /ptb-xl-a-large-publicly-available-electrocardiography-dataset-1.0.1 au format WaveForm DataBase (WFDB) avec une fréquence d'échantillonnage de 500 Hz. Afin qu'on puisse utiliser ses données, nous avons choisi la version sous-échantillonnées des données de forme d'onde à une fréquence d'échantillonnage de 100 Hz.

Toutes les métadonnées pertinentes sont stockées dans ptbxl_database.csv chaque ligne représente une série chronologique identifié par ECG_id. Il contient 28 colonnes (caractéristiques) qui peuvent être classées en :

- **Identifiants** : Chaque enregistrement est identifié par un ECG_id. Le patient correspondant est codé via patient_id. Les chemins vers l'enregistrement d'origine (500 Hz) et une version sous-échantillonnée de l'enregistrement (100 Hz) sont stockés dans filename_hret filename_lr.
- **Métadonnées** générales telles que l'âge, le sexe, la taille, le poids, l' infirmière, le site, l'appareil et la date d'enregistrement.
- **Relevés ECG** : les composants de base sont scp_codes (Déclarations SCP-ECG sous forme de dictionnaire avec des entrées de la forme statement: likelihood, où la probabilité est définie sur 0 si elle est inconnue) et report (chaîne de rapport).

- **Métadonnées du signal** : qualité du signal comme le bruit (static_noiseet burst_noise), dérives de la ligne de base (baseline_drift) et d'autres artefacts tels que electrodes_problems. Nous fournissons également extra_beats pour compter les systoles supplémentaires et le stimulateur cardiaque pour les modèles de signaux indiquant un stimulateur cardiaque actif.
- **Plis de Cross validation**: 10-fraction (fold) train-test splits (strat_fold) obtenu par échantillonnage stratifié tout en respectant les affectations des patients, c'est-à-dire que tous les dossiers d'un patient particulier ont été affectés au même pli. Les disques des plis 9 et 10 ont subi au moins une évaluation humaine et sont donc d'une qualité d'étiquette particulièrement élevée. Nous proposons donc d'utiliser les plis 1 à 8 comme jeu d'apprentissage, le pli 9 comme jeu de validation et le pli 10 comme jeu de test.

Toutes les informations relatives au schéma d'annotation utilisé sont stockées dans un scp_statements.csv. Physionet fournit des informations secondaires supplémentaires telles que la catégorie à laquelle chaque déclaration peut être attribuée (diagnostic, forme et/ou rythme).

Avant de modéliser les données évolutives par les séries chronologiques, nous devons faire dans un premier temps, un prétraitement de données.

3. Prétraitement de données

Le dataset utilisé PTB-XL n'est pas préparé pour l'analyse automatique ; il contient des données bruitées, des classes non équilibrées avec une énorme masse de données non seulement pour chaque patient mais encore dans chaque dérivation d'ECG. Les performances des méthodes d'analyse de données dépendent fortement sur la quantité et la qualité des données d'apprentissage, plus de donnés avec qualité plus de précision et de bon résultat.

3.1. Nettoyage de données

Le nettoyage de données (Data Cleaning) est un processus qui vise à réduire le volume de données. Il identifie et corrige les données altérées, inexactes ou non pertinentes. Cette étape fondamentale du traitement des données améliore la cohérence, fiabilité et valeur des données.

Dans notre cas, nous avons effectué les opérations suivantes :

- Suppression de données manquantes (indiquées par la valeur NULL, NAN et INF).
- Suppression des ECG dans lesquels la probabilité de classification était inférieure à 100%.

- Eliminer les sous-classes supplémentaires, dont des échantillons sont inférieure à 20.

3.2. Réduction de la dimensionnalité

Au cours de la phase de préparation du données, la fréquence d'échantillonnage de 100 Hz a été sélectionnée pour l'étude, avec 10 s comme durée pour les données ECG obtenue après le nettoyage. La réduction de dimensionnalité vise à réduire le nombre de caractéristiques dans les données d'apprentissage. Ceci est effectué après le nettoyage et avant la modélisation et le traitement.

La technique la plus populaire pour la réduction des dimensions est l'Analyse en Composantes Principales (ACP) (Principal Component Analysis (PCA) en anglais). Elle permet d'explorer des jeux de données multidimensionnels constitués de variables quantitatives.

3.3. Equilibrage des classes

Une des difficultés principales rencontrée avec le jeu de données PTB-XL est le déséquilibre des classes. Un jeu de données est déséquilibré si les catégories de classification, ou classes, ne sont pas représentées de manière approximativement égale (voir le table IV.1).

Nombre d'enregistrements	Classes	Description
7185	NORME	ECG normal
3232	CD	Perturbation de la conduction
3064	STTC	Changement ST/T
2936	MI	Infarctus du myocarde
815	HYP	Hypertrophie

Table IV.1. Déséquilibre des classes de PTB-XL

Dans notre travail, nous nous focalisons sur la stratégie de sous échantillonnage qui vise à rééquilibrer directement les données par la diminution de nombre d'échantillons à traiter. Le choix de cette stratégie est motivé par la taille et le volume très important de données. Après le rééquilibrage on obtient 4 classes de diagnostique qui a des échantillons supérieurs à 2000 :

1 : ‘ECG NORMAL

2 : ‘Perturbation de la conduction’

3 :’Changement ST/T’

4 :'Infarctus du myocarde'

4. Modélisation et stockage de données ECG par les séries chronologiques

Comme nous avons indiqué dans ce travail que notre proposition sert à modéliser les données évolutives par les séries chronologiques en utilisant un système de gestion de données dédié appelé InfluxDB. Les travaux existants dans ce contexte sont basés sur des modèles mathématiques décrivant les séries chronologiques.

Afin de concevoir et modéliser les données PTB-XL par les séries chronologiques, nous devons savoir la structure de base de données de InfluxDB, apprendre l'utilisation de InfluxDB sous le manque de documentations concernant l'analyse de données stockées dans InfluxDB.

Une série chronologique est représentée sous la forme suivante :

```
<measurement>[,<tag_key>=<tag_value>[,<tag_key>=<tag_value>]]
<field_key>=<field_value>[,<field_key>=<field_value>] [<timestamp>]
```

La base de données stocke des mesures (measurements). Dans notre cas, il concerne les signaux ECG de 12 dérivations pour chaque patient durant 10 secondes. Chaque mesure est composée de :

- key : pour chaque ligne de mesure. Dans notre dataset, on a besoin d'une seule clé indiquant l'identifiant de patient.
- Une ou plusieurs clés de champs (field keys) sont définies dans une ligne pour présenter les propriétés ou les caractéristiques de dataset PTB-XL.
- Tag : pour annoter une série chronologique. Il représente les propriétés pour l'indexation (comme les indexes des bases de données).
- Timestamp : le temps d'évolution de données au format yyyy-mm-dd T hh:mm:ssZ

La base de données de séries chronologiques contient des données recueillies à des périodes adjacentes et qui sont ordonnées chronologiquement. Une série chronologique exprimant une données ECG doit être composé des changement cyclique (C) suivant les battements de cœur mesurés (Voir le chapitre 01) et probablement des Tendance (T) et Variation irrégulière (I) dans le cas de l'existence des troubles cardiaques.

Avec le client influx, nous avons créé des séries chronologiques des points dans la base PTB-XL par la conversation de fichier PTBXL-database.CSV au pandas data frame (par python) ensuite chaque ligne est transformée en format JSON interprétable par le moteur Influx qui a son tour transfert les fichiers JSON à Influxdb par la fonction client.write_points en temps réel.
En raison de volume important de données, la figure IV.2 illustre une partie des tag/key de modèle de séries chronologiques implémenté par InfluxDB.

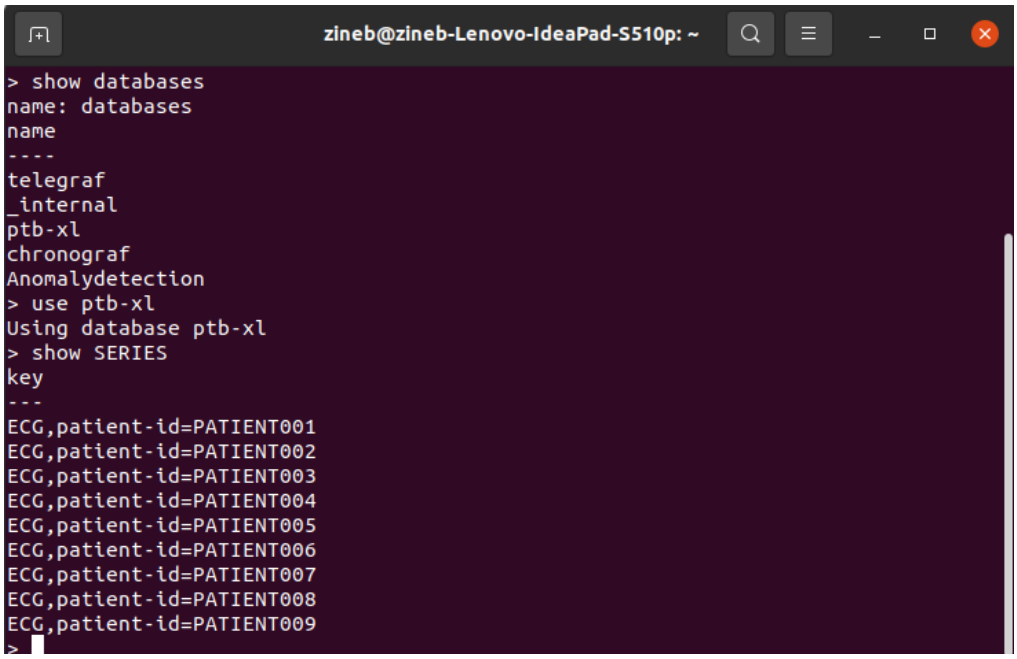


Figure IV.2. Les Tag Key de notre base de données InfluxDB

Les Fields keys de 12 dérivations de measurment ECG sont présentées comme suit :

Figure IV.3. Les Fields Key de measurment ECG

5. Visualisation et manipulation de données

La visualisation de données est une forme de graphique qui capture notre attention et attire notre regard sur le message communiqué. En observant un tracé, nous pouvons rapidement identifier les tendances et valeurs évolutives.

Après avoir élaboré la modélisation et le stockage de données évolutives d'ECG par InfluxDB, la deuxième partie de contribution sert à analyser visuellement le changement de données, détecter les anomalies, des pics et de faire l'interrogation. Pour cela, nous avons choisi deux framework les plus récents dédiées au traitement de séries chronologiques : Chronograf et Grafana.

5.1. Visualisation par Chronograf

Chronograf permet de voir rapidement les données stockées dans InfluxDB et créer des requêtes et des alertes robustes.

Chronograf propose une solution complète de dashboarding pour visualiser les données. Plus de 20 tableaux de bord prédéfinis sont disponibles pour nous permettre de démarrer très rapidement. Dans le cas de notre travail, les données ECG modélisées par les séries chronologiques sont illustrées dans la figure suivante :

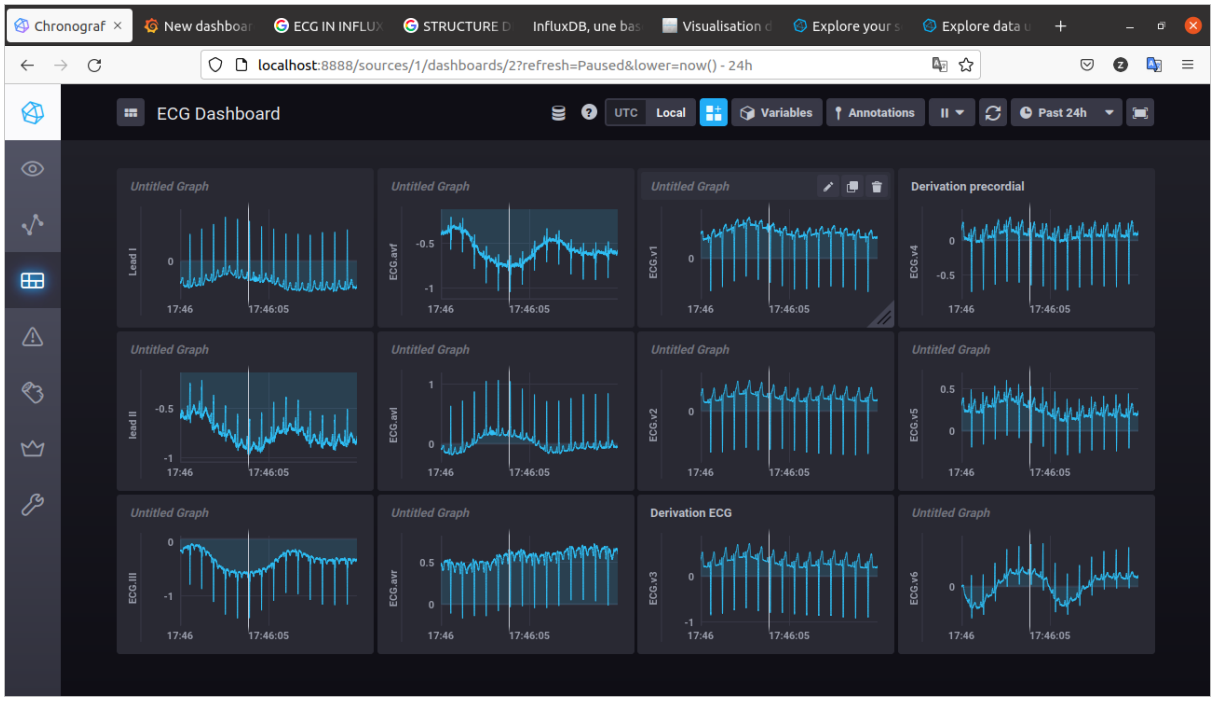


Figure IV.4. Les dérivations de patient 001 sous Chronograf

5.2. Visualisation par Grafana

Grafana est une plateforme open source de monitoring, analyse et visualisation des données systèmes en temps réel[5]. L'objectif de cette solution est de présenter facilement et de façon intuitive une grande quantité de données issues de sources différentes.

Grafana permet de tracer des données historiques, temporelles et évolutives et de créer des tableaux de bord (dashboard) en temps réel pour toutes les métriques écrites dans la base de données. Grafana s'exécute sur un ordinateur (Lenovo Processeur core i7, RAM 8 Go avec 1 Téra de disque dure avec le système d'exploitation Ubuntu 20.4) et on peut accéder aux serveurs InfluxDB et aux ordinateurs à partir desquels nous souhaitons surveiller.

La figure IV.5 montre notre Dashboard de surveillance de signaux ECG du patient « Patient001» qui est compose de 3 panneaux ces panneaux représente les trois groupes de dérivation.

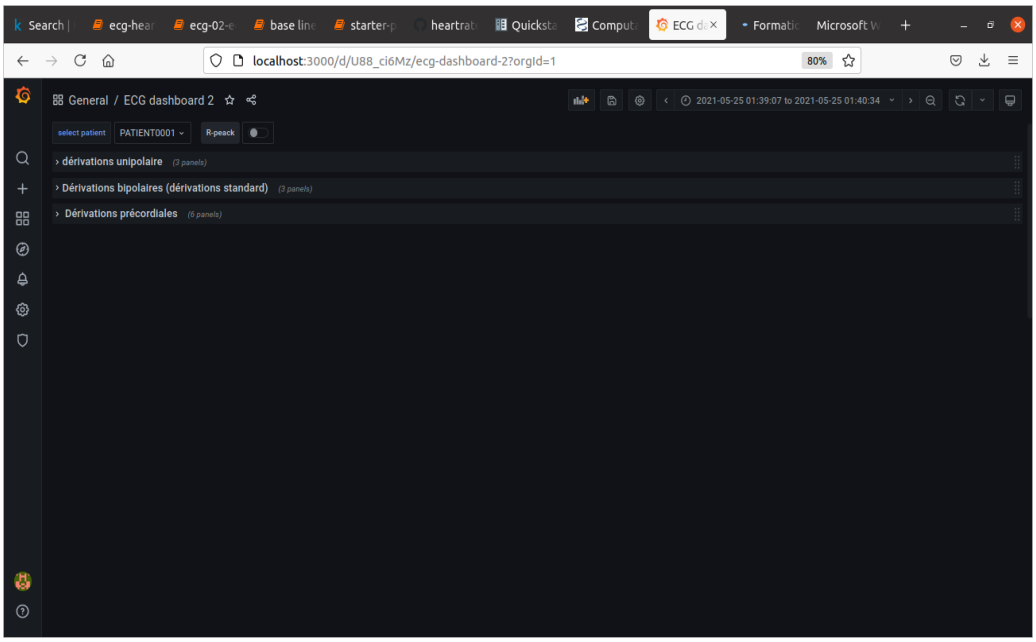


Figure IV.5. Les dérivations de patient 001 sous Grafana

La figure IV.6. Illustre les différentes dérivations présentées sous forme des séries chronologiques de patient 001.

[5] https://www.syloe.com/glossaire/grafana/

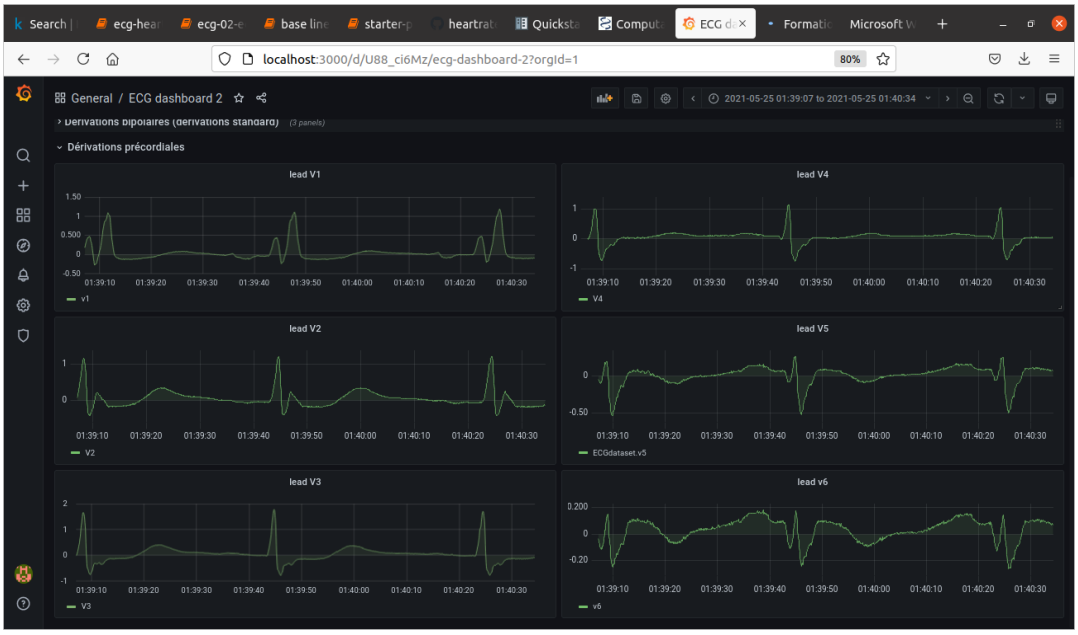


Figure IV.6. Les séries chronologiques des dérivations sous Grafana

6. Conclusion

Dans ce chapitre nous avons présenté la conception et l'implémentation de modèle de représentation de données évolutives par les séries chronologiques. Ces données avec stockage efficace dans un système de gestion de séries chronologiques à large échelle. L'analyse descriptive de ce type de données ce n'est pas une tâche simple à réaliser. Une préparation de données prend beaucoup temps notamment pour les données qui contient plus de 252.000.000 (1000*12*21000). Le modèle de données obtenu peut être utilisé non seulement pour l'analyse de données mais aussi pour la migration de données évolutives vers les séries chronologiques.

Conclusion Générale

Dans ce mémoire, nous avons abordé un des principaux axes de recherche de data mining qui concerne l'étude exploratoire de données en vue d'une classification supervisée. Cette étude est appliquée aux données évolutives qui sont de structure complexe à cause de changement continuelle de leurs valeurs. Les données d'ECG de diagnostic de maladies cardiovasculaires sont un très bon exemple de données évolutives. Ces données sont vues comme des séquences de valeurs présentant les ondes électriques des impulsions cardiaques.

Les séries chronologiques représentent un modèle de données le plus adéquats pour représenter et étudier les données évolutives. Le stockage de données ECG sous forme des séries chronologiques améliore la description et l'exploration de données en termes de composants tels que les measurements, les fields et les tags et aussi de manipulation.

L'objectif principal de ce travail était d'étudier les données évolutives dans le sens de faire une analyse descriptive (ou exporatoire) et une visualisation graphique de données modélisées par les séries chronologiques.

Le premier objectif de cette étude était de faire une revue de la littérature sur les données évolutives, les séries chronologiques, les données ECG, ainsi que les différents types d'analyse de données modélisées par les séries chronologiques.

Le second objectif était d'apprendre le fonctionnement des systèmes et frameworks utilisés : InfluxDB, Chronograf, Grafana et la programmation par le langage Python.

La nouveauté de notre travail par rapport les travaux existants sur l'analyse de données évolutives en séries chronologiques est qu'il permet de représenter les données par un modèle efficace géré par un système de gestion de base de données des séries chronologiques (InfluxDB) plutôt que d'utiliser des simples fichiers CSV, et l'analyse descriptive performante par la visualisation graphique de données via des frameworks spécialisés plutôt que d'utiliser des packages de python (par exemple) qui ne sont pas toujours applicable aux données qui changent des valeurs au cours du temps.

Les difficultés rencontrées dans notre travail sont principalement dans la phase de préparation et la modélisation de données. En fait, les données évolutives sont très complexes notamment les données ECG où chaque patient doit avoir 12 séries chronologiques présentant les 12 dérivations (ou leads) où chaque série a une longueur de 1000 valeurs. L'utilisation de données ECG de 12 dérivations donne un meilleur diagnostic et interprétation par contre leur stockage et leurs analyses sont complexes, nécessitant des machines puissantes.

Nous avons essayé de faire un modèle neuronal RNN-LSTM pour la classification supervisée de données ECG mais nous préférons de faire des améliorations de performance dans les futurs travaux afin d'effectuer une analyse complète de fouille de données. De plus, nous voudrons faire un prototype IoT en utilisant des capteurs biomédicaux comme AD8232 pour le suivi du rythme cardiaque, etc. et de carte Arduino (nous avons déjà achetés ces capteurs) afin d'effectuer une analyse automatique en temps réel.

Bibliographies

[1] Aggarwal, C. C., Philip, S. Y., Han, J., & Wang, J. (2003, January). A framework for clustering evolving data streams. In Proceedings 2003 VLDB conference (pp. 81-92). Morgan Kaufmann.

[2] M. Tahmassebpour and A. M. Otaghvari, "Increase efficiency big data in intelligent transportation system with using IoT integration cloud", J. Fundam. Appl. Sci., vol. 8, pp. 2443-2461, 2016

[3] Balani, Naveen.Enterprise IOT: A Definitive Handbook. ISBN 1518790860

[4] Time Series Database vs. Common Database Technologies for IOT, https://www.alibabacloud.com

[5] Ahlame Douzal-Chouakria. Contribution à l'analyse de données temporelles. Machine Learning [stat.ML]. Université Joseph-Fourier - Grenoble I, 2012.

[6] Joshi, P., Massaron, L., and Hearty, J. Python: Real World Machine Learning. PacktPublishing, 2017

[7] Time Series Databases and InfluxDB, Syeda Noor Zehra Naqvi (000455274) Sofia Yfantidou (000456361) December 17, 2017

[8] Table Store Time Series Data Storage — Architecture ByZhaofeng Zhou https://alibaba-cloud.medium.com Consulté le 14/05/2021

[9] Frederic Sur Ecole des Mines de Nancy Series chronologiques - Décomposition d'une chronique-ttps://members.loria.fr/FSur/enseignement/modprev/ Consulté le 14/05/2021

[10] M.-C.Viano 'Maîtrise d'Économétrie Cours de Séries Temporelles' Années1999 à 2004

[11] Aghabozorgi, Saeed; Shirkhorshidi, Ali S.; Wah, Teh Y. (2015). "Time-series clustering – A decade review". Information Systems. Elsevier. 53: 16–38.

[12] Keogh, Eamonn J. (2003). "On the need for time series data mining benchmarks". Data Mining and KnowledgeDiscovery. Kluwer. 7: 349–371

[13] Agrawal, Rakesh; Faloutsos, Christos; Swami, Arun (October 1993). "Efficient SimilaritySearch In Sequence Databases". Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms. International Conference on Foundations of Data Organization and Algorithms. 730. pp. 69–84.

[14] TABLEAU SOFTWARE,. (n.d.). Time Series Analysis: Definition, Types, Techniques, and When It's Used. Retrieved from https://www.tableau.com/.

[15] Faisal, S., Sarwar, M., Shahzad, K., Sarwar, S., Jaffry, W., & Yousaf, M. M. (2017). Temporal and evolving data warehouse design. Scientific Programming, 2017.

[16] Da Silva, A. (2007). Analyzing the evolution of Web usage data. Monde des Util. Anal.Données, 36, 75-84.

Références

[17] Aggoune, A., Bouramoul, A., & Kholladi, M. K. (2017). Mediation system for dealing with semantic problems in databases. International Journal of Data Mining, Modelling and Management, 9(2), 99-121.

[18] Aggoune, A., Bouramoul, A., & Kholladi, M. K. (2016, March). Big data integration: A semantic mediation architecture using summary. In 2016 2nd International Conference onAdvanced Technologies for Signal and Image Processing (ATSIP) (pp. 21-25). IEEE.

[19] Navani, D., Jain, S., & Nehra, M. S. (2017, December). The internet of things (IoT): A study of architectural elements. In 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) (pp. 473-478). IEEE.

[20] Fu, T. C., Chung, F. L., Ng, V., & Luk, R. (2001). Pattern discovery from stock time series using self-organizing maps. In Workshop Notes of KDD2001 Workshop on Temporal Data Mining (Vol. 1).

[21] Diop L, Diop CT, Giacometti A, Li D, Soulet A (2019) Sequential pattern sampling with norm-based utility. Knowl Inf Syst. https://doi.org/10.1007/s10115-019-01417-3

[22] Keogh, E., Chu, S., Hart, D., & Pazzani, M. (2004). Segmenting time series: A survey and novel approach. In Data mining in time series databases (pp. 1-21).

[23] Duri Schmidt, Angelika Kotz Dittrich, Werner Dreyer, and Robert W. Marti.Time series,A neglected issue in temporal database research ? In Recent Advances in Temporal Databases, Proceedings of the International Workshop on Temporal Databases, pages 214–232, 1995.

[24] T. Dunning and E. Friedman. Time Series Databases : New Ways to Store and Access Data. O'Reilly Media, Inc., 2015.

[25] Wang, J., Zhang, Y., Gao, Y., & Xing, C. (2013, July). PLSM: a highly efficient LSM-tree index supporting real-time big data analysis. In 2013 IEEE 37th Annual Computer Software and Applications Conference (pp. 240-245). IEEE.

[26] https://db-engines.com/en/ranking/time+series+dbms . Consulté le 04/04/2021

[27] Jensen, S. K., Pedersen, T. B., & Thomsen, C. (2017). Time series management systems:A survey. IEEE Transactions on Knowledge and Data Engineering, 29(11), 2581-2600.

[28] https://www.aptech.com/blog/introduction-to-the-fundamentals-of-time-series-data-and-analysis/ Consulté le 15/05/2021.

[29] https://itfeature.com/time-series-analysis-and-forecasting/components-of-time-series Consulté le 01/06/2021.

[30] Python vs R : le duel https://moncoachdata.com Consulté le 17/08/2021

[31] https://blog.mediprostore.com/definition-ecg-electrocardiographie/ Consulté le 17/08/2021

Références

[32] Flandrin, P. Rilling, G. and Gonçalvès, P., "Une extension bivariée pour la Décomposition Modale Empirique - Application à des bruits blancs complexes", colloque GRETSI-Troyes, 11-14 septembre (2007).

[33] Flandrin, P. Gonçalves, P. and Rilling, G. " Empirical Mode Decomposition and its Algorithms ",IEEE-EURASIP Workship on Nonlinear Signal and Image Processing NSIP-03, Grado(I),(2003).

[34] https://fr.my-ekg.com/generalites-ecg/derivations-ecg.html Consulté le 17/08/2021

[35] https://www.maxisciences.com/holter-ecg Consulté le 17/08/2021

[36] Attia, ZI, Kapa, S., Lopez-Jimenez, F. et al. Dépistage de la dysfonction contractile cardiaque à l'aide d'un électrocardiogramme basé sur l'intelligence artificielle. Nat Med **25,** 70-74 (2019). https://doi.org/10.1038/s41591-018-0240-2

[37] Kang.H and Choi.S 2014. Bayesian common spatial patterns for multi-subject EEG classification. Neural Networks

[38] Q. Yao, R. Wang and X. Fan et al. 'Multi-class Arrhythmia detection from 12-lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network' Information Fusion 53 (2020) 174–182

[39] Ebrahimi, M. Loni and M. Daneshtalab et al. / Expert Systems with Applications: X 7 "A review on deep learning methods for ECG arrhythmia classification" (2020)

[40] Anake Pomprapaa , Waqar Ahmeda et al. Deep Learning of Arrhythmia Analysis Based on Convolutional Neural Network

[41] Sun, C.; Chen, C.; Li, W.; Fan, J.; and Chen, W. 2019. A Hierarchical Neural Network for Sleep Stage Classification Based on Comprehensive Feature Learning and Multi-Flow Sequence Learning. IEEE Journal of Biomedical and HealthInformatics .

[42] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature 521, 436–444, https://doi.org/10.1038/nature14539

[43] Hong, S., Zhou, Y., Shang, J., Xiao, C., and Sun, J.(2020). Opportunities and challenges of deep learning methods for electrocardiogram data: a systematic review. Comput. Biol. Med. 103801

[44] Ismail Fawaz, H., Forestier, G., Weber, J. et al. Deep learning for time series classification: a review. Data Min Knowl Disc 33, 917–963 (2019). https://doi.org/10.1007/s10618-019-00619-1

[45] Papernot N, McDaniel P (2018) Deep k-nearest neighbors: towards confident, interpretable and robust deep learning. arXiv:1803.04765

[46] LeCun Y, Bottou L, Orr GB, Müller KR (1998b) Étai arrière efficace. In : Montavon G (ed) Réseaux de neurones : trucs du métier. Springer, Berlin, pp 9-50 Return to ref 1998b in article.

[47] Filippo Maria Bianchi, Simone Scardapane, Sigurd Løkse, Robert Jenssen. Reservoir computing approaches for representation and classification of multivariate time series.

Références

[48] Ismail Fawaz, H., Lucas, B., Forestier, G. et al. InceptionTime: Finding AlexNet for time series classification. Data Min Knowl Disc 34, 1936–1962 (2020).

[49] Haoyan Xu Ziheng Duan, Yunsheng Bai et al. Multivariate Time Series Classification with Hierarchical Variational Graph Pooling (2020)

[50] Zhang, X.; Gao, Y.; Lin, J.; and Lu, C.-T. 2020. TapNet: Multivariate Time Series Classification with Attentional Prototypical Network. In AAAI, 6845–6852

[51] Saadatnejad, S., Oveisi, M., Hashemi, M., 2019. Lstm-based ecg classification for continuous monitoring on personal wearable de vices. IEEE journal of biomedical and health informatics.

[52] Jensen, S. K., Pedersen, T. B., & Thomsen, C. (2017). Time series management systems: A survey. IEEE Transactions on Knowledge and Data Engineering, 29(11), 2581-2600

[53] Ouadi Beya. Analyse et reconnaissance de signaux vibratoires : contribution au traitement et à l'analyse de signaux cardiaques pour la télémédecine. Université de Bourgogne, 2014.

[54] https://cardiologs.com/blog/ai-systems-powered-by-deep-learning/ Consulté le 25/07/2021

[56] Time Series Classification with Deep Learning | by Marco Del Pra | Towards Data Science

[57] T. TOHARUDIN ET AL " Employing long short-term memory and Facebook prophet model in air temperature forecasting" 2020

[58] T. Mahmud, SA Fattah et M. Saquib, "DeepArrNet: Une architecture CNN profonde efficace pour la détection et la classification automatiques des arythmies à partir des battements ECG débruités", dans IEEE Access , vol. 8, p. 104788-104800, 2020, doi : 10.1109/ACCESS.2020.2998788.

[59] X. Xu et H. Liu, "Classification des battements de cœur ECG à l'aide de réseaux de neurones convolutifs",dans IEEE Access , vol. 8, p. 8614-8619, 2020, doi : 10.1109/ACCESS.2020.2964749.

[60] Wang, L.; Zhou, X. Detection of Congestive Heart Failure Based on LSTM-Based Deep Network via Short-Term RR Intervals. Sensors **2019**, 19, 1502. https://doi.org/10.3390/s19071502

[61] H. M. Lynn, S. B. Pan and P. Kim, "A Deep Bidirectional GRU Network Model for Biometric Electrocardiogram Classification Based on Recurrent Neural Networks," in IEEE Access, vol. 7, pp. 145395-145405, 2019, doi: 10.1109/ACCESS.2019.2939947.

[62] Klosowski, G.; Rymarczyk, T.; Wójcik, D.; Skowron, S.; Cieplak, T.; Adamkiewicz, P. L'utilisation des moments temps-fréquence comme entrées du réseau LSTM pour la classification des signaux ECG. Électronique **2020** , 9 , 1452. https://doi.org/10.3390/electronics9091452

[63] https://www.sante-sur-le-net.com/maladies/cardiologie/fibrillation-atriale

[64] https://m.20-bal.com/pravo/9419/index.html

[65] Soon S , Svavarsdottir H , Downey C , et al Dispositifs portables pour la surveillance à distance des signes vitaux en ambulatoire : un aperçu du domaine Innovations BMJ 2020 ; 6 : 55-71.

[66] Renard, X. (2017). Time series representation for classification: a motif-based approach (Doctoral dissertation, Université Pierre et Marie Curie-Paris VI).

[67] Polat, K., & Güneş, S. (2007). Detection of ECG Arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine. Applied Mathematics and Computation, 186(1), 898-906.