

Marie GUYOMARD

Post-doctorante en Machine Learning

Laboratoire SESSTIM, INSERM, IRD

Université d'Aix-Marseille

✉ marie.guyomard@univ-amu.fr

1. PARCOURS UNIVERSITAIRE ET DIPLÔMES

DOCTORAT Thèse de Doctorat en Automatique, Traitement du Signal et des Images

Du 01/10/2020 📍 Laboratoire I3S, Université Côte d'Azur, France

Au 29/11/2023 J'ai débuté ma thèse le 1er Octobre 2020 et obtenu mon diplôme de Doctorat le 29 Novembre 2023. La thèse s'intitule « *Régression Logistique à Noyau Neural Explicable : Application à la Médecine de Précision* ». Elle a été co-encadrée par **Lionel Fillatre**, Professeur en Apprentissage Statistique à Polytech-Nice Sophia-Antipolis, et **Nicolas Glaichenhaus**, Professeur de Biologie à l'Université Côte d'Azur. L'objectif principal de ma thèse a été de développer une méthode de classification adaptée aux contraintes rencontrées dans l'application médicale. Nous nous sommes concentrés sur l'interprétabilité des réseaux de neurones et leur fiabilité d'apprentissage au travers de leur linéarisation. L'article [CI-4] résume les différents fondements théoriques et contributions de ma thèse. Davantage de détails sont disponibles dans mon manuscrit de thèse [TH-1].

MASTER Diplôme d'ingénieur en Data Sciences de l'ENSAI

2017-2020 📍 Rennes, France

L'École Nationale de la Statistique et de l'Analyse de l'Information (ENSAI) est la première école française qui donne accès aux multiples fonctions de l'ingénierie statistique, grâce à un enseignement approfondi en statistique et en informatique. Au cours de ma dernière année d'étude à l'Ensaï, j'ai décidé de me spécialiser dans la recherche en intégrant la filière « *Génie Statistique* ».

ERASMUS Master of Statistical Science for the Life and Behavioural Sciences

2019 📍 Université de Leiden, Pays-Bas

Durant mon cursus à l'Ensaï j'ai eu l'opportunité de réaliser un échange Erasmus. J'ai choisi d'intégrer pendant un semestre le Master de bio-statistiques de l'Université de Leiden aux Pays-Bas afin de développer mes compétences en méthodes statistiques adaptées au domaine de la recherche en santé.

CPGE Classe Préparatoire aux Grandes Écoles ENS Cachan D2

2014-2017 📍 Paris, France

2. EXPÉRIENCE POST-DOCTORALE

AI4Elderly : L'intelligence artificielle pour l'étude de parcours de soins des patients âgés, atteints de cancer

📍 SESSTIM, Université d'Aix Marseille, France

📅 01/01/2024 - À ce jour

Mon post doctorat est encadré par **Raquel Ureña**, Maîtresse de Conférence en Intelligence Artificielle à l'Université d'Aix Marseille. Ma recherche post-doctorale s'applique à des données longitudinales provenant de dossiers médicaux électroniques (*Electronic Health Records*, EHR), et plus particulièrement de la base du **Système National des Données de Santé (SNDS)**. L'objectif principal de ce projet est de développer un outil d'aide à la décision pour les oncologues afin de personnaliser les thérapies des personnes âgées atteintes d'un cancer. Pour répondre à cet objectif, plusieurs outils méthodologiques sont étudiés, notamment les méthodes de *Representation Learning* [CI-3].

3. ACTIVITÉS D'ENSEIGNEMENTS

3.1. Vacations

Data Valorization

📍 Master 1 Mathématiques Appliquées et Modélisation, Polytech Nice-Sophia

📅 2020, 2021 et 2022

Ce module introduit des concepts fondamentaux d'apprentissage statistique : maximum de vraisemblance, test naïf de Bayes, régression logistique ou encore les arbres de décision. Les applications pratiques ont été implémentées sous Rstudio.

JAVA Programming

📍 Licence 3 Mathématiques Appliquées et Modélisation, Polytech Nice-Sophia

📅 2020, 2021 et 2022

Ce module combine une introduction aux systèmes d'exploitation Linux (gestion des droits, administration, installation de logiciels sous Ubuntu) et une initiation à la programmation orientée objet (POO) en Java. Les notions enseignées incluaient la création de classes (instances, constructeur, méthodes), l'héritage, la gestion des erreurs, la création de packages et la documentation avec Doxygen.

Big Data

📍 Master 2 Mathématiques Appliquées et Modélisation, Polytech Nice-Sophia

📅 2020, 2021 et 2022

Ce module vise à introduire des outils pour la gestion et l'analyse de grandes bases de données. Les étudiants ont été formés aux technologies suivantes : stockage en HDFS (Hadoop Distributed File System), gestion des ressources avec YARN, utilisation de frameworks comme Spark (MapReduce), HBase (NoSQL) et MongoDB. Les implémentations pratiques étaient réalisées en Scala.

Programming and Languages Coding

📍 Master 2 Ingénierie de la Santé, Université de Pharmacologie d'Aix-Marseille

📅 2024

Ce module aborde une introduction pratique au langage de programmation Python et introduit les concepts fondamentaux de l'apprentissage automatique.

3.2. Jury de projets et de stages de fin d'étude

Jury de stages de fin d'étude

2021 Master de Mathématiques Appliquées et Modélisation (MAM) de Polytech-Nice (Université Côte d'Azur).

Jury d'évaluation de projets de fin d'étude

2022 & 2023 Master de Mathématiques Appliquées et Modélisation (MAM) de Polytech-Nice (Université Côte d'Azur).

3.3. Missions de supervision et d'encadrement

Co-supervision d'un ingénieur d'étude

2024 Housseem Eddine Azzag a travaillé sur le développement d'un package en R afin de faciliter le pré-traitement des données du SNDS ; caractériser les patients atteints d'un cancer ; identification des cancers et des traitements reçus, identification de comorbidités, représentation des parcours thérapeutiques dans le temps à l'aide d'un Diagramme de Sankey.

Co-encadrement de stagiaires

2024 *Predicting Hepatocellular Carcinoma in Individuals with Chronic Hepatitis B : Comparison of Survival and Classification Machine Learning Algorithms.*

Stage de fin d'étude de 6 mois du Master Digital Solutions for Pharmaceutical Sciences (DIGI-PHARM), AMU.

2022 *Intelligible model in machine learning : Explainable Boosting Machines.*

Stage de 3 mois, étudiant en Master 1 Mathématiques Appliquées et Modélisation, Polytech-Nice.

2021 *Machine Learning applied to medical data.*

Stages de 3 mois, deux étudiants en Master 1 Mathématiques Appliquées et Modélisation, Polytech-Nice.

4. PUBLICATIONS

4.1. Manuscrit de Thèse

[TH-1] Marie GUYOMARD. « Régression logistique à noyau neural explicable : application à la médecine de précision ». In : *Thèse.fr* (2023).

4.2. Journaux Internationaux

[J-2] Cyprien GILET, Marie GUYOMARD, Sébastien DESTERCKE et Lionel FILLATRE. « Softmin discrete minimax classifier for imbalanced classes and prior probability shifts ». In : *Machine Learning* 113.2 (2024), p. 605-645.

4.3. Conférences Internationales à Comité de Relecture

- [CI-3] **Marie GUYOMARD**, Anne-Déborah BOUHNİK, Louis TASSY et Raquel URENA. « Encoding breast cancer patients' medical pathways from reimbursement data using representation learning : a benchmark for clustering tasks ». In : *2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE. 2024, p. 120-125.
- [CI-4] **Marie GUYOMARD**, Susana BARBOSA et Lionel FILLATRE. « Kernel logistic regression approximation of an understandable ReLU neural network ». In : *International Conference on Machine Learning (ICML)*. PMLR. 2023, p. 12268-12291.
- [CI-5] **Marie GUYOMARD**, Susana BARBOSA et Lionel FILLATRE. « Understandable relu neural network for signal classification ». In : *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2023, p. 1-5.
- [CI-6] Cyprien GILET, **Marie GUYOMARD**, Susana BARBOSA et Lionel FILLATRE. « Multiclass Minimax Learning for Deep Neural Networks ». In : *2023 31st European Signal Processing Conference (EUSIPCO)*. IEEE. 2023, p. 1265-1269.
- [CI-7] Cyprien GILET, **Marie GUYOMARD**, Susana BARBOSA et Lionel FILLATRE. « Adjusting decision trees for uncertain class proportions ». In : *ECML/PKDD 2020 Tutorial and Workshop on Uncertainty in Machine Learning*. 2020.

4.4. Conférences Nationales à Comité de Relecture

- [CN-8] **Marie GUYOMARD**, Anne-Déborah BOUHNİK, Tassy LOUIS et Raquel UREÑA. « Representation Learning pour la codification des parcours thérapeutiques de patientes atteintes de cancer du sein à partir de données de remboursement : un benchmark pour des tâches de clustering ». In : *Atelier Intelligence Artificielle Centrée sur les Données - Extraction et Gestion des Connaissances*. 2025.
- [CN-9] Cyprien GILET, **Marie GUYOMARD**, Sébastien DESTERCHE et Lionel FILLATRE. « Classifieur minimax discret randomisé pour la détection de classes rares et la présence de probabilités a priori imprécises ». In : *GRETSI'23-XIXème Colloque Francophone de Traitement du Signal et des Images*. 2023.
- [CN-10] **Marie GUYOMARD**, Susana BARBOSA et Lionel FILLATRE. « Approximation d'un Réseau de Neurones ReLU interprétable par une Régression Logistique à Noyau ». In : *54es Journées de Statistique de la SFdS (JDS)*. 2023.
- [CN-11] Cyprien GILET, **Marie GUYOMARD**, Sébastien SDESTERCHE et Lionel FILLATRE. « Apprentissage d'un classifieur minimax randomisé pour risques d'erreur par classe déséquilibrés et probabilités a-priori incertaines ». In : *54es Journées de Statistique de la SFdS (JDS)*. 2023.
- [CN-12] **Marie GUYOMARD**, Susana BARBOSA et Lionel FILLATRE. « Adaptive splines-based logistic regression with a ReLU neural network ». In : *Les Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM)*. 2022.
- [CN-13] **Marie GUYOMARD**, Susana BARBOSA et Lionel FILLATRE. « Régression logistique à base de splines adaptatives avec un réseau de neurones ReLU ». In : *GRETSI'22-XXVIIIème Colloque Francophone de Traitement du Signal et des Images*. 2022.
- [CN-14] Cyprien GILET, **Marie GUYOMARD**, Susana BARBOSA et Lionel FILLATRE. « Apprentissage Minimax pour les Réseaux de Neurones ». In : *GRETSI'22-XXVIIIème Colloque Francophone de Traitement du Signal et des Images*. 2022.
- [CN-15] **Marie GUYOMARD**, Cyprien GILET, Susana BARBOSA et Lionel FILLATRE. « Sur l'équivalence entre la régression logistique à base de splines et l'apprentissage profond ». In : *Congrès des Jeunes Chercheuses et Chercheurs en Mathématiques Appliquées (CJ-CMA)*. 2021.

[CN-16] **Marie GUYOMARD**, Cyprien GILET, Susana BARBOSA et Lionel FILLATRE. « Réseaux de Neurones Convolutifs avec Apprentissage Minimax pour des Proportions par classe incertaines et déséquilibrées ». In : *ORASIS 2021*. 2021.

4.5. Congrès Médicaux

[CM-17] Rodolphe ANTY, **Marie GUYOMARD**, Schiappa RENAUD, Marwin A. FARRUGIA, Cyprien GILET, Dann J OUIZEMAN, Jocelyn GAL, Emmanuel CHAMOREY, Stéphanie PATOURAUX, Louisa VONGHIA, Albert TRAN, Antonio IANNELLI, Sven FRANCQUE, Jérôme BOURSIER, Lionel FILLATRE et Philippe GUAL. « Diagnostic de la MASH fibrosante à partir de paramètres cliniques et biologiques simples sélectionnés et mis en modèles par intelligence artificielle : le score FibMASH ». In : *95e Journées Scientifiques de l'AFEF*. 2024.

[CM-18] **Marie GUYOMARD**, Dann OUIZEMAN, Renaud SCHIAPPA, Cyprien GILET, Jocelyn GAL, Emmanuel CHAMOREY, Stéphanie PATOURAUX, Thierry PICHE, Albert TRAN, Philippe GUAL et al. « Diagnostic non invasif de la NASH fibrosante à l'aide de l'intelligence artificielle ». In : *87e Journées Scientifiques de l'AFEF*. 2020.