UNIVERSITE D'ABOMEY CALAVI
(BENIN)

THE **ABDUS SALAM** INTERNATIONAL
CENTER FOR THEORETICAL
PHYSICS (ITALY)

INSTITUT DE MATHEMATIQUES ET DE SCIENCES PHYSIQUES

Centre d'Excellence Africain en Sciences Mathématiques Informatique et Applications

Adresse e-mail : secretariat@imsp-uac.org

THÈSE DE DOCTORAT

FILIÈRE : TIC

SPÉCIALITÉ : INTELLIGENCE ARTIFICIELLE

ANNÉE ACADÉMIQUE : 2024-2025

Thème

Nez électronique et intelligence artificielle pour la caractérisation sensorielle des tomates locales

Rédigée par :

LIONEL VINCENT DE PAUL O. AFFOGNON

Jury :

Président :

M. MOHAMED GIBIGAYE, PROFESSEUR TITULAIRE, UNIVERSITÉ D'ABOMEY-CALAVI, BÉNIN

Rapporteurs :

M. PIERRE GOUTON, PROFESSEUR TITULAIRE, UNIVERSITÉ BOURGOGNE EUROPE, FRANCE

M. ARNAUD AHOUANDJINOU, MAÎTRE DE CONFÉRENCES, UNIVERSITÉ D'ABOMEY-CALAVI, BÉNIN

M. GAOUSSOU CAMARA, MAÎTRE DE CONFÉRENCES, UNIVERSITÉ ALIOUNE DIOP, SÉNÉGAL

Examinateur :

MME. HÉLÈNE KIRCHNER, PROFESSEURE TITULAIRE, INSTITUT NATIONAL DE RECHERCHE EN SCIENCES ET TECHNOLOGIES DU NUMÉRIQUE, FRANCE

Directeur de thèse :

M. EUGÈNE C. EZIN, PROFESSEUR TITULAIRE, UNIVERSITÉ D'ABOMEY-CALAVI, BÉNIN

Co-directeur de thèse :

M. CHÉRIF DIALLO, PROFESSEUR TITULAIRE, UNIVERSITÉ GASTON BERGER, SÉNÉGAL

Soutenu le 03 avril 2025

DÉDICACE

A

ma maman, **Chantal**

mon papa, **Gervais**

mon frère, **Oswald**

REMERCIEMENTS

Cette page m'a demandé un effort considérable, car je craignais de passer à côté de certains remerciements importants.

Je souhaiterais exprimer ma profonde gratitude et mes remerciements les plus sincères à mon Directeur de thèse, le Professeur Eugène C. EZIN, qui a été pour moi bien plus qu'un simple professeur. Il m'a guidé et soutenu tout au long de ma formation académique, depuis mes années de licence et de master à l'Institut de Formation et de Recherche en Informatique (IFRI) jusqu'à la thèse à l'Institut de Mathématiques et de Sciences Physiques (IMSP). Sa patience, son dévouement et ses conseils avisés ont été inestimables et ont contribué grandement à forger en moi une personnalité. Qu'il trouve ici l'expression de ma filiale gratitude. Merci pour la disponibilité, la gentillesse et les nombreuses relectures.

Ma profonde reconnaissance va également à toute l'administration de l'IMSP, avec une mention spéciale pour le Professeur Carlos OGOUYANDJOU, Directeur de l'Institut, pour son précieux soutien et sa contribution inestimable à la réussite de cette thèse.

Je remercie tout aussi chaleureusement, le Professeur Chérif DIALLO qui a également été un encadrant précieux, me prodiguant ses connaissances et sa bienveillance. Merci pour la disponibilité, la gentillesse et les nombreuses relectures.

J'exprime ma reconnaissance à l'endroit du Professeur Michel DOSSOU pour ses conseils avisés et son soutien durant mon cursus académique. Merci pour la gentillesse et le soutien.

Je suis particulièrement reconnaissant envers le Docteur Hervé B. OLOU, avec qui j'ai partagé tant de choses et qui a été un véritable ami tout au long de ces années de thèse. Merci pour les moments de rires, de succès, de peine, d'entraide et les nombreuses relectures.

Je suis reconnaissant envers mon oncle Serge AVAHOUIN, qui a toujours été un soutien indéfectible pour moi. Merci pour la persévérance.

Je tiens à remercier Faouz BASSABI pour son soutien constant et sa présence en tant que frère aîné. Je suis fier d'avoir eu la chance de croiser son chemin.

Je tiens à exprimer ma sincère gratitude envers mes chers camarades de licence, de master et de thèse, dont le soutien, la collaboration et l'amitié ont été des éléments essentiels de mon

parcours académique. Les idées, les discussions et les encouragements ont enrichi ma réflexion et m'ont aidé à progresser tout au long de ce cheminement. Les heures passées à travailler ensemble, à échanger nos connaissances et à surmonter les défis académiques ont renforcé notre lien et ont contribué à forger des souvenirs inoubliables. Votre amitié a été une source de réconfort et de motivation, et je suis honoré de vous avoir eu comme collègues et amis pendant cette période de ma vie. Merci pour votre soutien indéfectible et pour avoir été des collaborateurs exceptionnels dans cette quête du savoir.

Je voudrais exprimer mes remerciements au groupe OYANA pour les moments de détente passés ensemble. Merci pour ces moments et la solidarité.

Ce travail a été rendu possible grâce au soutien du Digital Science and Technology Network (DSTN) financé par l'Institut de Recherche pour le Développement (IRD) et l'Agence Française de Développement (AFD). Je tiens à exprimer ma sincère reconnaissance envers les différents agents de ces organisations avec lesquels j'ai eu l'opportunité d'échanger. Leur contribution a été précieuse à chaque étape de cette recherche.

Je tiens à remercier avec gratitude YoupiLab pour leur aide essentielle dans la mise en place du nez électronique pour nos expériences.

Tout ceci n'aurait pas été possible sans la bienveillance de mes parents. Qu'ils trouvent ici l'expression de ma profonde gratitude.

SOMMAIRE

Dédicace **ii**

Remerciements **iii**

Sommaire **v**

Résumé **viii**

Abstract **x**

Sigles et abréviations **xi**

Glossaire **xviii**

Liste des figures **xix**

Liste des tableaux **xxii**

Introduction générale **1**

1 Généralités sur l'apprentissage automatique **6**

Introduction 6

1.1 Fondements de l'apprentissage automatique 6

1.2 Catégories de l'apprentissage automatique 7

1.3 Limites de l'apprentissage automatique 21

Conclusion 22

2 Approches et outils du contrôle de qualité **23**

Introduction 23

2.1 Historique et évolution du contrôle de qualité 23

2.2 Typologie des méthodes de contrôle qualité 26

2.3 Contrôle de qualité dans quelques secteurs 29

2.4 Technologies modernes de contrôle de qualité des aliments 34
2.5 Impact environnemental et économique 41
Conclusion 41

3 Nez électronique et traitement de données 42
Introduction 42
3.1 Technologies du nez électronique 43
3.2 Analyse des données du nez électronique 44
3.3 Discussion 60
Conclusion 62

4 Nez électronique expérimental 64
Introduction 64
4.1 Étude des applications du nez électronique 65
4.2 Conception et spécifications techniques du nez électronique 68
4.3 Discussion sur l'optimisation et la validation du dispositif 71
4.4 Protocole expérimentale de collecte des données sensorielles de la tomate 75
Conclusion 84

5 État des tomates : approches supervisées 86
Introduction 86
5.1 Approches d'apprentissage automatique pour les tomates 86
5.2 Random Forest 88
5.3 k-Nearest Neighbors 89
5.4 Multi-Layer Perceptron 89
5.5 Support Vector Machine 90
5.6 Métriques d'évaluation 91
5.7 Matrice de confusion 92
5.8 Pré-traitement et nettoyage des données 92
5.9 Evaluation des performances des modèles 93
5.10 Analyse des résultats et comparaison avec les méthodes existantes 99
Conclusion 99

6 Prédiction de la durée de vie des tomates 101
Introduction 101
6.1 Application des séries temporelles dans l'agriculture 101
6.2 Pré-traitement et nettoyage des données 103
6.3 Long Short-Term Memory 103
6.4 Gated Recurrent Unit 105
6.5 Configuration du modèle 106

6.6 Résultats expérimentaux 108
6.7 Discussion 110
Conclusion 111

Conclusion générale et perspectives **112**

Liste des publications **114**

Bibliographie **140**

Table des matières **141**

RÉSUMÉ

La qualité des produits alimentaires est un enjeu majeur de santé publique, particulièrement dans les régions où les méthodes de conservation et de contrôle sont limitées. Au Bénin, les tomates constituent une denrée essentielle, mais leur conservation pose des défis significatifs. Dans ce contexte, la mise en place d'un nez électronique innovant représente une solution prometteuse pour améliorer la sécurité alimentaire en permettant une évaluation précise et continue de la qualité des tomates. Le nez électronique conçu pour cette étude est composé de huit capteurs de gaz (MQ135, MQ136, TGS822, TGS2600, TGS2602, TGS2603, TGS2610 et TGS2611), ainsi que d'un capteur de température et d'humidité (DHT22). Les tomates locales, représentées par les variétés *Tounvi* et *Akikon*, ont été utilisées pour les besoins de cette étude. Une collecte de données a été réalisée sur une période de plusieurs semaines pour capturer les données volatiles à différents stades de maturation et de détérioration. Les données recueillies ont été utilisées pour entraîner des modèles d'apprentissage supervisé. Parmi les algorithmes testés, tels que les forêts aléatoires (random forest), les k plus proches voisins (k-NN), les perceptrons multicouches (MLP) et les machines à vecteurs de support (SVM), l'algorithme de forêts aléatoires s'est distingué par ses performances supérieures avec une précision de 94% pour la prédiction de l'état des tomates. De plus, des réseaux de neurones récurrents avancés, tels que Long Short-Term Memory (LSTM) et Gated Recurrent Unit (GRU), ont été employés pour analyser les données séquentielles et prédire le nombre de jours restants avant la détérioration totale des tomates. Ces modèles ont montré des performances remarquables, avec des erreurs quadratiques moyennes respectives de 0,000586 pour LSTM et 0,000505 pour GRU. Les résultats obtenus démontrent que le nez électronique, combiné à des modèles d'apprentissage automatique et à des réseaux de neurones récurrents, offre une solution précise et fiable pour prédire la qualité des tomates. Cette approche permet non seulement de déterminer les périodes de consommation sécuritaire, mais aussi de réduire les pertes post-récolte en optimisant les chaînes d'approvisionnement. Ainsi, la mise en place d'un tel système intelligent représente une avancée significative pour la sécurité alimentaire et la valorisation des produits agricoles locaux au Bénin. Cette recherche ouvre également la voie à de futures applications des technologies de détection et de l'intelligence artificielle dans le secteur

agroalimentaire.

Mots clés : Nez électronique, Sécurité alimentaire, Qualité des tomates, Apprentissage automatique, Apprentissage profond.

ABSTRACT

The quality of food products is a major public health concern, particularly in regions where preservation and control methods are limited. In Benin, tomatoes are an essential commodity, but their preservation presents significant challenges. In this context, the implementation of an innovative electronic nose represents a promising solution to improve food safety by enabling precise and continuous evaluation of tomato quality. The designed electronic nose for this study comprises eight gas sensors (MQ135, MQ136, TGS822, TGS2600, TGS2602, TGS2603, TGS2610, and TGS2611), as well as a temperature and humidity sensor (DHT22). Local tomatoes, represented by the varieties *Tounvi* and *Akikon*, were used for the purposes of this study. Data collection was carried out over several weeks to capture volatile compounds at different stages of ripening and deterioration. The collected data were used to train supervised learning models. Among the algorithms tested, such as random forests, k-nearest neighbors (k-NN), multilayer perceptrons (MLP), and support vector machines (SVM), the random forest algorithm stood out with superior performance, achieving an accuracy of 94% for predicting the state of the tomatoes. Additionally, advanced recurrent neural networks, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), were employed to analyze sequential data and predict the number of days remaining before the tomatoes' complete deterioration. These models demonstrated remarkable performance, with mean squared errors of 0.000586 for LSTM and 0.000505 for GRU, respectively. The results obtained demonstrate that the electronic nose, combined with machine learning models and recurrent neural networks, offers a precise and reliable solution for predicting tomato quality. This approach not only determines safe consumption periods but also reduces post-harvest losses by optimizing supply chains. Thus, the implementation of such an intelligent system represents a significant advancement for food safety and the valorization of local agricultural products in Benin. This research also paves the way for future applications of detection technologies and artificial intelligence in the agri-food sector.

Keywords : Electronic nose, Food safety, Tomato quality, Machine learning, Deep learning.

SIGLES ET ABRÉVIATIONS

k-NN

k-Nearest Neighbors. vi, viii, x, 50, 56, 57, 87, 89, 93, 96, 97, 98, 99, 143

AFD

Agence Française de Développement. iv, 1

ANN

Artificial Neural Network. 50, 54, 58, 86

ART

Adaptive Resonance Theory. 57

AS9100

Aerospace Standard 9100. 29

BAW

Bulk Acoustic Wave. 44, 45, 48

BPNN

Back Propagation Neural Network. 58

BRC

British Retail Consortium. 33

BYOL

Bootstrap Your Own Latent. 16

CA

Cluster Analysis. 50, 52

CD

Contrôle Destructif. 26, 28, 142

CND

Contrôle Non Destructif. 26, 28, 142

CNN

Convolutional Neural Network. xix, 19, 20, 21, 59, 101, 102, 111

COV

Composés Organiques Volatils. 46, 66, 68, 69

CP

Conducting Polymer. 46, 49, 66

CP-ANN

Counterpropagation Artificial Neural Network. 58

CSV

Comma-Separated Values. 72, 79, 92

DARTS

Differentiable Architecture Search. 59

DBSCAN

Density-Based Spatial Clustering of Applications with Noise. 10

DEEE

Déchets d'Équipements Électriques et Électroniques. 30

DFA

Discriminant Function Analysis. 44, 52

DL

Deep Learning. 17

DNN

Deep Neural Network. 59

DSTN

Digital Science and Technology Network. iv, 1

EFSIS

Service européen d'inspection de la sécurité des aliments. 33

EMA

European Medicines Agency. 32

EOS

Electronic Olfactory Sensor. 65, 66, 67

FAO

Food and Agriculture Organization. 26, 75

FCC

Federal Communications Commission. 30

FCM

Fuzzy C-means. 57

FDA

Food and Drug Administration. 32

FDA

Fisher Discriminant Analysis. 86, 87

FedAvg

Federated Averaging. 17

FedProx

Federated Proximal. 17

FIS

Fuzzy Inference Systems. 57

FL

Fuzzy Logic. 50, 57

FNN

Fuzzy Neural Network. 57

FST

Fuzzy Set Theory. 57

GC

Gas Chromatography. 44

GFSI

Global Food Safety Initiative. 33

GPL

Gaz de Pétrole Liquéfié. 69

GRNN

Generalized Regression Neural Network. 58

GRU

Gated Recurrent Unit. vi, viii, x, xx, xxi, 3, 5, 19, 20, 102, 105, 106, 107, 108, 109, 110, 111, 112, 144

HACCP

Hazard Analysis and Critical Control Points. 26, 33

HPLC

High-Performance Liquid Chromatography. 34

IA

Intelligence Artificielle. 21, 101

IFRI

Institut de Formation et de Recherche en Informatique. iii

IFS

International Featured Standards. 33

IMSP

Institut de Mathématiques et de Sciences Physiques. iii

IRD

Institut de Recherche pour le Développement. iv, 1

IRM

Imagerie par Résonance Magnétique. 39

ISO

International Organization for Standardization. 25, 29, 30, 32, 33

LCD

Liquid Crystal Display. 69

LDA

Linear Discriminant Analysis. 44, 51

LSSVR

Least-Squares Support Vector Regression. 54

LSTM

Long Short-Term Memory. vi, viii, x, xx, 3, 5, 19, 20, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 143

LVQ

Learning Vector Quantization. 58, 86

MAE

Mean Absolute Error. xx, xxi, 107, 109, 110

MDP

Markov Decision Process. 12

ML

Machine Learning. 6, 17

MLP

Multi-Layer Perceptron. vi, viii, x, xx, 59, 86, 87, 89, 90, 93, 96, 97, 98, 99, 143

MoCo

Momentum Contrast. 16

MOS

Metal-Oxide-Semiconductor. 43, 47, 48, 49, 65, 66, 67, 68

MOSFET

Metal-Oxide-Semiconductor Field-Effect Transistor. 43, 45, 46

MS

Mass Spectrometry. 44

MSE

Mean Squared Error. xx, xxi, 107, 109, 110

NADCAP

National Aerospace and Defense Contractors Accreditation Program. 29

OMS

Organisation Mondiale de la Santé. 26

PARC

Pattern Recognition and Classification. 53

PCA

Principal Component Analysis. 10, 11, 44, 50, 51

PEN

Portable Electronic Nose. 64, 65, 66, 67

pH

Potentiel Hydrogène. 34, 87

PLS

Partial Least Squares. 44, 51, 52

PLS-DA

Partial Least Squares Discriminant Analysis. 52

PLSR

Partial Least Squares Regression. 52

PNN

Probabilistic Neural Network. 58

QMB

Quartz Micro-Balance. 66

RBF

Radial Basis Function. 58, 86, 90

RF

Random Forest. vi, xx, 87, 88, 99, 143

RFID

Radio Frequency Identification. 69, 70, 77

RL

Reinforcement Learning. 12

RNN

Recurrent Neural Network. 19, 20, 21, 102, 103, 105

RoHS

Restriction of Hazardous Substances. 30

SAW

Surface Acoustic Wave. 44, 45, 47, 48, 49, 66

SimCLR

Simple Framework for Contrastive Learning. 16

SLDA

Stepwise Linear Discriminant Analysis. 51

SOM

Self-Organizing map. 58

SPC

Statistical Process Control. 29

SVC

Support Vector Classifier. 86, 87

SVM

Support Vector Machine. vi, viii, x, 3, 50, 54, 87, 90, 91, 93, 96, 97, 98, 99, 112, 143

SVR

Support Vector Regression. 54, 90

TCN

Temporal Convolutional Network. 102

TDNN

Time Delay Neural Network. 58

TQM

Total Quality Management. 24

t-SNE

t-Distributed Stochastic Neighbor Embedding. 10, 11, 12

USDA

United States Department of Agriculture. 33

YSB

Yellow Stem Border. 101

GLOSSAIRE

CE

Marquage européen qui certifie qu'un produit respecte les exigences de sécurité, de santé publique et de protection de l'environnement définies par l'Union européenne. 30

LISTE DES FIGURES

1.1 Principaux types d'apprentissage automatique [46] 15
1.2 Applications des principaux types d'apprentissage automatique [1] 15
1.3 Taxonomie de l'Intelligence Artificielle [2] 18
1.4 Différence entre l'apprentissage profond et l'apprentissage automatique traditionnel [72] 18
1.5 Comparaison des performances de l'apprentissage profond à celles de l'homme [72] 18
1.6 Exemple d'architecture CNN pour la classification d'images [72] 20

2.1 NeOse Advance[1] 36
2.2 Alpha MOS Fox-Gemini[2] 36
2.3 Spironose[3] 37
2.4 Portable Electronic Nose[4] 37
2.5 zNose 4200[5] 37
2.6 Cyranose 320[6] 37
2.7 Langue électronique [137] 38
2.8 TS-5000Z [138] 38
2.9 Alpha MOS ASTREE[7] 39
2.10 Langue électronique de Shanghai Bosin[8] 39
2.11 X-TRACT[9] 40
2.12 Compass optical sorter[10] 40
2.13 X36 Series[11] 40
2.14 FoodScan™ [12] 40

3.1 Structure des réseaux neuronaux artificiels [3] 59
3.2 Production scientifique annuelle de 1993 à 2023 60
3.3 Production scientifique par pays de 1993 à 2023 61
3.4 Pays les plus cités de 1993 à 2023 61

4.1 Circuit du nez électronique expérimental 70
4.2 Nez électronique expérimental 71
4.3 Échantillons pour la validation du dispositif 71
4.4 Exemple de données sur le vide provenant du nez électronique proposé 72
4.5 Échantillon de données sur la tomate provenant du nez électronique proposé 73
4.6 Échantillon de données sur la pomme verte provenant du nez électronique proposé 73
4.7 Échantillon de données sur la banane provenant du nez électronique proposé 74
4.8 Échantillon de données sur l'orange provenant du nez électronique proposé 74
4.9 Échantillon de données sur l'oignon provenant du nez électronique proposé 75
4.10 Part de la production de tomates par région | Moyenne 1994 - 2021 76
4.11 Échantillons de tomates 78
4.12 Données d'échantillon de la variété de tomate *Tounvi* provenant du nez électronique proposé 79
4.13 Données d'échantillon de la variété de tomate *Akikon* provenant du nez électronique proposé 80
4.14 Variation du capteur TGS2603 81
4.15 Variation du capteur TGS2610 81
4.16 Variation du capteur TGS2611 82
4.17 Variation du capteur TGS2600 82
4.18 Variation du capteur TGS2602 82
4.19 Variation du capteur MQ136 83
4.20 Variation du capteur MQ135 83
4.21 Variation du capteur TGS822 83
4.22 Variation de l'humidité 84
4.23 Variation de la température 84

5.1 Random Forest [4] 88
5.2 Structure de base du MLP [5] 90
5.3 Méthodologie pour la prédiction de la qualité des tomates à l'aide de données de nez électronique 94
5.4 Matrices de confusion sur l'ensemble de données normal 97
5.5 Matrices de confusion sur l'ensemble de données sous-échantillonné 98
5.6 Matrices de confusion sur l'ensemble de données sur-échantillonné 98

6.1 Architecture d'une unité LSTM [6] 105
6.2 Architecture d'une unité GRU [7] 106
6.3 Architecture du modèle proposé 108
6.4 Courbes d'apprentissage du modèle basé sur LSTM : MSE et MAE pour l'entraînement et la validation 109

6.5 Courbes d'apprentissage du modèle basé sur GRU : MSE et MAE pour l'entraînement et la validation 109

LISTE DES TABLEAUX

3.1 Types et mécanismes des capteurs de gaz couramment utilisés dans les nez électroniques . . . 45
3.2 Résumé des avantages et inconvénients des types de capteurs de nez électronique . 46
3.3 Modèle de nez électronique expérimental . . . 48
3.4 Modèle de nez électronique commercial . . . 49

4.1 Résumé de certaines applications du nez électronique . . . 67
4.2 Capteurs du nez électronique expérimental . . . 69
4.3 Durée de vie des tomates . . . 78

5.1 Matrice de confusion pour un problème de classification multi-classe . . . 92
5.2 Plages de valeurs des caractéristiques . . . 92
5.3 Critères de classification des tomates . . . 93
5.4 Comparaison des performances sur l'ensemble de données normal . . . 96
5.5 Comparaison des performances sur l'ensemble de données sous-échantillonné . . . 96
5.6 Comparaison des performances sur l'ensemble de données sur-échantillonné . . . 97

6.1 Évaluation des performances des modèles basés sur LSTM et GRU . . . 110

INTRODUCTION GÉNÉRALE

Face aux enjeux cruciaux liés à la qualité des produits alimentaires, il est impératif d'étudier des approches novatrices pour garantir la sécurité des consommateurs, maintenir la confiance dans les produits locaux et préserver les traditions culinaires [1, 2]. Les technologies de contrôle traditionnelles ont connu des succès, mais elles présentent des limites dans l'évaluation objective de la qualité sensorielle subtile des produits, ainsi que dans la prévention des risques pour les testeurs humains. Le "nez électronique" offre une solution prometteuse, mais son potentiel reste largement sous-exploité dans le contexte de la caractérisation des produits locaux [3, 4]. La combinaison de cette technologie avec des techniques de l'intelligence artificielle peut révolutionner la manière de percevoir et de garantir la qualité des produits. Une telle approche répond à un besoin pressant de renforcer la sécurité alimentaire [5]. En outre, elle assure la préservation des traditions locales et consolide la confiance des consommateurs dans les produits du terroir [6, 7]. Cette étude se justifie par le potentiel réel qu'elle offre pour collecter des informations susceptibles d'améliorer la qualité des produits locaux, tout en minimisant les risques pour la santé des testeurs et en renforçant la durabilité de la chaîne d'approvisionnement alimentaire locale.

1- Contexte

C'est dans un cadre similaire à celui décrit ci-dessus, que la technologie du "nez électronique" a émergé comme un outil prometteur largement utilisé pour relever ces défis. Les capacités de cette technologie à évaluer la qualité olfactive des produits alimentaires sans les altérer offrent un potentiel considérable [8]. L'utilisation de capteurs et d'analyseurs électroniques pour détecter les signatures odorantes des produits alimentaires représente une avancée significative [9, 10]. De plus, l'analyse olfactive par un nez électronique permet une évaluation un peu plus objective, éliminant la subjectivité associée aux panels sensoriels humains. Ce progrès technologique contribue également à prévenir les risques pour la santé des testeurs en identifiant rapidement les produits avariés [11, 12].

Cette recherche a été initiée en 2021 à l'initiative du DSTN, financé par l'IRD et l'AFD, démon-

trant ainsi l'engagement de ces institutions à explorer des solutions novatrices visant à améliorer la qualité des produits locaux, en particulier en Afrique. Cette collaboration unique offre une plateforme essentielle pour repenser les méthodes traditionnelles d'évaluation des produits locaux et établir des normes de qualité fiables et objectives. Cette recherche étudie l'intégration du nez électronique avec des techniques de l'intelligence artificielle pour caractériser la qualité des produits locaux, en se concentrant sur un exemple significatif, la tomate, tout en ouvrant la voie à des innovations potentiellement révolutionnaires pour la garantie de la qualité des produits alimentaires locaux.

2- Problématique

La qualité des produits de consommation est essentielle au bien-être, à la confiance des consommateurs et au respect des traditions locales. Pourtant, assurer cette qualité présente des défis considérables, notamment pour les produits alimentaires en Afrique. Souvent dépourvus de chaînes de suivi bien définies, la consommation de ces produits peut comporter des risques pour la santé et l'environnement en l'absence de mécanismes de prévention et de contrôle [13]. Les méthodes conventionnelles, comme les panels sensoriels humains, sont parfois insuffisantes pour évaluer avec précision les caractéristiques sensorielles subtiles, telles que l'arôme des aliments [9, 10]. De plus, ces méthodes peuvent parfois altérer la qualité des produits lors de leur évaluation [14]. Par ailleurs, diverses sources indiquent que les dispositifs commerciaux de nez électronique peuvent être onéreux, avec des prix variant généralement entre 5 000 $ et plus de 100 000 $, selon les modèles et les fonctionnalités proposés [15, 16, 17, 18]. Ces coûts constituent un frein majeur à leur adoption dans les laboratoires à ressources limitées. Face à ces défis, il est crucial d'explorer des solutions technologiques innovantes pour garantir la qualité des produits alimentaires, tout en préservant la santé des consommateurs, la sécurité alimentaire, la traçabilité, les traditions locales et en minimisant l'impact sur l'environnement. Cette situation souligne l'importance d'une approche intégrée, combinant les capacités de l'électronique et de l'intelligence artificielle, pour caractériser avec un peu plus précision les produits locaux en Afrique et en garantir la qualité [8].

3- Objectifs et hypothèses de l'étude

Le principal objectif de cette recherche est de développer un nez électronique dédié à la prédiction de la qualité des produits locaux, en mettant particulièrement l'accent sur la tomate. Afin d'atteindre cet objectif, cette recherche se décline en une série d'objectifs spécifiques, à savoir :

— Concevoir un dispositif électronique pour la collecte de données.
— Mettre en place des modèles d'apprentissage automatique pour analyser les données collectées afin de prédire la qualité de la tomate.
— Tester l'efficacité du nez électronique en utilisant quelques variétés de tomates locales.

Pour atteindre ces objectifs, plusieurs hypothèses ont été formulées :

— Le nez électronique, grâce à ses capteurs, peut détecter des variations subtiles dans les composés volatils émis par les tomates, ce qui permet de différencier les différentes étapes de détérioration.

— Les modèles d'apprentissage automatique, tels que les Support Vector Machine (SVM) et les forêts aléatoires, sont capables d'analyser les données volatiles collectées afin de prédire la qualité des tomates.

— Les réseaux de neurones récurrents notamment Long Short-Term Memory (LSTM) et Gated Recurrent Unit (GRU) capturent les dépendances temporelles dans les données séquentielles pour fournir des prédictions sur l'état des tomates et leur date de détérioration.

— L'utilisation du nez électronique améliore significativement la précision et la fiabilité de la surveillance de la qualité des tomates locales, par rapport aux méthodes traditionnelles de contrôle de qualité.

4- Contributions scientifiques

Les résultats de ces recherches ont été présentés dans diverses publications scientifiques et lors d'événements académiques tels que des conférences, des workshops et des colloques. Les principales contributions découlant de ces travaux sont :

— Affognon, L., Diallo, A., Diallo, C., & Ezin, E. C. (2023). *Supervised Learning Models for Tomato Quality Prediction using Electronic Nose Data*. https://doi.org/10.1109/ICECET58911.2023.10389485. Cet article présente une étude comparative de différents modèles d'apprentissage supervisé en combinaison avec la technologie du nez électronique pour identifier la qualité de la tomate. La classification a été effectuée en fonction des différents états répertoriés pendant la phase de collecte de données. Cette recherche vise à évaluer l'efficacité des modèles d'apprentissage supervisé dans la prédiction de la qualité des tomates à partir de données du nez électronique.

— Affognon, L., Diallo, A., Diallo, C., Ezin, E.C. (2023). *Electronic Nose Architecture for Tomato Data Collection*. In : Arai, K. (eds) Proceedings of the Future Technologies Conference (FTC) 2023, Volume 3. FTC 2023. Lecture Notes in Networks and Systems, vol 815. Springer, Cham. https://doi.org/10.1007/978-3-031-47457-6_9. Cet article présente l'architecture du nez électronique expérimental, en utilisant la tomate comme cas d'étude. La collecte de données, effectué tout au long de la vie d'une tomate, de la récolte à la détérioration, y est détaillé. De plus, le jeu de données correspondant a été rendu publique.

— Affognon, L., Diallo, A., Diallo, C., & Ezin, E. C. (2023). *Design of an Experimental Electronic Nose for Data Collection for Food Quality*. IEEE EUROCON 2023 - 20th International Conference on Smart Technologies, Torino, Italy, 2023, pp. 406-411. https://doi.org/10.1109/EUROCON56442.2023.10199067. Cet article présente la conception et le développement du nez électronique expérimental utilisé au cours de ces travaux. Conçu pour une utilisation

générale, cet article met en exergue sa capacité à réagir aux différents produits locaux, démontrant ainsi sa polyvalence dans la collecte de données.

— Affognon, L., Diallo, A., Diallo, C. & Ezin, E. C. *A Survey on Statistical and Machine Learning Algorithms Used in Electronic Noses for Food Quality Assessment.* SN COMPUT. SCI. **4**, 590 (2023). https://doi.org/10.1007/s42979-023-02052-0. Cet article, présente une revue détaillée et une analyse des algorithmes statistiques et d'apprentissage automatique qui sont utilisés dans le contexte du nez électronique pour l'évaluation de la qualité des aliments. Les opportunités et les défis spécifiques à l'application de cette technologie en Afrique ont été mis en lumière.

5- Organisation du document

Le document est organisé en 6 chapitres :

- Le chapitre 1 offre une introduction aux concepts fondamentaux de l'apprentissage automatique (*machine learning*), une composante essentielle de l'intelligence artificielle. Il présente les principes de base, les types d'apprentissage (supervisé, non supervisé, renforcement), ainsi que les principaux algorithmes utilisés dans le domaine. De plus, une attention particulière est accordée à l'apprentissage profond (*deep learning*). Les fondements de cette approche sont expliqués, suivis d'une présentation des principales architectures de réseaux de neurones et d'exemples d'applications variées pour chaque type d'apprentissage.
- Le chapitre 2 présente les généralités du contrôle de qualité, avec une attention particulière portée aux outils utilisés dans ce domaine, spécialement dans le secteur alimentaire. Il offre une vue d'ensemble des pratiques courantes de contrôle de qualité, en soulignant les technologies et méthodes employées pour garantir la sécurité et la qualité des produits alimentaires. L'accent est mis sur les différentes techniques, depuis les approches traditionnelles jusqu'aux innovations modernes, pour fournir une compréhension complète des mécanismes de contrôle de qualité dans l'industrie agroalimentaire.
- Le chapitre 3 est une contribution scientifique. Il met en lumière la technologie du nez électronique et les techniques de classification qui lui sont associées. Il offre une analyse de cette innovation technologique, en détaillant son fonctionnement et ses applications spécifiques. Il fournit des explications détaillées sur les composants, les capteurs et les algorithmes qui rendent le nez électronique efficace pour l'évaluation de la qualité des produits alimentaires. En examinant de près les méthodes de classification utilisées, ce chapitre illustre la manière dont ces techniques permettent d'analyser et d'interpréter les données collectées par le nez électronique.
- Le chapitre 4 est également une contribution scientifique. Il détaille le processus de mise en place du nez électronique utilisé dans les expériences. Il met en avant les différents capteurs et outils technologiques employés pour développer ce dispositif. Ce chapitre offre

une vue complète des étapes de conception et d'implémentation, soulignant la sélection et l'intégration des composants essentiels. Une fois le nez électronique opérationnel, diverses collectes de données ont été réalisées pour valider l'efficacité du dispositif. Par la suite, une collecte de données spécifiques sur les tomates locales a été effectuée et rendue publique, démontrant la robustesse et la fiabilité de notre approche technologique.

- Le chapitre 5 est une contribution scientifique non négligeable qui s'ajoute aux précédentes. Il exploite les données obtenues grâce au nez électronique et utilise des algorithmes d'apprentissage supervisé pour prédire l'état des tomates. En s'appuyant sur les données collectées, ce chapitre démontre la manière dont les modèles de machine learning peuvent analyser les profils volatils des fruits pour anticiper leur état de fraîcheur et de détérioration. Cela illustre l'application concrète des techniques d'intelligence artificielle dans l'évaluation et la surveillance de la qualité des produits agricoles, en mettant en avant les prédictions un peu plus précises obtenues grâce à l'interaction entre le nez électronique et les algorithmes.
- Le chapitre 6 est une dernière contribution scientifique. Il utilise les données collectées par le nez électronique et applique un modèle basé sur les réseaux de neurones récurrents (LSTM ou GRU) pour prédire les jours restants avant la détérioration des tomates. Grâce à ce modèle avancé, des prédictions un peu plus précises ont été réalisées afin d'évaluer la performance et la fiabilité de l'approche. Ce chapitre met en évidence la capacité des modèles LSTM et GRU à anticiper la durée de conservation des tomates, offrant ainsi un outil précieux pour la gestion et la planification de la consommation des produits agricoles.

CHAPITRE 1

GÉNÉRALITÉS SUR L'APPRENTISSAGE AUTOMATIQUE

Introduction

L'apprentissage automatique ou *Machine Learning*, est une branche de l'intelligence artificielle qui permet aux systèmes informatiques d'apprendre et de s'améliorer à partir de données sans être explicitement programmés. Cet outil puissant joue un rôle clé dans divers domaines, tels que la vision par ordinateur, la reconnaissance vocale, et la bioinformatique [19].
Ce chapitre explore les concepts fondamentaux de l'apprentissage automatique, ses catégories principales, les méthodes couramment utilisées, les étapes de développement, ainsi que ses défis et perspectives futures.

1.1 Fondements de l'apprentissage automatique

L'apprentissage automatique repose sur des principes mathématiques, statistiques, et informatiques. Les concepts clés incluent : les bases mathématiques et statistiques, les bases informatiques et les paradigmes de modélisation.

1.1.1 Les bases mathématiques et statistiques

— **Probabilité et statistique :** Ces outils sont essentiels pour modéliser l'incertitude et analyser les données. Par exemple, les modèles bayésiens exploitent les probabilités conditionnelles pour faire des prédictions [20].
— **Algèbre linéaire :** L'algèbre linéaire est utilisée pour représenter et manipuler les données, comme les matrices et les vecteurs, qui sont omniprésents dans des modèles comme les réseaux neuronaux [19].
— **Optimisation :** L'entraînement des modèles repose souvent sur des techniques d'optimisation pour minimiser une fonction de coût, comme la descente de gradient [21].

1.1.2 Les bases informatiques

— **Structures de données et algorithmes :** Une compréhension approfondie des structures comme les arbres, les graphes, et les tables de hachage est cruciale pour concevoir des algorithmes efficaces [22].

— **Complexité computationnelle :** La faisabilité des algorithmes dépend de leur efficacité en termes de temps et d'espace, surtout avec les grands ensembles de données [23].

1.1.3 Les paradigmes de modélisation

— **Modélisation paramétrique vs non paramétrique :** Les modèles paramétriques, comme la régression logistique, supposent une forme spécifique pour la fonction, tandis que les modèles non paramétriques, comme les forêts aléatoires, sont plus flexibles [20].

— **Apprentissage en ligne vs par lot :** L'apprentissage en ligne met à jour les modèles au fur et à mesure que de nouvelles données arrivent, contrairement à l'apprentissage par lot, qui entraîne les modèles sur des ensembles de données fixes [24].

Ces bases constituent le socle sur lequel reposent les différentes catégories et algorithmes d'apprentissage automatique, explorés dans les sections suivantes.

1.2 Catégories de l'apprentissage automatique

L'apprentissage automatique repose sur trois grandes catégories de paradigmes qui constituent ses fondements théoriques et pratiques : l'apprentissage supervisé, non supervisé et par renforcement. Ces paradigmes se distinguent par la manière dont ils exploitent les données et leurs objectifs respectifs. Les sous-sections 1.2.1, 1.2.2 et 1.2.3 se concentrent sur ces trois paradigmes majeurs, qui constituent les fondements nécessaires pour comprendre et explorer les techniques modernes d'apprentissage automatique. Les Figures 1.1 et 1.2, fournissent une vue d'ensemble synthétique des concepts, méthodes et applications abordés dans ces sous-sections. Elles servent à éclairer les relations entre ces paradigmes et à illustrer leurs caractéristiques distinctives et leurs points communs.

1.2.1 Apprentissage supervisé

L'apprentissage supervisé est une sous-catégorie de l'apprentissage automatique dans laquelle un modèle est entraîné sur un ensemble de données étiquetées. Ce type d'apprentissage repose sur la relation entre les entrées et les sorties pour effectuer des prédictions ou des classifications sur de nouvelles données. Il est largement utilisé dans divers domaines, notamment la vision par ordinateur, la reconnaissance vocale et le traitement du langage naturel [19].

Principe de fonctionnement

L'objectif principal de l'apprentissage supervisé est de construire une fonction $f : X \to Y$, où X représente l'ensemble des caractéristiques d'entrée et Y les étiquettes correspondantes. Pendant la phase d'entraînement, un ensemble de données d'entraînement $\mathscr{D} = \{(x_i, y_i)\}_{i=1}^{n}$ est utilisé, où chaque $x_i \in X$ et $y_i \in Y$ représentent une paire entrée-sortie. Le modèle est optimisé pour minimiser une fonction de perte $\mathscr{L}(f(x), y)$, qui mesure la différence entre les prédictions $f(x)$ et les véritables étiquettes y [19].

Méthodes courantes

L'apprentissage supervisé peut être subdivisé en deux grandes catégories :

— **Régression :** où les sorties Y sont continues. Par exemple, la prédiction du prix d'un bien immobilier sur la base de ses caractéristiques [25].

— **Classification :** où les sorties Y appartiennent à un ensemble fini de classes. Par exemple, la reconnaissance des chiffres manuscrits [26].

Étapes de l'apprentissage supervisé

L'apprentissage supervisé repose sur une série d'étapes méthodologiques bien définies, allant de la collecte des données étiquetées jusqu'au déploiement du modèle. Ces étapes sont essentielles pour garantir la qualité des données, optimiser les performances du modèle et assurer sa généralisation sur des données non vues, comme détaillé ci-dessous.

1. **Collecte des données étiquetées :** Les données doivent être représentatives du problème à résoudre. Elles sont recueillies et chaque observation est associée à une étiquette ou une sortie cible, ce qui permet au modèle d'apprendre la relation entre les caractéristiques d'entrée et les résultats attendus [27]. Par exemple, pour une tâche de classification d'images, les données doivent contenir une variété d'images couvrant toutes les classes cibles.
2. **Prétraitement :** Les données brutes sont souvent transformées pour améliorer leur qualité, par des étapes comme la normalisation, l'encodage des catégories ou la gestion des valeurs manquantes [27].
3. **Séparation des ensembles :** Les données sont généralement divisées en trois ensembles : un ensemble d'entraînement (pour entraîner le modèle), un ensemble de validation (pour ajuster les hyperparamètres), et un ensemble de test (pour évaluer la performance du modèle après l'entraînement) [27].
4. **Interprétation des résultats**
   - **Entraînement du modèle :** L'entraînement consiste à ajuster les paramètres du modèle pour minimiser une fonction de perte, telle que l'erreur quadratique moyenne pour les problèmes de régression ou l'entropie croisée pour la classification [27].

- **Validation et ajustement des hyperparamètres :** Le modèle est évalué sur l'ensemble de validation pour ajuster les hyperparamètres (comme le taux d'apprentissage, la taille des mini-batchs, etc.) afin d'optimiser ses performances [27].
- **Évaluation sur l'ensemble de test :** Une fois l'entraînement terminé, les performances du modèle sont évaluées sur un ensemble de test pour vérifier sa capacité à généraliser sur des données non vues et obtenir une estimation réaliste de sa performance dans un contexte réel [27].

5. **Déploiement et surveillance :** Le modèle peut être déployé dans un environnement de production, où ses performances peuvent être surveillées et ajustées en fonction des nouvelles données ou de l'évolution des conditions.

Algorithmes courants

Plusieurs algorithmes sont utilisés dans l'apprentissage supervisé, notamment :

— **Régression linéaire :** Utilisée pour prédire une variable continue à partir d'une ou plusieurs variables explicatives [21]. Par exemple prédire une note à un examen en fonction du nombre d'heures de révisions.
— **Régression logistique :** Adaptée aux problèmes de classification binaire, elle prédit la probabilité qu'une observation appartienne à une classe donnée [20].
— **Machines à vecteurs de support :** Conçues pour maximiser la marge entre différentes classes dans les données [22].
— **Arbres de décision :** Algorithmes simples et interprétables qui partitionnent les données en fonction de critères successifs [23].
— **Réseaux neuronaux artificiels :** Capables de capturer des relations complexes entre les entrées et les sorties grâce à une architecture multi-couches [24].

Applications

L'apprentissage supervisé est utilisé dans divers domaines :

— **Classification d'images :** Par exemple, dans les systèmes de diagnostic médical pour identifier des anomalies sur des radiographies [28].
— **Reconnaissance vocale :** Comme dans les assistants virtuels tels que Siri et Google Assistant.
— **Prédiction financière :** Pour anticiper les variations de prix des actions à partir des données historiques [29].
— **Détection de fraude :** Pour identifier des transactions suspectes dans les systèmes bancaires [21].

Limites et défis

Malgré son succès, l'apprentissage supervisé présente des défis importants :

— **Dépendance aux données étiquetées :** La collecte et l'annotation des données peuvent être coûteuses et chronophages.
— **Surapprentissage :** Les modèles complexes risquent de s'adapter trop étroitement aux données d'entraînement, ce qui diminue leur capacité de généralisation [19].

1.2.2 Apprentissage non supervisé

L'apprentissage non supervisé est une approche de l'apprentissage automatique dans laquelle un algorithme analyse des données sans étiquettes préalables, cherchant à identifier des structures ou des motifs sous-jacents. Contrairement à l'apprentissage supervisé, où un modèle apprend à partir d'exemples étiquetés, l'objectif principal ici est de découvrir des regroupements ou des relations latentes au sein des données [30].

Concepts de base

L'apprentissage non supervisé repose sur la capacité d'un modèle à organiser les données en catégories significatives en fonction de leurs caractéristiques intrinsèques. Les deux tâches les plus courantes dans ce domaine sont :

— **Le clustering** : qui consiste à regrouper les données en sous-ensembles homogènes, appelés clusters. Les algorithmes comme *k-means* [31], *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)* [32] et les méthodes hiérarchiques [33] sont parmi les plus utilisés.
— **La réduction de dimensionnalité** : qui vise à projeter les données dans un espace de plus petite dimension tout en conservant l'essentiel de l'information. Les techniques courantes incluent l'Analyse en Composantes Principales (PCA : Principal Component Analysis) [34] et le *t-Distributed Stochastic Neighbor Embedding (t-SNE)* [35].

Clustering

Le clustering est une méthode qui regroupe les observations en fonction de leur similarité. Les algorithmes de clustering sont largement utilisés dans des domaines variés comme la biologie, le marketing et la vision par ordinateur. Les principales catégories de méthodes de clustering incluent :

— **Clustering partitionnel :** Divise les données en un nombre prédéfini de clusters. *k*-means, par exemple, minimise la variance intra-cluster [31].
— **Clustering basé sur la densité** : Identifie des clusters de formes arbitraires en utilisant des mesures de densité. DBSCAN est un exemple clé [32].
— **Clustering hiérarchique :** Construit une hiérarchie de clusters imbriqués, souvent représentée sous forme de dendrogramme [33].

Réduction de dimensionnalité

La réduction de dimensionnalité est essentielle lorsque les données possèdent un grand nombre de variables, ce qui peut rendre les analyses complexes et coûteuses en termes de calcul. Parmi les techniques les plus utilisées :

— **PCA :** Une méthode linéaire qui projette les données sur des axes orthogonaux maximisant la variance [34].

— **t-SNE :** Une méthode non linéaire particulièrement adaptée à la visualisation de données complexes [35].

Étapes de l'apprentissage non supervisé

1. **Collecte des données :** Les données sont collectées sans étiquettes. L'objectif est de trouver des structures cachées ou des relations dans ces données.
2. **Prétraitement des données :** Comme dans l'apprentissage supervisé, les données brutes sont souvent transformées pour améliorer leur qualité. Cela peut inclure la normalisation, la réduction de dimensionnalité ou la gestion des valeurs manquantes.
3. **Choix du modèle :** Un modèle d'apprentissage non supervisé est choisi en fonction du type de structure recherché, comme le clustering ou la réduction de dimensionnalité.
4. **Apprentissage du modèle :** Le modèle apprend à partir des données en identifiant des structures, des patterns ou des groupes.
5. **Évaluation du modèle :** L'évaluation en apprentissage non supervisé peut être plus complexe, car il n'y a pas de labels. On peut utiliser des mesures comme la cohésion intra-cluster, la séparation inter-cluster ou des techniques comme la validation croisée.
6. **Interprétation des résultats :** Les résultats de l'apprentissage non supervisé doivent être interprétés pour comprendre la structure des données. Cela peut inclure la visualisation des clusters ou l'analyse des principales composantes.
7. **Déploiement et surveillance :** Le modèle peut être implémenté dans un environnement de production, avec une surveillance continue de ses performances et des ajustements effectués en fonction des nouvelles données ou des changements de conditions.

Applications courantes

L'apprentissage non supervisé est largement utilisé dans de nombreux domaines, notamment :

— **Analyse de segmentation de marché :** Identifier des groupes de clients avec des comportements similaires pour un ciblage marketing efficace [36, 37].

— **Détection d'anomalies :** Repérer des comportements inhabituels, par exemple, des transactions frauduleuses ou des pannes dans des systèmes techniques [23].

— **Recommandations :** Découvrir des relations implicites dans les données des utilisateurs pour générer des recommandations personnalisées [36].

— **Compression des données :** Réduire la taille des données pour les stocker ou les transmettre plus efficacement [19].

Défis spécifiques

— **Évaluation des résultats :** Contrairement à l'apprentissage supervisé, il est difficile d'évaluer directement les performances car il n'existe pas de "vérité terrain" pour valider les résultats [38].
— **Sensibilité aux paramètres :** Les algorithmes comme k-means ou t-SNE nécessitent de choisir des paramètres critiques (comme k, perplexité), ce qui peut influencer fortement les résultats [33].
— **Scalabilité :** Gérer des ensembles de données massifs reste un défi pour de nombreux algorithmes [32].

1.2.3 Apprentissage par renforcement

L'apprentissage par renforcement (RL) est une branche de l'intelligence artificielle qui traite de la manière dont un agent peut apprendre à prendre des décisions optimales en interagissant avec un environnement dynamique. Contrairement à l'apprentissage supervisé, où un modèle est formé à partir de données étiquetées, l'apprentissage par renforcement repose sur un système d'essais et erreurs pour maximiser une récompense cumulative. Cette approche est inspirée des mécanismes d'apprentissage comportementaux observés dans la nature [39].

Concepts de base

Les éléments fondamentaux de l'apprentissage par renforcement incluent :
— **Agent :** L'entité qui prend des actions dans l'environnement.
— **Environnement :** Le système dans lequel évolue l'agent et qui répond à ses actions.
— **État :** Une représentation des informations actuelles de l'environnement.
— **Action :** Une décision ou un choix effectué par l'agent.
— **Récompense :** Un retour quantitatif reçu après l'exécution d'une action, servant de signal pour guider l'apprentissage.
— **Politique (Policy) :** Une stratégie qui détermine les actions à entreprendre en fonction de l'état actuel.
— **Valeur (Value) :** Une estimation du retour futur attendu à partir d'un état ou d'une action donnée.

Cadre mathématique

Le cadre mathématique de l'apprentissage par renforcement est souvent formalisé à l'aide des processus de décision markoviens (MDP) [40]. Un MDP est défini par un 5-uplet (S, A, P, R, γ), où :

— S est l'ensemble des états possibles.
— A est l'ensemble des actions possibles.
— $P(s'|s,a)$ est la probabilité de transition vers l'état s' en prenant l'action a à partir de l'état s.
— $R(s,a)$ est la récompense immédiate associée à l'état s et à l'action a.
— $\gamma \in [0,1]$ est le facteur d'actualisation, qui pondère l'importance des récompenses futures.

Méthodes et algorithmes

Les approches d'apprentissage par renforcement peuvent être divisées en plusieurs catégories principales :

— **Méthodes basées sur la valeur :** Elles cherchent à estimer la valeur des états ou des paires état-action. L'algorithme *Q-learning* [39] est un exemple emblématique.
— **Méthodes basées sur la politique :** Ces méthodes optimisent directement la politique sans passer par une estimation explicite des valeurs. Les algorithmes comme *Policy Gradient* [41] en sont des exemples.
— **Méthodes hybrides :** Elles combinent les approches basées sur la valeur et celles basées sur la politique, comme dans l'algorithme *Actor-Critic* [42].

Étapes de l'apprentissage par renforcement

1. **Définition de l'environnement :** L'environnement est défini avec ses états possibles, les actions disponibles, et les récompenses ou pénalités associées à chaque transition d'état.
2. **Définition de l'agent :** L'agent est le système qui interagit avec l'environnement, en prenant des décisions (actions) en fonction des états observés et recevant des récompenses ou pénalités.
3. **Choix d'une politique :** L'agent suit une politique qui détermine l'action à entreprendre pour chaque état. Cette politique peut être déterministe ou stochastique.
4. **Exploration vs exploitation :** L'agent choisit entre explorer de nouvelles actions (exploration) ou exploiter les actions qui ont déjà donné les meilleures récompenses (exploitation).
5. **Apprentissage de la fonction de valeur ou de la politique :** L'agent apprend une fonction de valeur qui évalue la récompense attendue pour chaque état ou il apprend directement une politique optimale pour chaque état à l'aide d'algorithmes comme le *Q-learning*.
6. **Mise à jour des connaissances :** Après chaque interaction avec l'environnement, l'agent met à jour sa politique ou sa fonction de valeur en fonction des récompenses reçues, en utilisant des techniques comme la mise à jour de Bellman ou les gradients.
7. **Évaluation de la performance :** L'agent est évalué pour s'assurer que la politique ou la fonction de valeur apprise permet de maximiser la récompense à long terme. Cela peut inclure la mesure de la récompense cumulative ou le suivi des progrès de l'agent au fil du temps.

8. **Déploiement et amélioration continue :** Une fois l'agent entraîné, il peut être déployé dans un environnement réel pour prendre des décisions autonomes. L'agent peut continuer à apprendre et à s'améliorer en interagissant avec l'environnement.

Applications

L'apprentissage par renforcement a connu un succès retentissant dans divers domaines :

— **Jeux :** AlphaGo [43] a démontré des performances surhumaines dans le jeu de Go.

— **Robotique :** L'apprentissage par renforcement est utilisé pour apprendre des comportements complexes à des robots, tels que la marche et la manipulation d'objets.

— **Optimisation :** Des applications incluent l'optimisation des réseaux de communication ou la gestion de l'énergie.

Défis

Les défis les plus courants de l'apprentissage par renforcement, regroupent :

— **Exploration vs exploitation :** Le compromis entre exploration (essayer de nouvelles actions) et exploitation (utiliser les actions connues pour maximiser les récompenses) est un problème classique en apprentissage par renforcement, où un mauvais équilibre peut nuire à la performance de l'agent [39].

— **Efficacité de l'échantillonnage :** L'apprentissage par renforcement nécessite souvent un grand nombre d'interactions avec l'environnement pour apprendre des politiques efficaces, ce qui peut être coûteux en temps et en ressources. Ce problème est particulièrement aigu dans les environnements complexes ou en temps réel [44].

— **Retard dans la récompense :** Dans de nombreux environnements, les actions d'un agent peuvent avoir des effets différés sur les récompenses. Cela complique l'attribution des récompenses et rend l'apprentissage plus difficile, en particulier pour les tâches à long terme [39].

— **Espace d'état et d'action élevé :** Dans des environnements avec un grand nombre d'états et d'actions, l'apprentissage devient complexe et nécessite une grande capacité de calcul. Des techniques comme la réduction de dimensionnalité ou les approches basées sur les modèles peuvent être nécessaires, mais elles augmentent la difficulté de l'apprentissage [45].

La Figure 1.2 présente les différentes applications associées aux principales catégories d'apprentissage automatique. La Figure 1.1 montre les approches principales, telles que la classification et la régression pour l'apprentissage supervisé, le regroupement pour l'apprentissage non supervisé, ainsi que l'apprentissage par renforcement. Les points et triangles colorés représentent les données d'apprentissage, tandis que les étoiles jaunes indiquent les prédictions possibles.

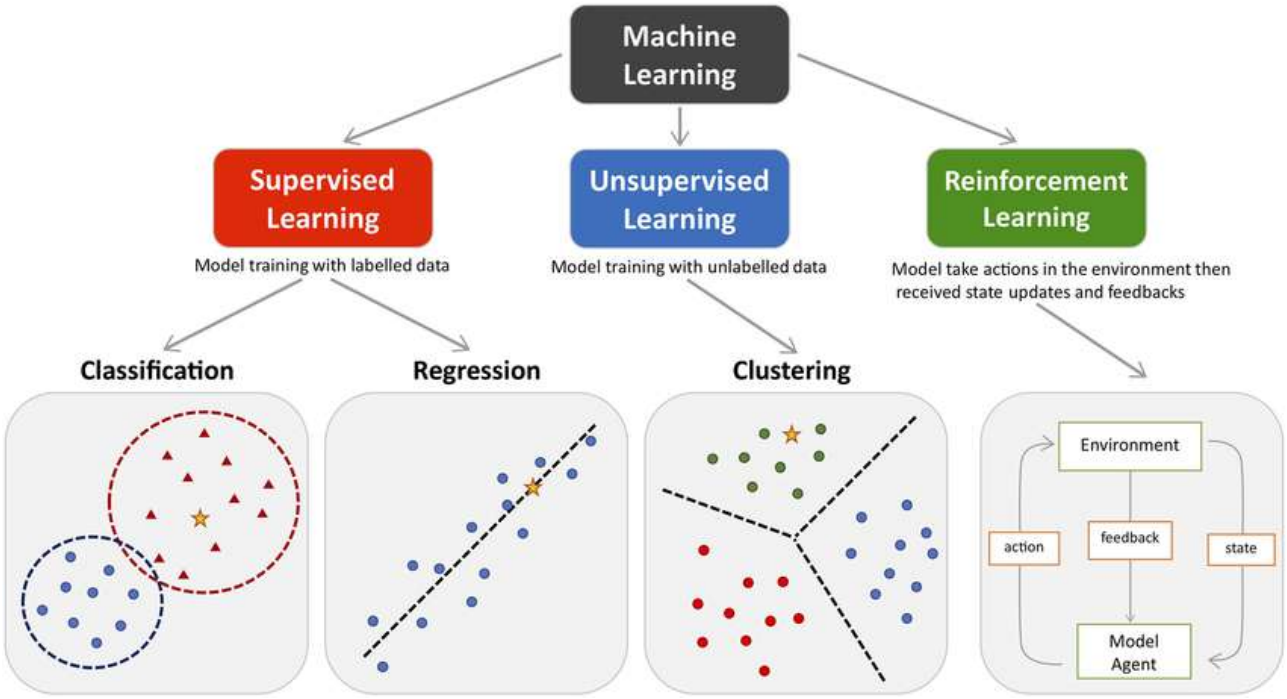


FIGURE 1.1 – Principaux types d'apprentissage automatique [46]

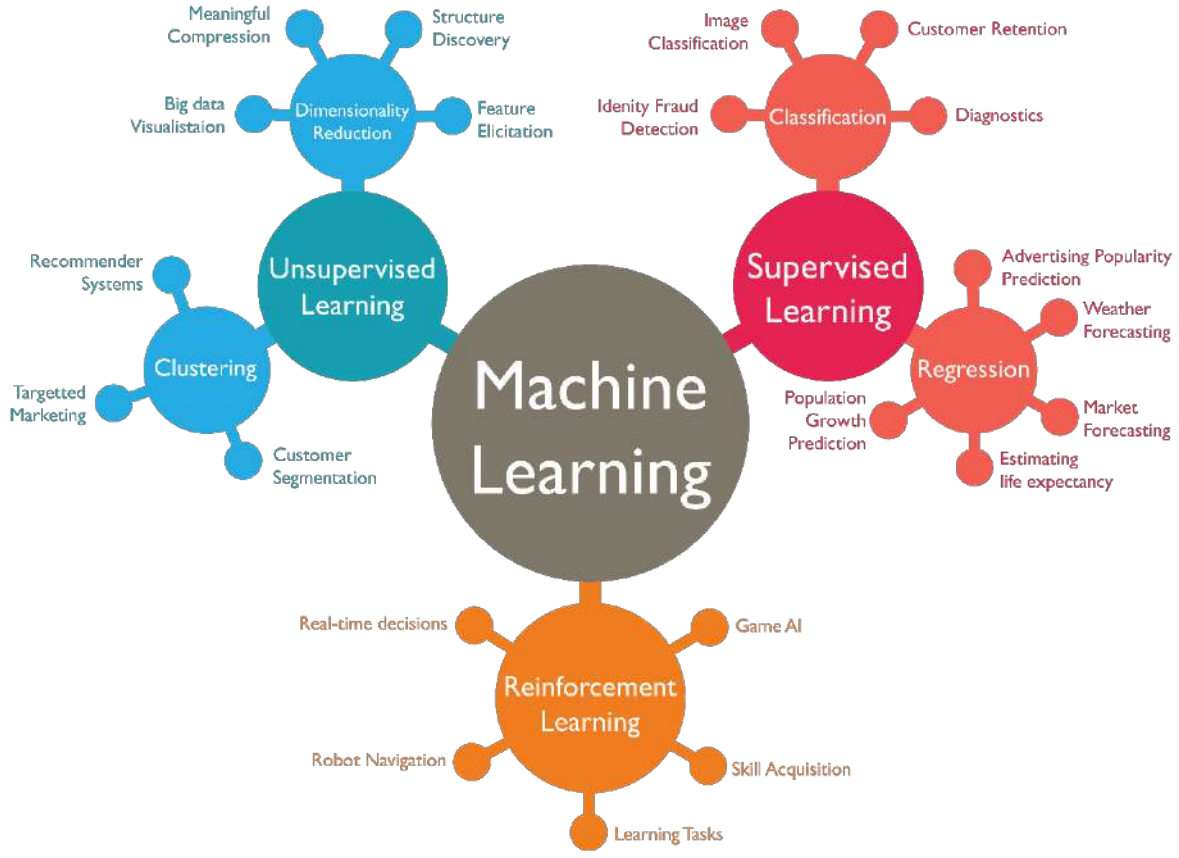


FIGURE 1.2 – Applications des principaux types d'apprentissage automatique [1]

1. Source : `https://blog.clevy.io/fr/le-machine-learning-decrypte-3-3-types-dapprentissage-et-limites`, consultée le 04 décembre 2024.

Outre les paradigmes majeurs, l'apprentissage automatique propose également des approches complémentaires telles que l'apprentissage semi-supervisé, auto-supervisé et fédéré. Bien que moins universelles, ces méthodes répondent à des besoins spécifiques, comme l'optimisation des données étiquetées, l'automatisation de la génération des labels ou la préservation de la confidentialité des données. Ces approches émergentes, souvent issues des paradigmes principaux, illustrent la richesse et l'évolution constante du domaine. Leur présentation permettra de compléter le panorama des méthodes disponibles.

1.2.4 Apprentissage semi-supervisé

L'apprentissage semi-supervisé constitue une solution intermédiaire entre l'apprentissage supervisé et non supervisé. Il exploite un ensemble de données partiellement étiquetées pour entraîner des modèles, permettant ainsi de réduire les coûts liés à l'étiquetage manuel tout en améliorant la performance du modèle.

Principe : L'objectif est de tirer parti des données non étiquetées en les combinant avec un petit ensemble de données étiquetées. Les techniques courantes incluent les méthodes basées sur le *self-training*, où un modèle supervise son propre entraînement, et les méthodes *graph-based*, qui exploitent les relations structurelles des données [47].

Algorithmes courants : Semi-Supervised GANs [48], FixMatch [49], Mean Teacher [50].

Applications : Ce paradigme est particulièrement utile dans les domaines où l'acquisition de labels est coûteuse, comme la reconnaissance d'images médicales [51], le traitement du langage naturel et la détection d'anomalies.

Limites : Une limitation réside dans la sensibilité à la qualité des données non étiquetées, qui peuvent introduire du bruit si elles contiennent des informations incorrectes ou ambiguës.

1.2.5 Apprentissage auto-supervisé

L'apprentissage auto-supervisé est une approche émergente qui permet à un modèle de générer ses propres labels à partir des données brutes, supprimant ainsi le besoin d'un étiquetage manuel.

Principe : Cette méthode repose sur la conception de tâches prétextes, où le modèle est entraîné sur une tâche auxiliaire avant d'être transféré sur la tâche principale. Par exemple, en vision par ordinateur, on peut entraîner un modèle à prédire l'orientation d'une image ou à compléter des parties manquantes [52, 53].

Algorithmes courants : SimCLR [54], BYOL [55], MoCo [56].

Applications : L'apprentissage auto-supervisé est particulièrement adapté à des contextes avec de grandes quantités de données non étiquetées, comme la reconnaissance d'images, le traitement d'images satellites et l'analyse génomique.

Limites : Les tâches prétextes doivent être soigneusement conçues pour capturer des représentations utiles. Un mauvais choix peut limiter les performances sur la tâche principale.

1.2.6 Apprentissage fédéré

L'apprentissage fédéré permet de construire des modèles globaux à partir de données réparties sur plusieurs appareils ou organisations, tout en préservant la confidentialité des données [57].

Principe : Au lieu de centraliser les données, chaque appareil ou noeud d'un réseau entraîne localement un modèle, et seul un résumé des poids ou gradients est partagé pour l'agrégation au niveau global.

Algorithmes courants : FedAvg [58], FedProx [59], Split Learning [60].

Applications : Ce paradigme est couramment utilisé dans les secteurs sensibles, comme les soins de santé pour le diagnostic médical, les applications mobiles pour la personnalisation, et les domaines financiers [61].

Limites : L'apprentissage fédéré fait face à des défis tels que l'hétérogénéité des données entre les nœuds, les contraintes computationnelles locales et les questions de sécurité liées aux attaques par inférence.

1.2.7 Apprentissage profond

L'apprentissage profond ou *Deep Learning*, est un sous-ensemble de l'apprentissage automatique (voir Figure 1.3) qui utilise des modèles hiérarchiques appelés réseaux de neurones profonds. Ces modèles, inspirés de la structure du cerveau humain, sont particulièrement efficaces pour traiter de grandes quantités de données complexes et non structurées, telles que des images, du texte ou des signaux audio [21]. L'apprentissage profond permet de réaliser l'apprentissage et la classification en une seule fois contrairement aux algorithmes présentés jusque là (voir Figure 1.4). En outre, il est devenu un algorithme de *machine learning* extrêmement populaire ces dernières années, grâce à l'essor et à l'évolution significative du domaine du *big data* [62, 19]. Il est toujours en développement continu en ce qui concerne les nouvelles performances pour plusieurs tâches de ML [63, 64, 65, 66] et a simplifié l'amélioration de nombreux domaines d'apprentissage [67, 68], tels que la détection d'objets [69, 70] et la reconnaissance d'images [65, 71]. Récemment, les performances des DL ont dépassé les performances humaines dans des tâches telles que la classification d'images (voir Figure 1.5).

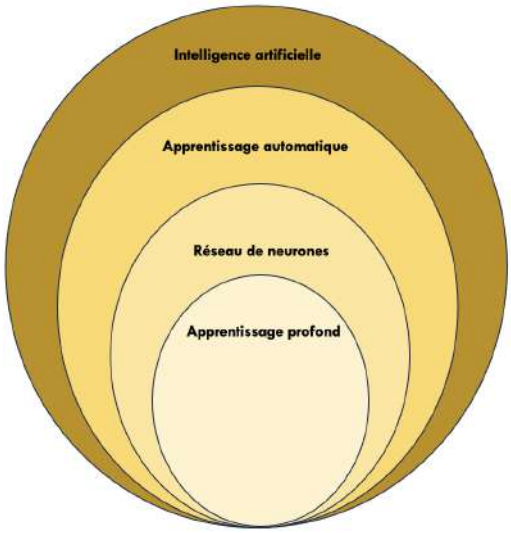


FIGURE 1.3 – Taxonomie de l'Intelligence Artificielle [2]

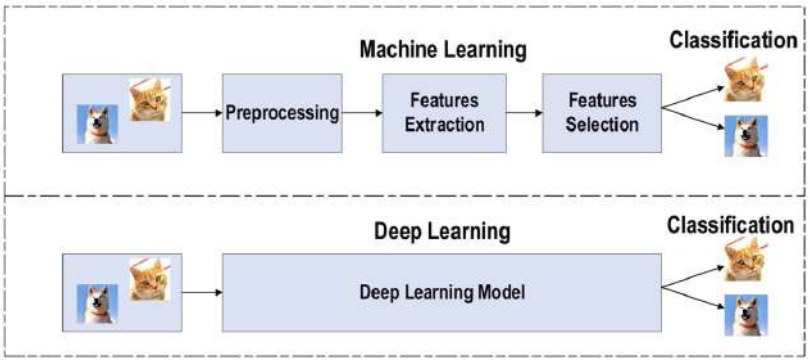


FIGURE 1.4 – Différence entre l'apprentissage profond et l'apprentissage automatique traditionnel [72]

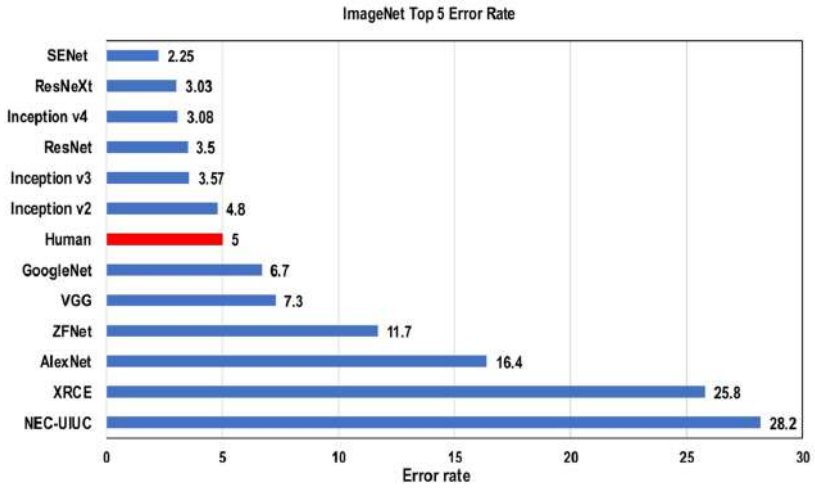


FIGURE 1.5 – Comparaison des performances de l'apprentissage profond à celles de l'homme [72]

2. Inspirée de : `https://datascientest.com/intelligence-artificielle-definition`, consultée le 30 décembre 2024.

Principe de fonctionnement

L'apprentissage profond repose sur des couches multiples de neurones artificiels interconnectés. Chaque couche extrait des caractéristiques de plus en plus abstraites des données d'entrée, permettant aux modèles de capturer des relations complexes. Les réseaux les plus courants incluent les Recurrent Neural Network (RNN) et les Convolutional Neural Network (CNN).

Recurrent Neural Network

Les RNN sont une classe particulière de réseaux neuronaux conçus pour traiter des séquences de données. Contrairement aux réseaux neuronaux traditionnels, qui supposent des entrées indépendantes, les réseaux récurrents utilisent des connexions récurrentes pour intégrer l'historique des données. Cette caractéristique les rend adaptés aux problèmes temporels ou séquentiels [73, 74, 75].
Les RNN traitent les données en chaînes temporelles, où chaque entrée dépend non seulement des données actuelles mais aussi de celles qui ont été traitées précédemment. Cela permet aux RNN de mémoriser des informations contextuelles importantes à chaque étape du traitement.

Structure et fonctionnement : Un RNN est constitué d'un ensemble de neurones, où chaque neurone est connecté non seulement aux autres neurones dans le même niveau, mais aussi à lui-même à travers un état caché (*hidden state*). Cela signifie que l'information traitée à une étape donnée influence les calculs de l'étape suivante [76].

1. **Propagation dans le temps :** Les RNN traitent les données en itérant à travers la séquence. À chaque itération, l'état caché du réseau est mis à jour en fonction de l'entrée actuelle et de l'état caché précédent.
2. **Mémoire de long terme :** Les RNN traditionnels ont des difficultés à se souvenir de l'information sur des longues périodes, ce qui est connu sous le nom de problème du gradient qui disparaît (*vanishing gradient*). Pour résoudre ce problème, des variantes comme les Long Short-Term Memory (LSTM) et les Gated Recurrent Unit (GRU) ont été introduites, permettant aux réseaux de mieux capturer des dépendances à long terme [76, 77].

Applications : Les RNN et leurs variantes sont particulièrement utilisés dans des domaines où les relations séquentielles sont cruciales, tels que :

— **Traitement du langage naturel :** Modélisation des traductions automatiques, analyse de sentiments, et génération de texte [78].
— **Analyse des séries temporelles :** Prédiction de séries temporelles financières, analyse des stocks, prévision météo [79].
— **Reconnaissance vocale :** Traitement du signal audio dans des systèmes comme les assistants vocaux [80].

Limites : Malgré leur capacité à traiter des séquences, les RNN classiques ont une portée limitée en raison du problème du gradient qui disparaît, rendant difficile l'apprentissage de dépen-

dances à long terme. Les architectures LSTM et GRU sont utilisées pour pallier ce problème, mais elles restent coûteuses en termes de calcul [81].

Convolutional Neural Network

Les CNN sont une classe de réseaux neuronaux spécialement conçus pour traiter des données sous forme de grilles, comme des images ou des vidéos, où les pixels sont disposés en matrices [26]. Contrairement aux RNN, qui traitent des séquences de données, les CNN exploitent la spatialité des données pour extraire des caractéristiques locales et hiérarchiques.
Le concept clé derrière un CNN est la convolution, une opération mathématique qui permet au réseau d'appliquer des filtres (ou noyaux) à des régions locales de l'image pour en extraire des caractéristiques telles que des bords, des textures ou des formes [26].

Structure et fonctionnement : Un CNN se compose généralement de plusieurs couches successives qui effectuent des transformations de plus en plus complexes sur les données d'entrée. Les étapes clés sont :

1. **Convolution :** Chaque couche de convolution applique un filtre qui glisse sur l'image pour extraire des caractéristiques locales (par exemple, détecter les bords) [65].
2. **Poolling (Sous-échantillonnage) :** Après la convolution, un processus de sous-échantillonnage (pooling) est souvent appliqué pour réduire la dimensionnalité des données tout en préservant les informations importantes. Le pooling typique inclut la max-pooling, où la valeur maximale d'une région est sélectionnée, ce qui aide à réduire la sensibilité aux petites variations dans l'image [82].
3. **Couches entièrement connectées (Fully Connected layers) :** Une fois les caractéristiques extraites à travers les couches convolutives et de pooling, les données sont souvent passées à des couches entièrement connectées, qui servent à effectuer des prédictions finales, comme la classification [83].

Un exemple d'architecture CNN pour la classification d'images est illustré à la Figure 1.6.

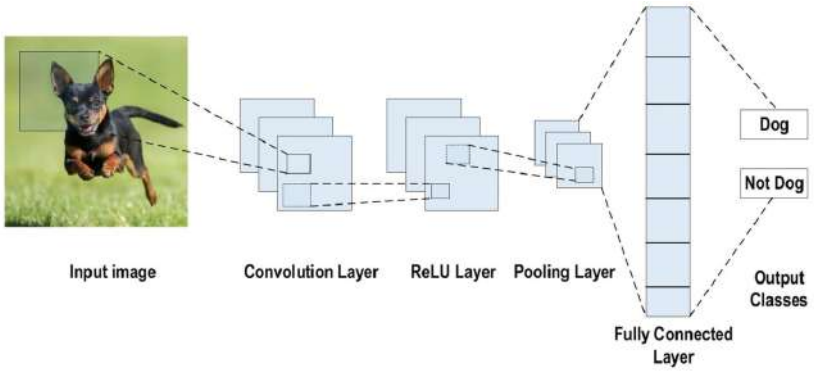


FIGURE 1.6 – Exemple d'architecture CNN pour la classification d'images [72]

Applications : Les CNN sont extrêmement populaires dans des tâches où la détection de motifs et de structures visuelles est essentielle :

— **Reconnaissance d'images :** Identification et classification d'objets dans des images (par exemple, diagnostic médical, reconnaissance faciale) [84].

— **Vision par ordinateur :** Segmentation d'images, détection d'objets, reconnaissance de scènes [85].

— **Analyse vidéo :** Suivi d'objets dans des vidéos, analyse des comportements humains [86].

— **Jeux vidéo et simulations :** Amélioration de l'IA dans les jeux par l'analyse visuelle [44].

Avantages : Les CNN ont un grand avantage par rapport aux réseaux traditionnels dans le traitement d'images ou de données structurées en grilles, car ils peuvent extraire des caractéristiques locales avec une très faible complexité de calcul. De plus, l'utilisation de convolutions réduit drastiquement le nombre de paramètres, rendant l'entraînement de grands réseaux plus efficace [71].

Limites :

— **Nécessité de grandes quantités de données annotées :** Les CNN, comme les RNN, nécessitent de vastes ensembles de données pour entraîner les modèles de manière efficace [87].

— **Puissance de calcul :** Les CNN exigent des ressources matérielles considérables, en particulier pour le traitement des images à haute résolution ou de vastes ensembles de données [88].

1.3 Limites de l'apprentissage automatique

L'apprentissage automatique constitue un outil puissant, notamment grâce à sa capacité à générer des modèles à partir de données, y compris lorsque celles-ci sont imprécises. Cependant, cet avantage peut également se transformer en vulnérabilité.

Pour être pertinent, cet outil nécessite une quantité significative de données, qui sont aujourd'hui de plus en plus accessibles, notamment grâce au développement du *big data*. Néanmoins, l'augmentation du volume de données à traiter engendre une demande accrue en ressources matérielles, ce qui peut considérablement allonger la durée d'entraînement des modèles. Par exemple, en avril 2019, une machine a surpassé des champions du jeu vidéo Dota 2. Pour atteindre cet exploit, elle a dû consacrer dix mois à apprendre le jeu, dans une version simplifiée [89].

Un autre défi majeur réside dans la relation entre l'apprentissage automatique et les données d'entrée. En cas de biais dans les données fournies, ces biais seront inévitablement reproduits dans les résultats du programme. Par exemple, en 2014, Amazon a introduit un algorithme dédié à la présélection automatique des curriculum vitae dans son processus de recrutement. Cependant, l'expérience a été abandonnée après que l'entreprise a constaté que cet outil discriminait les candidatures féminines pour les postes de développeur. Cette discrimination provenait des données d'entraînement, constituées majoritairement de profils masculins, reflétant une sous-représentation des femmes dans cette profession [90].

Ainsi, bien qu'automatique, l'apprentissage automatique demeure intrinsèquement lié à ses concepteurs et à la qualité des données utilisées. S'il peut reproduire, dans une certaine mesure, la capacité humaine d'apprentissage, il est également susceptible d'amplifier les biais et les faiblesses inhérents aux données ou aux processus qui l'ont créé.

Conclusion

L'apprentissage automatique est un domaine en pleine expansion, riche en opportunités et en défis. Ce chapitre a présenté les bases théoriques et pratiques, les différentes méthodes, ainsi que leurs applications variées dans des domaines comme la vision par ordinateur, la reconnaissance vocale, et bien d'autres. En explorant les défis techniques et éthiques, il devient évident que ce domaine nécessite non seulement des avancées techniques mais aussi une réflexion approfondie sur son impact sociétal.
Dans le chapitre suivant, nous approfondirons les généralités sur le contrôle de qualité, une étape cruciale pour assurer la fiabilité et la précision des modèles, ainsi que leur application dans des environnements réels.

CHAPITRE 2

APPROCHES ET OUTILS DU CONTRÔLE DE QUALITÉ

Introduction

Le contrôle de qualité est un aspect essentiel de la production et de la fourniture de biens et services. Il s'assure que les produits répondent aux normes de qualité définies et satisfont les attentes des consommateurs [91].
Ce chapitre examine les méthodes, les technologies et les pratiques couramment utilisées dans le contrôle de qualité, ainsi que leur évolution et leur importance dans divers secteurs industriels. Il explore également les concepts fondamentaux du contrôle de qualité, en mettant en lumière les principes de gestion, les outils statistiques et les approches modernes, comme l'intégration de l'intelligence artificielle et des technologies numériques. Cette exploration vise à fournir une base solide pour comprendre les défis actuels et les opportunités dans l'amélioration continue de la qualité des produits et services.

2.1 Historique et évolution du contrôle de qualité

Le contrôle de qualité remonte à l'Antiquité, époque où les artisans utilisaient diverses techniques pour garantir la satisfaction de leurs clients. Au Moyen Âge, les guildes et les corporations imposaient des normes strictes à leurs membres afin de maintenir la qualité des produits. Cette approche collective visait à protéger la réputation des métiers tout en assurant des standards élevés pour les consommateurs [92].

La révolution industrielle des XVIIIe et XIXe siècles a marqué un tournant décisif pour le contrôle de qualité. Avec l'essor de la production de masse, des méthodes plus systématiques sont devenues nécessaires. Les usines ont commencé à employer des inspecteurs pour vérifier la qualité des produits avant leur sortie de l'usine. Cette période a vu les premières tentatives d'intégration de pratiques de contrôle de qualité à grande échelle, mais elles étaient souvent rudimentaires et limitées à des inspections finales [92].

L'émergence de la notion de qualité remonte aux années 1920 et 1930, marquées par des avancées significatives grâce aux travaux de Walter A. Shewhart. Travaillant chez Bell Telephone Laboratories, Shewhart a introduit les cartes de contrôle en 1924 et a développé des méthodes statistiques pour contrôler la qualité des processus de production, un concept approfondi dans son livre "Economic Control of Quality of Manufactured Product" (1931). Ce courant de pensée, né au sein de Bell Telephone aux USA, était une réponse aux graves problèmes de non-qualité rencontrés par l'entreprise. Shewhart, mathématicien de formation, a proposé l'idée d'un contrôle organisé des produits et des services en utilisant les statistiques. La vision de l'organisme, fondée sur l'organisation scientifique du travail, imposait que ce contrôle soit confié à des services distincts de ceux en charge de la production. Ainsi, sont nés les services qualité, chargés de vérifier la conformité des produits livrés. Cependant, ces idées n'ont pas connu immédiatement des développements considérables [92].

La Seconde Guerre mondiale a favorisé la diffusion des principes de Shewhart dans l'ensemble de l'industrie américaine. La demande accrue de production à partir de 1941 a conduit à la diffusion et la généralisation des techniques de contrôle statistique. Durant cette période, des figures comme W. Edwards Deming et Joseph Juran ont joué un rôle crucial. Deming a introduit des méthodes statistiques pour améliorer la production militaire américaine et a aidé à transformer l'industrie japonaise dans les années 1950, avec ses ouvrages "Out of the Crisis" et "The New Economics for Industry, Government, Education". Joseph Juran, quant à lui, a popularisé le concept de "gestion de la qualité totale" Total Quality Management (TQM), en mettant l'accent sur la responsabilité de la direction et sur la formation continue [92, 93].

Dès le début des années 1950, un événement marquant dans l'histoire de la qualité se produit au Japon, où celle-ci devient un paramètre essentiel dans la reconstruction de l'industrie japonaise, complètement démantelée après la guerre. L'introduction de la notion de qualité au Japon est orchestrée par des experts américains, mandatés par le général MacArthur pour contribuer à cette reconstruction. Les idées de Deming, alors peu connu aux USA, rencontrent un écho remarquable auprès des industriels nippons. Par la suite, dans les années 1960, émerge le concept de qualité totale, porté par des experts japonais soulignant la nécessité de transformer les modes d'organisation et les systèmes de pensée afin d'améliorer encore davantage la qualité. Les principes fondateurs de la qualité totale, tels que la participation des personnels et la prévention, voient alors le jour [92, 93].

Simultanément, Kaoru Ishikawa, un pionnier japonais, développe des outils essentiels pour la gestion de la qualité, tels que les cercles de qualité et le diagramme d'Ishikawa (ou diagramme en arêtes de poisson), destinés à identifier et analyser les causes des problèmes de qualité. Son ouvrage "What is Total Quality Control? The Japanese Way" influence largement la manière dont le contrôle de qualité est perçu et pratiqué [92, 93].

Pendant ce temps, en Occident, de la fin des années 1950 à la fin des années 1970, la qualité est principalement vue sous l'angle classique du contrôle qualité et du "bon produit", se diffusant dans de nombreux organismes. C'est également durant cette période que naissent les principes de

l'assurance qualité, un concept selon lequel la confiance dans la qualité d'un produit doit être démontrée à tout moment à travers des documents écrits et archivés. Cette approche systématique et documentée vise à garantir la qualité requise tout au long du cycle de vie du produit [92, 93].

Un changement important dans la gestion de la qualité se situe à la fin des années 1970 aux USA et en Europe. Il s'explique par plusieurs facteurs clés [92, 93] :

— La crise industrielle qui pèse de plus en plus lourdement sur les sociétés occidentales.

— L'arrivée massive de produits japonais dans des domaines stratégiques, particulièrement l'automobile, qui surpassent les produits occidentaux en termes de qualité et représentent une menace extrême pour l'industrie occidentale.

Face à ces défis, de grands groupes, notamment de grandes entreprises américaines, prennent conscience de la nécessité de changer leur approche de la qualité et initient des démarches de qualité totale. Ces démarches se fondent sur des modes d'action favorisant la participation et la mobilisation des personnels [92, 93].

Pendant les années 1980, le mouvement vers la qualité totale prend de l'ampleur, et de nombreuses entreprises de biens et de services s'engagent dans cette voie. La qualité totale repose sur l'implication de tous les employés dans le processus de qualité, visant à améliorer continuellement les produits et services. Cette période marque un tournant majeur, avec des entreprises adoptant des pratiques de gestion qui intègrent pleinement les principes de la qualité totale, mettant l'accent sur la prévention des défauts, la participation de tous les niveaux de l'organisation et l'amélioration continue [92, 93].

Les années 1980 et 1990 ont vu la publication des normes ISO 9000 en 1987, qui ont standardisé les pratiques de gestion de la qualité à l'échelle mondiale. Ces normes fournissent un cadre pour la mise en place de systèmes de gestion de la qualité, devenant une référence internationale pour les entreprises. En établissant des standards clairs et unifiés, les normes ISO 9000 ont permis une plus grande cohérence et fiabilité dans la production de biens et services à travers le monde [92, 93].

Plus récemment, des méthodologies comme Six Sigma, popularisées par Motorola et General Electric, et Lean Manufacturing, dérivées du système de production Toyota, ont révolutionné le contrôle de qualité. Six Sigma combine des techniques statistiques avancées et des principes de gestion pour réduire les défauts et améliorer l'efficacité des processus. Lean Manufacturing se concentre sur l'élimination des gaspillages et l'amélioration continue, en engageant tous les niveaux de l'organisation. Ces approches ont transformé les pratiques de gestion de la qualité, les rendant plus rigoureuses et axées sur des résultats mesurables, promouvant ainsi l'excellence opérationnelle et la satisfaction client [92, 93].

Ainsi, l'évolution de la gestion de la qualité en Occident, des années 1950 aux années 1990, montre une progression continue vers des méthodes plus intégrées et systématiques, répondant aux défis industriels et à la concurrence internationale. De l'assurance qualité à la qualité totale, puis à la standardisation mondiale et aux méthodologies avancées, la gestion de la qualité est devenue un élément central de la stratégie des entreprises modernes [92, 93].

Quant au contrôle de qualité des aliments, il trouve ses origines dans l'Antiquité. À cette époque, des civilisations comme les Égyptiens, les Grecs et les Romains utilisaient des méthodes rudimentaires pour garantir la sécurité de leurs produits alimentaires [94]. Cependant, ce n'est qu'à partir du XIXe siècle, avec la révolution industrielle, que des avancées majeures ont été réalisées. L'industrialisation a entraîné une production alimentaire de masse, nécessitant des mesures plus rigoureuses pour prévenir les contaminations et assurer la qualité des produits [95].

En 1906, les États-Unis ont adopté le Pure Food and Drug Act, une législation pionnière visant à réguler les aliments et les médicaments, marquant un tournant dans le contrôle de qualité des aliments [96]. En Europe, des standards similaires ont été établis, avec des règlements de plus en plus stricts pour protéger les consommateurs [97].

L'ère moderne a vu l'introduction de normes internationales comme celles définies par le Codex Alimentarius, créé en 1963 par la FAO et l'OMS, pour harmoniser les critères de qualité et de sécurité alimentaire à l'échelle mondiale [98]. Parallèlement, des systèmes comme l'*Hazard Analysis and Critical Control Points (HACCP)* ont été développés dans les années 1960 pour prévenir les risques sanitaires tout au long de la chaîne de production alimentaire [99].

Avec l'avènement des technologies numériques et de l'intelligence artificielle, le contrôle de qualité des aliments a encore évolué. Des dispositifs sophistiqués et des algorithmes de *machine learning* permettent aujourd'hui une analyse plus précise et en temps réel des produits alimentaires, promettant une sécurité et une qualité accrues [100]. Cette évolution continue reflète l'engagement permanent envers la protection des consommateurs et l'amélioration des normes alimentaires à l'échelle mondiale.

2.2 Typologie des méthodes de contrôle qualité

Dans le domaine du contrôle qualité, les méthodes d'évaluation des produits et des processus se divisent principalement en deux catégories : le Contrôle Non Destructif (CND) et le Contrôle Destructif (CD).

2.2.1 Contrôle Non Destructif (CND)

Le CND regroupe un ensemble de techniques permettant d'évaluer la conformité et les propriétés d'un produit sans altérer son intégrité physique ou fonctionnelle. Ces méthodes sont essentielles pour détecter des défauts ou des anomalies tout en préservant l'utilisabilité du produit testé. Elles sont largement utilisées dans des industries où la sécurité et la fiabilité sont primordiales, telles que l'aéronautique, l'automobile et la construction.

Parmi les techniques courantes du CND, on trouve :

— **Contrôle par examen visuel :** est effectué avant toute autre méthode CND, selon la norme **NF EN 13018/A1**. Il repose sur l'observation directe à l'œil nu pour détecter les anomalies de surface. Bien qu'il fasse partie des contrôles CND, il se distingue par l'absence de technique

spécifique, nécessitant seulement des conditions d'observation optimisées [101, 102].

— **Contrôle par ressuage :** est une technique d'essai non destructif permettant de détecter les discontinuités débouchantes sur la surface d'une pièce. Prolongement logique de l'examen visuel, elle consiste à faire pénétrer un liquide pénétrant (coloré ou fluorescent) dans les défauts de la pièce par capillarité. Après un temps de repos, la surface est nettoyée et recouverte d'un révélateur, qui absorbe le liquide pénétrant. En cas de discontinuité, le révélateur met en évidence le liquide infiltré, permettant à l'opérateur de visualiser les défauts sous forme d'indications colorées ou fluorescentes. Cette technique est compatible avec de nombreux matériaux, à l'exception des plus poreux [103, 104].

— **Contrôle par magnétoscopie :** est une technique de contrôle non destructif utilisée pour détecter des discontinuités de surface, débouchant en surface ou sous-jacentes (jusqu'à quelques millimètres de profondeur dans certaines conditions), exclusivement sur matériaux ferromagnétiques. Ce contrôle par champ magnétique, l'un des plus anciens avec le ressuage, permet de localiser ces discontinuités grâce à la détection de perturbations du champ magnétique sur la pièce examinée [103, 105].

— **Contrôle par ultrasons :** consiste à propager des ondes ultrasonores à travers un matériau et à analyser leur interaction avec celui-ci pour détecter des discontinuités ou des défauts. La fréquence des ondes, choisie en fonction des caractéristiques du matériau, influence la sensibilité de la détection : des ondes de haute fréquence (par exemple 5 MHz) sont plus sensibles, tandis que des ondes de basse fréquence (par exemple 2 MHz) conviennent mieux aux surfaces grossières. Ce contrôle peut être effectué en transmission ou en réflexion, avec des dispositifs d'émission et de réception séparés ou confondus, et varie selon le type d'ondes et leur inclinaison [103, 106].

— **Contrôle par radiographie :** est une méthode de contrôle non destructif qui consiste à obtenir une image de la densité de matière d'un objet en le traversant avec un rayonnement électromagnétique X ou gamma. Similaire à la radiographie médicale, cette technique utilise un faisceau de rayons X pour irradier un film placé à l'arrière de la surface à contrôler, permettant ainsi d'inspecter l'intérieur de la pièce sans l'endommager. Elle peut être appliquée sur tous les types de matériaux [103, 107].

— **Contrôle par courants de Foucault :** utilise des courants de Foucault induits dans des matériaux conducteurs par un champ magnétique variable. La répartition de ces courants dépend du champ magnétique, de la géométrie et des propriétés électriques du matériau. En cas d'anomalie, ces courants sont perturbés, modifiant l'impédance apparente du capteur. L'analyse de cette variation permet de détecter l'anomalie, comparée à celle d'une pièce de référence. La variation est affichée sous forme de courbes de "Lissajous", en amplitude et phase [108].

Ces techniques permettent de préserver l'intégrité des matériaux, détecter précocement les défauts, réduire les coûts de maintenance et améliorer la sécurité en assurant la fiabilité des structures critiques.

2.2.2 Contrôle Destructif (CD)

Le CD, quant à lui, implique des tests qui endommagent ou détruisent partiellement ou totalement l'échantillon analysé. Ces méthodes sont utilisées pour déterminer les limites de performance, la résistance et d'autres caractéristiques mécaniques ou chimiques des matériaux et des produits [109].

Les tests destructifs courants incluent :

— **Essais de traction :** évaluentz la réaction d'un matériau lorsqu'il est étiré jusqu'à la rupture, mesurant sa résistance à la traction, son allongement et sa ductilité. Utilisé principalement dans les secteurs automobile et aérospatial, il permet de garantir que les métaux et alliages résistent aux contraintes opérationnelles [109].

— **Essais d'impact :** mesurent la capacité d'un matériau à absorber l'énergie lors d'un choc soudain. Les tests Charpy et Izod sont les plus courants, et cette méthode est cruciale pour évaluer la résistance des matériaux utilisés dans la construction, comme les poutres et supports en acier [109].

— **Essais de dureté :** mesurent la résistance d'un matériau à la déformation par indentation. Les méthodes Brinell, Rockwell et Vickers sont les plus courantes. Ces tests sont essentiels dans des secteurs comme la fabrication et le génie civil pour vérifier la résistance à l'usure des matériaux tels que le béton et les métaux [109].

— **Essais de fatigue :** soumettent les matériaux à des charges cycliques répétées pour déterminer leur durée de vie avant la défaillance. Ce test destructif est crucial dans l'industrie aérospatiale, où des composants comme les pales de turbines doivent résister à des contraintes constantes sur de longues périodes [109].

— **Essais de corrosion :** servent à mesurer la capacité d'un matériau à résister à la corrosion dans des conditions environnementales particulières, comme l'exposition à l'eau salée ou aux solutions acides. Cette épreuve est indispensable dans des domaines tels que la marine, le pétrole et le gaz, ainsi que les infrastructures, où la ténacité à la corrosion est primordiale pour la sûreté et la pérennité [109].

— **Essais de fluage :** évaluent la déformation progressive des matériaux exposés à une pression stable dans le temps, habituellement à des températures hautes. Ce genre de test est crucial dans les domaines de la production d'énergie et de l'aérospatiale, où les matériaux sont soumis à des conditions prolongées de haute température [109].

Bien que ces tests entraînent la destruction de l'échantillon, ils fournissent des informations essentielles sur les performances et la sécurité des produits, garantissant ainsi leur conformité aux normes et spécifications requises.

La sélection entre CND et CD dépend des objectifs de l'évaluation, de la nature du produit et des exigences spécifiques de l'industrie concernée.

2.3 Contrôle de qualité dans quelques secteurs

2.3.1 Industrie automobile

L'industrie automobile impose des normes de qualité très élevées. Chaque composant utilisé dans la production de véhicules doit satisfaire des exigences strictes afin de garantir fiabilité et sécurité. L'automobile occupe une place essentielle dans la vie quotidienne, en offrant des solutions de mobilité tout en s'intégrant profondément aux habitudes et au mode de vie contemporain [110]. La qualité d'un véhicule dépend étroitement des pièces qui le composent, de leur assemblage et de leur fonctionnement harmonieux. Étant donné que la majorité des composants provient de sous-traitants, l'industrie applique des normes rigoureuses tout au long de la chaîne d'approvisionnement. L'identification des processus critiques et leur optimisation sont essentielles. Le contrôle statistique des processus (SPC : Statistical Process Control) permet de surveiller et de maîtriser la production, tandis que des audits internes quotidiens contribuent à éliminer les risques pour la qualité [110].

2.3.2 Industrie aéronautique

Dans le domaine de l'aéronautique, la qualité est essentielle pour assurer le bon déroulement des processus, la conformité des biens ou services, et la protection des occupants. Ce secteur est particulièrement strict en raison des défis liés à la sûreté et à la fiabilité des avions. Les compagnies aériennes doivent respecter des normes internationales telles que l'AS9100 (Aerospace Standard 9100), l'ISO (International Organization for Standardization) 9001, ainsi que des normes spécifiques comme le NADCAP (National Aerospace and Defense Contractors Accreditation Program). Ces exigences couvrent l'ensemble de la chaîne de production : de la conception à la fabrication, y compris les contrôles, assemblages et l'entretien [111].

L'évitement des défauts est crucial dans l'industrie aéronautique, car même une défaillance mineure peut entraîner des conséquences graves. Ainsi, tous les intervenants, des fournisseurs aux producteurs, doivent garantir une qualité irréprochable. Les pièces subissent des tests destructifs et non destructifs, comme les rayons X, les ultrasons et d'autres méthodes non invasives, permettant d'examiner les composants sans les endommager. Chaque pièce doit être parfaitement traçable, avec une documentation complète. L'industrie repose sur un réseau complexe de fournisseurs [111].

L'industrie est fortement régulée, et les entreprises doivent sans cesse s'adapter aux nouvelles exigences des autorités de sécurité aérienne. Chaque nouveau produit ou modification doit être validé par les autorités, ce qui peut entraîner des délais de certification et des coûts élevés. Les changements dans les processus ou matériaux nécessitent également validation et approbation, ce qui impose des contraintes supplémentaires sur la flexibilité et la réactivité des processus de production. Bien que la qualité soit primordiale, l'industrie fait face à une pression constante pour réduire les délais de production [111].

2.3.3 Industrie électronique

L'industrie de l'électronique grand public est soumise à des normes rigoureuses en matière de sécurité et de marquage de conformité, telles que la directive *Restriction of Hazardous Substances (RoHS)*, les régulations *Federal Communications Commission (FCC)*, *Déchets d'Équipements Électriques et Électroniques (DEEE)* et *CE*. De plus, elle doit répondre à d'importantes attentes de la part des clients. La vérification que les produits et leurs composants respectent les normes de qualité et les spécifications souhaitées est rendue plus complexe par la diversité des usines de production et des fournisseurs [112].

Les inspections de produits électroniques sont conçues pour évaluer non seulement la durabilité et la qualité de fabrication du produit, mais aussi ses performances et sa fonctionnalité. Elles peuvent être réalisées de la manière suivante [112] :

Inspection avant production : Elle garantit que le fabricant d'électronique grand public peut démarrer la production dans les délais prévus et que toutes les spécifications du produit ainsi que les exigences de qualité sont respectées au plus haut niveau.

Inspection en cours de production : Les produits sont inspectés pendant la production afin de s'assurer qu'ils sont conformes aux normes spécifiées, ce qui permet d'identifier et de résoudre rapidement toute source de défaut.

Inspection avant expédition : Inspection d'un échantillon de produits finis pour vérifier qu'ils répondent aux exigences et aux normes de qualité, y compris l'inspection visuelle, l'évaluation de l'aspect général, de la fabrication, de la fonction, de la sécurité et d'autres spécifications précisées.

Inspection pièce par pièce : Évaluation de chaque article du lot de production, en prenant en compte une série de variables telles que la fonctionnalité, la sécurité, la qualité de fabrication, etc.

Supervision du chargement : Inspection du processus d'emballage et de chargement pour s'assurer que les produits sont emballés et expédiés conformément aux normes convenues.

Vérification de l'assemblage correct : Vérification que tous les composants du produit sont correctement assemblés et montés à l'aide d'un formulaire de données sur les composants.

2.3.4 Industrie textile

L'industrie du textile, un secteur générant des milliards, revêt une grande importance dans l'économie d'un pays. Cependant, malgré les revenus qu'elle génère, un contrôle inadéquat de la qualité peut rapidement nuire à sa réputation, en raison d'un taux élevé de défauts et de produits de moindre qualité. Ces problèmes entraînent des coûts considérables, tant sur le plan financier qu'environnemental. Heureusement, des programmes et réglementations de certification de qualité, tels que la série ISO 9000, permettent aux entreprises de maintenir des normes élevées dans

l'industrie textile et de la mode [113]. Pour évaluer la qualité des vêtements et des matériaux de fabrication, plusieurs considérations doivent être prises en compte, notamment [113] :

— Les différentes nuances de couleurs ;
— Les fils et étamines lâches ;
— La qualité des fibres ;
— Les encoches apparentes ;
— Les défauts visibles du tissu ; etc.

Il existe plusieurs procédures de contrôle qualité pour les vêtements et les textiles, qui doivent être réalisées de manière optimale à différentes étapes de la production : avant, pendant et après celle-ci, afin d'assurer la qualité à chaque étape de la fabrication.

Contrôle qualité des vêtements en pré-production : Le contrôle qualité des vêtements en pré-production est effectué avant le début de la production. Il comprend l'inspection des échantillons de tissus et de vêtements avant leur découpe ou leur couture. Ce processus vise à s'assurer que le tissu respecte les normes requises en matière de [113] :

— solidité des couleurs ;
— texture ;
— propriétés techniques ;
— durabilité ;
— détection de fils lâches ou de trous.

Contrôle qualité des vêtements pendant la production : L'inspection en cours de la production est essentielle pour garantir que les vêtements sont fabriqués conformément aux attentes et aux spécifications des clients. Ces inspections, généralement effectuées lorsque 15 à 20% de la marchandise est achevée, peuvent inclure plusieurs méthodes, notamment [113] :

— l'inspection visuelle (évaluation de la découpe, assemblage des pièces et des coutures) ;
— les prises de mesures ;
— les essais destructifs.

Contrôle qualité des vêtements finis (inspection avant expédition) : Le contrôle qualité des vêtements finis est effectué avant l'expédition de la marchandise au client, une fois qu'au moins 80% de la commande est emballée pour l'envoi. Ce processus permet de détecter d'éventuels défauts et de minimiser le risque de réclamations clients. L'inspection comprend généralement [113] :

— une vérification visuelle des vêtements pour identifier tout défaut apparent ;
— la vérification de l'étiquetage ;
— le comptage des articles dans le lot de production.

2.3.5 Industrie pharmaceutique

Le contrôle de qualité dans le secteur de l'industrie pharmaceutique est essentiel pour garantir la sécurité, l'efficacité et la conformité des médicaments. Ce processus rigoureux englobe une série

de procédures et de normes destinées à surveiller et maintenir la qualité des produits pharmaceutiques. Il s'étend tout au long du cycle de vie des produits, de la recherche et du développement à la production, la distribution et la commercialisation. Les entreprises pharmaceutiques appliquent des systèmes de gestion de la qualité comme les bonnes pratiques de fabrication et les bonnes pratiques de laboratoire, qui sont régis par des organismes de réglementation tels que la *Food and Drug Administration (FDA)* aux États-Unis et l'*European Medicines Agency (EMA)* en Europe [114].

Le contrôle de qualité doit être réalisé dans un environnement confiné afin d'assurer la fiabilité des résultats. Cette condition limite les risques de faux négatifs ou de faux positifs, qui pourraient retarder la validation des produits pharmaceutiques. Pour éviter ces résultats erronés, l'utilisation d'isolateurs de contrôle de stérilité s'est largement imposée. Ces équipements permettent de mener les tests dans un environnement séparé des polluants particulaires, chimiques ou microbiologiques, garantissant ainsi la fiabilité et la rapidité des tests de stérilité réalisés [114].

Le contrôle de qualité en pharmaceutique se déroule généralement en laboratoire, utilisant des équipements adéquats comme des isolateurs. Les tests peuvent porter sur divers domaines : biochimiques, physico-chimiques, immunologiques, biologiques, microbiologiques, immuno-hématologiques, etc. Les objectifs de ces tests sont multiples : vérifier et confirmer la qualité des médicaments produits, détecter d'éventuels défauts de fabrication, évaluer la dangerosité potentielle des produits, et mettre en place des actions préventives ou correctives. Des tests exhaustifs, des inspections régulières et des audits rigoureux sont effectués pour s'assurer que chaque lot de médicament répond aux normes de qualité et de sécurité définis par les réglementations nationales et internationales [114].

La traçabilité est également cruciale, permettant de suivre chaque étape de la production et de réagir rapidement en cas de problèmes de qualité. Grâce à ces pratiques strictes de contrôle de qualité, l'industrie pharmaceutique est capable de fournir des médicaments sûrs et efficaces, protégeant ainsi la santé des patients et maintenant la confiance du public [114].

2.3.6 Industrie manufacturière

Le contrôle de qualité dans l'industrie manufacturière est un élément fondamental pour assurer la conformité, la fiabilité et la performance des produits finis. Ce processus englobe diverses techniques et méthodes visant à surveiller et à améliorer continuellement la production. Dès le début du 20^{e} siècle, des pionniers comme Frederick Winslow Taylor ont introduit des principes de gestion scientifique pour optimiser les processus de fabrication [115]. Plus tard, Walter A. Shewhart, avec ses cartes de contrôle introduites en 1924, a apporté des avancées significatives en matière de contrôle statistique des processus [116]. Ces méthodes ont été raffinées et popularisées par W. Edwards Deming et Joseph Juran, qui ont joué un rôle crucial dans la mise en œuvre de pratiques de qualité totale, particulièrement après la Seconde Guerre mondiale [117, 118].

L'application des normes ISO 9000, établies en 1987, a standardisé les pratiques de gestion de la qualité à l'échelle mondiale, fournissant un cadre systématique pour l'amélioration continue

(ISO, 2015). Plus récemment, des méthodologies telles que Six Sigma, développée par Motorola, et Lean Manufacturing, dérivée du système de production de Toyota, ont révolutionné le contrôle de qualité en combinant des techniques statistiques avancées et des principes de gestion allégée pour réduire les défauts et augmenter l'efficacité [119, 120].

Les entreprises manufacturières modernes utilisent des systèmes de gestion de la qualité (QMS) intégrés, qui incluent des audits réguliers, des inspections visuelles, et des tests de performance pour s'assurer que chaque produit répond aux spécifications. Par exemple, l'utilisation de l'inspection par rayons X permet de détecter des défauts internes invisibles à l'œil nu, garantissant ainsi un niveau de qualité élevé. La traçabilité est également cruciale dans ce secteur, permettant de suivre chaque composant du produit tout au long de la chaîne d'approvisionnement et de production, facilitant ainsi un rappel rapide et efficace en cas de défaillance [121, 122].

2.3.7 Industrie alimentaire

Le contrôle de qualité dans le secteur de l'industrie alimentaire est crucial pour garantir la sécurité, la conformité et la satisfaction des consommateurs. Même si les aliments sont produits dans divers lieux et environnements, les normes sont généralement très similaires en ce qui concerne le contrôle de la qualité, grâce aux normes internationales comme le système *Hazard Analysis and Critical Control Points (HACCP)*. Créé par l'*United States Department of Agriculture (USDA)*, le *HACCP* est l'exigence la plus courante que les producteurs de produits alimentaires doivent remplir. Ce système identifie, évalue et contrôle les risques biologiques, chimiques et physiques dans la chaîne d'approvisionnement alimentaire [123]. D'autres normes supplémentaires sont souvent exigées lors de l'exportation de produits alimentaires :

— *British Retail Consortium (BRC)* : Association professionnelle britannique du commerce de détail, également considérée comme une norme internationale.
— *International Featured Standards (IFS)* : Norme établie pour l'évaluation des fabricants de produits alimentaires.
— *Global Food Safety Initiative (GFSI)* : Projet dont l'objectif est d'harmoniser les normes internationales de sécurité des aliments et de diminuer les exigences d'audit de divers fournisseurs.
— *Service européen d'inspection de la sécurité des aliments (EFSIS)* : En jouant le rôle d'un tremplin pour atteindre les exigences de la norme BRC complète.

Les normes de l'industrie alimentaire se sont resserrées au cours des dernières décennies, incitant les producteurs à améliorer leurs processus de sécurité et de qualité. Des audits réguliers, des tests microbiologiques et des inspections visuelles sont effectués pour assurer que les produits répondent aux normes définis par les réglementations nationales et internationales. Par exemple, la mise en place d'un système d'inspection par rayons X sur la chaîne de production permet de satisfaire, et même de surpasser, les exigences des grandes chaînes de distribution mondiales et locales [123].

Les autorités et les grands détaillants ont défini plusieurs exigences en matière de sécurité et de contrôle de la qualité des aliments. La traçabilité des produits est également essentielle pour pouvoir réagir rapidement en cas de détection de défauts ou de contaminations, protégeant ainsi les consommateurs et maintenant la confiance dans les produits alimentaires. Grâce à ces pratiques rigoureuses de contrôle de qualité, l'industrie alimentaire peut fournir des produits sûrs et de haute qualité, répondant aux exigences des clients et des régulateurs [123].

Les méthodes traditionnelles de contrôle de qualité des aliments reposent principalement sur des techniques physiques, chimiques et biologiques. Les analyses physico-chimiques, telles que la mesure de l'humidité, du pH, de la teneur en cendres et des analyses spectrophotométriques, sont largement utilisées pour déterminer la composition des aliments et détecter les contaminants [124]. Par exemple, la chromatographie en phase gazeuse et la chromatographie liquide haute performance (HPLC) permettent d'identifier et de quantifier les composés chimiques présents dans les aliments [125].

Les méthodes microbiologiques, quant à elles, sont cruciales pour détecter et identifier les micro-organismes pathogènes dans les produits alimentaires. Les techniques telles que la culture en milieu sélectif, les tests de dénombrement bactérien et les tests immunologiques sont couramment utilisées pour assurer la sécurité microbiologique des aliments [126]. Ces méthodes permettent de surveiller la contamination microbienne et de prévenir les épidémies de maladies d'origine alimentaire.

En outre, les panels sensoriels humains jouent un rôle essentiel dans l'évaluation des caractéristiques organoleptiques des aliments, telles que le goût, l'odeur, la texture et l'apparence. Ces panels sont composés de dégustateurs entraînés qui évaluent les produits selon des critères spécifiques et standardisés [9]. Bien que subjectives, ces évaluations sont indispensables pour garantir que les aliments répondent aux attentes des consommateurs en termes de qualité sensorielle.

Malgré leur efficacité, ces méthodes traditionnelles présentent certaines limites. Elles peuvent être coûteuses financièrement, chronophages et nécessiter un équipement sophistiqué ainsi qu'un personnel qualifié. De plus, certaines techniques peuvent altérer la qualité des aliments lors de l'analyse [125]. Face à ces défis, de nouvelles approches technologiques, telles que l'utilisation de capteurs électroniques et de l'intelligence artificielle, émergent pour améliorer et compléter les méthodes traditionnelles de contrôle de qualité des aliments.

2.4 Technologies modernes de contrôle de qualité des aliments

Les technologies modernes de contrôle de la qualité des aliments ont considérablement évolué, permettant une surveillance plus précise et efficace tout au long de la chaîne de production. Parmi les innovations clés, on trouve la blockchain, ainsi que des dispositifs tels que le nez électronique, la langue électronique, et les systèmes d'imagerie et de vision par ordinateur. Ces technologies contribuent à assurer la sécurité alimentaire, à optimiser la traçabilité et à garantir la qualité des produits.

2.4.1 Blockchain

La blockchain est une technologie de stockage et de transmission d'informations, transparente, sécurisée et fonctionnant sans organe central de contrôle (définition de Blockchain France). Malgré son apparence complexe, elle correspond simplement à un type de base de données sécurisé et distribué, utilisé pour l'enregistrement des transactions sur les ordinateurs de tous les participants. Contrairement aux bases de données traditionnelles, où les données peuvent être modifiées ou manipulées, la blockchain garantit qu'une fois une information ajoutée, elle ne peut plus être modifiée [127].

Dans le domaine du contrôle de la qualité des aliments, la blockchain offre une solution innovante pour améliorer la traçabilité et la transparence tout au long de la chaîne d'approvisionnement. Chaque étape du processus, de la production à la distribution, peut être enregistrée de manière sécurisée, garantissant ainsi l'intégrité des informations et prévenant les fraudes. Ce système permet également de s'assurer que les produits respectent les standards de qualité et de sécurité, car toutes les informations sont accessibles et vérifiables en temps réel par les acteurs de la chaîne, des producteurs aux consommateurs [128].

L'un des principaux avantages de la blockchain dans ce contexte est sa capacité à suivre en temps réel l'origine des produits et les conditions dans lesquelles ils ont été stockés ou transportés. Par exemple, une étude menée par IBM Food Trust a démontré l'efficacité de la blockchain dans le suivi des produits alimentaires, en permettant une traçabilité instantanée des produits. Cette technologie réduit ainsi le temps nécessaire pour identifier l'origine d'une contamination en cas de rappel de produit [129]. De plus, Walmart a intégré la blockchain dans ses processus pour suivre la provenance de certains produits alimentaires et a considérablement réduit le délai nécessaire pour retracer l'origine d'un produit en cas de problème sanitaire, améliorant ainsi la sécurité alimentaire [130].

La blockchain facilite également la gestion des certifications de qualité, telles que celles relatives aux produits biologiques ou durables. Par exemple, Everledger, une entreprise spécialisée dans l'application de la blockchain, a collaboré avec des producteurs de vin pour garantir l'authenticité et la qualité des bouteilles en enregistrant chaque étape de la production sur un registre décentralisé [131].

Il découle de ces exemples que la blockchain renforce la confiance des consommateurs en leur offrant la possibilité de vérifier, via des applications mobiles, les informations concernant la chaîne de production et les conditions de stockage des aliments qu'ils achètent, assurant ainsi un meilleur contrôle de la qualité tout au long du cycle de vie du produit.

2.4.2 Nez électronique

Le nez électronique est une technologie moderne révolutionnaire qui imite le sens de l'odorat humain pour identifier et distinguer différentes odeurs, jouant ainsi un rôle crucial dans le contrôle de qualité. Composé de capteurs chimiques capables de détecter les composés volatils

présents dans l'air, ces capteurs sont connectés à un système informatique qui analyse les données pour reconnaître les odeurs. Un détecteur de fumée capable non seulement de repérer la fumée, mais aussi de reconnaître des milliers d'odeurs distinctes, telles que celles du café frais, du pain grillé ou même de certaines maladies, pourrait exister. C'est précisément ce que fait un nez électronique, rendant cette technologie indispensable dans de nombreux secteurs.

Dans l'industrie alimentaire, le nez électronique assure la qualité et la fraîcheur des produits. Il peut détecter les signes précoces de détérioration avant même que des narines humaines ne s'en rendent compte, garantissant ainsi que seuls les produits les plus frais atteignent les consommateurs [3]. De plus, dans le domaine médical, cette technologie moderne est utilisée pour diagnostiquer des maladies à travers l'analyse de l'haleine des patients, permettant de repérer des conditions telles que le diabète ou le cancer à un stade précoce [132].

Les nez électroniques sont également employés dans l'industrie de la sécurité pour détecter des substances dangereuses ou illégales. Ils peuvent repérer des traces de drogues ou d'explosifs dans les bagages des voyageurs, améliorant ainsi les procédures de sécurité dans les aéroports [133].

Les Figures 2.1 à 2.6 présentent quelques nez électroniques créés par différentes équipes de recherche et entreprises innovantes.



FIGURE 2.1 – NeOse Advance[1]



FIGURE 2.2 – Alpha MOS Fox-Gemini[2]

1. Source : https://www.cea.fr/presse/Pages/actualites-communiques/ntic/aryballe-propose-un-nez-electronique-universel.aspx, consultée le 20 novembre 2024.

2. Source : https://www.instrument-solutions.com/en/technology/organoleptic-analysis/e-nose/attachment/alpha-mos-fox-gemini-e-nose, consultée le 20 novembre 2024.



FIGURE 2.3 – Spironose[3]



FIGURE 2.4 – Portable Electronic Nose[4]



FIGURE 2.5 – zNose 4200[5]



FIGURE 2.6 – Cyranose 320[6]

2.4.3 Langue électronique

La langue électronique est une technologie moderne qui imite le sens du goût humain pour analyser et identifier les saveurs de différentes substances liquides. Composée de capteurs chimiques qui réagissent aux diverses saveurs (sucré, salé, acide, amer et umami), cette technologie est connectée à un système informatique avancé qui interprète les données pour reconnaître les profils gustatifs. Un sommelier électronique capable de distinguer et d'évaluer la complexité des

3. Source : `https://www.francetvinfo.fr/replay-radio/le-billet-vert/le-nez-electronique-l-appareil-qui-permettrait-de-detecter-des-maladies-a-travers-l-haleine-du-patient_4756781.html`, consultée le 20 novembre 2024.

4. Source : `https://airsense.com/en/products/portable-electronic-nose`, consultée le 20 novembre 2024.

5. Source : `https://www.ebay.com/itm/204633642513`, consultée le 20 novembre 2024.

6. Source : `https://www.sensigent.com/cyranose-320.html`, consultée le 20 novembre 2024.

vins, des boissons et même des solutions chimiques avec une précision inégalée. C'est ce que permet la langue électronique, rendant cette technologie essentielle pour le contrôle de qualité dans de nombreuses industries.

Dans l'industrie alimentaire et des boissons, la langue électronique est utilisée pour garantir la constance et la qualité des produits. Par exemple, elle peut vérifier que chaque lot de jus de fruits ou de boissons gazeuses possède le même goût, évitant ainsi les variations qui pourraient déplaire aux consommateurs [134]. Cette technologie permet aussi de développer de nouveaux produits en évaluant et en ajustant leurs profils de saveurs avant leur mise sur le marché.

Dans l'industrie pharmaceutique, la langue électronique joue un rôle crucial dans le contrôle de la qualité des médicaments liquides. Elle assure que les médicaments ont des goûts acceptables pour les patients, notamment pour les sirops destinés aux enfants, tout en garantissant que chaque lot respecte les normes strictes de saveur et de sécurité [135].

De plus, cette technologie est utilisée dans le domaine environnemental pour surveiller la qualité de l'eau. La langue électronique peut détecter des contaminants ou des variations de goût dans l'eau potable, assurant ainsi que l'eau distribuée aux consommateurs est toujours sûre et agréable à consommer [136].

Les Figures 2.7 à 2.10 présentent quelques langues électroniques créés par différentes équipes de recherche et entreprises innovantes.

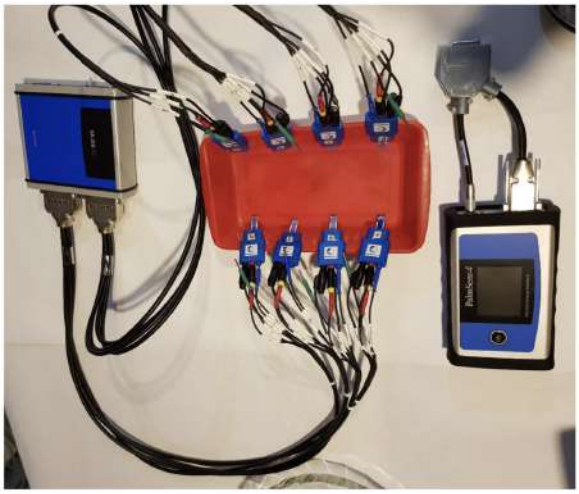


FIGURE 2.7 – Langue électronique [137]



FIGURE 2.8 – TS-5000Z [138]



FIGURE 2.9 – Alpha MOS ASTREE[7]



FIGURE 2.10 – Langue électronique de Shanghai Bosin[8]

2.4.4 Imagerie et vision par ordinateur

L'imagerie et la vision par ordinateur sont des technologies modernes qui utilisent des caméras et des algorithmes informatiques pour analyser visuellement des objets ou des scènes, de manière similaire à la perception humaine. Ces technologies permettent aux machines de "voir" et de "comprendre" les images afin de détecter des anomalies, de mesurer des dimensions ou de vérifier la conformité des produits. Par exemple, une machine capable de vérifier l'intégrité de chaque produit qui sort d'une chaîne de production avec une précision et une rapidité surpassant celles de l'œil humain. C'est précisément ce que permet la vision par ordinateur, rendant cette technologie essentielle pour le contrôle de qualité dans divers secteurs industriels.

Dans l'industrie manufacturière, l'imagerie et la vision par ordinateur sont utilisées pour inspecter les produits en temps réel. Par exemple, dans la fabrication de circuits imprimés, cette technologie peut détecter des minuscules défauts, tels que des soudures imparfaites ou des composants mal alignés, garantissant ainsi que seuls les produits conformes et de haute qualité sont expédiés aux clients [139].

Dans l'industrie alimentaire, la vision par ordinateur peut vérifier la qualité des fruits et légumes en détectant les imperfections de surface, les variations de couleur ou les dommages physiques. Cette automatisation permet d'assurer une qualité constante et d'éliminer les produits défectueux avant qu'ils n'atteignent les consommateurs [140].

En médecine, l'imagerie et la vision par ordinateur jouent un rôle crucial dans l'analyse des images médicales, telles que les radiographies et les IRM. Les algorithmes peuvent détecter des

7. Source : https://www.fourni-labo.fr/produit/astree-langue-electronique, consultée le 20 novembre 2024.

8. Source : https://fr.made-in-china.com/co_065b058624f10088/product_Electronic-Tongue-Fully-Automatic-Electronic-Tongue_uohgshyesg.html, consultée le 20 novembre 2024.

anomalies subtiles, aidant les médecins à diagnostiquer des affections plus rapidement et avec une précision accrue [141].

Ces technologies sont également utilisées dans l'industrie automobile pour vérifier l'assemblage correct des pièces et s'assurer que les véhicules répondent aux normes de sécurité les plus strictes. La vision par ordinateur permet de détecter les écarts ou les défauts de fabrication qui pourraient passer inaperçus lors d'une inspection humaine.

Les Figures 2.11 à 2.14 présentent quelques dispositifs d'imagerie et de vision par ordinateur créés par différentes équipes de recherche et entreprises innovantes.



FIGURE 2.11 – X-TRACT[9]



FIGURE 2.12 – Compass optical sorter[10]



FIGURE 2.13 – X36 Series[11]



FIGURE 2.14 – FoodScan™ [12]

9. Source : https://www.recyclingproductnews.com/product/3213/x-tract, consultée le 20 novembre 2024.

10. Source : https://www.snackandbakery.com/articles/111699-key-technology-ppm-technologies-consolidate, consultée le 20 novembre 2024.

11. Source : https://www.mt.com/ca/fr/home/products/Product-Inspection_1/safeline-x-ray-inspection/X36.html, consultée le 20 novembre 2024.

12. Source : https://www2.fossanalytics.com/l/507171/2019-01-14/cvp68b/507171/168485/FoodScan_2_Dairy_Solution_Brochure_EN.pdf, consultée le 20 novembre 2024.

2.5 Impact environnemental et économique

Le contrôle de qualité a un impact significatif à la fois sur l'environnement et l'économie, particulièrement dans les secteurs industriels. D'un point de vue environnemental, il contribue à réduire les émissions de gaz à effet de serre et de polluants atmosphériques en optimisant les processus de production. Cela permet de minimiser le gaspillage de ressources naturelles et de limiter la production de déchets, conformément aux objectifs de la hiérarchie des déchets de l'Union européenne, qui vise à prévenir leur génération et à promouvoir leur recyclage plutôt que leur enfouissement ou leur incinération [142, 143].

Sur le plan économique, l'intégration de méthodologies avancées de contrôle, telles que les outils de gestion de l'impact environnemental, peut engendrer des économies substantielles. Ces outils permettent de mieux anticiper les coûts liés aux externalités, comme les dommages environnementaux ou les impacts sur la santé humaine. Par exemple, les coûts des pollutions industrielles en Europe, liés notamment à des émissions de particules fines et de dioxyde de soufre, sont estimés entre 277 et 433 milliards d'euros annuels [144].

Ces efforts s'inscrivent dans le cadre d'initiatives telles que le "Green Deal" européen et le "Plan d'action zéro pollution", qui visent à aligner croissance économique et protection de l'environnement, tout en améliorant les performances industrielles grâce à des technologies plus propres [144, 142].

Conclusion

Le contrôle de qualité est une composante cruciale pour garantir la satisfaction des consommateurs et la sécurité des produits. L'évolution des méthodes et technologies, telles que le nez électronique et la langue électronique, a permis d'améliorer considérablement la précision et l'efficacité des processus de contrôle de qualité. Cependant, des défis subsistent, et l'innovation continue est nécessaire pour répondre aux exigences croissantes des consommateurs et des régulateurs.

Le chapitre suivant mettra l'accent sur la technologie du nez électronique en détaillant les outils, les applications et les techniques de classification utilisés.

CHAPITRE 3

NEZ ÉLECTRONIQUE ET TRAITEMENT DE DONNÉES

Introduction

Dans l'industrie technologique, les normes internationales jouent un rôle clé dans la conception de biens fiables et utilisables au marché mondial. Les experts peuvent concevoir des articles en s'appuyant sur les exigences de performance définies par ces normes afin de maximiser la fiabilité et la sécurité des nouveaux produits. Par exemple, les normes ont été essentielles dans le développement de produits liés aux sens de la vue, de l'ouïe et du toucher. Cependant, il n'existe pas encore de normes officielles pour les dispositifs électroniques qui utilisent les sens du goût et de l'odorat, contrairement à ceux relatifs au toucher, à la vue et à l'ouïe [11].

L'odorat a joué un rôle essentiel tout au long de l'histoire dans de nombreuses professions, notamment la médecine, la parfumerie, les arts culinaires, l'agriculture, la propreté et l'assainissement [11, 145]. La qualité sensorielle des produits était la base sur laquelle les marchands les commercialisaient. Par exemple, l'évaluation de la qualité des produits agricoles et alimentaires est encore souvent réalisée en fonction des préférences des consommateurs, par inspection visuelle et appréciation subjective [12]. Dans le même article, il est également noté que chaque produit agricole dégage des arômes uniques composés de mélanges distincts [12]. De plus, conformément à [146], il existe une corrélation directe entre l'arôme et la qualité du produit, attribuable aux préférences individuelles de chaque consommateur pour un parfum spécifique.

Actuellement, des experts humains spécialisés dans la description et l'évaluation des sensations gustatives et olfactives sont utilisés dans de nombreux environnements industriels comme "outil de mesure" pour l'assurance qualité et le développement de produits. Bien que les panels d'odeur et de goût fournissent des informations sensorielles précieuses, leur utilisation est limitée [11, 12]. Premièrement, des panélistes humains qualifiés ne sont pas toujours disponibles lorsque cela est nécessaire, et il peut être dangereux d'examiner des échantillons tels que des médicaments contaminés ou des odeurs de gaz nocifs. Deuxièmement, la diversité entre les panélistes est influencée par des facteurs tels que l'héritage génétique et les variations acquises du sens du

goût et de l'odorat. Troisièmement, en raison de l'épuisement sensoriel, qui peut compromettre la fiabilité et la validité des résultats, le nombre d'échantillons qu'un humain peut évaluer simultanément reste limité. Ainsi, il est essentiel de développer des outils d'évaluation rapide et précise des odeurs et des goûts qui peuvent compléter les panels sensoriels humains [11].

Sur la base des travaux mentionnés où, plusieurs technologies ont été développées en combinant l'intelligence artificielle et l'électronique, donnant naissance à des dispositifs appelés langues électroniques et nez électroniques [147, 148, 149].

Ce chapitre examine les travaux existants sur les technologies de nez électronique. Une attention particulière est accordée aux dispositifs actuellement disponibles, suivie d'une analyse des techniques de traitement des données appliquées à ces systèmes.

3.1 Technologies du nez électronique

La technologie du nez électronique a considérablement progressé depuis son apparition en 1982, couvrant à la fois des domaines techniques et des applications commercialement viables [146]. Un nez électronique est un dispositif de détection électronique intelligent non destructif capable de détecter, distinguer et catégoriser des variétés d'odeurs en simulant le système olfactif humain [150]. Le nez électronique classique était constitué d'un ordinateur doté d'un algorithme de reconnaissance de motifs, d'un ensemble de capteurs à gaz, et d'un convertisseur analogique-numérique [151]. En résumé, un nez électronique se compose d'un ensemble de capteurs chimiques, d'un dispositif d'échantillonnage de vapeur, et d'une technique de traitement des signaux permettant de classifier les réponses des capteurs chimiques. Les nez électroniques ne sont pas seulement spécifiques à un type de produit chimique volatil, mais ont également une large sélectivité [152]. Grâce cette technologie, une nouvelle méthode permet désormais d'évaluer de manière non destructive la qualité des produits agricoles et alimentaires [12]. Aujourd'hui, de nombreuses personnes et industries utilisent les nez électroniques.

Les nez électroniques peuvent être classés en deux catégories : les nez électroniques expérimentaux et les nez électroniques commerciaux. Les premiers sont conçus pour des activités de recherche, tandis que les seconds sont fabriqués et vendus par des entreprises. Parmi les nez électroniques commerciaux, on trouve le Fox 2000, le Fox 3000 et le Fox 4000 (Alpha MOS, France), l'Airsense PEN2 et l'Airsense PEN3 (Airsense Analytics, Allemagne), le Cyranose 320 (Cyrano Sciences Inc., États-Unis), et le zNose (Electronic Sensor Technology, États-Unis). En revanche, les nez électroniques expérimentaux sont configurés pour répondre à des besoins spécifiques, selon les capteurs sélectionnés.

Le Tableau 3.1 de l'étude [3] présente une liste complète des différentes technologies de nez électronique disponibles. Actuellement, quatre technologies principales sont utilisées dans les nez électroniques. Parmi celles-ci, deux types de capteurs fonctionnent à haute température : les semi-conducteurs à oxyde métallique (MOS : Metal-Oxide-Semiconductor) et les transistors à effet de champ à semi-conducteur à oxyde métallique (MOSFET : Metal-Oxide-Semiconductor Field-

Effect Transistor). Ces capteurs présentent un faible taux de transfert d'une mesure à l'autre et sont moins sensibles à l'humidité. Les deux autres technologies, les capteurs à ondes acoustiques de masse (BAW : Bulk Acoustic Wave) et les capteurs à ondes acoustiques de surface (SAW : Surface Acoustic Wave), sont constituées de cristaux piézoélectriques et de polymères organiques conducteurs. Ces capteurs fonctionnent à basse température et sont appelés capteurs de froid [153].

Les différents types de capteurs de nez électronique présentent des avantages et des inconvénients. Ceux-ci sont liés à leur durée de réponse et de récupération, leur sensibilité, leur gamme de détection, leurs contraintes opérationnelles, leur taille physique, leur susceptibilité à l'inactivation par certains agents toxiques, ainsi qu'à d'autres limites spécifiques à chaque type de capteur. La technologie qui sous-tend le principe de détection et les types de gaz analysés est étroitement liée aux catégories et types de capteurs de nez électronique. Le Tableau 3.2 de l'étude [3] offre un aperçu des principaux avantages et inconvénients associés à chaque type de capteur.

Selon l'expérience, un nez électronique peut comporter un ou plusieurs capteurs. Le nombre de capteurs utilisés dans les nez électroniques expérimentaux et commerciaux à des fins spécifiques est indiqué dans les Tableaux 3.3 et 3.4.

Le nombre de capteurs utilisés dans un travail de recherche dépend des substances spécifiques à détecter et varie d'une expérience à l'autre. Wilson et Baietto ont remarqué que les capteurs à oxydes métalliques sont utilisés dans la plupart des travaux de recherche sur les nez électroniques [3].

Avant le développement des technologies de nez électronique décrites plus haut, la chromatographie en phase gazeuse (GC : Gas Chromatography) et la spectrométrie de masse (MS : Mass Spectrometry) étaient principalement utilisées pour analyser les produits chimiques volatils. Le coût de mise en œuvre de ces technologies est supérieur à celui de l'implémentation d'un nez électronique [171]. De plus, ce dernier est léger, rapide, robuste et portable. Pour cette raison, la chromatographie en phase gazeuse et la spectrométrie de masse sont utilisées de manière sporadique ou en complément avec la technologie du nez électronique [172, 173].

3.2 Analyse des données du nez électronique

Pour produire des informations utiles, les sorties numériques générées par les capteurs de nez électronique doivent être évaluées et interprétées [174]. Lors de la configuration d'un nez électronique, le choix de la technique utilisée dépend des données collectées et des objectifs à atteindre. Une liste actualisée des méthodes d'analyse utilisées est présentée ci-dessous. Ces méthodes sont généralement classées en paramétriques ou non paramétriques, et supervisées ou non supervisées [175]. Elles peuvent être réparties en deux groupes :

— Les techniques d'analyse de motifs statistiques qui comprennent l'analyse en composantes principales (PCA : Principal Component Analysis), l'analyse discriminante linéaire (LDA : Linear Discriminant Analysis), la régression partielle des moindres carrés (PLS : Partial Least Squares), l'analyse de fonction discriminante (DFA : Discriminant Function Analysis), et

TABLEAU 3.1 – Types et mécanismes des capteurs de gaz couramment utilisés dans les nez électroniques

Type de capteur	Matériel sensible	Principe de détection
Capteurs acoustiques : microbalance à cristal de quartz (QCM) & ondes acoustiques de surface et de masse (SAW, BAW)	Couches de films organiques ou inorganiques	Changement de masse (décalage de fréquence)
Calorimétrique ; perle catalytique (CB)	Pellistor	Variation de température ou de chaleur (provenant de réactions chimiques)
Capteurs catalytiques à effet de champ (MOSFET)	Métaux catalytiques	Variation du champ électrique
Capteurs colorimétriques	Colorants organiques	Changement de couleur, absorbance
Capteurs à polymères conducteurs	Polymères conducteurs modifiés	Variation de résistance
Capteurs électrochimiques	Electrolytes solides ou liquides	Variation de courant ou de tension
Capteurs de fluorescence	Détecteur sensible à la fluorescence	Émissions de lumière fluorescente
Capteurs infrarouges	Détecteur sensible aux IR	Absorption du rayonnement infrarouge
Oxydes métalliques semiconducteurs (MOS, Taguchi)	Oxydes métalliques semiconducteurs dopés (SnO2, GaO)	Variation de résistance
Capteurs optiques	Photodiode, sensible à la lumière	Modulation de la lumière, changement optiques

TABLEAU 3.2 – Résumé des avantages et inconvénients des types de capteurs de nez électronique

Types de capteurs	Avantages	Inconvenients
Calorimétrique ou perle catalytique (CB)	Temps de réponse et de récupération rapide, haute spécificité pour les composés oxydés	Fonctionnement à haute température, sensible uniquement aux composés contenant de l'oxygène
Capteurs catalytiques à effet de champ (MOSFET)	Capteur de petite taille, coûts d'exploitation peu élevés	Nécessite un contrôle environnemental, dérive de la ligne de base, faible sensibilité à l'ammoniac et au dioxyde de carbone
Capteurs à polymère conducteur (CP)	Fonctionnement à température ambiante, sensibles à de nombreux COV, temps de réponse court, divers revêtements de capteurs, peu coûteux, résistance à l'empoisonnement des capteurs	Sensibles à l'humidité et à la température, les capteurs peuvent être surchargés par certains analytes, la durée de vie des capteurs est limitée.
Capteurs électrochimiques (EC)	Fonctionnement à température ambiante, faible consommation d'énergie, très sensible à divers COV	Taille encombrante, sensibilité limitée aux gaz simples ou de faible poids moléculaire

Types de capteurs	Avantages	Inconvenients
Capteurs MOS	Très haute sensibilité, plage de détection limitée, temps de réponse et de récupération rapides pour les composés de faible poids moléculaire (mais pas pour les composés de poids élevé)	Fonctionnement à haute température, consommation d'énergie élevée, empoisonnement par le soufre et les acides faibles, revêtements de capteurs limités, sensible à l'humidité, précision médiocre
Capteurs optiques	Très haute sensibilité, capacité à identifier des composés individuels dans des mélanges, capacités de détection multi-paramètres	Systèmes de capteurs complexes, plus coûteux à utiliser, faible portabilité en raison des composants optiques délicats et des composants électriques
Microbalance à cristal de quartz (QCM)	Bonne précision, large gamme de revêtements de capteurs, haute sensibilité	Circuits complexes, faible rapport signal/bruit, sensible à l'humidité et à la température
Ondes acoustiques de surface (SAW)	Haute sensibilité, bon temps de réponse, divers revêtements de capteurs, petits, peu coûteux, sensibles à pratiquement tous les gaz	Circuits complexes, sensibles à la température, spécificité des groupes d'analytes affectée par le revêtement du capteur en film polymère

TABLEAU 3.3 – Modèle de nez électronique expérimental

Objectif	Nombre de capteurs	Technologie	Référence
Distinguer le lait contaminé	6	capteurs BAW	[154]
Séparation du lait UHT et du lait pasteurisé; séparation par niveaux de rancidité	4 ou 5	capteurs MOS	[155, 156]
Classification du vin	8	capteurs MOS	[157]
Classification du vin	8	capteurs SAW	[158]

TABLEAU 3.4 – Modèle de nez électronique commercial

Modèle	Nombre de capteurs	Technologie	Référence
Fox 2000 (Alpha MOS, France)	6	capteurs MOS	[159, 160]
Fox 3000 (Alpha MOS, France)	12	capteurs MOS	[161, 162]
Fox 4000 (Alpha MOS, France)	18	capteurs MOS	[163, 164]
PEN2 (Airsense Analytics, Germany)	10	capteurs MOS	[165, 166]
PEN3 (Airsense Analytics, Germany)	10	capteurs MOS	[167]
Cyranose 320 (Cyrano Sciences Inc., États-Unis)	32	capteurs CP	[168]
EOS 835 (Sacmi Imola S.C., Italy)	12	capteurs MOS	[169]
zNose (Electronic Sensor Technology, États-Unis)	1	capteurs SAW	[170]

l'analyse de clusters (CA : Cluster Analysis).

— Les techniques d'analyse de motifs intelligentes qui incluent la machine à vecteurs de support (SVM : Support Vector Machine), le k plus proches voisins (k-NN : k-Nearest Neighbors), la logique floue (FL : Fuzzy Logic), et le réseau de neurones artificiels (ANN : Artificial Neural Network).

3.2.1 Techniques statistiques

La chimiométrie, un domaine des mathématiques appliquées, est aujourd'hui reconnue comme le berceau des premiers développements et applications des techniques statistiques classiques utilisant des modèles de probabilité [175]. Cette section présente certaines techniques mathématiques qui peuvent être utilisées dans l'analyse multi-composants des odeurs.

3.2.1.1 Analyse en composantes principales

Principal Component Analysis (PCA) est une approche mathématique qui permet de réduire la dimensionnalité des données tout en préservant la majorité de leurs variations [34]. La réduction de la dimensionnalité est cruciale pour le nez électronique car elle réduit considérablement le nombre de processus nécessaires au post-traitement lorsque l'on travaille avec une dimension réduite [176]. PCA, qui fait partie de la famille des méthodes de réduction de dimension, est particulièrement utile lorsque les données sont volumineuses (c'est-à-dire comportent de nombreuses variables), vastes (c'est-à-dire comportent de nombreuses observations par variable) et fortement corrélées. L'objectif est de trouver un ensemble de caractéristiques plus petit qui représente fidèlement les données d'origine dans un sous-espace de dimension inférieure, en perdant le moins d'informations possible. Les données peuvent être résumées et les informations sur les différences individuelles peuvent être extraites à l'aide de PCA et de techniques apparentées [177]. Les échantillons peuvent ensuite être représentés graphiquement, ce qui permet de comparer visuellement les similitudes et les différences entre les échantillons et de déterminer les groupes d'échantillons [178].

Les étapes de PCA [179] sont les suivantes :

1. normaliser l'ensemble de données ;
2. calculer la matrice de covariance pour les caractéristiques de l'ensemble de données ;
3. calculer les valeurs propres et les vecteurs propres de la matrice de covariance ;
4. trier les valeurs propres et les vecteurs propres correspondants ;
5. prendre k valeurs propres et former une matrice de vecteurs propres ;
6. transformer la matrice originale.

3.2.1.2 Analyse discriminante linéaire

Une technique fondamentale d'analyse des données appelée analyse discriminante linéaire (LDA : Linear Discriminant Analysis) a été mise au point pour identifier un sous-espace dont les dimensions sont inférieures à celles de l'échantillon de données d'origine. Les points de données du problème initial peuvent être séparés dans ce sous-espace en fonction des mesures statistiques de la moyenne et de la variance [180]. Il s'agit d'identifier l'hyperplan de projection qui maximise la séparation entre les moyennes projetées des classes et minimise la variance au sein des classes. Ces deux objectifs peuvent être atteints de manière similaire à PCA en résolvant un problème de valeurs propres, le vecteur propre correspondant définissant l'hyperplan d'intérêt. Cet hyperplan peut être utilisé pour la catégorisation, la réduction de la dimensionnalité et l'évaluation de la pertinence des caractéristiques fournies. La principale distinction entre LDA et PCA est que LDA classifie les données, tandis que PCA se concentre davantage sur la réduction des dimensions des caractéristiques [181]. L'analyse discriminante linéaire par étapes (SLDA), l'une des trois variantes de LDA, en est une illustration [182].

Voici une liste des cinq étapes générales pour effectuer une analyse discriminante linéaire après avoir préparé le jeu de données [183] :

1. Calculer les vecteurs moyens $d-dimensionnels$ pour les différentes classes de l'ensemble de données ;
2. Calculer les matrices de dispersion (entre les classes et à l'intérieur des classes) ;
3. Calculer les vecteurs propres $(e_1, e_2, \ldots, e_d)$ et les valeurs propres correspondantes $(\lambda_1, \lambda_2, \ldots, \lambda_d)$ pour les matrices de dispersion ;
4. Trier les vecteurs propres par valeurs propres décroissantes et choisir k vecteurs propres avec les plus grandes valeurs propres pour former une matrice $\mathscr{W}$ de dimension $d \times k$ (où chaque colonne représente un vecteur propre) ;
5. Utiliser cette matrice de vecteurs propres $d \times k$ pour transformer les échantillons dans le nouveau sous-espace. Ceci peut être résumé par la multiplication matricielle : $\mathscr{Y} = \mathscr{X} \times \mathscr{W}$, où $\mathscr{X}$ est une matrice de $n \times d-dimension$ représentant les n échantillons, et $\mathscr{Y}$ sont les échantillons transformés de $n \times k-dimension$ dans le nouveau sous-espace.

3.2.1.3 Moindres carrés partiels

La méthode des moindres carrés partiels (PLS : Partial Least Squares) est une technique permettant de construire des modèles prédictifs lorsque les composantes sont nombreuses et fortement colinéaires. Il convient de noter que la prévision des résultats prime sur la compréhension des relations sous-jacentes entre les variables. Par exemple, la méthode PLS n'est généralement pas adaptée à l'élimination des variables ayant un faible impact sur la réponse. La PLS peut néanmoins être un outil utile lorsque l'objectif est la prédiction et qu'il n'est pas nécessaire de restreindre le nombre de facteurs évalués. En chimiométrie, l'analyse PLS est fréquemment utilisée,

en particulier lorsque le nombre de variables indépendantes est bien supérieur à celui des points de données [184]. La régression PLS, souvent appelée régression des moindres carrés partiels, est le modèle des moindres carrés partiels le plus largement utilisé. Les autres modèles de la famille PLS sont dérivés du modèle PLSR (Partial Least Squares Regression), tels que l'analyse discriminante des moindres carrés partiels (PLS-DA : Partial Least Squares Discriminant Analysis) [185, 186].

Les étapes de l'algorithme PLSR "standard" ou "traditionnel" sont les suivantes. Les variables **X** et **y** sont supposées être centrées sur la moyenne et standardisées respectivement [187].

1. Commencez par définir : $\mathbf{X_0} = X$, et $\mathbf{y_0} = y$;
2. Répéter pour $h = 1, \ldots, r = rank(\mathbf{X})$:
   (a) Commencez avec les poids : $\tilde{\mathbf{w}}_h = \mathbf{X}_{h-1}^{\top}\mathbf{y}_{h-1} / \mathbf{y}_{h-1}{}^{\top}\mathbf{y}_{h-1}$
   (b) Normaliser les poids : $\mathbf{w}_h = \tilde{\mathbf{w}}_h / \parallel \tilde{\mathbf{w}}_h \parallel$
   (c) Calculer la composante PLS : $\mathbf{z}_h = \mathbf{X}_{h-1}\mathbf{w}_h / \mathbf{w}_h{}^{\top}\mathbf{w}_h$
   (d) Régresser la réponse $\mathbf{y}_h$ onto $\mathbf{z}_h$: $b_h = \mathbf{y}_h{}^{\top}\mathbf{z}_h / \mathbf{z}_h{}^{\top}\mathbf{z}_h$
   (e) Régresser toutes les entrées $\mathbf{x}_j$ onto $\mathbf{z}_h$: $\mathbf{v}_h = \mathbf{X}_{h-1}{}^{\top}\mathbf{z}_h / \mathbf{z}_h{}^{\top}\mathbf{z}_h$
   (f) Déflater les prédicteurs (résiduels) : $\mathbf{X}_h = \mathbf{X}_{h-1} - \mathbf{z}_h\mathbf{v}_h^{\top}$
   (g) Déflater la réponse (résiduel) : $\mathbf{y}_h = \mathbf{y}_{h-1} - b_h\mathbf{z}_h$

3.2.1.4 Analyse discriminante fonctionnelle

Lorsque les conditions d'une régression linéaire sont remplies, l'analyse de la fonction discriminante (DFA : Discriminant Function Analysis) peut être appliquée. À l'aide d'une équation de prédiction discriminante, les instances sont divisées en groupes afin d'évaluer les variations au sein des groupes ou entre eux. Elle permet de déterminer la méthode la plus efficace pour différencier les groupes d'éléments, ainsi que de rejeter les variables ayant peu de rapport avec les distinctions entre les groupes. L'analyse discriminante se déroule en deux étapes. La première étape, le test F, est utilisée pour déterminer la significativité du modèle discriminant dans son ensemble. Lorsque le test F indique que le modèle est significatif, la deuxième étape peut être réalisée. La moyenne de groupe de chaque variable indépendante est comparée à la moyenne individuelle avant d'être utilisée pour catégoriser la variable dépendante [182].

3.2.1.5 Analyse de clusters

Cluster Analysis (CA) permet de classer les éléments décrits par de multiples caractéristiques en groupes relativement homogènes. Le développement de clusters, ou d'ensembles d'objets pour lesquels deux objets du même ensemble sont plus similaires que n'importe quel objet d'un autre ensemble, est l'aspect le plus crucial du CA. Comme les grappes ne se croisent pas, aucun élément ne peut appartenir à plus d'une classe. L'espacement entre les grappes est calculé avec précision et peut être exprimé de différentes manières. Il existe deux types de techniques de regroupement : les

techniques non hiérarchiques, qui déplacent les objets d'une grappe à l'autre dans le but de trouver la meilleure solution pour le critère donné, et les techniques hiérarchiques, qui permettent de générer une hiérarchie de grappes sur la base de la distance entre celles-ci. Les approches hiérarchiques et non hiérarchiques se réfèrent respectivement à l'agglomération et au regroupement par k-means [182].

La procédure décrite ci-dessous peut être divisée en une analyse de regroupement classique [188].

1. **Représentation des motifs** (comprenant éventuellement l'extraction et/ou la sélection des caractéristiques) : L'objectif est de déterminer quelles caractéristiques seront utilisées pour le regroupement. Celles-ci doivent encoder autant d'informations que possible.
2. **Mesure de similarité :** Définition d'une mesure de proximité des motifs adaptée au domaine des données : la mesure de similarité calcule le degré de similarité entre deux motifs ou points de données. Dans la plupart des cas, il est essentiel de s'assurer que toutes les caractéristiques choisies contribuent de manière égale au calcul de la mesure de proximité et qu'aucune caractéristique ne prédomine sur les autres.
3. **Algorithme de clustering :** L'algorithme de regroupement est choisi à cette étape. Il doit conduire à la définition d'une stratégie de regroupement efficace pour la collection de données étudiée. Pour un traitement rapide, l'algorithme de clustering est également crucial.
4. **Evaluation du résultat :** Étant donné que les groupes définis par les algorithmes de regroupement ne sont pas connus a priori, une certaine forme d'évaluation est nécessaire avant que les données ne soient finalement réparties. Les critères et procédures appropriés doivent être appliqués afin de déterminer si les résultats d'un algorithme de regroupement sont exacts. Ces méthodes, regroupées sous le terme générique de "méthodes de validité des grappes", visent à évaluer quantitativement les résultats des algorithmes de regroupement. Elles répondent à des questions telles que : "combien de grappes y a-t-il dans l'ensemble de données ?", "le schéma de regroupement obtenu est-il adapté à cet ensemble de données ?", "existe-t-il un meilleur partitionnement pour cet ensemble de données ?" et "quelle est la cohérence ou la robustesse des grappes lors du rééchantillonnage des données ?"
5. **Représentation graphique :** Pour faciliter l'interprétation, les résultats des grappes doivent être représentés graphiquement et résumés d'une manière ou d'une autre.

3.2.2 Techniques intelligentes

Lors de la caractérisation des données du nez électronique, il est souvent préférable d'utiliser une méthode de reconnaissance des formes (PARC) plus puissante, capable de traiter des données non linéaires et présentant des avantages supplémentaires par rapport aux méthodes plus traditionnelles, notamment la capacité d'apprendre, de s'auto-organiser, de se généraliser et de tolérer le bruit. Il existe toujours un attrait intellectuel à travailler sur des systèmes PARC d'inspiration physiologique qui imitent le cerveau humain en apprenant à partir de modèles, lorsque

l'objectif est de construire un nez électronique qui ressemble au système olfactif humain. L'intérêt récent pour l'apprentissage guidé par les données a conduit à la création de techniques d'inspiration biologique [175]. Un examen approfondi de plusieurs d'entre elles est présenté dans la section suivante.

3.2.2.1 Machine à vecteurs de support

Support Vector Machine (SVM) est une méthode de classification et de régression qui évite automatiquement le surajustement des données tout en maximisant la précision de la prédiction. Il fait partie de la famille des classificateurs linéaires généralisés [189]. La machine à vecteurs de support (SVM) de l'algorithme d'apprentissage supervisé décrit un modèle mathématique capable de faire des prédictions. À l'instar des ANN, l'approche SVM utilise un ensemble d'apprentissage pour affecter des objets à des classes particulières. SVM est un substitut des ANN car il permet, par exemple, une meilleure classification. L'idée fondamentale qui sous-tend les SVM est la construction d'un hyperplan idéal avec la marge la plus large possible qui diviserait les données des classes opposées. La technique SVM suppose que l'approche optimale pour diviser les deux classes d'ensembles de données est un hyperplan à marge maximale. La marge est la séparation entre les vecteurs de support et l'hyperplan. Ces ensembles de données soutiennent les vecteurs de support, qui sont des hyperplans divisant deux classes de points de données. La régression vectorielle de support (SVR) et la régression vectorielle de support par moindres carrés (LSSVR) sont deux variantes de cette technique qui viennent s'ajouter à la machine vectorielle de support conventionnelle [182].

Voici quelques façons mathématiques de comprendre le fonctionnement du classificateur SVM [190].

1. L'algorithme SVM prédit les classes. L'une des classes est identifiée comme étant 1, tandis que l'autre est identifiée comme étant -1.
2. Une fonction de perte, également connue sous le nom de fonction de perte de la charnière, est utilisée et ajustée pour trouver la marge maximale.

   Fonction hinge loss :

   $$C(x, y, f(x)) = \begin{cases} 0 & \text{si } yf(x) \geq 1, \\ 1 - yf(x) & \text{sinon} \end{cases} \tag{3.1}$$

   où :
   — x est le vecteur de caractéristiques de l'exemple de données.
   — y est l'étiquette de classe réelle de l'exemple, prenant les valeurs $+1$ ou -1.
   — $f(x)$ est la fonction de décision ou score prédictif produit par le modèle SVM pour l'entrée x.
3. Cette fonction de perte peut également être considérée comme une fonction de coût, dont le coût est nul lorsqu'aucune classe n'est prédite à tort. L'erreur/la perte est alors évaluée

si ce n'est pas le cas. Le problème de la situation actuelle est qu'il existe un conflit entre la maximisation de la marge et la perte qui résulte d'une telle maximisation à l'extrême. L'ajout d'un paramètre de régularisation fonde théoriquement ces concepts.

Fonction de perte :

$$\min \lambda \| w \|^2 + \sum_{i=1}^{n} \left(1 - y_i \langle x_i, w \rangle\right)_+ \tag{3.2}$$

où :

— w est le vecteur des poids ou coefficients du modèle.
— λ est le paramètre de régularisation, qui équilibre la maximisation de la marge et la minimisation des erreurs.
— $\| w \|^2$ est la norme euclidienne au carré du vecteur des poids, qui pénalise les modèles complexes.
— n est le nombre d'exemples de données dans l'ensemble d'entraînement.
— y_i est l'étiquette de classe réelle de l'exemple i, avec $y_i \in \{-1, +1\}$.
— x_i est le vecteur de caractéristiques de l'exemple i.
— $\langle x_i, w \rangle$ est le produit scalaire entre x_i et w.
— $(1 - y_i \langle x_i, w \rangle)_+$ est la fonction hinge loss, où la notation $(\cdot)_+$ indique que l'on prend le maximum entre 0 et l'expression à l'intérieur.

4. Les poids sont optimisés en calculant les gradients à l'aide d'idées de calcul avancées, à savoir les dérivées partielles, comme c'est le cas pour la majorité des problèmes d'optimisation.

Gradients :

$$\frac{\partial}{\partial w_k} \left(\lambda \| w \|^2\right) = 2 \lambda w_k \tag{3.3}$$

où :

— w_k est le k-ème composant du vecteur des poids w.
— λ est le paramètre de régularisation.
— $\| w \|^2$ est la norme euclidienne au carré du vecteur des poids.
— $\frac{\partial}{\partial w_k}$ est l'opérateur de dérivée partielle par rapport au poids w_k.

$$\frac{\partial}{\partial w_k} \left(1 - y_i \langle x_i, w \rangle\right) = \begin{cases} 0 & \text{si } y_i \langle x_i, w \rangle \geq 1, \\ -y_i x_{ik} & \text{sinon} \end{cases} \tag{3.4}$$

où :

— y_i est l'étiquette de classe réelle de l'exemple i, prenant les valeurs +1 ou −1.
— x_{ik} est le k-ème composant du vecteur de caractéristiques x_i de l'exemple i.
— $\langle x_i, w \rangle$ est le produit scalaire entre le vecteur de caractéristiques x_i et le vecteur des poids w.
— $\frac{\partial}{\partial w_k}$ est l'opérateur de dérivée partielle par rapport au poids w_k.

5. Lorsqu'il n'y a pas d'erreur de classification, les gradients sont simplement mis à jour à l'aide du paramètre de régularisation ; en cas de mauvaise classification, la fonction de perte est également utilisée.

   Mise à jour des gradients en l'absence de mauvaise classification :

$$w = w - \alpha(2\lambda w) \tag{3.5}$$

   où :
   — w est le vecteur des poids du modèle.
   — α est le taux d'apprentissage (learning rate), qui détermine la taille des pas lors de la mise à jour des poids.
   — λ est le paramètre de régularisation, qui contrôle l'importance de la régularisation par rapport à la perte.
   — $2\lambda w$ est le gradient de la fonction de régularisation $\lambda \| w\|^2$ par rapport aux poids w.

   Mise à jour des gradients en cas de mauvaise classification :

$$w = w - \alpha\left[(2\lambda w) + (y_i x_i - 2\lambda w)\right] \tag{3.6}$$

   où :
   — w est le vecteur des poids du modèle.
   — α est le taux d'apprentissage (learning rate), qui détermine la taille des pas lors de la mise à jour des poids.
   — λ est le paramètre de régularisation, qui contrôle l'importance de la régularisation par rapport à la perte.
   — y_i est l'étiquette de classe réelle pour l'exemple i.
   — x_i est le vecteur de caractéristiques de l'exemple i.
   — $y_i x_i$ représente la contribution de l'exemple i à la perte lorsque la classification est incorrecte.
   — $2\lambda w$ est le gradient de la fonction de régularisation par rapport aux poids.

6. Lorsqu'il n'y a pas d'erreur de classification, les gradients sont simplement mis à jour à l'aide du paramètre de régularisation ; en cas de mauvaise classification, la fonction de perte est également utilisée.

3.2.2.2 k plus proches voisins

L'algorithme k-Nearest Neighbors (k-NN) est l'un des algorithmes qui effectue la description de la fonction d'objectif du classificateur après que celui-ci a été formé, plutôt qu'avant de classer un objet. Le principe fondamental qui guide le fonctionnement du classificateur est qu'un objet appartenant à une classe est entouré d'autres objets de cette même classe [191]. L'objet à classer est comparé à tous les objets de l'ensemble d'apprentissage et les k objets présentant la plus

grande similarité sont choisis dans ce groupe pour effectuer la classification. La distance euclidienne est souvent utilisée pour évaluer le degré de similitude entre deux vecteurs de caractéristiques; les éléments qui sont séparés par la plus petite distance euclidienne sont jugés proches. Parmi les k voisins choisis, l'objet est classé dans la classe représentée par le plus grand nombre de ces objets. Lorsque le même nombre de voisins appartient à plus d'une classe, la classe dont les voisins sont les plus proches est choisie. Pour obtenir la meilleure catégorisation pour un ensemble de données donné, le paramètre k est choisi de manière empirique [182].

L'algorithme suivant peut être utilisé pour décrire le fonctionnement du k-NN [192] :

1. Charger les données.
2. Initialiser le nombre k de voisins.
3. Pour chaque exemple dans les données :
   (a) Calculer la distance entre l'exemple de la requête et l'exemple actuel à partir des données.
   (b) Ajouter la distance et l'indice de l'exemple à une collection ordonnée.
4. Trier la collection ordonnée de distances et d'indices du plus petit au plus grand (par ordre croissant) en fonction des distances.
5. Choisir les k premières entrées de la collection triée.
6. Obtenir les étiquettes des k entrées sélectionnées.
7. Si régression, renvoie la moyenne des k étiquettes.
8. Si classification, renvoie le mode des k étiquettes.

3.2.2.3 Logique floue

La théorie des ensembles flous (FST : Fuzzy Set Theory) introduite par Zadeh [193] fournit un mécanisme pour traiter les ambiguïtés linguistiques et le langage imprécis utilisé par les gens (par exemple *chaud, froid, lent, assez lent*). Un ensemble dont les limites ne sont pas clairement définies est appelé ensemble flou. Des tentatives d'utilisation de fonctions floues pour identifier les odeurs ont été faites, et la logique floue (FL : Fuzzy Logic) a été appliquée à l'analyse des modèles de nez électroniques. Dans le domaine de l'olfaction automatique, où les échantillons d'odeurs sont décrits par des descripteurs olfactifs tels que poivré, fleuri ou sucré, ainsi que par des caractéristiques d'intensité telles que assez, très ou puissant, FST est donc attrayante [194]. Dans cette théorie, plusieurs algorithmes existent, tels que les systèmes d'inférence flous (FIS : Fuzzy Inference Systems) [195], les réseaux neuronaux flous (FNN : Fuzzy Neural Network) [196], les c-moyens flous (FCM : Fuzzy C-means) [197], la théorie de la résonance adaptative (ART : Adaptive Resonance Theory) [198], et l'ARTMAP flou [199]. Ils sont utilisés dans l'analyse des données obtenues par les sens électroniques.

Pour mettre en place un système FL, il faut prendre les mesures suivantes [200] :

1. Définir les variables et les termes linguistiques.

2. Construire des fonctions d'appartenance pour eux.
3. Construire les règles de la base de connaissances.
4. Convertir des données brutes en ensembles de données floues à l'aide de fonctions d'appartenance (fuzzification).
5. Évaluer les règles de la base de règles (moteur d'inférence).
6. Combiner les résultats de chaque règle (moteur d'inférence).
7. Convertir les données de sortie en valeurs non floues (défuzzification).

3.2.2.4 Réseau de neurones artificiels

Un modèle informatique d'inspiration biologique appelé réseau de neurones artificiels (ANN : Artificial Neural Network) est composé de milliers de neurones artificiels couplés par des coefficients (poids) qui représentent la structure neuronale [201]. Les neurones artificiels sont les composants fondamentaux des réseaux neuronaux artificiels. Chaque neurone reçoit un poids initial, qui est ensuite modifié par un processus itératif ou "d'apprentissage". La rétropropagation de l'erreur (BPNN : Back Propagation Neural Network) et ses formes modifiées sont les techniques les plus souvent utilisées pour la formation des ANN [174]. La capacité d'apprentissage permet aux réseaux neuronaux de répondre de manière appropriée à des signaux non anticipés par le concepteur, ce qui constitue l'un de ses principaux avantages par rapport à d'autres techniques de traitement des données. Un réseau neuronal peut être appliqué à de nombreux modèles différents sans nécessiter de changements majeurs, contrairement aux techniques mathématiques et aux algorithmes. Les avantages susmentionnés ne sont accessibles que lorsqu'une méthode d'apprentissage appropriée est utilisée. Plusieurs types d'ANN, dont la fonction de base radiale (RBF : Radial Basis Function), le réseau de neurones artificiels à contre-propagation (CP-ANN : Counterpropagation Artificial Neural Network), le réseau neuronal à régression généralisée (GRNN : Generalized Regression Neural Network), le réseau neuronal à retardement (TDNN : Time Delay Neural Network), le réseau neuronal probabiliste (PNN : Probabilistic Neural Network), la carte auto-organisatrice (SOM : Self-Organizing map) et le réseau de neurones artificiels à quantification vectorielle d'apprentissage (LVQ : Learning Vector Quantization), sont appliqués à l'analyse des données obtenues par les sens électroniques [182]. Lorsqu'il s'agit d'évaluer des échantillons inconnus par rapport à une variété de références connues, un ANN correctement entraîné donne d'excellents résultats.

Le réseau de neurones artificiels fonctionne de la manière suivante [202] :

1. Les données sont transmises à la couche cachée avec des poids appliqués. Il peut y avoir d'innombrables couches cachées. Les entrées $x_1, x_2, x_3, \ldots, x_n$ sont transmises (voir Figure 3.1).
2. Les neurones constituent chaque couche cachée. Chaque neurone est connecté pour recevoir toutes les entrées.

3. La couche cachée est l'endroit où tous les calculs sont effectués une fois que les entrées ont été transmises (ovale bleu dans la Figure 3.1). Le calcul dans les couches cachées est divisé en deux parties énumérées ci-dessous :

   (a) Toutes les entrées sont d'abord multipliées par leurs poids respectifs. Le poids est le gradient ou le coefficient de chaque variable. Il démontre la puissance de l'entrée spécifique. Une variable de biais est incluse après l'attribution des poids. La présence d'un biais permet de s'assurer que le modèle s'ajuste aux données de la manière la plus précise possible.

   $$Z_1 = \sum_{i=1}^{n} x_i w_i + b \tag{3.7}$$

   où w_i sont les poids attribués aux entrées x_i, et b est le biais.

   (b) L'équation linéaire 3.7 est ensuite soumise à la fonction d'activation dans la phase suivante. Avant que l'entrée ne soit transmise à la couche suivante de neurones, elle subit un changement non linéaire connu sous le nom de fonction d'activation. La fonction d'activation est cruciale pour favoriser la non-linéarité dans le modèle.

4. Chaque couche cachée suit l'ensemble de la procédure décrite au point 3. On passe à la dernière couche, la couche de sortie, qui fournit la sortie finale après avoir traversé chaque couche cachée. Le processus décrit ci-dessus est appelé propagation vers l'avant.

5. L'erreur, ou la différence entre la sortie réelle et la sortie attendue, est calculée après avoir reçu les prédictions de la couche de sortie. Si l'erreur est importante, des mesures sont prises pour la minimiser et, dans le même but, une rétro-propagation est effectuée.

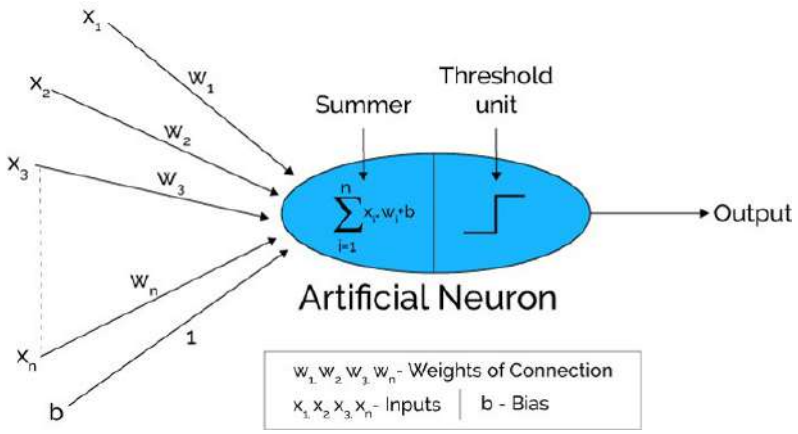


FIGURE 3.1 – Structure des réseaux neuronaux artificiels [1]

Les réseaux neuronaux profonds (DNN) sont également utilisés dans les applications de nez électronique pour améliorer les performances de détection et de classification des odeurs [203, 204] avec des algorithmes tels que les réseaux neuronaux convolutionnels (CNN), les sniffNets, la recherche d'architecture différentiable (DARTS) et les MLP profonds.

1. Source : `https://www.xenonstack.com`, consultée le 22 novembre 2024.

3.3 Discussion

Pour cette enquête, les expressions de recherche et mots-clés *("nose* sensor*" or "e-nose*" or "electronic nose*")* ont été utilisés dans le champ *Topic* le 21 avril 2023 pour extraire des données. Cette recherche a abouti à un total de 9,014 résultats liés aux nez électroniques. Les résultats obtenus ont ensuite été exportés sous forme de fichiers texte. Les activités de l'étude ont été réalisées avec l'aide de Bibliometrix, un outil R dédié l'analyse bibliométrique et à l'exploration des co-citations [205].

La production scientifique sur les nez électroniques a débuté en 1975 et a augmenté chaque année jusqu'à dépasser les 100 articles par an en 1998. Le nombre le plus élevé de publications a été enregistré en 2022 avec 799 publications. Depuis 1993, l'intérêt pour les nez électroniques croît à un taux annuel moyen de 11,44% (voir Figure 3.2).

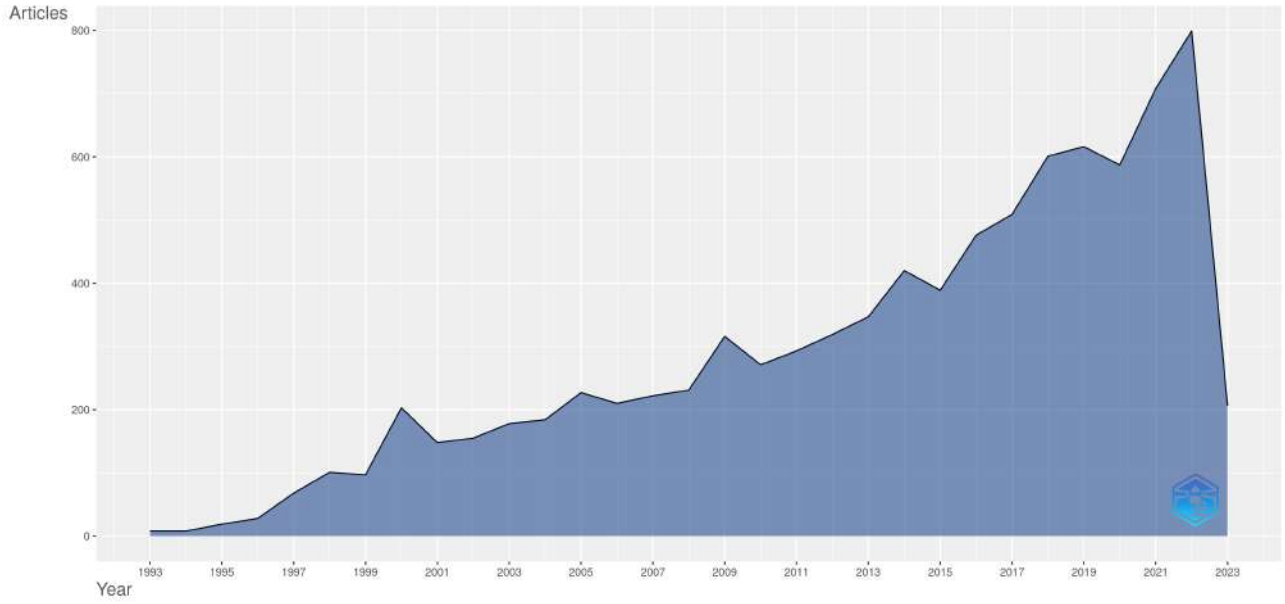


FIGURE 3.2 – Production scientifique annuelle de 1993 à 2023

Les Figures 3.3 et 3.4 illustrent respectivement la production scientifique des pays et les pays cités. La Chine se classe première avec la plus grande production scientifique (4.925 articles), suivie de l'Italie (2.219 articles), des États-Unis (1.938 articles), du Royaume-Uni (962 articles) et de l'Espagne (944 articles). En termes de citations, la Chine est également en tête avec 38.164 citations, suivie des États-Unis (32.848 citations), de l'Italie (22.884 citations), du Royaume-Uni (13.556 citations) et de l'Espagne (10.687 citations).

Ces données montrent que la Chine joue un rôle majeur dans la production mondiale de publications sur les nez électroniques, comme en témoigne le nombre élevé de citations reçues, soulignant ainsi leur pertinence dans ce domaine. On observe également que les États-Unis, l'Italie et l'Espagne investissent considérablement dans la production de publications de haute qualité, fréquemment citées. En revanche, le nombre de contributions et de citations provenant des pays africains reste relativement faible.

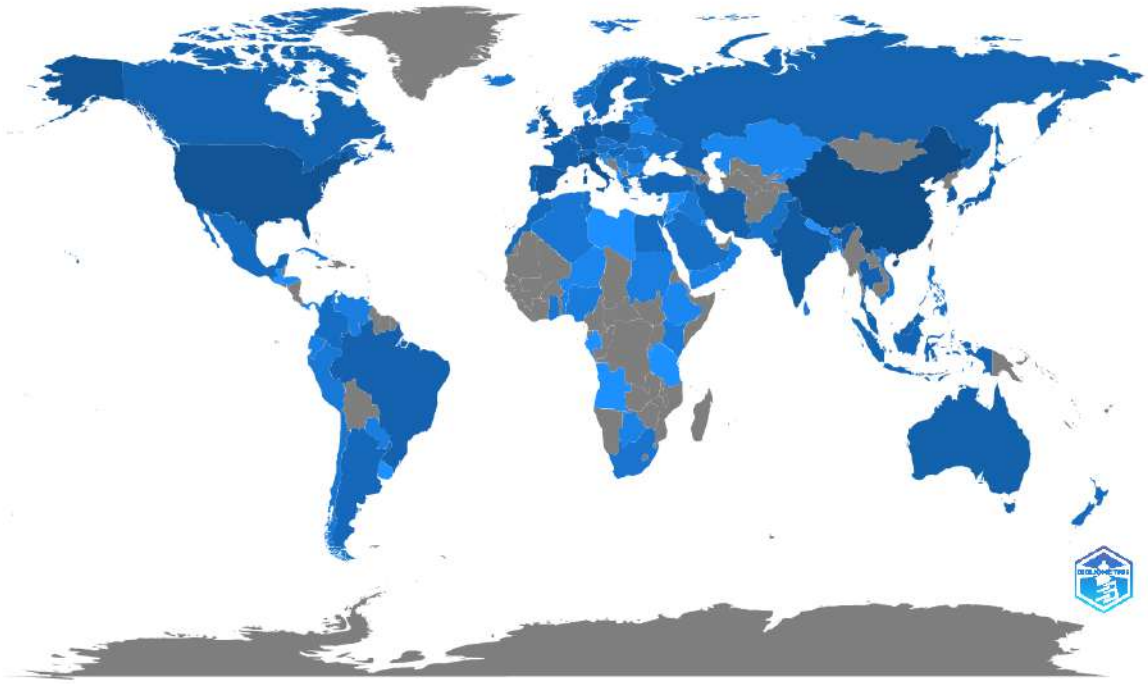


FIGURE 3.3 – Production scientifique par pays de 1993 à 2023

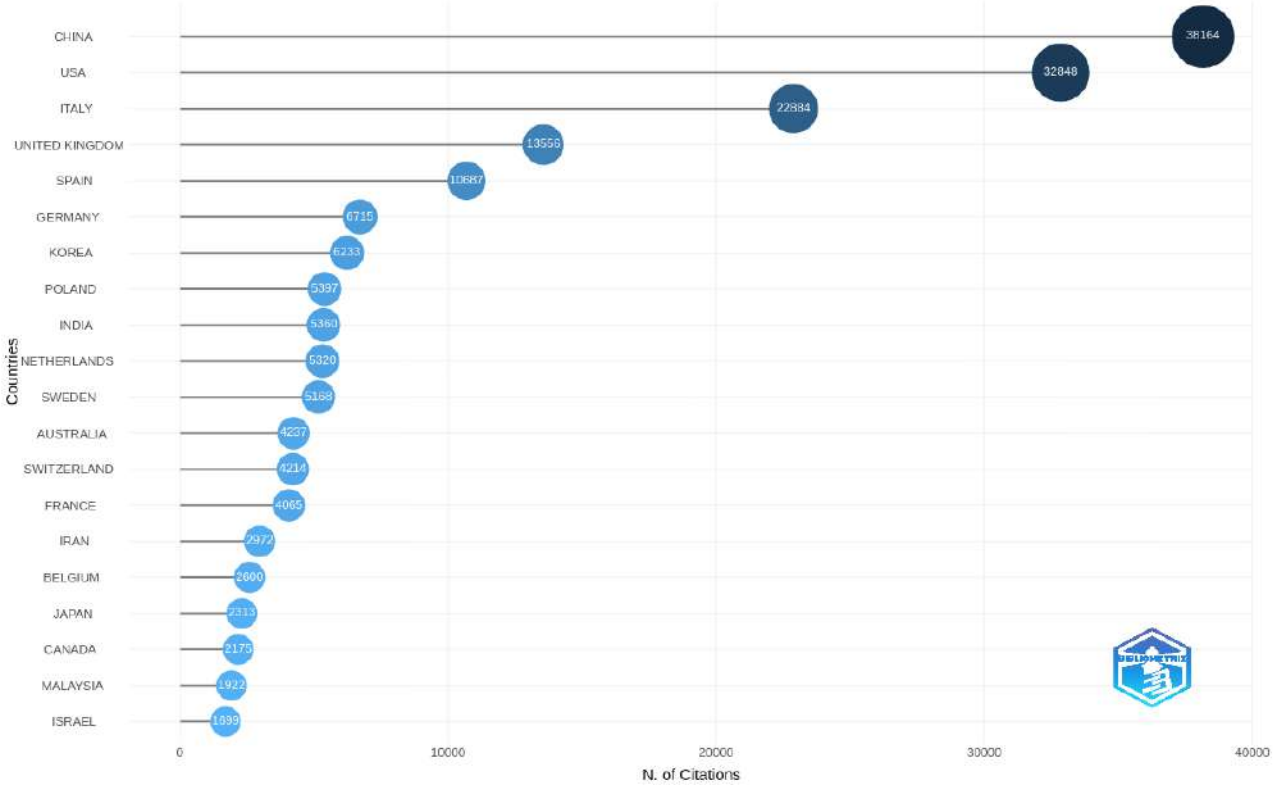


FIGURE 3.4 – Pays les plus cités de 1993 à 2023

Bien que la recherche sur la technologie du nez électronique connaisse un essor significatif à l'échelle mondiale grâce aux nombreuses études menées, elle reste encore peu développée en

Afrique. Par ailleurs, des statistiques générales sur les défis alimentaires en Afrique sont présentées [206] :

— En 2021, 20,2% de la population africaine, soit 278 millions de personnes, souffraient de la faim. Entre 2019 (l'année précédant la pandémie de COVID-19) et 2020, le nombre de personnes touchées a augmenté de 35 millions, suivi de 15 millions supplémentaires en 2021, représentant une hausse totale de 50 millions en seulement deux ans. L'Afrique est la région où la faim a connu la plus forte augmentation, tant en pourcentage qu'en nombre de personnes, depuis 2019.

— En 2021, plus d'un tiers (278 millions de personnes) des 768 millions de personnes sous-alimentées dans le monde résidaient en Afrique. Au niveau régional, bien que les niveaux mondiaux d'insécurité alimentaire modérée et grave soient restés constants, la situation diffère en Afrique. Ce continent affiche les niveaux de prévalence les plus élevés pour les deux degrés de gravité de l'insécurité alimentaire et est la région où cette dernière a le plus augmenté entre 2020 et 2021. Selon les estimations, 322 millions d'Africains souffraient d'une insécurité alimentaire sévère en 2021, soit 21,5 millions de personnes de plus qu'en 2020 et 58 millions de plus qu'en 2019, avant la pandémie de COVID-19. En 2021, l'Afrique comptait plus d'un tiers de la population mondiale en situation d'insécurité alimentaire sévère.

Au regard des données précédemment mentionnées et des résultats de recherche sur l'utilisation du nez électronique décrits dans les sections précédentes, il est recommandé de focaliser les efforts sur l'amélioration des défis alimentaires en Afrique grâce à des technologies comme le nez électronique, combinées à d'autres méthodes pour renforcer leur efficacité. À cette fin, il serait pertinent de mener des recherches visant à développer des nez électroniques à faible coût, accessibles au grand public et aux centres de contrôle de la qualité.

Conclusion

La recherche dans le domaine de la sécurité alimentaire, basée sur les technologies de nez électronique utilise soit un nez électronique expérimental, soit un nez électronique commercial. Différents types de capteurs, chacun doté d'un mécanisme de détection des composants volatils, sont généralement utilisés pour construire un nez électronique. Les capteurs sélectionnés pour les nez électroniques expérimentaux sont choisis en fonction des objectifs de l'étude, de leur coût et de leur disponibilité. Les nez électroniques commerciaux sont proposés par des entreprises avec des procédures spécifiques. Le nombre de capteurs utilisés dans un nez électronique varie en fonction des objectifs des fabricants. De même, plusieurs expériences peuvent utiliser le même nez électronique. Des techniques statistiques et intelligentes d'analyse de modèles sont utilisées pour l'analyse des données provenant d'un nez électronique. Elles permettent de détecter et de classer les odeurs.

Bien que la technologie du nez électronique soit largement utilisée pour les échantillons solides et

liquides dans le domaine de la sécurité alimentaire, elle reste moins répandue en Afrique pour relever les défis liés à la sécurité alimentaire. L'Afrique demeure un continent où les produits locaux sont nombreux et font l'objet de très peu de contrôles, si ce n'est leur apparence et leur odeur. Les prochains axes de recherche pourraient inclure l'exploration de l'utilisation de l'apprentissage supervisé pour l'évaluation de la qualité des aliments et leur classification. Étant donné que la qualification d'un aliment en termes de *excellent, bon, mauvais, pire, etc.* est floue par nature, l'utilisation de systèmes d'inférence flous, voire de réseaux neuronaux flous, pour relever ces défis pourrait être étudiée. D'autres recherches pourrait porter sur des algorithmes d'apprentissage non supervisé et d'apprentissage par renforcement ou, plus généralement, des outils d'apprentissage profond pour améliorer la précision du nez électronique pour les produits locaux et le regroupement d'aliments. Enfin, la combinaison de la technologie du nez électronique avec la vision par ordinateur ou la langue électronique peut contribuer à améliorer l'évaluation de la qualité des aliments. Une fois le meilleur processus d'évaluation de la qualité des aliments identifié, la fabrication d'un nez électronique à faible coût peut contribuer à diffuser cette technologie auprès des familles pour faciliter son adoption.

Le chapitre suivant traitera des aspects techniques liés à la conception du nez électronique utilisé pour la suite des travaux.

CHAPITRE 4

NEZ ÉLECTRONIQUE EXPÉRIMENTAL

Introduction

Avec l'évolution constante de la technologique, l'émulation des capacités sensorielles humaines représente un défi stimulant. Parmi les technologies émergentes, les capteurs biologiques et chimiques, les prothèses sensorielles, ainsi que la réalité virtuelle et augmentée jouent un rôle central. Parmi celles-ci, le nez électronique, qui imite le sens olfactif humain pour détecter et identifier des substances chimiques ou des odeurs, ouvrant ainsi de nouvelles perspectives dans des domaines variés tels que la santé, l'industrie, l'agriculture et l'environnement.

Un nez électronique est un dispositif de détection intelligent et non destructif qui imite le système olfactif humain afin d'identifier, de séparer et de catégoriser des variétés d'odeurs. Il est destiné à distinguer un ensemble de composés dans son ensemble. Il n'est pas conçu pour nommer des substances spécifiques. Il est composé d'une variété de capteurs non sélectifs sensibles à des composés cibles [150, 207]. Aujourd'hui, de nombreuses personnes utilisent cette technologie. Ses applications incluent l'évaluation, le contrôle et la sécurité des aliments tels que les fruits et légumes [208], les produits laitiers [207], le café [209], le thé [210], le riz [211], les produits de boulangerie [212], les épices [213], le vin [149], la bière [214] et les viandes [215]. Un nez électronique commercial ou expérimental est utilisé à plusieurs fins. Tandis que le nez expérimental est configuré pour répondre aux besoins identifiés sur la base de capteurs achetés, le nez commercial est acheté auprès d'entreprises. De nombreux nez électroniques commerciaux sont mentionnés dans la littérature, notamment *Fox 2000*, *Fox 3000* et *Fox 4000* (Alpha MOS, France), *Airsense PEN2* et *Airsense PEN3* (Airsense Analytics, Allemagne), *Cyranose 320* (Cyrano Sciences Inc., États-Unis) et *zNose* (Electronic Sensor Technology, États-Unis). Pour certaines applications, les nez électroniques peuvent être associés à d'autres techniques telles que la vision par ordinateur [216, 217], la spectrométrie de masse et/ou la chromatographie en phase gazeuse [218, 219], la langue électronique [149, 209] et le panel sensoriel entraîné [220].

Malheureusement, les résultats des essais, qu'ils impliquent le nez électronique seul ou en

association avec d'autres technologies, sont tenus confidentiels pour l'instant. En outre, les expériences n'ont pas été menées dans un environnement d'Afrique de l'Ouest. Par ailleurs, la réalisation d'un nez électronique expérimental est moins coûteuse par rapport aux autres méthodes mentionnées ci-dessus. Ce chapitre est consacré à l'étude de la création d'un nez électronique expérimental en utilisant des matériaux disponibles sur le marché local.

4.1 Étude des applications du nez électronique

Les nez électroniques ont été le sujet de nombreux travaux de recherche. Le Tableau 4.1 fournit un résumé de plusieurs applications des nez électroniques sur différents aliments et produits de 2006 à 2022. Ces travaux de recherche visaient à comparer, détecter, évaluer ou surveiller des produits consommables dans différents contextes.

Les éléments suivants fournissent une brève description des recherches qui ont été menées sur divers aliments et produits, en mettant l'accent sur le nez électronique. Chaque application utilise soit un nez électronique expérimental soit commercial.

— Les expériences réalisées dans les études [166, 168, 170, 221] portent sur différentes variétés de pommes provenant de Nouvelle-Zélande, de Belgique et des États-Unis. Les applications incluent la détermination de la maturité, la détection de défauts [168, 221], l'identification du pourrissement [170], et l'étude des effets des agents anti-microbiens [166] sur les pommes. *Cyranose 320, zNose* et *PEN2 (Portable Electronic Nose)* sont les nez électroniques utilisés dans ces expériences sur les pommes.

— Les études décrites dans les articles [222, 223, 224, 225, 226] ont porté sur différents types de tomates. Les tomates utilisées dans les expériences étaient dans deux états différents : sous forme de purée et fraîches. Les recherches ont examiné la maturité et la qualité des tomates [222, 223], l'effet des variétés sur la durée de conservation [224], l'impact du processus de déshydratation [225], et la détection précoce de la contamination microbienne [226]. *Libra Nose*, le dispositif *Airsense, PEN2* et *EOS (Electronic Olfactory Sensor) 835* sont les nez électroniques utilisés dans ces expériences sur les tomates.

— Plusieurs études ont été menées sur l'utilisation du nez électronique pour détecter les produits contrefaits [227] et déterminer le degré de torréfaction des grains de café [165]. *PEN2* et certains nez électroniques expérimentaux basés sur 12 capteurs d'oxydes métalliques à semi-conducteurs (MOS : Metal-Oxide-Semiconductor) sont les nez électroniques utilisés dans ces expériences sur le café.

— Le nez électronique expérimental avec respectivement 14 et 16 capteurs MOS, ainsi qu'un nez commercial *PEN3* ont été utilisés pour évaluer le niveau d'oxygène et la teneur en phénol [228], détecter le vieillissement du vin rouge [229] et mesurer l'évolution continue du vin [230] au cours des recherches menées.

— Le nez électronique *PEN2* et le nez électronique basé sur 8 capteurs MOS ont été utilisés pour identifier différents échantillons de riz [231], détecter diverses maladies des plantes

de riz [232], ainsi que classer différents types de riz [233].

— Wang et al. [234] ont utilisé un nez électronique pour prédire les comptages bactériens dans le porc réfrigéré, et Tian et al. [235] ont analysé l'adultération du porc dans le mouton haché en utilisant un nez électronique. Les nez électroniques *Fox 4000* et *PEN2* ont été utilisés dans ces expériences.

Cyranose 320 est un dispositif de nez électronique qui utilise un réseau de capteurs de polymères conducteurs (CP) pour détecter et identifier les Composés Organiques Volatils (COV) dans l'air. Le dispositif se compose de 32 capteurs internes en polymères composites à base de noir de carbone, qui peuvent fonctionner à température ambiante. Ces capteurs fonctionnent en mesurant les changements de résistance provoqués par l'adsorption des COV sur leur surface [168, 170].

zNose est un dispositif qui utilise la chromatographie en phase gazeuse couplée à un capteur à ondes acoustiques de surface (SAW) pour détecter et identifier les COV dans l'air. La colonne capillaire sert à séparer les différents COV en fonction de leurs propriétés physiques et chimiques, tandis que le capteur SAW mesure la présence et la concentration des COV lorsqu'ils traversent la colonne [170].

Portable Electronic Nose 2 est un dispositif de nez électronique qui utilise un réseau de 10 capteurs MOS pour détecter et identifier les COV dans l'air. Les capteurs à base d'oxydes métalliques sont des semi-conducteurs qui présentent une grande sensibilité aux COV et peuvent détecter une large gamme de composés volatils. Le réseau de capteurs fonctionne en mesurant les changements de résistance électrique provoqués par l'adsorption des COV sur leurs surfaces [166].

Libra Nose est un dispositif de nez électronique développé par Technobiochip et l'Université de Rome "Tor Vergata" en Italie. Il utilise des capteurs de microbalance à quartz (QMB) revêtus de métallo-porphyrines modifiées et de composés apparentés. Ces capteurs sont très sensibles aux COV et peuvent détecter une large gamme de composés volatils. Les composés métallo-porphyrines utilisés dans le revêtement du capteur offrent une grande sélectivité vis-à-vis de certains COV [224].

Le dispositif *Airsense* se compose d'une section d'échantillonnage, où l'air est aspiré dans le dispositif et passe sur les capteurs, et d'un réseau de 10 capteurs MOS hautement sensibles aux COV. Le dispositif comprend également un logiciel d'enregistrement des données et d'analyse statistique multivariée [225].

Electronic Olfactory Sensor (EOS) 835 est un système de nez électronique avec un ensemble pneumatique utilisé pour l'échantillonnage dynamique et comprend une pompe, une électrovanne et un débitmètre électronique. La chambre de capteur, ayant un volume interne de 35 cc, est thermo-régulée pour garantir que les capteurs fonctionnent dans leur plage de température optimale. La carte électronique contrôle les éléments chauffants des capteurs et mesure les couches de détection [169].

Portable Electronic Nose 3 est un dispositif de nez électronique qui utilise un réseau de 10 capteurs MOS tempérés pour détecter et identifier les COV dans l'air. Le dispositif se compose d'une

TABLEAU 4.1 – Résumé de certaines applications du nez électronique

Application	**Objectif**	**Nez électronique**	**Référence**
Pomme	Détection de la maturité et des défauts de la pomme	*Cyranose 320*	[168]
	Détection de l'altération des pommes et profils volatils	*Cyranose 320, zNose*	[170]
	Détection de la détérioration des pommes coupées et non coupées	*Cyranose 320, zNose*	[221]
	Étude de l'effet des antimicrobiens naturels pour prolonger la durée de conservation des pommes	*PEN2*	[166]
Tomate	Discrimination du degré de maturité, de la durée de conservation, des variétés	*Libra Nose*	[224]
	Suivi du processus de déshydratation	*Airsense Device :* 10 *MOS sensors*	[225]
	Évaluation de la maturité des tomates, suivi de la durée de conservation, prédiction des caractéristiques qualitatives des fruits	*PEN2*	[222, 223]
	Évaluation de la contamination microbienne dans les tomates transformées	*EOS 835*	[226]
Café	Détermination de la contrefaçon du café	12 *MOS sensors*	[227]
	Prédiction du degré de torréfaction du café	*PEN2*	[165]
Vin	Évaluation du niveau d'oxygène et du contenu phénolique dans les vins rouges	14 *MOS sensors*	[228]
	Détection du vieillissement du vin rouge	*PEN3*	[229]
	Mesure en continu de l'évolution du vin	16 *MOS sensors*	[230]
Riz	Discrimination des différents types de dommages chez les plants de riz	*PEN2*	[232]
	Classification des échantillons de riz	8 *MOS sensors*	[233]
	Identification précoce des échantillons de riz moisi	8 *MOS sensors*	[231]
Viande	Prévision des comptages bactériens dans la viande de porc réfrigérée	*Fox 4000*	[234]
	Analyse de l'adultération du porc dans le mouton haché	*PEN2*	[235]

petite chambre abritant le réseau de capteurs. Ces derniers sont tempérés, ce qui signifie que leur température est contrôlée pour garantir qu'ils fonctionnent dans la plage optimale pour détecter les COV [229].

Fox 4000 est un dispositif de nez électronique qui comprend un dispositif d'échantillonnage, une unité de détection avec un réseau de capteurs, et un logiciel de reconnaissance de motifs pour l'enregistrement et l'analyse des données. 18 capteurs MOS composent ce réseau de capteurs. Les capteurs sont répartis en trois systèmes [234].

Les remarques suivantes peuvent être mentionnées après l'analyse des travaux de recherche précédents :

— les nez électroniques utilisent principalement des capteurs MOS ;
— les données des recherches précédemment mentionnées ne sont pas disponibles publiquement ;
— le nez électronique, qu'il soit expérimental ou commercial, peut comporter un ou plusieurs capteurs ;
— il n'existe aucune recherche utilisant le nez électronique pour évaluer la qualité des cultures dans un contexte africain, et en particulier au Bénin.

4.2 Conception et spécifications techniques du nez électronique

4.2.1 Réseau de capteurs

Le réseau de capteurs de cette étude (Tableau 4.2) utilise différents capteurs. Le capteur DHT22 est couramment utilisé dans diverses applications, telles que la surveillance de l'environnement, la domotique et les stations météorologiques, où la gestion de la température et de l'humidité sont importants. Pour la détection de gaz inflammables et dangereux dans l'air, les capteurs de gaz MOS sont des capteurs de gaz abordables et fréquemment utilisés. Les capteurs MOS utilisés sont le MQ135, le MQ136, le TGS822, le TGS2600, le TGS2602, le TGS2603, le TGS2610 et le TGS2611. Ces capteurs ont été choisis en raison de leur large gamme de sélectivité aux gaz, car une odeur est souvent perçue comme une combinaison de gaz.

MQ135 : Ce capteur de gaz à semi-conducteur à oxyde métallique peut identifier l'ammoniac (NH_3), les oxydes d'azote (NO_x), l'alcool, le benzène, la fumée, le dioxyde de carbone (CO_2) et d'autres produits chimiques. Il est fréquemment utilisé dans des dispositifs qui surveillent la pollution, régulent la qualité de l'air et détectent les fuites de gaz.

MQ136 : Ce capteur de gaz est conçu pour détecter le sulfure d'hydrogène (H_2S). Il est fréquemment utilisé dans les secteurs miniers et industriels où le sulfure d'hydrogène (H_2S) est présent.

TGS822 : Il s'agit d'un capteur de gaz à semi-conducteur à oxyde métallique sensible au monoxyde de carbone (CO). Il est couramment utilisé dans les détecteurs de fuites de gaz, les systèmes de surveillance des gaz industriels et les alarmes de monoxyde de carbone.

TGS2600 : Ce capteur de gaz est conçu pour détecter la présence de méthane (CH_4). Il est fréquemment utilisé dans les systèmes de détection de fuites de gaz, les pipelines de gaz naturel et d'autres applications où le gaz méthane est présent.

TGS2602 : Il s'agit d'un capteur de gaz de COV constitué de semi-conducteurs à oxyde métallique. Il est fréquemment utilisé dans les moniteurs de qualité de l'air intérieur, les détecteurs de fuites de gaz et les systèmes de gestion de la qualité de l'air.

TGS2603 : Ce capteur de gaz est conçu pour identifier le dioxyde d'azote (NO_2). Il est fréquemment utilisé dans des dispositifs qui surveillent la pollution, régulent la qualité de l'air et détectent les fuites de gaz.

TGS2610 : Ce capteur de gaz à semi-conducteur à oxyde métallique est utilisé pour détecter le monoxyde de carbone (CO) dans l'air. Il est fréquemment employé dans les alarmes de monoxyde de carbone, les systèmes de surveillance des gaz industriels et les détecteurs de fuites de gaz.

TGS2611 : Ce capteur de gaz est conçu pour détecter l'ammoniac (NH_3). Il est fréquemment utilisé dans des applications agricoles, les systèmes de réfrigération et d'autres environnements contenant de l'ammoniac.

TABLEAU 4.2 – Capteurs du nez électronique expérimental

Capteur	**Application principale**
MQ135	Ammoniac, oxyde d'azote, alcools, composés aromatiques, sulfure et fumée
MQ136	Gaz de sulfure d'hydrogène, vapeur organique comprenant du soufre
TGS822	Solvants organiques à la vapeur
TGS2600	Sensibilité élevée aux contaminants totaux de l'air
TGS2602	Haute sensibilité aux COV, à l'ammoniac et au sulfure d'hydrogène
TGS2603	Sensibilité élevée aux odeurs d'amines et de séries de soufre
TGS2610	GPL, méthane, hydrogène, isobutane, propane
TGS2611	Méthane, isobutane, méthanol
DHT22	Température et humidité

4.2.2 Microcontrôleur

Les données provenant des capteurs sont collectées par le microcontrôleur Arduino Mega 2560, qui a été programmé à cet effet. L'ESP32 utilise Internet pour stocker les données. Le microcontrôleur Arduino :

— gère l'identification de chaque bocal grâce à des étiquettes Radio Frequency Identification (RFID) et à un lecteur RFID ;
— commande le système de ventilation ; et
— propose une interface utilisateur via un écran Liquid Crystal Display (LCD).

La Figure 4.1 montre le circuit du nez électronique conçu.

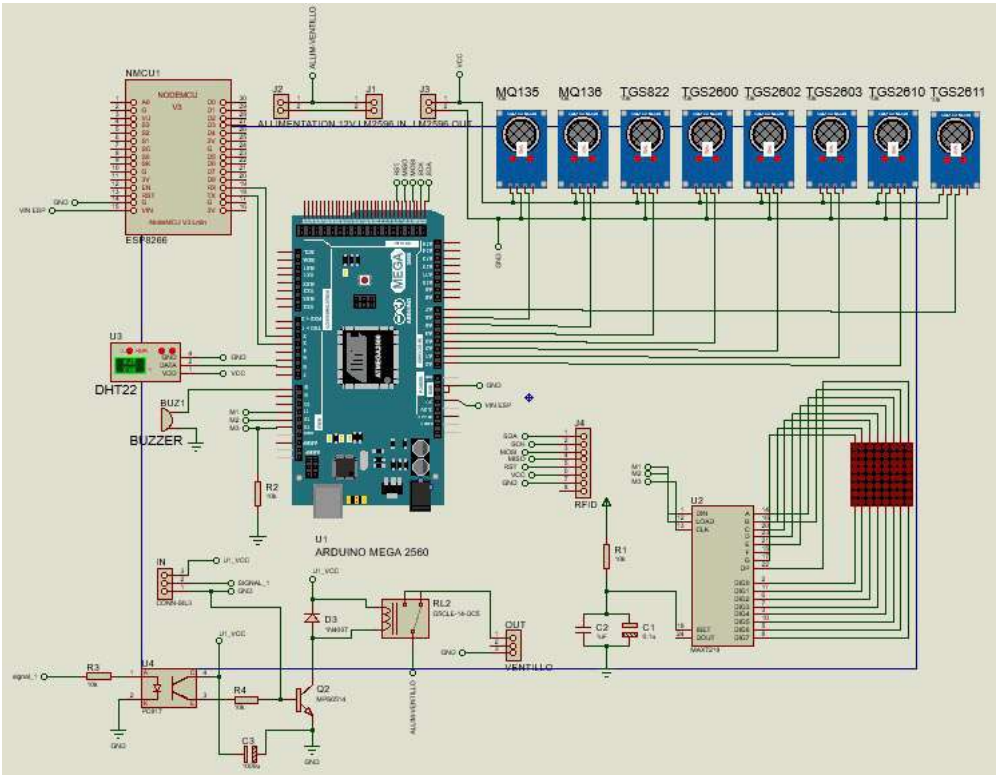
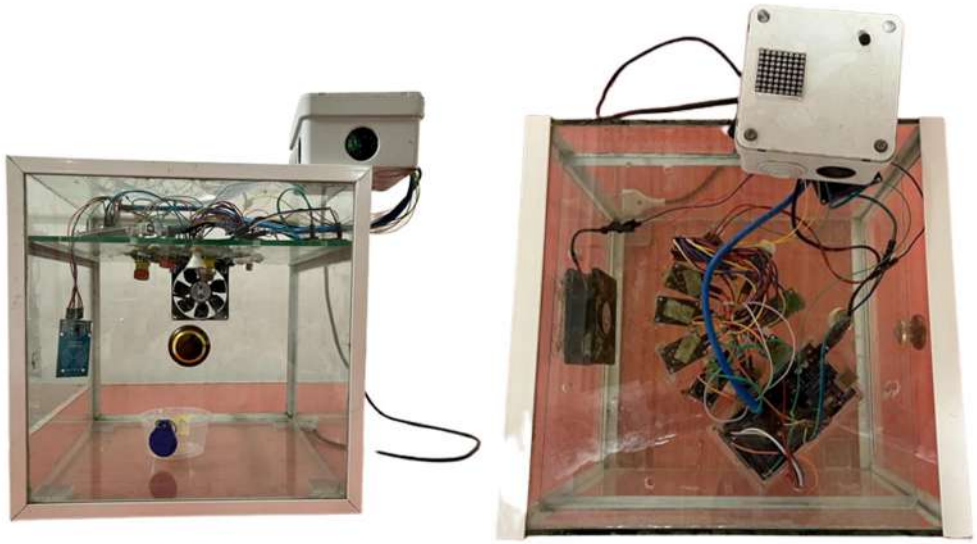


FIGURE 4.1 – Circuit du nez électronique expérimental

4.2.3 Description

Le nez électronique utilisé pour les expériences est présenté à la Figure 4.2. Il est composé de récipients, d'une chambre de réaction, d'un système de nettoyage et d'un système de contrôle. Chaque récipient est équipé d'une étiquette RFID pour l'identification du fruit de la variété en question. La chambre de réaction offre un environnement relativement stable pour permettre au réseau interne de capteurs d'interagir avec la composition gazeuse. Le réseau de capteurs est composé de huit éléments, à savoir les capteurs de gaz à semi-conducteur à oxyde métallique MQ135, MQ136, TGS822, TGS2600, TGS2602, TGS2603, TGS2610, et TGS2611, ainsi qu'un capteur de température et d'humidité DHT22 (Tableau 4.2). Le système de nettoyage utilise un souffleur pour ventiler la chambre de réaction. Le système de contrôle est chargé de superviser l'ensemble de l'expérience, y compris l'échantillonnage et le nettoyage après l'ajout de chaque récipient dans la chambre de réaction.

Un cube en verre de 30 cm sert de dispositif complet de collecte de données. Il possède une étagère située à 5 cm en dessous du haut du dispositif. Le trou central de cette étagère mesure 2,5 cm de diamètre, et il y a deux trous plus petits de chaque côté. Les divers trous créés sont utilisés pour faire passer les jumpers de l'appareil vers le haut de l'étagère, permettant ainsi l'agencement du microcontrôleur Arduino Mega 2560, de l'ESP32, etc. Les capteurs ont été fixés au bas de l'étagère. Après la fermeture du dispositif, des précautions spéciales ont été prises pour éviter que des substances volatiles ne s'évaporent de celui-ci.

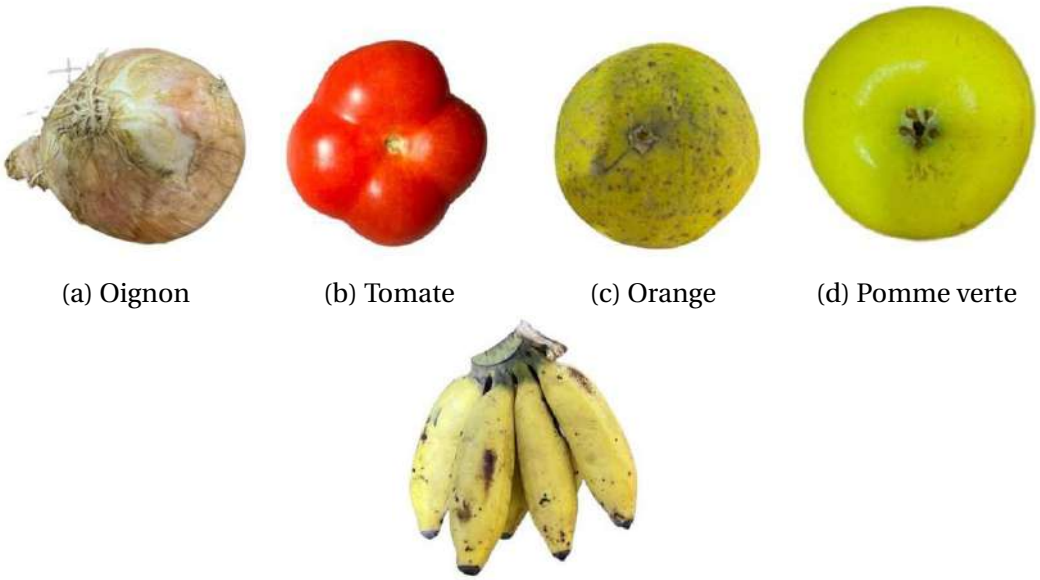


(a) Vue de face (b) Vue du dessus

FIGURE 4.2 – Nez électronique expérimental

4.3 Discussion sur l'optimisation et la validation du dispositif

Pour mettre en avant la capacité du nez électronique à collecter des informations à partir de diverses sources, similaire à un nez biologique, une expérience a été menée avec une variété de produits acquis sur le marché béninois sans sélection préalable particulière. Les éléments utilisés comprennent une tomate, une pomme verte, une banane, une orange et un oignon, comme illustré dans la Figure 4.3.



(a) Oignon (b) Tomate (c) Orange (d) Pomme verte

(e) Banane

FIGURE 4.3 – Échantillons pour la validation du dispositif

Les données issues des capteurs de gaz et du capteur de température et d'humidité permettent de déterminer les substances volatiles émises par différents produits, ainsi que l'humidité et la

température de l'environnement de mesure. Les capteurs de gaz génèrent des valeurs numériques dans la plage de 0 à 1.023, tandis que le capteur de température et d'humidité fournit des lectures de température de $-40\ ^{\circ}C$ à $+80\ ^{\circ}C$ et des lectures d'humidité de 0 à 100%. Les nombres entiers entre 0 et 1.023 renvoyés par les capteurs de gaz sont des lectures analogiques correspondant à la concentration de gaz détectée. Bien que la relation précise puisse varier en fonction du type de capteur et de la configuration, la valeur numérique renvoyée est proportionnelle à la concentration de gaz observée. Ces données numériques représentent simplement la quantité de gaz présente dans l'air au niveau du capteur. Le signal du capteur de gaz peut également être influencé par des variables telles que la ventilation, la température et l'humidité de la zone environnante, d'où la nécessité d'un calibrage constant pour assurer la fiabilité des données.

Les données du fichier CSV comprennent une colonne temporelle et 8 colonnes de gaz, chacune correspondant au niveau de sensibilité des capteurs de gaz à différentes substances. Le fichier inclut également des colonnes pour l'humidité et la température, fournissant des informations sur l'environnement des mesures. Un exemple de fichier CSV est présenté dans les Figures 4.4 à 4.9.

La Figure 4.4 montre les données recueillies lorsque le nez électronique est en cours d'utilisation sans échantillon. Ces données sont comparées à celles des Figures 4.5 à 4.9. Les valeurs du nez électronique en présence et en l'absence d'échantillon diffèrent, prouvant que les capteurs réagissent en présence des échantillons utilisés.

Time	MQ135	MQ136	TGS822	TGS2600	TGS2610	TGS2602	TGS2603	TGS2611	Temperature	Humidity
2023-04-03 15:51:06	499	673	262	223	201	151	104	189	32.20	77.90
2023-04-03 15:51:11	501	668	249	198	198	160	105	174	32.20	77.90
2023-04-03 15:51:16	516	665	256	230	181	149	120	169	32.20	77.90
2023-04-03 15:51:22	505	756	245	236	193	162	104	171	32.30	77.90
2023-04-03 15:51:26	507	746	239	225	189	160	112	167	32.30	77.80
2023-04-03 15:51:31	507	749	231	221	184	160	101	167	32.30	77.80
2023-04-03 15:51:36	522	677	217	222	168	146	116	173	32.30	77.80
2023-04-03 15:51:42	521	661	218	207	165	146	118	171	32.30	77.80
2023-04-03 15:51:47	524	663	211	221	163	146	114	164	32.30	77.80
2023-04-03 15:51:51	512	762	221	238	176	157	99	162	32.30	77.70
2023-04-03 15:51:58	510	753	205	221	161	144	98	175	32.40	77.70
2023-04-03 15:52:02	526	665	203	207	157	145	113	168	32.40	77.60
2023-04-03 15:52:07	523	754	203	243	155	145	110	160	32.40	77.60
2023-04-03 15:52:12	512	668	199	200	155	142	96	172	32.40	77.60
2023-04-03 15:52:18	511	183	212	214	167	157	97	159	32.40	77.50
2023-04-03 15:52:22	512	128	205	203	160	152	99	173	32.40	77.50
2023-04-03 15:52:27	512	114	208	223	162	157	97	157	32.40	77.40
2023-04-03 15:52:32	525	116	193	203	149	145	110	155	32.50	77.40
2023-04-03 15:52:37	523	113	191	240	149	142	109	156	32.50	77.40
2023-04-03 15:52:42	518	114	208	208	162	158	96	156	32.50	77.40
2023-04-03 15:52:47	512	115	206	211	155	157	94	155	32.50	77.30
2023-04-03 15:52:54	521	113	201	229	158	158	102	154	32.50	77.30
2023-04-03 15:52:57	511	113	200	208	155	154	97	156	32.50	77.30
2023-04-03 15:53:03	523	114	187	217	142	143	105	151	32.50	77.20
2023-04-03 15:53:07	510	114	198	214	152	154	94	150	32.50	77.20
2023-04-03 15:53:13	508	113	197	203	150	153	93	157	32.50	77.20
2023-04-03 15:53:17	507	116	198	224	150	155	92	148	32.60	77.20
2023-04-03 15:53:23	510	126	183	204	137	141	110	157	32.60	77.10
2023-04-03 15:53:32	520	119	183	213	153	144	101	149	32.60	77.10

FIGURE 4.4 – Exemple de données sur le vide provenant du nez électronique proposé

Time	MQ135	MQ136	TGS822	TGS2600	TGS2610	TGS2602	TGS2603	TGS2611	Temperature	Humidity
2023-01-11 15:36:56	462	83	72	205	200	144	152	148	34.80	67.50
2023-01-11 15:37:02	497	76	119	188	187	151	151	149	34.90	67.40
2023-01-11 15:37:07	486	89	179	192	163	130	156	147	34.90	67.50
2023-01-11 15:37:12	494	74	197	173	152	140	137	147	34.90	67.40
2023-01-11 15:37:18	479	84	210	203	158	124	133	126	34.90	67.40
2023-01-11 15:37:22	477	84	209	212	153	123	128	124	34.90	67.40
2023-01-11 15:37:28	475	88	219	202	137	121	138	129	34.90	67.30
2023-01-11 15:37:32	473	84	205	176	146	122	123	119	35.00	67.40
2023-01-11 15:37:38	472	90	214	170	132	120	132	122	35.00	67.40
2023-01-11 15:37:42	487	74	201	167	142	119	121	119	35.00	67.50
2023-01-11 15:37:48	468	87	209	171	140	121	122	116	35.00	67.50
2023-01-11 15:37:52	468	87	202	199	143	120	119	118	35.00	67.50
2023-01-11 15:37:58	475	75	195	209	127	129	133	127	35.00	67.40
2023-01-11 15:38:02	471	74	195	210	127	128	131	127	35.10	67.40
2023-01-11 15:38:08	462	73	202	206	127	127	129	125	35.10	67.30
2023-01-11 15:38:13	461	82	206	193	125	119	129	122	35.10	67.30
2023-01-11 15:38:18	460	88	204	184	125	117	125	113	35.10	67.30
2023-01-11 15:38:23	472	75	190	177	123	134	120	122	35.10	67.20
2023-01-11 15:38:29	471	74	191	190	127	134	117	116	35.10	67.20
2023-01-11 15:38:33	470	74	190	213	129	135	119	115	35.10	67.20
2023-01-11 15:38:39	456	88	199	180	135	120	118	111	35.20	67.30
2023-01-11 15:38:43	462	73	188	160	122	128	129	124	35.20	67.30
2023-01-11 15:38:48	465	74	190	167	124	131	132	127	35.20	67.30
2023-01-11 15:38:54	452	85	198	162	134	120	118	110	35.20	67.30
2023-01-11 15:38:58	467	74	188	177	120	131	118	117	35.20	67.30
2023-01-11 15:39:04	449	76	192	185	122	127	129	122	35.20	67.30
2023-01-11 15:39:08	465	74	187	157	125	132	118	114	35.20	67.20
2023-01-11 15:39:14	448	82	189	202	131	117	118	112	35.30	67.30
2023-01-11 15:39:19	460	77	185	175	123	132	118	117	35.30	67.20
2023-01-11 15:39:24	459	76	185	179	121	131	122	126	35.30	67.20

FIGURE 4.5 – Échantillon de données sur la tomate provenant du nez électronique proposé

Time	MQ135	MQ136	TGS822	TGS2600	TGS2610	TGS2602	TGS2603	TGS2611	Temperature	Humidity
2023-01-19 18:29:24	373	90	100	189	89	65	76	83	33.80	72.10
2023-01-19 18:29:29	380	79	105	121	90	65	75	84	33.80	72.40
2023-01-19 18:29:34	373	84	119	183	78	64	87	100	33.80	72.50
2023-01-19 18:29:39	374	87	120	176	80	67	89	84	33.80	72.70
2023-01-19 18:29:44	388	79	108	166	81	79	86	97	33.90	72.90
2023-01-19 18:29:49	387	76	112	178	82	75	76	89	33.90	72.80
2023-01-19 18:29:55	387	78	109	166	79	76	85	97	33.90	72.80
2023-01-19 18:30:01	377	82	108	191	90	65	77	86	33.90	72.70
2023-01-19 18:30:04	389	76	111	144	79	77	87	95	33.90	72.70
2023-01-19 18:30:10	389	76	108	190	79	79	87	95	33.90	72.70
2023-01-19 18:30:15	388	78	108	159	78	78	80	93	34.00	72.70
2023-01-19 18:30:20	377	76	118	176	80	81	89	97	34.00	72.70
2023-01-19 18:30:25	392	76	110	167	79	78	88	95	34.00	72.70
2023-01-19 18:30:30	389	76	111	173	91	66	79	86	34.00	72.70
2023-01-19 18:30:35	384	85	121	176	79	65	91	98	34.00	72.70
2023-01-19 18:30:41	392	76	110	173	83	75	79	89	34.00	72.70
2023-01-19 18:30:45	394	76	111	176	78	77	82	93	34.00	72.70
2023-01-19 18:30:50	383	88	122	167	83	67	86	84	34.10	72.70
2023-01-19 18:30:55	384	85	122	198	79	66	93	98	34.10	72.70
2023-01-19 18:31:00	393	77	111	157	92	66	81	84	34.10	72.70
2023-01-19 18:31:05	384	86	122	178	79	66	93	98	34.10	72.60
2023-01-19 18:31:10	389	76	118	167	79	79	93	97	34.10	72.70
2023-01-19 18:31:16	399	76	112	175	80	80	91	94	34.10	72.70
2023-01-19 18:31:21	396	77	113	156	91	67	84	84	34.10	72.80
2023-01-19 18:31:25	401	76	113	188	79	78	91	96	34.10	72.80
2023-01-19 18:31:31	399	76	114	150	92	73	87	85	34.10	72.80
2023-01-19 18:31:36	405	77	113	187	81	79	96	97	34.20	72.90
2023-01-19 18:31:43	404	77	114	175	80	78	87	93	34.20	72.90
2023-01-19 18:31:46	404	78	115	152	91	76	88	85	34.20	73.00
2023-01-19 18:31:51	394	89	128	184	89	68	89	85	34.20	73.00

FIGURE 4.6 – Échantillon de données sur la pomme verte provenant du nez électronique proposé

Time	MQ135	MQ136	TGS822	TGS2600	TGS2610	TGS2602	TGS2603	TGS2611	Temperature	Humidity
2023-01-19 18:02:05	453	76	135	211	94	75	84	86	34.00	71.90
2023-01-19 18:02:09	451	76	136	151	94	69	84	86	34.00	72.00
2023-01-19 18:02:14	442	88	147	205	81	69	96	91	34.00	72.20
2023-01-19 18:02:19	443	89	149	160	84	69	96	86	34.00	72.20
2023-01-19 18:02:25	442	76	145	182	82	84	95	98	34.00	72.30
2023-01-19 18:02:29	443	87	147	156	83	69	97	87	34.00	72.40
2023-01-19 18:02:36	455	77	135	181	95	77	83	86	34.00	72.40
2023-01-19 18:02:39	443	87	149	170	82	72	98	87	34.10	72.40
2023-01-19 18:02:45	444	80	148	150	82	74	97	99	34.10	72.40
2023-01-19 18:02:49	449	76	144	204	82	83	95	98	34.10	72.40
2023-01-19 18:02:55	445	84	146	141	83	74	98	98	34.10	72.40
2023-01-19 18:03:00	448	78	144	181	83	83	95	97	34.10	72.40
2023-01-19 18:03:06	448	83	146	169	82	73	100	98	34.10	72.40
2023-01-19 18:03:10	444	88	151	151	91	71	89	87	34.20	72.40
2023-01-19 18:03:17	458	78	136	198	84	82	89	94	34.20	72.40
2023-01-19 18:03:20	447	75	149	163	84	84	100	98	34.20	72.40
2023-01-19 18:03:26	446	87	149	136	82	72	103	86	34.20	72.40
2023-01-19 18:03:30	458	76	140	200	96	81	89	86	34.20	72.50
2023-01-19 18:03:35	460	76	141	156	82	83	94	96	34.20	72.60
2023-01-19 18:03:41	456	78	147	161	82	85	103	97	34.20	72.50
2023-01-19 18:03:45	449	88	152	198	83	72	107	92	34.30	72.60
2023-01-19 18:03:51	450	82	150	164	82	79	106	99	34.30	72.50
2023-01-19 18:03:55	452	88	154	150	82	73	107	100	34.30	72.50
2023-01-19 18:04:03	453	88	155	203	90	73	96	88	34.30	72.50
2023-01-19 18:04:07	453	76	150	154	85	86	108	99	34.30	72.50
2023-01-19 18:04:11	454	86	155	175	86	74	108	85	34.40	72.60
2023-01-19 18:04:16	465	76	143	209	95	73	99	88	34.40	72.60
2023-01-19 18:04:21	457	87	153	134	83	73	114	88	34.40	72.60
2023-01-19 18:04:26	456	87	155	200	91	74	104	86	34.40	72.60
2023-01-19 18:04:31	458	88	153	155	83	74	115	86	34.40	72.70

FIGURE 4.7 – Échantillon de données sur la banane provenant du nez électronique proposé

Time	MQ135	MQ136	TGS822	TGS2600	TGS2610	TGS2602	TGS2603	TGS2611	Temperature	Humidity
2023-01-19 17:10:30	361	77	138	181	86	85	94	100	35.10	68.60
2023-01-19 17:10:35	367	87	158	172	93	77	92	92	35.10	68.80
2023-01-19 17:10:40	390	78	154	184	90	114	107	107	35.00	69.10
2023-01-19 17:10:45	392	89	173	185	101	118	126	99	35.00	69.20
2023-01-19 17:10:50	417	76	163	156	91	131	150	109	35.00	69.50
2023-01-19 17:10:55	416	77	162	207	107	125	149	102	35.00	69.70
2023-01-19 17:11:00	408	88	175	180	106	122	157	104	35.00	70.00
2023-01-19 17:11:08	413	91	173	161	107	123	161	104	35.00	70.20
2023-01-19 17:11:18	424	76	165	197	104	132	165	105	35.10	70.40
2023-01-19 17:11:22	448	81	177	189	96	144	178	123	35.10	70.50
2023-01-19 17:11:25	452	89	192	164	109	150	171	111	35.10	70.70
2023-01-19 17:11:31	461	88	214	208	106	149	187	116	35.10	70.90
2023-01-19 17:11:35	484	74	232	192	115	152	207	123	35.10	71.10
2023-01-19 17:11:41	494	76	243	178	103	162	240	136	35.20	71.30
2023-01-19 17:11:45	508	76	236	173	103	169	250	136	35.20	71.50
2023-01-19 17:11:52	506	93	240	185	115	167	239	122	35.20	71.50
2023-01-19 17:11:55	512	89	234	220	115	186	240	121	35.20	71.70
2023-01-19 17:12:00	527	727	221	206	105	223	251	131	35.20	71.90
2023-01-19 17:12:06	518	688	234	191	110	235	246	119	35.20	72.00
2023-01-19 17:12:12	522	656	239	167	116	257	247	118	35.20	72.20
2023-01-19 17:12:16	538	650	229	176	109	284	249	121	35.20	72.30
2023-01-19 17:12:21	534	784	230	217	118	290	251	118	35.30	72.40
2023-01-19 17:12:26	537	651	241	163	106	307	267	131	35.20	72.50
2023-01-19 17:12:33	552	757	235	201	106	323	266	133	35.30	72.60
2023-01-19 17:12:36	552	657	249	183	112	318	260	118	35.30	72.80
2023-01-19 17:12:42	558	658	245	168	119	325	260	122	35.30	72.90
2023-01-19 17:12:46	576	733	246	194	107	342	269	130	35.30	72.90
2023-01-19 17:12:52	566	797	254	212	120	338	266	123	35.30	73.00
2023-01-19 17:12:56	570	775	256	206	121	345	269	124	35.30	73.10
2023-01-19 17:13:03	575	749	269	197	109	350	284	132	35.30	73.20

FIGURE 4.8 – Échantillon de données sur l'orange provenant du nez électronique proposé

Time	MQ135	MQ136	TGS822	TGS2600	TGS2610	TGS2602	TGS2603	TGS2611	Temperature	Humidity
2023-01-19 17:36:25	386	82	196	176	92	116	106	89	34.30	71.20
2023-01-19 17:36:29	393	77	197	175	88	126	108	95	34.30	71.30
2023-01-19 17:36:34	384	92	195	163	92	116	108	89	34.30	71.50
2023-01-19 17:36:40	402	77	198	167	82	129	112	99	34.30	71.60
2023-01-19 17:36:44	393	90	197	205	91	118	118	90	34.30	71.70
2023-01-19 17:36:49	405	77	197	148	99	133	119	98	34.30	71.80
2023-01-19 17:36:54	409	78	197	181	88	143	124	111	34.30	72.20
2023-01-19 17:37:00	401	79	207	197	90	144	139	115	34.40	73.60
2023-01-19 17:37:04	415	75	200	209	101	134	140	102	34.40	75.20
2023-01-19 17:37:09	409	85	214	188	89	133	166	118	34.40	76.40
2023-01-19 17:37:14	421	76	204	181	100	131	165	105	34.40	77.40
2023-01-19 17:37:20	416	78	213	177	88	142	178	118	34.50	77.90
2023-01-19 17:37:24	428	87	216	190	88	126	181	120	34.50	78.20
2023-01-19 17:37:30	451	77	206	211	103	125	168	108	34.50	78.10
2023-01-19 17:37:35	468	76	204	205	92	136	173	119	34.50	78.40
2023-01-19 17:37:39	474	76	204	184	105	129	158	108	34.60	78.40
2023-01-19 17:37:45	481	74	204	150	106	128	155	109	34.60	78.60
2023-01-19 17:37:50	484	78	202	175	109	132	152	111	34.60	78.70
2023-01-19 17:37:56	477	84	212	200	97	128	162	123	34.60	78.70
2023-01-19 17:38:00	495	77	202	150	99	138	154	121	34.60	78.70
2023-01-19 17:38:06	484	78	210	172	100	139	160	126	34.60	78.70
2023-01-19 17:38:10	496	76	201	196	115	133	151	114	34.70	78.70
2023-01-19 17:38:16	500	77	201	201	103	137	152	124	34.60	78.70
2023-01-19 17:38:20	502	76	203	178	103	137	154	125	34.70	78.60
2023-01-19 17:38:25	492	89	214	163	103	127	168	118	34.70	78.60
2023-01-19 17:38:30	496	89	207	215	114	126	156	119	34.70	78.50
2023-01-19 17:38:35	509	77	202	195	117	134	155	120	34.70	78.50
2023-01-19 17:38:41	499	90	201	171	115	126	159	119	34.70	78.50
2023-01-19 17:38:46	502	87	213	166	105	126	169	135	34.70	78.60
2023-01-19 17:38:51	506	89	203	197	115	127	159	122	34.70	78.50

FIGURE 4.9 – Échantillon de données sur l'oignon provenant du nez électronique proposé

4.4 Protocole expérimentale de collecte des données sensorielles de la tomate

4.4.1 Tomate en lumière

L'une des cultures les plus largement cultivées dans le monde est la tomate. Pour les petits et moyens cultivateurs commerciaux, c'est une source cruciale de vitamines et une culture génératrice de revenus importants. Sur une estimation de 3,9 millions d'hectares de terres, environ 105 millions de tonnes de tomates fraîches ont été produites dans le monde en 2001 [236]. Selon les données de la FAO figurant dans la Figure 4.10, la production de tomates est répartie sur tous les continents, avec 54,9% en Asie, 17,3% en Amérique, 15,5% en Europe, 11,9% en Afrique et 0,3% en Océanie. En 2020, le Bénin a produit 261.103 tonnes de tomates fraîches sur une surface récoltée de 39.429 hectares. La plus grande production, soit 400.572 tonnes générées sur 45.630 hectares récoltés, a été enregistrée en 2014 [237].

La tomate est membre de la famille des *Solanaceae*. D'autres espèces bien connues de cette famille comprennent la pomme de terre, le tabac, le poivron et l'aubergine [236]. Après la pomme de terre et la patate douce, la tomate est le troisième légume le plus largement cultivé dans le monde et le deuxième légume le plus populaire à l'échelle mondiale [238]. Par conséquent, la culture de la tomate représente une préoccupation économique mondiale.

La production de tomates reste chaotique au Bénin malgré l'importance de la culture, avec

une longue période de pénurie d'octobre à mai et une courte période d'abondance de juin à septembre [239]. Étant donné que les tomates ne se conservent pas très longtemps à l'état frais et à température ambiante, d'importantes pertes post-récolte sont observées pendant les saisons de surproduction [240]. La teneur élevée en eau du fruit (plus de 85%), contribuant à son apparence de fragilité et de périssabilité, est un autre facteur de ces pertes [241]. Par conséquent, en raison du manque de mesures de conservation efficaces, plus de 40% de la production du pays est gaspillée pendant la période d'abondance [239].

La méthode de transformation/préservation la plus populaire pour réduire ces pertes post-récolte est la transformation artisanale des tomates en purée au Bénin, comme dans de nombreux autres pays d'Afrique de l'Ouest [242]. Pour évaporer l'eau présente dans la tomate pendant cette transition, une grande quantité d'énergie (charbon ou gaz combustible) doit être utilisée [241].

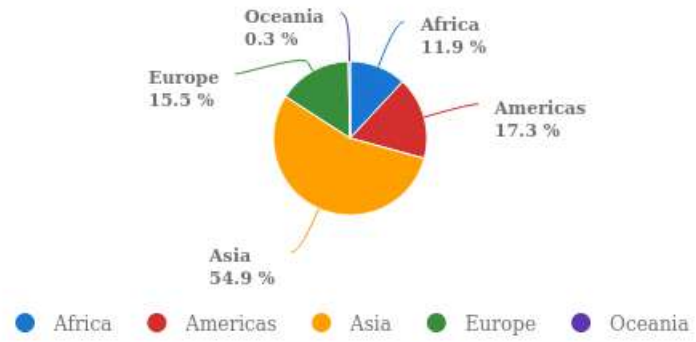


FIGURE 4.10 – Part de la production de tomates par région | Moyenne 1994 - 2021 [1]

La sélection de la tomate pour la poursuite des expériences découle des raisons précédemment énoncés.

4.4.2 Espèces de tomates

Les tomates se déclinent en plusieurs variétés, des locales aux variétés améliorées (importées), sur les marchés du Bénin. Les cultivars locaux, caractérisés par la petite taille des fruits, sont les plus largement cultivés. Il s'agit de *Tounvi*, *Akikon* et *Ahougbo*. De plus, il existe des variantes locales telles que *Gbogan*, *Kêkêfo*, *Ouaga* et *Sonafel* qui ont de gros fruits. Elles se distinguent les unes des autres par leur forme arrondie et le nombre de lobes qu'elles possèdent (plus de 6 lobes). Ces variétés sont cultivées seulement partiellement, et en raison des pertes post-récolte considérables (plus de 40%) observées, elles deviennent de plus en plus rares [243]. Les caractéristiques des variétés de tomates rencontrées sont les suivantes [239] :

Tounvi **:** est le type qui prédomine dans le sud du Bénin. Il existe deux variétés : l'une a des fruits plus gros, légèrement plus arrondis et plus densément lobés (3 à 4 lobes), tandis que l'autre a des fruits plus petits, aplatis et moins sévèrement lobés (2 lobes).

1. Source : FAOSTAT (20 mars 2023)

Akikon **:** se distingue par ses fruits allongés, lisses et d'un rouge foncé. Souvent, ces fruits ont un seul lobe et un col vert. Il s'agit d'une variété de tomate "de luxe" qui est environ deux fois plus chère que *Tounvi*.

Gbogan **:** est reconnaissable par ses fruits fortement nervurés, aplatis, énormes, d'un rouge-orange, et par son placenta bien développé, qui a 6 à 8 lobes. Ce type est proposé sur les marchés en quantités moins uniformes. De plus, les fruits de *Gbogan* sont moins adaptés au stockage à température ambiante que ceux de *Tounvi* et *Akikon*.

Ces types sont rejoints par un nombre croissant de variétés de meilleure qualité en provenance d'Europe, d'Amérique et d'Asie (*Tropimech*, *Petomech*, *Caraïbe*, *Mongal*, *Tima*, *Roma*, *Cobra*, *Padma*, *Platinum*, *Thorgal*, etc.). Leur grande taille, leur lobe solitaire et des rendements élevés de 15 à 30 tonnes/ha servent de traits distinctifs [244]. Ils ont une chair ferme et une couleur rouge vive. Ils sont excellents pour la conservation en raison de leur peau épaisse et charnue. Cependant, bien qu'ils soient excellents pour la conservation, ils ne conviennent pas à toutes les saisons et sont sensibles aux maladies et aux ravageurs, raison pour laquelle les agriculteurs béninois ne les ont pas encore pleinement adoptés [243]. La bactériose, qui diminue le rendement des variétés locales, est plus difficile à contracter pour les variétés *Padma*, *Platinum* et *Thorgal* [245].

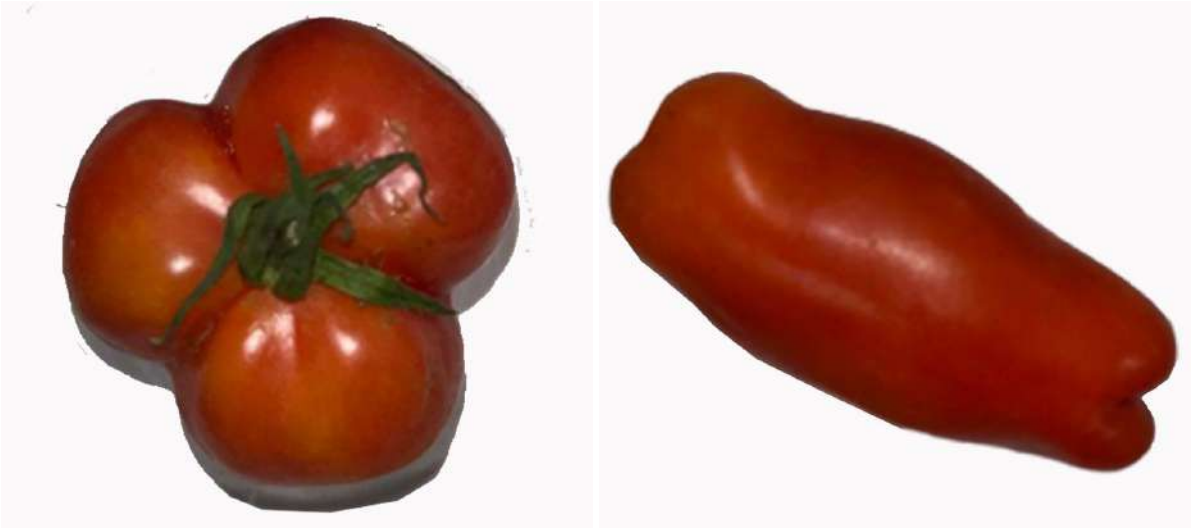
Pott et al. dans [246] rapportent que, pendant la maturation, les tomates libèrent principalement de l'éthylène ainsi qu'une gamme de composés volatils, incluant des alcools, des aldéhydes et des esters, qui sont caractéristiques de différentes variétés. Klee et Tieman dans [247] soulignent que, bien que les proportions de certains composés puissent varier légèrement d'une variété à l'autre, les profils volatils responsables de l'arôme et de la détérioration sont généralement conservés. De plus, Beckles [248] appuie cette conclusion en notant que l'éthylène joue un rôle central dans le processus de maturation de toutes les variétés de tomates. Sur la base de ces observations, il est raisonnable de conclure que les variétés utilisées dans cette étude sont adéquates pour atteindre les objectifs expérimentaux.

En raison du moment de l'achèvement du nez électronique, les variétés locales *Tounvi* et *Akikon* (Figure 4.11) dans le cadre de cette étude. Celles-ci viennent d'être récoltées et ont été livrées au marché de Dantokpa au Bénin, où elles ont été mises en vente.

4.4.3 Collecte de données

Suite à l'acquisition des fruits et avant la collecte, chaque fruit est placé dans un bocal en plastique transparent équipé d'une étiquette RFID. La collecte a été réalisée selon les étapes ci-dessous :

1. Allumer le nez électronique.
2. S'assurer que l'ESP32 est connecté à l'internet.
3. Ouvrir le cube en verre (chambre de réaction).
4. Prendre un bocal et le passer devant le lecteur RFID pour identifier la variété présente.



(a) Variété locale *Tounvi* (b) Variété locale *Akikon*

FIGURE 4.11 – Échantillons de tomates

5. Placer le bocal à l'intérieur du cube en verre, à au moins 5 cm en dessous des capteurs.
6. Fermer le cube en verre (chambre de réaction).
7. Attendre 20 minutes pendant que les données sont stockées en ligne, toutes les 5 secondes. Une fois terminé, le buzzer émet trois sons pour indiquer que la collecte est terminée.
8. À la suite de la phase de collecte, une phase de nettoyage d'une durée de 10 minutes a lieu pendant laquelle la chambre de réaction est ouverte et le bocal est retiré. Cela permet de nettoyer le circuit, de restaurer l'état initial des capteurs et d'éliminer toute substance volatile persistante de la chambre de réaction. Après les 10 minutes, le buzzer émet un son pour indiquer que le prochain échantillon peut être inséré dans le bocal.

Les étapes ci-dessus sont répétées chaque jour pour chaque tomate jusqu'à ce que sa détérioration soit visuellement constatée. Une fois les mesures quotidiennes prises, le nez électronique peut être éteint.

La collecte de données a eu lieu sur une période de 80 jours, du 11 février 2023 au 1er mai 2023, en raison du processus de maturation des tomates. Les durées de vie de chaque tomate sont indiquées dans le Tableau 4.3.

TABLEAU 4.3 – Durée de vie des tomates

Tomate	**Durée de vie (jours)**
Akikon 1, *Akikon* 2	20
Tounvi 5	30
Tounvi 1, *Akikon* 3	36
Tounvi 3, *Tounvi* 4, *Akikon* 5	50
Akikon 4	70
Tounvi 2	80

Les données présentées dans le Tableau 4.3 montrent une grande variabilité dans la durée de vie des tomates, malgré le fait qu'elles aient été dans le même état au début de la collecte, conservées dans un environnement identique et placées dans des bocaux différents, éloignés les uns des autres. Par exemple, les tomates *Akikon* 1 et *Akikon* 2 ont eu une durée de vie relativement courte, de seulement 20 jours, tandis que la tomate *Tounvi* 2 a survécu jusqu'à 80 jours. Cette différence pourrait être expliquée par des facteurs intrinsèques à chaque variété, tels que leur composition ou leur susceptibilité à la détérioration, malgré des conditions de stockage identiques. De plus, il est intéressant de noter que les tomates de la même variété, comme les différentes *Akikon* et *Tounvi*, ont montré des durées de vie variées, ce qui suggère que des variations au sein de la même variété peuvent également influencer la durée de conservation des fruits.

4.4.4 Dataset

Les travaux précédents ont permis la collecte et la mise en place d'un jeu de données, disponible à l'adresse suivante : `https://gitlab.com/liooonel/tomato-tounvi-and-akikon-pictures-and-e-nose-data.git`. Ce jeu de données comprend plusieurs fichiers CSV couvrant la durée de vie des deux variétés choisies (de la récolte jusqu'au moment où elles sont jugées impropres à la consommation). Le contenu de ces fichiers a été expliqué dans la section 4.3. Les Figures 4.12a et 4.13a présentent les données d'une bonne tomate *Tounvi* et *Akikon* respectivement, tandis que les Figures 4.12b et 4.13b montrent les données d'une tomate avariée *Tounvi* et *Akikon* respectivement.

Time	TGS2603	TGS2610	TGS2611	TGS2600	TGS2602	MQ136	MQ135	TGS822	Humidity	Temperature
2023-02-11 10:02:01	146	130	127	190	129	82	512	164	84.00	31.70
2023-02-11 10:02:06	132	125	139	200	140	82	496	206	83.60	31.80
2023-02-11 10:02:11	142	114	131	196	127	98	496	214	83.50	31.80
2023-02-11 10:02:17	140	111	131	206	126	97	494	219	83.60	31.80
2023-02-11 10:02:21	124	109	130	206	136	81	490	221	83.50	31.80
2023-02-11 10:02:26	121	107	126	197	136	84	488	217	83.40	31.90
2023-02-11 10:02:31	122	122	116	191	124	82	499	202	83.40	31.90
2023-02-11 10:02:36	131	106	118	198	123	97	483	207	83.30	31.90
2023-02-11 10:02:42	134	106	124	198	125	97	480	210	83.30	31.90
2023-02-11 10:02:47	117	115	114	201	138	83	489	197	83.20	31.90
2023-02-11 10:02:52	117	116	110	201	124	82	490	198	83.10	31.90
2023-02-11 10:02:57	116	101	125	194	135	84	482	211	83.10	32.00
2023-02-11 10:03:02	131	101	122	191	131	88	472	210	83.10	32.00
2023-02-11 10:03:07	124	105	109	199	124	96	471	193	83.00	32.00
2023-02-11 10:03:12	115	114	109	198	121	82	484	192	83.00	32.00
2023-02-11 10:03:17	132	100	122	194	131	93	467	206	83.00	32.00
2023-02-11 10:03:22	127	99	120	193	121	97	463	204	82.90	32.00
2023-02-11 10:03:27	115	99	122	186	133	84	470	208	82.90	32.10
2023-02-11 10:03:32	129	99	119	196	122	100	460	203	82.90	32.10
2023-02-11 10:03:37	114	115	110	200	120	92	464	190	82.80	32.10
2023-02-11 10:03:43	127	97	119	206	120	99	459	202	82.80	32.10
2023-02-11 10:03:48	115	110	107	206	120	83	471	189	82.80	32.10
2023-02-11 10:03:53	118	110	109	196	127	95	466	189	82.90	32.10
2023-02-11 10:03:58	114	110	109	197	120	82	469	189	82.70	32.10
2023-02-11 10:04:03	129	98	118	202	127	90	453	202	82.70	32.10
2023-02-11 10:04:08	114	107	106	195	134	83	463	188	82.60	32.20
2023-02-11 10:04:13	117	111	107	194	119	97	449	193	82.60	32.20
2023-02-11 10:04:18	113	95	123	187	132	82	461	187	82.60	32.20
2023-02-11 10:04:23	123	102	106	195	119	95	447	194	82.60	32.20
2023-02-11 10:04:28	113	111	106	194	120	94	445	189	82.60	32.20
2023-02-11 10:04:33	114	94	120	189	131	83	455	186	82.50	32.20
2023-02-11 10:04:38	127	96	119	181	128	84	444	200	82.50	32.20
2023-02-11 10:04:43	112	109	106	194	119	94	442	185	82.50	32.20
2023-02-11 10:04:48	126	94	117	193	130	82	443	201	82.50	32.20
2023-02-11 10:04:53	112	107	105	198	118	88	453	187	82.50	32.30
2023-02-11 10:04:58	112	93	123	192	131	84	453	187	82.40	32.30
2023-02-11 10:05:04	128	94	117	190	129	83	439	199	82.40	32.30
2023-02-11 10:05:09	128	92	118	191	132	83	447	187	82.40	32.30
2023-02-11 10:05:14	120	92	118	193	130	83	448	183	82.40	32.30

(a) Bonne tomate

Time	TGS2603	TGS2610	TGS2611	TGS2600	TGS2602	MQ136	MQ135	TGS822	Humidity	Temperature
2023-03-18 09:48:40	127	258	213	225	189	664	595	324	79.70	29.70
2023-03-18 09:48:44	125	245	185	280	192	791	573	299	79.90	29.80
2023-03-18 09:48:50	123	238	197	214	181	660	594	284	80.20	29.80
2023-03-18 09:48:56	139	221	178	265	181	763	585	291	80.30	29.80
2023-03-18 09:49:01	117	228	196	214	179	660	595	263	80.30	29.80
2023-03-18 09:49:05	132	234	174	254	193	761	575	270	80.30	29.80
2023-03-18 09:49:10	129	227	177	211	193	657	578	261	80.30	29.90
2023-03-18 09:49:15	132	205	175	266	176	761	593	264	80.20	29.90
2023-03-18 09:49:20	127	222	178	212	195	654	576	258	80.20	29.90
2023-03-18 09:49:26	113	201	187	248	177	752	592	239	80.20	29.90
2023-03-18 09:49:30	127	196	174	219	177	658	591	253	80.30	29.90
2023-03-18 09:49:36	126	195	171	251	192	756	582	249	80.30	29.90
2023-03-18 09:49:40	126	195	171	227	191	655	587	246	80.30	29.90
2023-03-18 09:49:45	107	193	185	240	171	743	593	230	80.40	30.00
2023-03-18 09:49:50	117	206	173	255	189	759	575	242	80.40	30.00
2023-03-18 09:49:55	120	204	169	212	189	655	581	239	80.40	30.00
2023-03-18 09:50:00	103	182	186	237	170	752	588	220	80.50	30.00
2023-03-18 09:50:05	121	178	188	258	188	761	589	218	80.50	30.00
2023-03-18 09:50:11	105	182	183	214	168	660	594	217	80.50	30.00
2023-03-18 09:50:16	105	197	165	233	185	744	572	233	80.50	30.00
2023-03-18 09:50:21	108	175	189	239	171	669	575	213	80.50	30.00
2023-03-18 09:50:26	117	193	163	210	182	654	588	227	80.50	30.00
2023-03-18 09:50:31	101	171	179	228	165	750	570	210	80.50	30.10
2023-03-18 09:50:37	118	172	163	248	167	761	587	207	80.50	30.10
2023-03-18 09:50:41	102	166	178	201	164	662	569	209	80.50	30.10
2023-03-18 09:50:46	100	169	179	238	175	762	572	218	80.50	30.10
2023-03-18 09:50:52	108	185	159	218	183	655	583	221	80.40	30.10
2023-03-18 09:50:56	117	166	173	229	163	662	590	206	80.40	30.10
2023-03-18 09:51:01	116	164	157	240	162	659	584	201	80.30	30.10
2023-03-18 09:51:06	116	164	174	199	165	246	567	200	80.30	30.10
2023-03-18 09:51:12	99	161	174	233	173	130	567	216	80.30	30.10
2023-03-18 09:51:25	110	179	156	218	164	112	583	197	80.30	30.10
2023-03-18 09:51:29	118	161	170	221	164	127	565	205	80.30	30.20
2023-03-18 09:51:32	114	177	154	219	160	111	581	196	80.20	30.20
2023-03-18 09:51:37	96	169	173	245	176	108	562	212	80.20	30.20
2023-03-18 09:51:42	101	155	168	199	164	131	563	209	80.20	30.20
2023-03-18 09:51:47	110	175	155	238	159	111	579	193	80.10	30.20
2023-03-18 09:51:52	102	153	168	206	160	128	560	210	80.20	30.20
2023-03-18 09:51:57	96	152	171	224	173	119	560	208	80.10	30.20

(b) Mauvaise tomate

FIGURE 4.12 – Données d'échantillon de la variété de tomate *Tounvi* provenant du nez électronique proposé

Time	TGS2603	TGS2610	TGS2611	TGS2600	TGS2602	MQ136	MQ135	TGS822	Humidity	Temperature
2023-02-11 12:35:16	64	76	99	174	94	85	209	94	72.00	33.70
2023-02-11 12:35:21	76	80	89	189	84	97	197	108	72.50	33.70
2023-02-11 12:35:26	74	77	100	216	96	84	197	104	72.70	33.70
2023-02-11 12:35:32	65	87	89	182	83	85	209	95	72.60	33.70
2023-02-11 12:35:36	77	78	102	167	98	85	199	103	72.50	33.60
2023-02-11 12:35:41	73	76	99	162	96	87	215	99	72.50	33.70
2023-02-11 12:35:49	66	87	91	167	84	98	199	96	72.60	33.70
2023-02-11 12:35:53	67	89	89	171	84	84	210	95	72.60	33.70
2023-02-11 12:35:58	67	89	89	175	95	87	214	97	72.60	33.70
2023-02-11 12:36:01	76	77	101	167	96	86	205	102	72.70	33.70
2023-02-11 12:36:06	72	85	89	166	88	97	203	102	72.60	33.70
2023-02-11 12:36:11	70	85	94	176	95	85	215	95	72.60	33.70
2023-02-11 12:36:16	69	87	95	179	95	84	216	97	72.70	33.70
2023-02-11 12:36:22	70	75	98	161	97	86	218	95	72.60	33.70
2023-02-11 12:36:27	71	90	89	173	95	85	217	96	72.70	33.70
2023-02-11 12:36:31	84	76	91	173	85	97	207	109	72.60	33.70
2023-02-11 12:36:38	71	86	91	172	87	100	206	97	72.60	33.80
2023-02-11 12:36:42	82	76	101	163	89	92	207	108	72.60	33.80
2023-02-11 12:36:47	73	91	88	166	95	86	219	95	72.60	33.80
2023-02-11 12:36:53	75	75	99	167	98	85	222	98	72.60	33.80
2023-02-11 12:36:57	73	86	88	179	85	99	210	95	72.60	33.80
2023-02-11 12:37:03	73	86	90	187	86	99	210	97	72.60	33.80
2023-02-11 12:37:07	74	88	91	187	85	86	219	96	72.60	33.80
2023-02-11 12:37:13	74	85	88	187	85	99	211	96	72.60	33.80
2023-02-11 12:37:18	87	76	90	177	86	98	210	109	72.60	33.80
2023-02-11 12:37:22	84	76	100	172	98	86	213	106	72.50	33.80
2023-02-11 12:37:28	77	88	90	176	86	98	215	96	72.50	33.80
2023-02-11 12:37:32	76	76	97	175	97	87	227	98	72.50	33.80
2023-02-11 12:37:38	76	87	88	181	85	84	221	98	72.50	33.80
2023-02-11 12:37:42	89	76	96	163	85	96	212	111	72.50	33.80
2023-02-11 12:37:48	77	84	90	174	86	98	215	98	72.40	33.80
2023-02-11 12:37:53	76	78	97	174	97	85	227	97	72.40	33.80
2023-02-11 12:37:58	79	91	91	175	95	85	227	97	72.40	33.90
2023-02-11 12:38:03	77	77	98	165	97	85	229	99	72.40	33.90
2023-02-11 12:38:08	83	77	99	162	98	86	229	97	72.40	33.90
2023-02-11 12:38:14	79	90	89	169	95	87	227	97	72.30	33.90
2023-02-11 12:38:18	77	79	96	176	98	84	227	98	72.30	33.90
2023-02-11 12:38:23	78	89	89	178	94	86	227	98	72.30	33.90
2023-02-11 12:38:29	84	76	101	162	98	87	231	100	72.30	33.90

(a) Bonne tomate

Time	TGS2603	TGS2610	TGS2611	TGS2600	TGS2602	MQ136	MQ135	TGS822	Humidity	Temperature
2023-03-02 08:47:22	86	134	131	219	165	99	567	154	84.00	31.40
2023-03-02 08:47:33	86	128	132	199	164	101	569	158	84.20	31.50
2023-03-02 08:47:36	94	117	134	206	146	110	563	168	84.60	31.50
2023-03-02 08:47:40	93	124	126	206	152	111	561	163	84.60	31.50
2023-03-02 08:47:43	85	133	125	194	154	101	577	157	84.50	31.50
2023-03-02 08:47:47	100	116	137	216	159	103	559	171	84.40	31.50
2023-03-02 08:47:52	88	124	126	199	166	100	572	156	84.30	31.50
2023-03-02 08:47:57	97	116	122	206	149	116	556	166	84.20	31.50
2023-03-02 08:48:05	103	114	138	201	159	100	556	169	84.10	31.50
2023-03-02 08:48:09	96	116	128	200	150	115	580	164	84.10	31.50
2023-03-02 08:48:12	97	113	133	213	147	117	558	166	83.90	31.60
2023-03-02 08:49:28	82	122	125	200	147	99	556	148	83.20	31.70
2023-03-02 08:54:20	90	131	133	196	145	112	490	86	83.50	31.80
2023-03-02 08:54:25	97	129	140	211	148	112	502	116	83.10	31.80
2023-03-02 08:54:30	95	116	144	205	151	105	506	139	83.30	31.90
2023-03-02 08:54:35	82	128	129	194	147	100	519	138	83.20	31.90
2023-03-02 08:54:41	92	115	131	201	142	116	503	158	83.20	31.90
2023-03-02 08:54:45	81	127	127	206	142	109	505	148	83.10	32.00
2023-03-02 08:54:54	93	110	134	208	140	116	502	161	82.90	32.00
2023-03-02 08:54:57	80	118	123	198	159	99	514	150	82.90	32.00
2023-03-02 08:55:01	83	121	121	191	141	111	501	153	82.80	32.00
2023-03-02 08:55:06	96	102	132	206	150	100	500	164	82.70	32.00
2023-03-02 08:55:14	80	116	119	200	141	111	500	144	82.60	32.00
2023-03-02 08:55:18	80	118	125	181	142	99	512	143	82.60	32.10
2023-03-02 08:55:21	80	113	119	194	158	101	509	142	82.50	32.10
2023-03-02 08:55:29	89	103	118	202	144	117	495	153	82.50	32.10
2023-03-02 08:55:33	79	115	120	182	140	103	506	141	82.40	32.10
2023-03-02 08:55:36	90	106	120	195	141	113	493	151	82.40	32.10
2023-03-02 08:55:45	79	116	119	211	141	112	496	147	82.30	32.10
2023-03-02 08:55:48	82	110	122	194	155	99	503	139	82.30	32.10
2023-03-02 08:55:51	80	111	120	179	159	99	503	141	82.20	32.20
2023-03-02 08:55:57	84	100	134	194	153	99	501	138	82.20	32.20
2023-03-02 08:56:02	93	101	125	217	143	115	497	154	82.20	32.20
2023-03-02 08:56:07	91	103	129	199	145	108	486	152	82.10	32.20
2023-03-02 08:56:12	79	110	120	178	143	99	503	138	82.10	32.20
2023-03-02 08:56:17	90	102	130	191	143	113	484	153	82.10	32.20
2023-03-02 08:56:25	77	118	118	214	141	114	484	147	82.00	32.20
2023-03-02 08:56:28	78	114	117	197	142	112	484	147	82.00	32.20
2023-03-02 08:56:32	79	114	118	179	143	106	492	140	82.00	32.20

(b) Mauvaise tomate

FIGURE 4.13 – Données d'échantillon de la variété de tomate *Akikon* provenant du nez électronique proposé

Les Figures 4.14 à 4.23 illustrent les signaux électroniques des tomates *Tounvi* et *Akikon* pour les capteurs TGS2603, TGS2610, TGS2611, TGS2600, TGS2602, MQ136, MQ135, TGS822, ainsi que pour l'humidité et la température. Selon la qualité de chaque tomate, des différences de signaux peuvent être observées. Les métriques dérivées, telles que la variance, l'écart type, le coefficient de variation et l'amplitude, offrent des informations sur cette variation. Les formules des métriques statistiques utilisées sont les suivantes :

1. Moyenne (également appelée moyenne arithmétique ou moyenne) :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4.1}$$

où $\bar{x}$ est la moyenne, n est le nombre de points de données, et x_i sont les points de données individuels.

2. Variance :

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{4.2}$$

où s^2 est la variance, n est le nombre de points de données, x_i sont les points de données individuels, et $\bar{x}$ est la moyenne.

3. Écart-type :

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{4.3}$$

où s est l'écart type, n est le nombre de points de données, x_i sont les points de données

individuels, et $\bar{x}$ est la moyenne.

4. Amplitude :

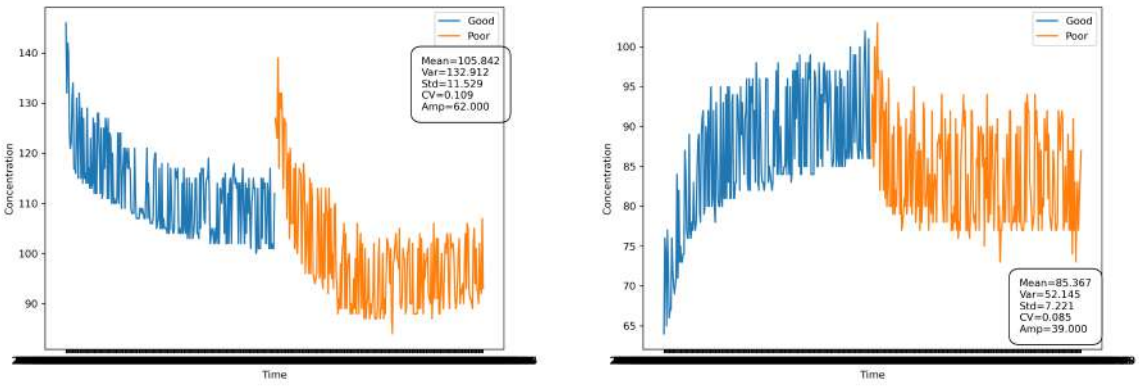
$$A = \max(x) - \min(x) \tag{4.4}$$

où A est l'amplitude, $\max(x)$ est la valeur maximale des points de données, et $\min(x)$ est la valeur minimale des points de données.

5. Coefficient de variation :

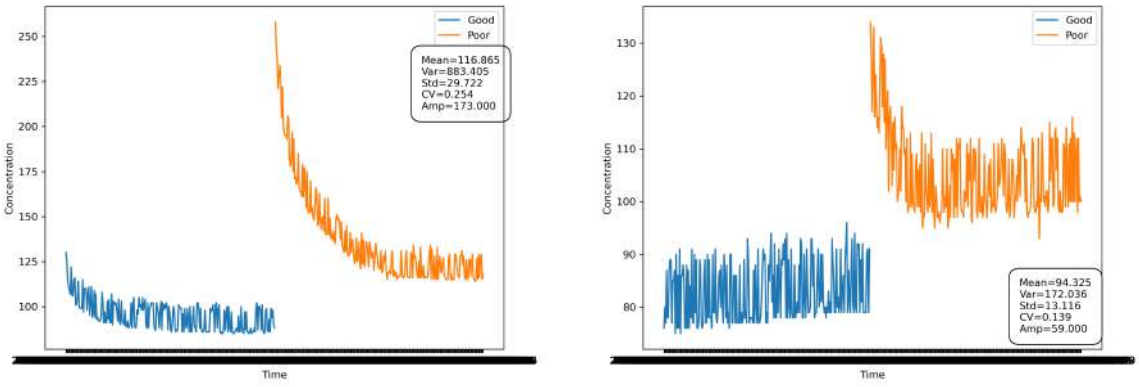
$$CV = \frac{s}{\bar{x}} \times 100\% \tag{4.5}$$

où CV est le coefficient de variation, s est l'écart type, et $\bar{x}$ est la moyenne.



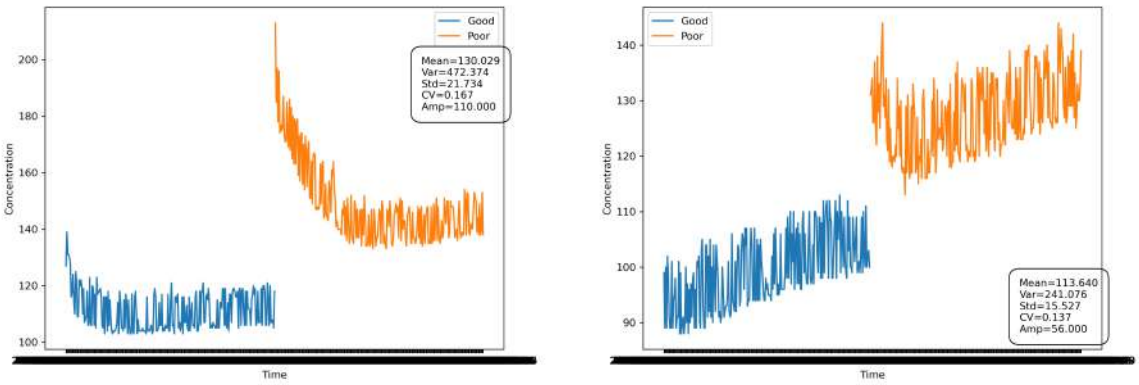
(a) Tomate *Tounvi* (b) Tomate *Akikon*

FIGURE 4.14 – Variation du capteur TGS2603



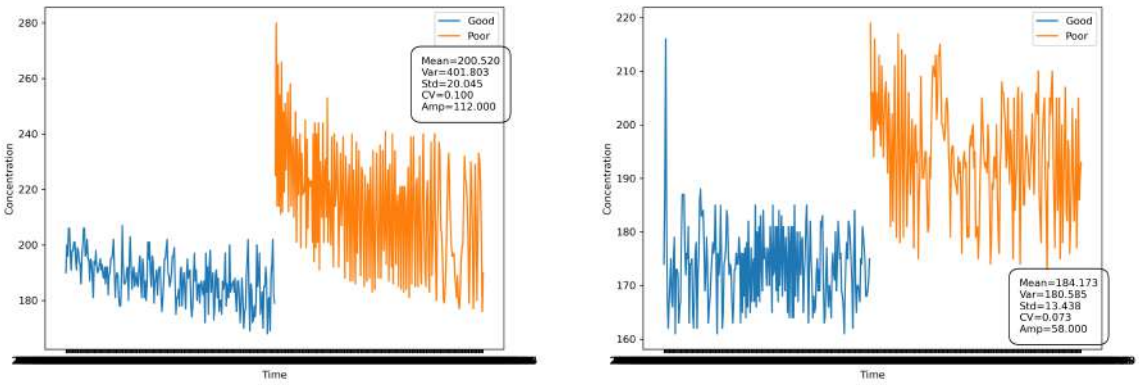
(a) Tomate *Tounvi* (b) Tomate *Akikon*

FIGURE 4.15 – Variation du capteur TGS2610



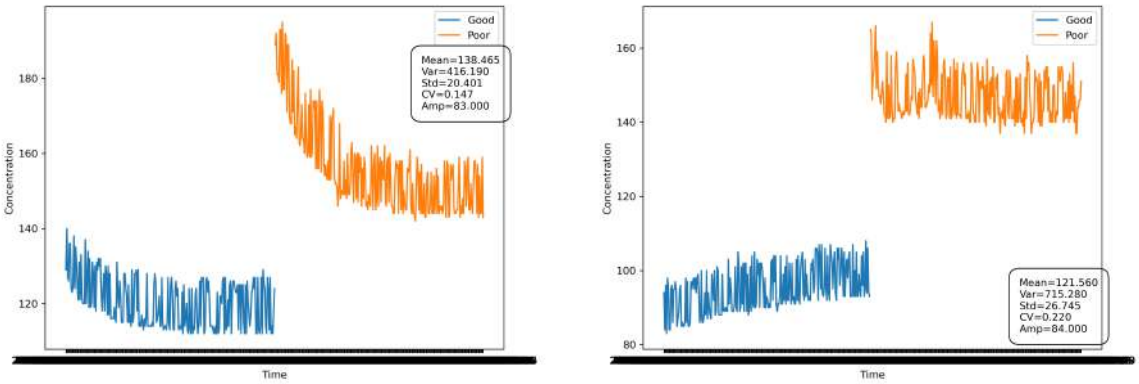
(a) Tomate *Tounvi* (b) Tomate *Akikon*

FIGURE 4.16 – Variation du capteur TGS2611



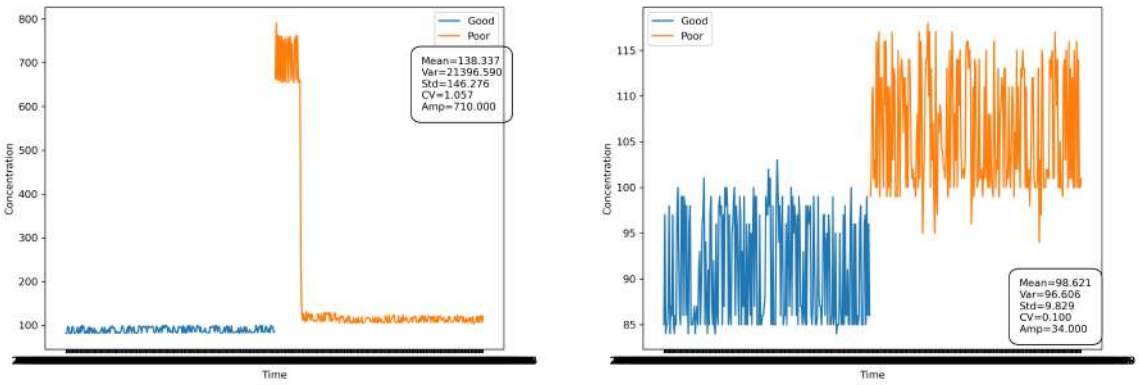
(a) Tomate *Tounvi* (b) Tomate *Akikon*

FIGURE 4.17 – Variation du capteur TGS2600



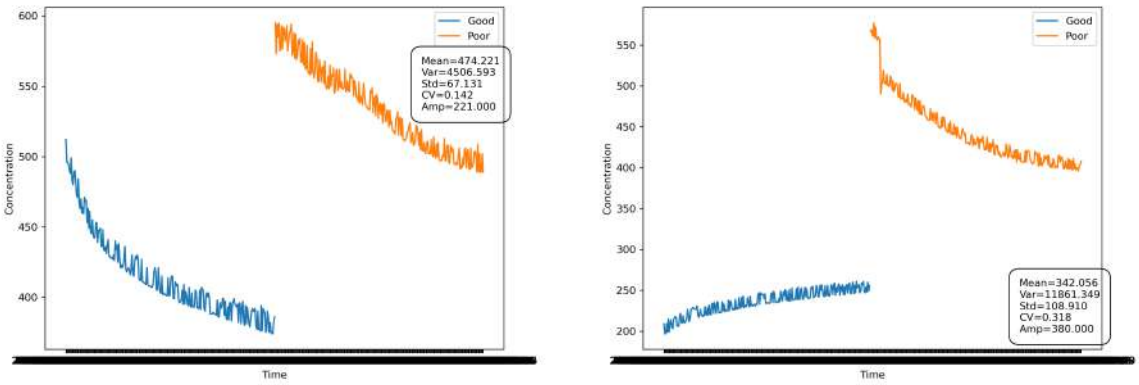
(a) Tomate *Tounvi* (b) Tomate *Akikon*

FIGURE 4.18 – Variation du capteur TGS2602



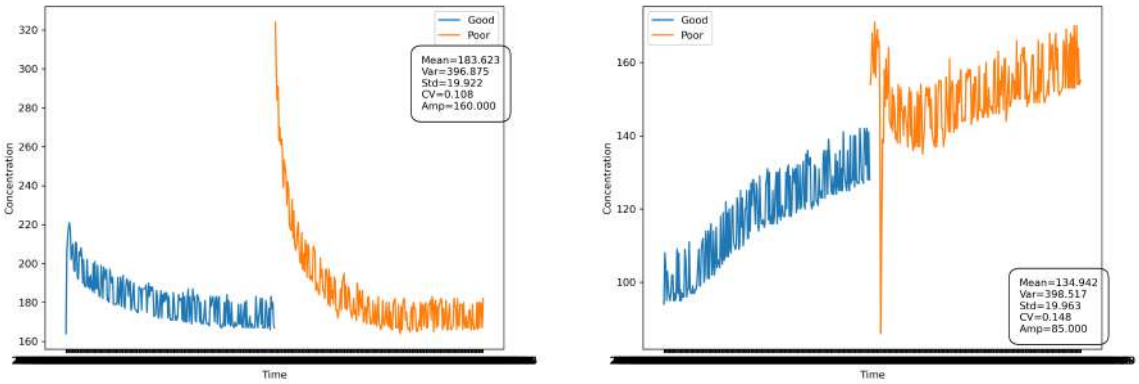
(a) Tomate *Tounvi* (b) Tomate *Akikon*

FIGURE 4.19 – Variation du capteur MQ136



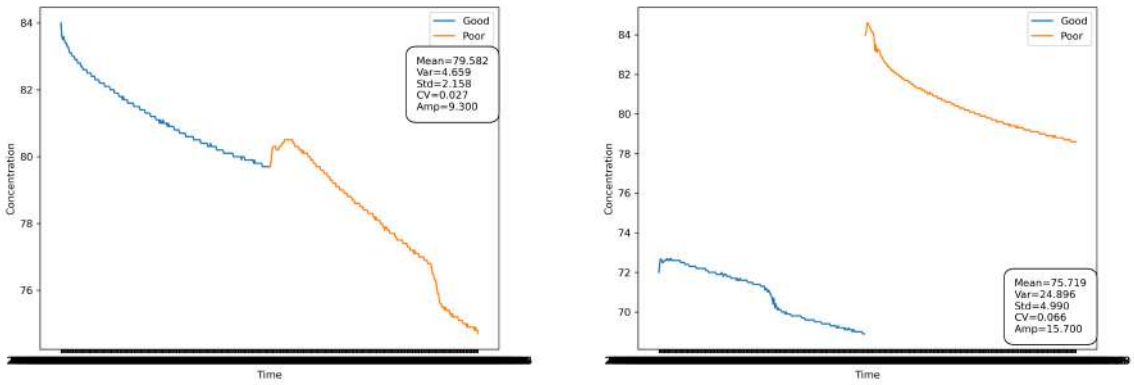
(a) Tomate *Tounvi* (b) Tomate *Akikon*

FIGURE 4.20 – Variation du capteur MQ135



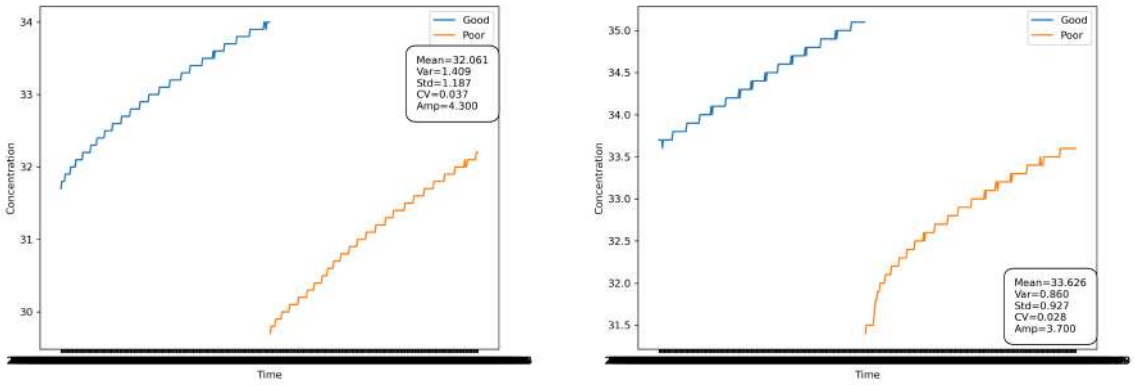
(a) Tomate *Tounvi* (b) Tomate *Akikon*

FIGURE 4.21 – Variation du capteur TGS822



(a) Tomate *Tounvi* (b) Tomate *Akikon*

FIGURE 4.22 – Variation de l'humidité



(a) Tomate *Tounvi* (b) Tomate *Akikon*

FIGURE 4.23 – Variation de la température

Les caractéristiques peuvent être classées de la moins dispersée à la plus dispersée à l'aide de la variance et de l'écart type. Cela pourrait conduire à une classification des caractéristiques, allant de celles qui sont les moins sensibles aux gaz générés par les tomates à celles qui le sont le plus. Les changements significatifs observés confirment l'efficacité du nez électronique expérimental.

Conclusion

La conception du nez électronique expérimental a été réalisée à l'aide de 8 capteurs de gaz ainsi que d'un capteur d'humidité et de température. Afin de garantir son bon fonctionnement, il a été utilisé pour collecter des données sur la tomate, la pomme verte, la banane, l'orange et l'oignon. En outre, grâce à un protocole de collecte de données, ce dispositif a permis de suivre le cycle de vie de deux variétés locales de tomates. Ces données rendues accessibles publiquement, seront utilisées pour classer la tomate en différents états tels que *très bon*, *bon*, *moyen* et *mauvais*, dans le but de préserver leur qualité, conformément aux connaissances d'experts.

Le chapitre suivant présente l'application des modèles d'apprentissage supervisé pour prédire l'état des tomates.

CHAPITRE 5

ÉTAT DES TOMATES : APPROCHES SUPERVISÉES

Introduction

La tomate est l'un des légumes les plus consommés dans le monde en raison de sa valeur nutritionnelle et de ses bienfaits pour la santé. La qualité des tomates est un facteur déterminant qui influence leur commercialisation et leur acceptation par les consommateurs [236, 238]. Traditionnellement, l'évaluation de la qualité des tomates repose sur des analyses sensorielles, une méthode subjective et chronophage. La technologie du nez électronique s'impose progressivement comme un outil efficace pour l'évaluation objective et rapide de la qualité des tomates. Grâce à cette technologie, de nombreuses données sont collectées, lesquelles sont utiles pour prédire la qualité des tomates. L'apprentissage automatique, en particulier l'apprentissage supervisé, semble être la meilleure approche pour traiter ces données.
Ce chapitre présente l'utilisation des données du nez électronique pour entraîner des modèles d'apprentissage supervisé dans le but de prédire l'état des tomates.

5.1 Approches d'apprentissage automatique pour les tomates

Différentes techniques d'apprentissage automatique sont utilisées dans la littérature pour des applications liées à la tomate. Ghaffari et al. [249] ont utilisé un perceptron multicouche (MLP), une quantification vectorielle d'apprentissage (LVQ) et une fonction à base radiale (RBF) basées sur des réseaux neuronaux artificiels (ANN) pour distinguer les plants de tomates sains des plants artificiellement infectés. Huang et al. [216] ont proposé une méthode non destructive pour évaluer la qualité des tomates. Ils ont utilisé l'analyse discriminante de Fisher (FDA) et le classificateur à vecteurs de support (SVC) pour la classification des tomates. Ils ont mis en place un nez électronique composé de 10 capteurs de gaz à semi-conducteur d'oxyde métallique. L'analyse en composantes principales a été utilisée pour réduire le nombre de variables et seules les quatre premières composantes principales, qui expliquent plus de 95% de l'information brute, ont été

conservées comme variables d'entrée pour les modèles. Les précisions obtenues sur les données de test sont respectivement de 68,12% et 75,36% pour le FDA et le SVC. Concepcion et al. [250] ont présenté un nez électronique composé de deux capteurs de gaz afin de répondre aux problèmes de catégorisation causés par des hypothèses subjectives humaines. Ils ont utilisé un réseau de neurones artificiels à perceptron multicouche (MLP) pour distinguer les différents états de la purée de tomate, en se basant uniquement les concentrations d'ammoniac et de méthane ainsi que les niveaux de pH. Une précision de 93,33% a été obtenue pour la catégorisation de la détérioration de la qualité de la purée de tomates en utilisant un MLP à trois couches avec 120 et 50 neurones artificiels sur les première et deuxième couches cachées, respectivement.

Dans la littérature, aucune étude ne semble avoir proposé une classification automatique de la qualité des tomates utilisant un ensemble de données couvrant l'intégralité de leur cycle de vie. Le travail présenté dans ce chapitre est basé sur des algorithmes classiques d'apprentissage supervisé. Plus précisément, la performance de quatre algorithmes, à savoir les forêts aléatoires (RF), les k plus proches voisins (k-NN), le perceptron multicouche (MLP) et les machines à vecteurs de support (SVM), est comparée pour la classification de la qualité des tomates en utilisant les données collectées par le nez électronique développé. Ces choix sont motivés par les raisons suivantes :

Random Forest : Cet algorithme est bien adapté aux données complexes et bruitées, souvent observées dans les systèmes comme le nez électronique. Il est robuste face aux surapprentissages grâce à son approche par agrégation de plusieurs arbres de décision. Il a démontré son efficacité dans de nombreuses applications de classification basées sur des capteurs chimiques, en raison de sa capacité à gérer des ensembles de données à hautes dimensions [251].

k-Nearest Neighbors : Il est un algorithme simple et interprétable, souvent utilisé comme base pour comparer d'autres modèles. Il est particulièrement performant lorsque les données ont des clusters bien séparés dans l'espace des caractéristiques, ce qui peut être le cas dans des systèmes comme le nez électronique [252].

Multi-Layer Perceptron : Les réseaux de neurones artificiels tels que le MLP sont particulièrement efficaces pour modéliser des relations non linéaires complexes, comme celles présentes dans les réponses des capteurs du nez électronique. De plus, ils sont largement utilisés dans la classification multi-classe grâce à leur capacité à apprendre des représentations internes pertinentes des données [253].

Support Vector Machine : Il est bien connu pour sa capacité à trouver une frontière de décision optimale entre les classes, même dans des espaces à haute dimension. Il est particulièrement utile dans le cas de données non linéaires, grâce à l'utilisation de fonctions noyau (kernel) adaptées. SVM a été fréquemment utilisé pour la classification dans des systèmes similaires et a montré une grande efficacité pour la détection d'états alimentaires via des capteurs [254].

Les nez électroniques produisent des données riches, souvent bruitées et multidimensionnelles. Les algorithmes choisis sont largement documentés pour leur efficacité dans l'analyse de ces types de données [3]. Ils offrent une bonne diversité en termes de complexité, de robustesse et de méthodes de prise de décision, permettant ainsi une évaluation complète des performances pour une classification multi-classe précise.

5.2 Random Forest

Le Random Forest est un algorithme d'apprentissage supervisé qui combine plusieurs arbres de décision pour améliorer la performance et la robustesse des prédictions. Chaque arbre est entraîné sur un échantillon aléatoire de données, et à chaque division, seul un sous-ensemble de caractéristiques est pris en compte [255]. En classification, la prédiction finale est obtenue par vote majoritaire, tandis qu'en régression, elle correspond à la moyenne des prédictions. La Figure 5.1 illustre ce fonctionnement. L'architecture du modèle repose sur une structure d'ensemble avec des arbres indépendants. Ses cas d'utilisation incluent la prédiction dans des domaines comme la médecine (diagnostic), la finance (analyse du risque), et l'agriculture (prédiction du rendement des cultures). Son avantage réside dans sa capacité à gérer les variables bruitées et redondantes, bien qu'il puisse être limité en termes d'interprétabilité [251].

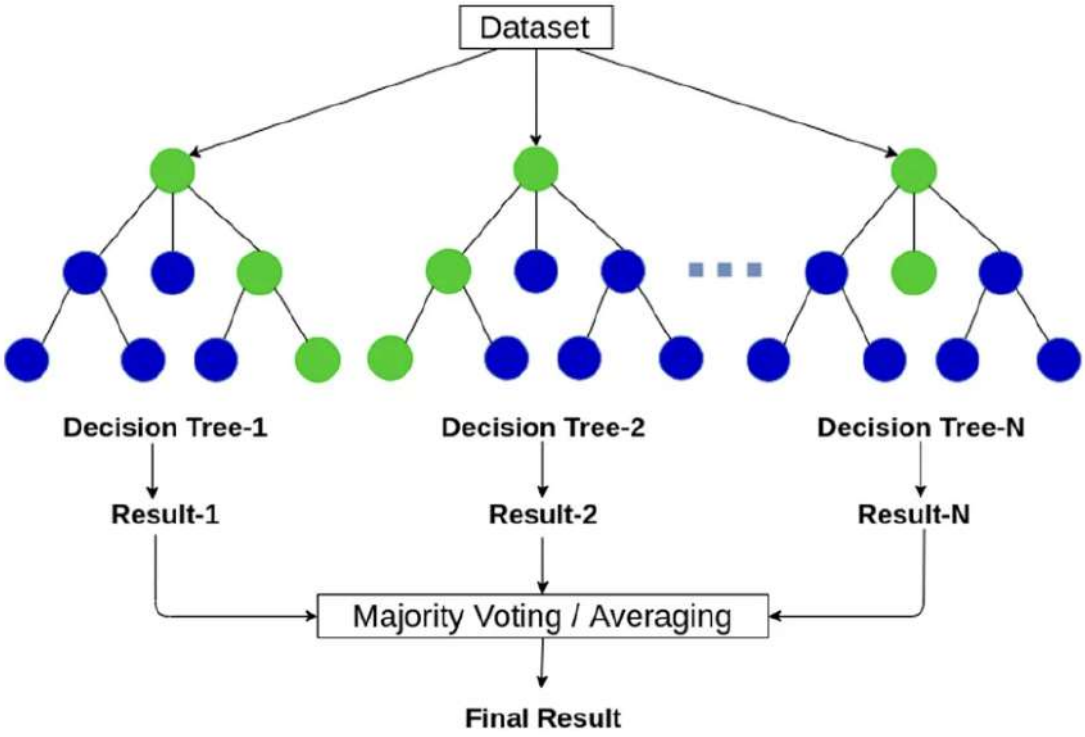


FIGURE 5.1 – Random Forest [1]

1. Source : https://medium.com/@jainvidip/understanding-random-forest-0ca15aaa443c, consultée le 23 novembre 2024.

5.3 k-Nearest Neighbors

L'algorithme k-NN classe une observation en fonction des k exemples les plus proches dans l'espace des caractéristiques, selon une métrique de distance comme la distance euclidienne [191]. Son architecture est simple, car il ne construit pas de modèle explicite mais conserve simplement l'ensemble des données. Les prédictions se basent uniquement sur les instances locales à l'observation cible. En classification, la finalité est d'attribuer à une nouvelle observation la classe majoritaire parmi les k voisins les plus proches. En régression, la prédiction correspond à la moyenne des valeurs cibles des voisins. Plus d'informations sont données à la sous-section 3.2.2.2. Utilisé en classification d'images, détection de fraudes ou systèmes de recommandation, k-NN est apprécié pour sa simplicité et son efficacité sur des petits ensembles de données. Toutefois, il devient inefficace pour des ensembles volumineux en raison du coût de calcul élevé des distances, et il est sensible à la présence de données déséquilibrées ou de bruit. Une normalisation préalable des variables est généralement requise pour garantir une performance optimale, surtout si les caractéristiques ont des échelles différentes [252].

5.4 Multi-Layer Perceptron

Le Multi-Layer Perceptron est une architecture de réseau de neurones composée de couches entièrement connectées, où chaque neurone reçoit des entrées de tous les neurones de la couche précédente. Chaque neurone applique une transformation linéaire suivie d'une fonction d'activation non linéaire. MLP peut être appliqué à des problèmes de classification binaire ou multi-classe [256]. Comme indiqué sur la Figure 5.2, un MLP comprend :

1. Une couche d'entrée dont le nombre de neurones correspond aux dimensions des caractéristiques des données.
2. Une ou plusieurs couches cachées, responsables de l'apprentissage des représentations complexes.
3. Une couche de sortie, où le nombre de neurones correspond au nombre de classes pour la classification ou à une unique sortie continue pour la régression.

Les poids des connexions sont ajustés pour minimiser une fonction de coût, comme l'entropie croisée pour la classification. Ce processus s'effectue via l'algorithme de rétropropagation, qui calcule les gradients par descente de gradient, souvent optimisée à l'aide de variantes comme Adam ou le gradient stochastique. Les MLP sont particulièrement efficaces pour des tâches complexes nécessitant une extraction de caractéristiques implicite, comme la reconnaissance vocale, la classification d'images, la prédiction de séries temporelles (par exemple, des données financières ou climatiques) et la classification de textes. Cependant, ils exigent des hyperparamètres bien calibrés (par exemple, la taille des couches, le taux d'apprentissage) et un volume significatif de données pour éviter le surapprentissage [257].

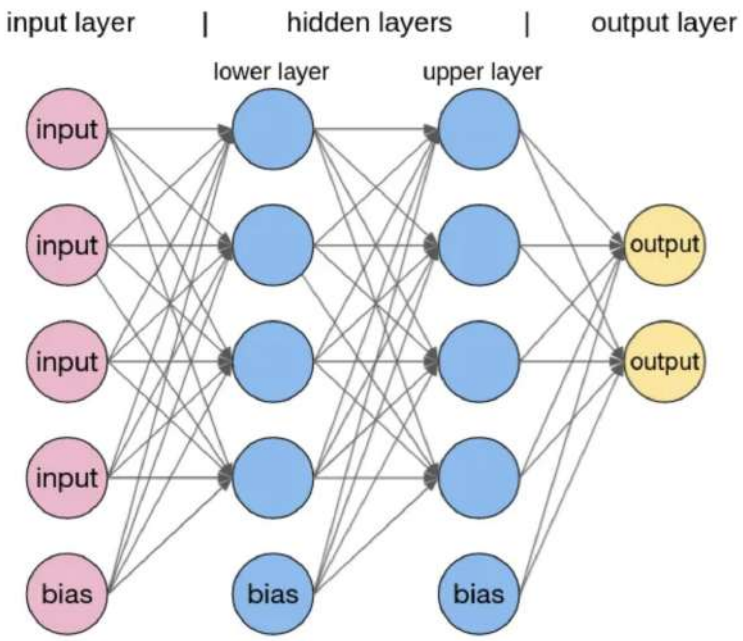


FIGURE 5.2 – Structure de base du MLP [2]

5.5 Support Vector Machine

Le SVM est un algorithme d'apprentissage supervisé conçu pour résoudre des problèmes de classification et de régression. L'idée principale est de trouver un hyperplan optimal qui sépare les différentes classes avec la plus grande marge possible. La marge est la distance entre l'hyperplan et les points d'entraînement les plus proches de chaque classe, appelés vecteurs de support. L'objectif est de maximiser cette marge pour améliorer la généralisation du modèle. La séparation optimale est obtenue en résolvant un problème d'optimisation mathématique [258]. Dans les cas où les données ne sont pas linéairement séparables, SVM utilise un concept appelé le kernel trick, qui permet de transformer les données dans un espace de dimension supérieure où elles deviennent séparables linéairement. Par exemple, le noyau gaussien (RBF) est largement utilisé. Cette transformation rend SVM particulièrement puissant pour des problèmes non linéaires tout en conservant sa capacité à maintenir un bon pouvoir de généralisation [254].

En classification, SVM cherche à classifier les données en trouvant l'hyperplan qui divise les classes avec la plus grande marge. En cas de données non linéaires, les noyaux permettent une séparation dans un espace de caractéristiques transformé. L'algorithme attribue la classe de la nouvelle observation en fonction de la classe du côté de l'hyperplan dans lequel elle se trouve. En régression (SVR), le but est de prédire des valeurs continues. Plutôt que de maximiser la marge entre les classes, le SVR pour régression tente de trouver une fonction qui s'adapte au mieux aux données tout en permettant certaines erreurs. Plus d'informations sont données à la sous-section

2. Source : `https://medium.com/codex/introduction-to-how-an-multilayer-perceptron-works-but-without-complicated-math-a423979897ac`, consultée le 23 novembre 2024.

3.2.2.1. SVM est largement utilisé dans des domaines comme la bio-informatique (par exemple, la classification de séquences d'ADN), la reconnaissance faciale, et l'analyse de sentiments dans les sciences sociales. Il est particulièrement efficace pour des données de haute dimension où d'autres algorithmes peuvent échouer. Cependant, il peut devenir computationalement coûteux lorsque l'on travaille avec de grands ensembles de données, en raison de la complexité de l'optimisation, surtout pour des noyaux non linéaires [254].

5.6 Métriques d'évaluation

Il est essentiel de définir des critères permettant de comparer les modèles. La performance des classificateurs doit être comparée en se basant sur des facteurs clés [259]. Les termes TP, TN, FP et FN sont couramment utilisés en classification pour décrire les résultats d'une prédiction. TP (Vrai Positif) désigne le nombre de cas positifs correctement identifiés, TN (Vrai Négatif) correspond au nombre de cas négatifs correctement identifiés, FP (Faux Positif) indique le nombre de cas négatifs incorrectement classés comme positifs, et FN (Faux Négatif) désigne les cas positifs incorrectement classés comme négatifs. Ces quatre termes sont utilisés pour calculer plusieurs métriques d'évaluation telles que l'exactitude, la précision, le rappel et le score F1, qui permettent d'évaluer la performance des modèles.

Le rappel est une métrique qui évalue la proportion de vrais positifs parmi l'ensemble des exemples réellement positifs. Il se calcule de la manière suivante :

$$\text{Rappel} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5.1}$$

Le rappel évalue la capacité du modèle à identifier tous les exemples positifs, même si cela implique d'identifier certains exemples négatifs comme positifs.

La précision mesure la justesse des prédictions positives d'un modèle de classification. Elle indique la proportion d'observations correctement classées comme positives parmi toutes celles qui ont été classées comme positives. En d'autres termes, elle évalue la capacité du modèle à éviter de classer incorrectement une observation comme positive. Elle est calculée comme suit :

$$\text{Précision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5.2}$$

La précision et le rappel sont utilisés pour calculer une autre métrique, le score F1, qui mesure la combinaison de ces deux indicateurs. Il est calculé comme suit :

$$\text{Score F1} = 2 \times \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} \tag{5.3}$$

L'exactitude mesure la proportion de prédictions correctes parmi toutes les prédictions effec-

tuées. Elle est calculée à l'aide de la formule suivante :

$$\text{Exactitude} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5.4}$$

5.7 Matrice de confusion

La matrice de confusion est un outil couramment utilisé pour évaluer la performance des modèles de classification. Dans les problèmes de classification multi-classe, il s'agit d'une matrice carrée qui affiche le nombre de prédictions correctes et incorrectes pour chaque classe. Les lignes de la matrice correspondent aux étiquettes de classe réelles, tandis que les colonnes représentent les étiquettes de classe prédites. Les éléments diagonaux indiquent le nombre de prédictions correctes, tandis que les éléments hors diagonale montrent le nombre de mauvaises classifications.

TABLEAU 5.1 – Matrice de confusion pour un problème de classification multi-classe

	Classe prédite 1	**Classe prédite 2**	$\cdots$	**Classe prédite n**
Classe réelle 1	C_{11}	C_{12}	$\cdots$	C_{1n}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
Classe réelle n	C_{n1}	C_{n2}	$\cdots$	C_{nn}

Ici, C_{ij} représente le nombre d'instances qui appartiennent réellement à la classe i et qui ont été classées comme appartenant à la classe j.

5.8 Pré-traitement et nettoyage des données

Les données collectées sur les variétés de tomates locales *Tounvi* et *Akikon*, dont les détails sont présentés dans la section 4.4, ont été utilisées pour cette expérience. Cette collecte a permis la création d'un ensemble de données structuré dans des fichiers CSV, comprenant 10 caractéristiques, comme illustré dans les Figures 4.12 et 4.13. Le Tableau 5.2 présente ces caractéristiques ainsi que leurs plages de valeurs respectives.

TABLEAU 5.2 – Plages de valeurs des caractéristiques

Caractéristiques	**Plage de valeurs**
MQ135, MQ136, TGS822, TGS2600, TGS2602, TGS2603, TGS2610, and TGS2611	$[0;1.023] \cap \mathbb{N}$
Température ($°C$)	$[-40;80]$
Humidité (%)	$[0;100]$

Les valeurs entières, comprises entre 0 et 1.023, renvoyées par les capteurs de gaz représentent des lectures analogiques de la concentration de gaz mesurée. Ces valeurs numériques sont proportionnelles à la concentration de gaz détectée.

Les échantillons ont été classés en différents états en utilisant les critères de classification des tomates après récolte observés pendant le stockage à température ambiante (voir Tableau 5.3). Cette classification, élaborée après observation de la manière dont les vendeurs au marché organisent leurs produits en tas et après examen des fruits tout au long de l'expérience, a été validée par des experts agronomes.

En conséquence, une nouvelle caractéristique cible a été définie, avec des valeurs appartenant à l'ensemble suivant $\{0;1;2;3\}$. Les stades de maturité, à savoir *très bon*, *bon*, *moyen* et *mauvais*, sont associés aux valeurs 0, 1, 2 et 3 respectivement. Par ailleurs, la taille de l'ensemble de données est de 98.907×11.

TABLEAU 5.3 – Critères de classification des tomates

Stade de maturité	**Caractéristiques du stade**
Très bon	Fruit ferme à 100%
Bon	Fruit ferme à 80% (ramollissement engagé)
Moyen	Les tissus de la paroi ramollissent fortement
Mauvais	La détérioration se produit à la paroi du fruit

Pour pré-traiter les données, un contrôle systématique a été mis en place afin de garantir leur conformité avec les exigences définies. Tout d'abord, les plages de valeurs acceptables ont été établies pour chaque paramètre mesuré, en tenant compte des caractéristiques spécifiques des données collectées (voir Tableau 5.2). Chaque colonne a été analysée pour vérifier que toutes les valeurs se situaient dans les intervalles spécifiés ou correspondaient aux ensembles d'options autorisées dans le cas des données catégoriques. Ce contrôle a permis de détecter et de signaler les écarts éventuels par rapport aux limites prédéfinies. Ensuite, une attention particulière a été portée à la cohérence des en-têtes des fichiers. Tous les fichiers ont été examinés pour s'assurer qu'ils partageaient une structure identique, évitant ainsi toute incohérence dans les analyses ultérieures. Par ailleurs, la présence de valeurs manquantes ou aberrantes a été évaluée. Il a été confirmé qu'aucune valeur manquante ni anomalie n'avait été détectée par les capteurs, rendant inutile toute imputation ou suppression de données. Ces vérifications approfondies ont permis de garantir l'intégrité et la fiabilité des données collectées, constituant ainsi une base solide pour les travaux ultérieurs.

5.9 Evaluation des performances des modèles

Un organigramme détaillé a été créé pour illustrer les étapes séquentielles de cette expérience. Il représente visuellement les différentes phases de la méthodologie, incluant la collecte des données, leur étiquetage en fonction du niveau de maturité, ainsi que le mélange des données, le sous-échantillonnage et le sur-échantillonnage. Plusieurs modèles d'apprentissage automatique, tels que Random Forest, k-NN, SVM et MLP, ont été entraînés sur les ensembles de données

sous-échantillonnés, sur-échantillonnés et normaux. Cette démarche a été adoptée pour traiter les déséquilibres observés dans la distribution des données. L'objectif était d'évaluer la performance des modèles sous différentes techniques d'échantillonnage. Ensuite, la performance de chaque modèle a été évaluée sur les trois types d'ensembles de données : sous-échantillonnés, sur-échantillonnés et normaux.

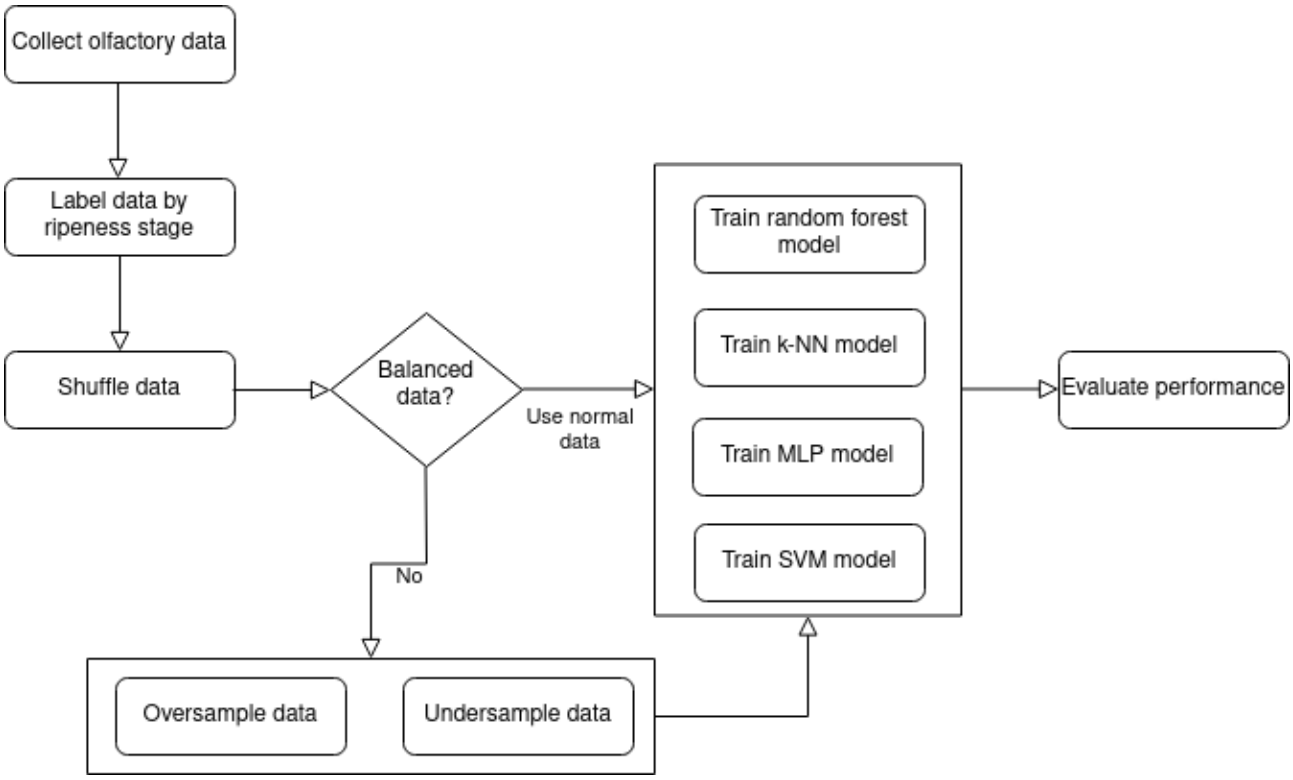


FIGURE 5.3 – Méthodologie pour la prédiction de la qualité des tomates à l'aide de données de nez électronique

Il y a 11.159, 13.910, 42.272, 31.566 observations pour les classes *très bon*, *bon*, *moyen* et *mauvais* respectivement. Le jeu de données présente une distribution déséquilibrée selon la caractéristique cible. Bien que les tomates de l'expérience aient été récoltées et conservées dans des conditions identiques, la durée de vie des fruits a présenté des variations notables (voir Tableau 4.3). Cette disparité peut être attribuée à plusieurs facteurs internes, souvent difficiles à quantifier. Même sous des conditions de stockage contrôlées, des différences subtiles dans la maturité initiale des tomates ou des caractéristiques biologiques cachées, comme des micro-dommages survenus lors de la récolte, peuvent influencer leur taux de dégradation [260, 261]. De plus, bien que l'environnement semble homogène, des microvariations dans les paramètres tels que la température ou l'humidité peuvent aussi avoir un impact sur la vitesse de détérioration des tomates [261, 262]. Ces résultats suggèrent que des facteurs biologiques complexes, en plus des conditions externes, jouent un rôle clé dans la conservation des tomates.

Pour remédier au biais potentiel dans l'entraînement des modèles, plusieurs méthodes auraient pu être envisagées. L'une de ces approches consiste à collecter des données équilibrées, en augmentant le nombre d'échantillons pour les classes sous-représentées. Une autre option

consiste à utiliser des techniques telles que le sur-échantillonnage et le sous-échantillonnage. La première option n'étant pas réalisable, le jeu de données devait couvrir l'ensemble de la durée de vie de la tomate sans aucune influence humaine. Dans ce contexte, il aurait été difficile de contrôler la durée de vie dans l'un des stades caractéristiques mentionnés dans le Tableau 5.3. Pour résoudre le problème du déséquilibre des classes, des techniques de sur-échantillonnage et de sous-échantillonnage ont été appliquées. Le sur-échantillonnage a permis d'augmenter le nombre d'instances dans la classe minoritaire, tandis que le sous-échantillonnage a permis de rééquilibrer la distribution en réduisant le nombre d'instances dans la classe majoritaire. Ces techniques ont pour objectif d'améliorer la précision et la fiabilité du modèle de classification, tout en évaluant simultanément les scénarios dans lesquels le modèle montre les meilleures performances. Ensuite, les différents algorithmes d'apprentissage supervisé ont été testés sur trois types de jeux de données :

Données normales : sans modification.

Données sous-échantillonnées : où la taille totale du jeu de données est de 44.636×11, avec 11.159 observations pour chacune des 4 classes dans la caractéristique cible.

Données sur-échantillonnées : où la taille totale du jeu de données est de 169.088×11, avec 42.272 observations pour chacune des 4 classes dans la caractéristique cible.

Chaque ensemble de données a été divisé en deux parties : 80% pour la phase d'entraînement et 20% pour la phase de test. Après l'entraînement des modèles sur l'ensemble d'entraînement, leur performance a été évaluée sur l'ensemble de test. Des matrices de confusion (Figures 5.4 à 5.6) et des rapports de classification (Tableaux 5.4 à 5.6) ont été calculés pour chaque modèle. Ces mesures d'évaluation ont permis de mieux comprendre les performances des classificateurs. Les colonnes de précision, de rappel et de score F1 indiquent les valeurs pour chaque classe, tandis que l'exactitude évalue la performance globale du modèle.

TABLEAU 5.4 – Comparaison des performances sur l'ensemble de données normal

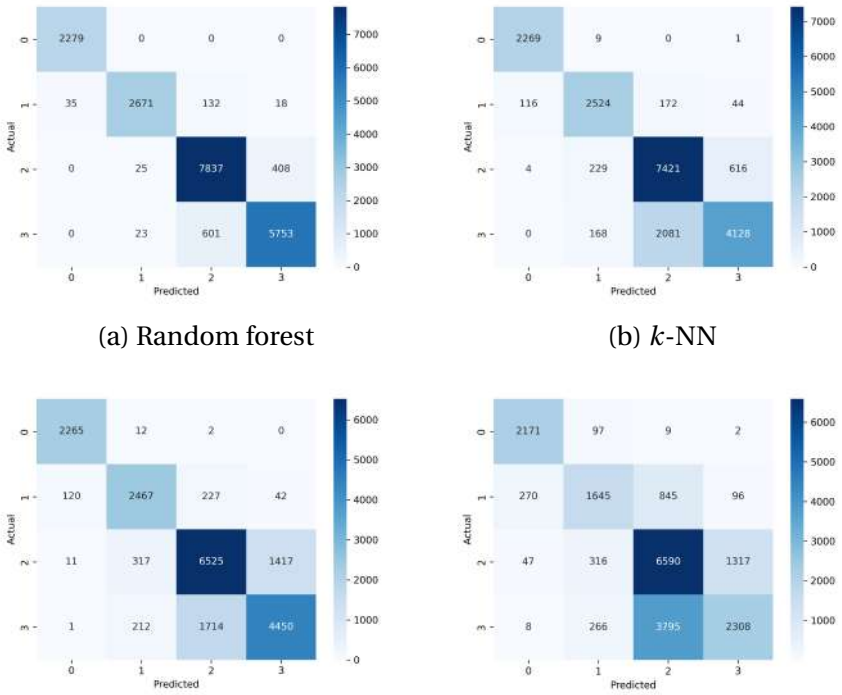
Algorithmes	**Classes**	**Précision**	**Rappel**	**Score F1**	**Exactitude**
Random Forest	*Très bon*	0,98	1,00	0,99	0,94
	Bon	0,98	0,94	0,96	
	Moyen	0,91	0,95	0,93	
	Mauvais	0,93	0,90	0,92	
k-NN	*Très bon*	0,95	1,00	0,97	0,83
	Bon	0,86	0,88	0,87	
	Moyen	0,77	0,90	0,83	
	Mauvais	0,86	0,65	0,74	
MLP	*Très bon*	0,94	0,99	0,97	0,79
	Bon	0,82	0,86	0,84	
	Moyen	0,77	0,79	0,78	
	Mauvais	0,75	0,70	0,72	
SVM	*Très bon*	0,87	0,95	0,91	0,64
	Bon	0,71	0,58	0,64	
	Moyen	0,59	0,80	0,68	
	Mauvais	0,62	0,36	0,46	

TABLEAU 5.5 – Comparaison des performances sur l'ensemble de données sous-échantillonné

Algorithmes	**Classes**	**Précision**	**Rappel**	**Score F1**	**Exactitude**
Random Forest	*Très bon*	0,99	1,00	0,99	0,92
	Bon	0,96	0,96	0,96	
	Moyen	0,86	0,84	0,85	
	Mauvais	0,86	0,87	0,86	
k-NN	*Très bon*	0,96	0,99	0,97	0,83
	Bon	0,86	0,91	0,88	
	Moyen	0,69	0,77	0,73	
	Mauvais	0,80	0,63	0,71	
MLP	*Très bon*	0,97	0,96	0,97	0,81
	Bon	0,88	0,85	0,87	
	Moyen	0,67	0,69	0,68	
	Mauvais	0,72	0,72	0,72	
SVM	*Très bon*	0,90	0,95	0,93	0,69
	Bon	0,72	0,73	0,73	
	Moyen	0,51	0,49	0,50	
	Mauvais	0,59	0,58	0,59	

TABLEAU 5.6 – Comparaison des performances sur l'ensemble de données sur-échantillonné

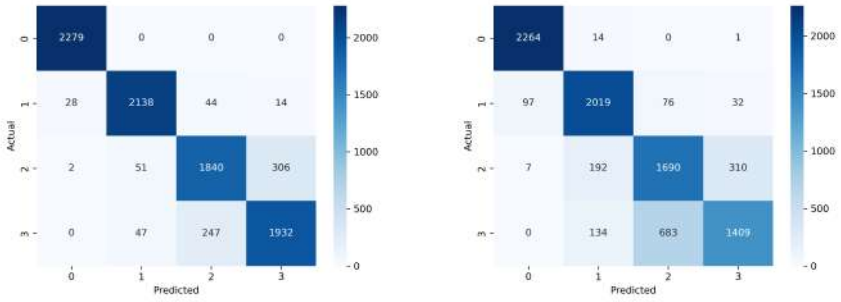
Algorithmes	**Classes**	**Précision**	**Rappel**	**Score F1**	**Exactitude**
Random Forest	*Très bon*	1,00	1,00	1,00	0,98
	Bon	0,99	1,00	1,00	
	Moyen	0,96	0,95	0,96	
	Mauvais	0,96	0,96	0,96	
k-NN	*Très bon*	0,99	1,00	0,99	0,90
	Bon	0,92	0,98	0,95	
	Moyen	0,80	0,83	0,82	
	Mauvais	0,88	0,78	0,83	
MLP	*Très bon*	0,98	0,99	0,99	0,85
	Bon	0,92	0,94	0,93	
	Moyen	0,80	0,64	0,71	
	Mauvais	0,73	0,84	0,78	
SVM	*Très bon*	0,90	0,96	0,93	0,69
	Bon	0,72	0,74	0,73	
	Moyen	0,52	0,50	0,51	
	Mauvais	0,61	0,58	0,59	



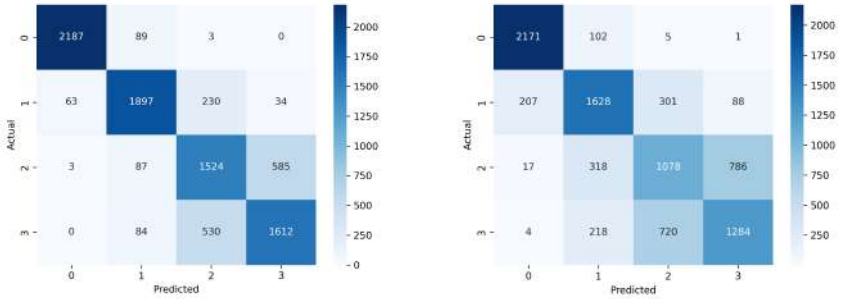
(a) Random forest (b) *k*-NN

(c) MLP (d) SVM

FIGURE 5.4 – Matrices de confusion sur l'ensemble de données normal

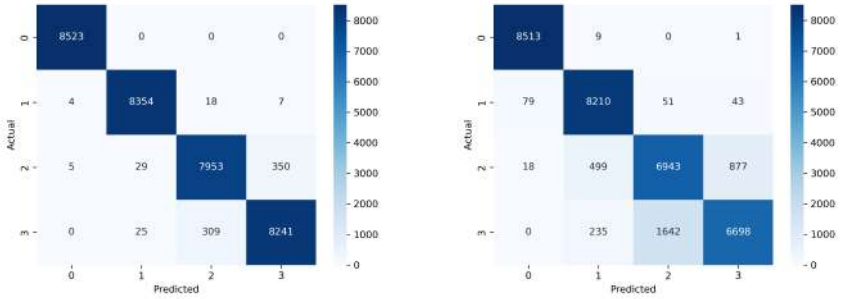


(a) Random forest (b) k-NN

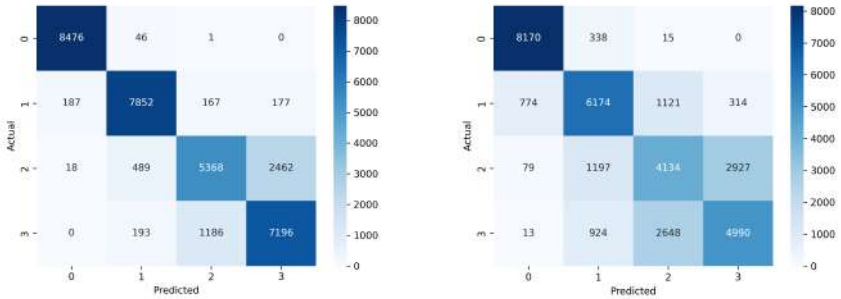


(c) MLP (d) SVM

FIGURE 5.5 – Matrices de confusion sur l'ensemble de données sous-échantillonné



(a) Random forest (b) k-NN



(c) MLP (d) SVM

FIGURE 5.6 – Matrices de confusion sur l'ensemble de données sur-échantillonné

5.10 Analyse des résultats et comparaison avec les méthodes existantes

Les expériences précédentes mentionnées dans la littérature ont été menées sur des données collectées à partir d'un nez électronique pendant des périodes spécifiques. En revanche, l'expérience menée a utilisé l'ensemble des données collectées, fournissant ainsi un jeu de données plus important pour les mêmes modèles de classification utilisés dans les expériences passées.

Indépendamment du type de données utilisé (normal, sur-échantillonné ou sous-échantillonné), l'algorithme de random forest a systématiquement surpassé k-NN et MLP, tandis que SVM a produit les résultats les moins satisfaisants. En termes de temps d'apprentissage, random forest et k-NN se sont révélés les plus rapides, suivis par MLP et SVM, qui nécessitent des temps d'apprentissage plus longs.

L'algorithme de random forest s'est distingué par ses performances supérieures par rapport aux autres méthodes d'apprentissage automatique pour plusieurs raisons. Tout d'abord, il est capable de traiter des ensembles de données volumineux et complexes, comme celui utilisé dans cette étude, comprenant 98.907 occurrences et 11 caractéristiques. Ensuite, il améliore la capacité de généralisation en limitant efficacement le sur-ajustement grâce à l'utilisation du *bagging* et à la sélection aléatoire des caractéristiques. De plus, les forêts aléatoires (RF) sont robustes face aux valeurs aberrantes et aux erreurs de mesure, qui peuvent avoir un impact significatif sur d'autres méthodes telles que k-NN ou SVM. Enfin, en combinant les prédictions de plusieurs arbres de décision au sein d'une méthode d'ensemble, elles capturent une variété de relations dans les données, ce qui renforce leurs performances prédictives. Ces caractéristiques expliquent les résultats supérieurs obtenus par l'algorithme des forêts aléatoires (RF) dans cette étude.

Les résultats ont révélé que la meilleure performance a été atteinte avec l'ensemble de données sur-échantillonné. Cela suggère qu'un ensemble de données équilibré, légèrement plus volumineux, pourrait encore améliorer les performances du modèle, bien que cela implique des temps d'apprentissage plus longs.

Conclusion

Cette expérimentation a démontré comment les données olfactives associées aux algorithmes d'apprentissage supervisé peuvent être exploitées pour classer de manière fiable les différents états de la tomate. L'algorithme des forêts aléatoires, appliqué aux données olfactives, a produit des résultats satisfaisants en comparaison avec d'autres approches supervisées.
En explorant des perspectives telles que l'analyse des séries temporelles, la calibration des capteurs, la validation des modèles pour différentes variétés de tomates et phases de croissance, ainsi que la réalisation d'évaluations comparatives basées sur des méthodologies reconnues, ce travail jette les bases d'une meilleure compréhension des techniques d'évaluation de la tomate. Il contri-

bue également au développement du domaine plus vaste de l'agriculture et des technologies alimentaires.

Dans le chapitre suivant, une approche différente sera explorée en considérant les données comme des séries temporelles afin de prédire le temps restant avant la détérioration des tomates.

CHAPITRE 6

PRÉDICTION DE LA DURÉE DE VIE DES TOMATES

Introduction

L'objectif de la sécurité alimentaire est de garantir un accès constant à une nourriture sûre. Cela implique de disposer d'une quantité suffisante d'aliments répondant aux besoins nutritionnels des individus. De plus, cette nourriture doit correspondre aux préférences alimentaires des personnes, leur permettant ainsi de mener une vie saine et active. Cela inclut également la garantie que les aliments soient produits, distribués et consommés de manière durable, sans compromettre la santé humaine, l'environnement ou les ressources naturelles. Dans cette optique, diverses techniques ont émergé, notamment l'agriculture de précision qui utilise des capteurs, des drones, des satellites, des services d'information géographique, l'informatique et l'intelligence artificielle (IA).
L'exploration des articles et études disponibles offre un aperçu des avancées significatives dans le domaine de la sécurité alimentaire et de l'agriculture de précision, en particulier celles qui utilisent des approches temporelles séquentielles.

6.1 Application des séries temporelles dans l'agriculture

Preeti et Bharti présentent dans [263], un projet visant à prévoir les rendements des cultures de blé en Inde en utilisant des images satellitaires disponibles publiquement. Le réseau profond proposé comprend trois modules. Le premier module est un CNN qui apprend à extraire des caractéristiques pertinentes des images. Le deuxième module est un réseau LSTM qui analyse la relation temporelle pendant la saison de croissance de la culture. Le troisième module est un réseau entièrement connecté qui fournit une prédiction finale du rendement des cultures. Ce modèle hybride CNN-LSTM a également été utilisé pour prédire la maladie de la bordure jaune de la tige (YSB) chez les plants de riz afin d'améliorer le rendement et la qualité de la culture [264]. Il a été développé en tenant compte des données historiques sur les ravageurs et des données météoro-

logiques de différentes régions de l'Inde. Maysa et al. [265] ont également utilisé une combinaison de CNN et LSTM pour estimer la superficie des cultures. De même, Seyed et al. [266], à la recherche d'une prédiction plus précise, ont proposé deux architectures. Le premier modèle comprend des CNN 2D, des connexions sautées et des LSTM avec attention. Le deuxième modèle comprend des CNN 3D, des connexions sautées et des ConvLSTM avec attention.

L'étude [267] illustre l'utilisation de méthodes d'apprentissage profond pour prévoir les rendements des cultures de blé, d'arachide et d'orge en Inde, chacune avec des caractéristiques de production annuelles uniques. Elle introduit une approche ayant pour objectif d'optimiser différents hyperparamètres dans les modèles LSTM. Deux variantes de LSTM, sLSTM et LSTM bidirectionnel (Bi-LSTM), ont été ajustées pour prédire les rendements des cultures. Sony et Tripty [268] ont également proposé un modèle basé sur les LSTM pour aider les agriculteurs à identifier le type de culture qui produira un bon rendement pour une saison particulière, en tenant compte du type de sol, de la fertilité du sol, des conditions climatiques, des précipitations et des conditions requises pour les graines individuelles. Dans [269], des RNN, y compris les modèles LSTM et GRU et leurs extensions, Bi-LSTM et GRU bidirectionnels (Bi-GRU), ont été mis en œuvre pour estimer les rendements des tomates en fonction des données climatiques, de la quantité d'irrigation et de la teneur en eau dans le profil du sol. Les ensembles de données agricoles sont des séries temporelles, et les prévisions agricoles reposent fortement sur des ensembles de données historiques.

La plupart des travaux se limitent à la prédiction des rendements basée sur un modèle pour une seule culture, mais Saeed et al. [270] ont mis en place YieldNet. Ce modèle prédit le rendement de plusieurs cultures (maïs et soja) et prend simultanément en compte l'interaction entre leurs rendements. Le travail de Liyun et al. [271] se distingue par la prédiction des rendements des cultures en serre en combinant deux réseaux pour le traitement de séquences temporelles : le réseau de convolution temporelle (TCN) [272] et le RNN. L'application d'une fonction d'optimisation améliorée pour une prédiction précise en utilisant LSTM [273] et une fonction de perte prenant en compte la variable de réponse multicible [270] a fait l'objet d'autres études.

Tous ces travaux se sont basés sur l'apprentissage automatique, en particulier l'apprentissage profond, avec des modèles conçus principalement pour prédire et estimer les rendements des cultures, classifier les cultures, détecter les maladies ou reconnaître les espèces sur différentes séquences temporelles. Cela inclut les réseaux RNN, LSTM et GRU. Les résultats de ces études ont montré que la combinaison de LSTM avec d'autres modèles tels que CNN et TCN donne de meilleurs résultats par rapport aux performances de chacun des modèles associés pris individuellement, ou avec des modèles d'apprentissage automatique conventionnels. Cependant, toutes ces études n'ont pas porté sur la compréhension de l'état d'une culture après la récolte dans le but d'une alimentation saine. Cela sera l'objectif principal de ce chapitre.

6.2 Pré-traitement et nettoyage des données

Les données mentionnées dans la section 5.8 seront traitées comme des séries temporelles puisqu'elles sont organisées chronologiquement de la date d'acquisition jusqu'à la date de dégradation complète. Cela a donné un total de 10 jeux de données, chacun correspondant à une variété de fruit spécifique. Par la suite, les données ont été rééchantillonnées pour refléter les mesures prises toutes les 15 minutes au lieu de la fréquence originale de toutes les 5 secondes. Le rééchantillonnage a été fait pour s'assurer que la granularité des données était suffisante pour l'analyse tout en maintenant l'efficacité des calculs. Une nouvelle caractéristique a été ajoutée à chaque ensemble de données. Il s'agit du nombre de jours restants avant que la récolte de tomates soit complètement endommagée. Il a été calculé comme suit :

$$\text{Jours restants} = \text{Date de fin de la collecte} - \text{Date de collecte} \tag{6.1}$$

Après le rééchantillonnage, il a été vérifié que les états étaient séquentiels, c'est-à-dire qu'ils suivaient l'ordre suivant : $0 \rightarrow 1 \rightarrow 2 \rightarrow 3$. De plus, il a été confirmé que les données conservaient leur ordre chronologique au sein des dates de collecte des données. Cette dernière vérification a permis de détecter des anomalies dans les jeux de données des tomates *Akikon* 1, 2 et 3.

Les différents ensembles de données individuels ont été concaténés en un seul dataframe final. Pour préserver l'identité de chaque tomate dans cet ensemble de données combiné, un identifiant a été ajouté au dataframe, permettant de distinguer les données correspondant à chaque tomate.

6.3 Long Short-Term Memory

LSTM, une variante des réseaux neuronaux récurrents (RNN), excelle dans la capture et l'apprentissage des dépendances à long terme [76]. Les réseaux LSTM sont constitués de cellules mémoire qui maintiennent un état interne et de trois types de portes : la porte d'oubli (f_t), la porte d'entrée (i_t) et la porte de sortie (o_t). Ces portes régulent le flux d'informations au sein du réseau, permettant de se souvenir ou d'oublier sélectivement les informations passées et d'intégrer de nouvelles informations. Cette architecture permet aux réseaux LSTM de capturer des dépendances à long terme en atténuant le problème du gradient évanescent et en facilitant l'apprentissage des patterns temporels sur de longues séquences. Chaque cellule LSTM effectue une série d'opérations pour mettre à jour son état interne en fonction de l'entrée actuelle et de l'état précédent, comme indiqué dans les équations suivantes :

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \tag{6.2}$$

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \tag{6.3}$$

$$\tilde{c}_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c) \tag{6.4}$$

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \tag{6.5}$$

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \tag{6.6}$$

$$h_t = o_t \times \tanh(c_t) \tag{6.7}$$

Ici :

— σ représente la fonction sigmoïde ;

— tanh désigne la fonction tangente hyperbolique ;

— W_f, W_i, W_c et W_o sont les matrices de poids pour les différentes portes et la nouvelle information ;

— b_f, b_i, b_c et b_o sont les vecteurs de biais correspondants ;

— h_{t-1} est l'état caché précédent ; et

— x_t est l'entrée actuelle.

La porte d'oubli, comme indiqué dans l'Equation 6.2, détermine les informations à éliminer de l'état précédent c_{t-1}. La porte d'entrée, définie par l'Equation 6.3, contrôle les nouvelles informations à ajouter à l'état, qui est mis à jour selon l'Equation 6.4. La porte de sortie, décrite dans l'Equation 6.6, décide quelles parties de l'état mis à jour sont exposées en tant que sortie, conformément à l'Equation 6.7. L'intégration des résultats des portes d'oubli et d'entrée est gérée par l'Equation 6.5, déterminant l'état de la cellule actuelle c_t. Grâce à ces mécanismes, les cellules LSTM peuvent efficacement apprendre et mémoriser des informations pertinentes sur de longues séquences. La Figure 6.1 illustre l'architecture d'une unité LSTM.

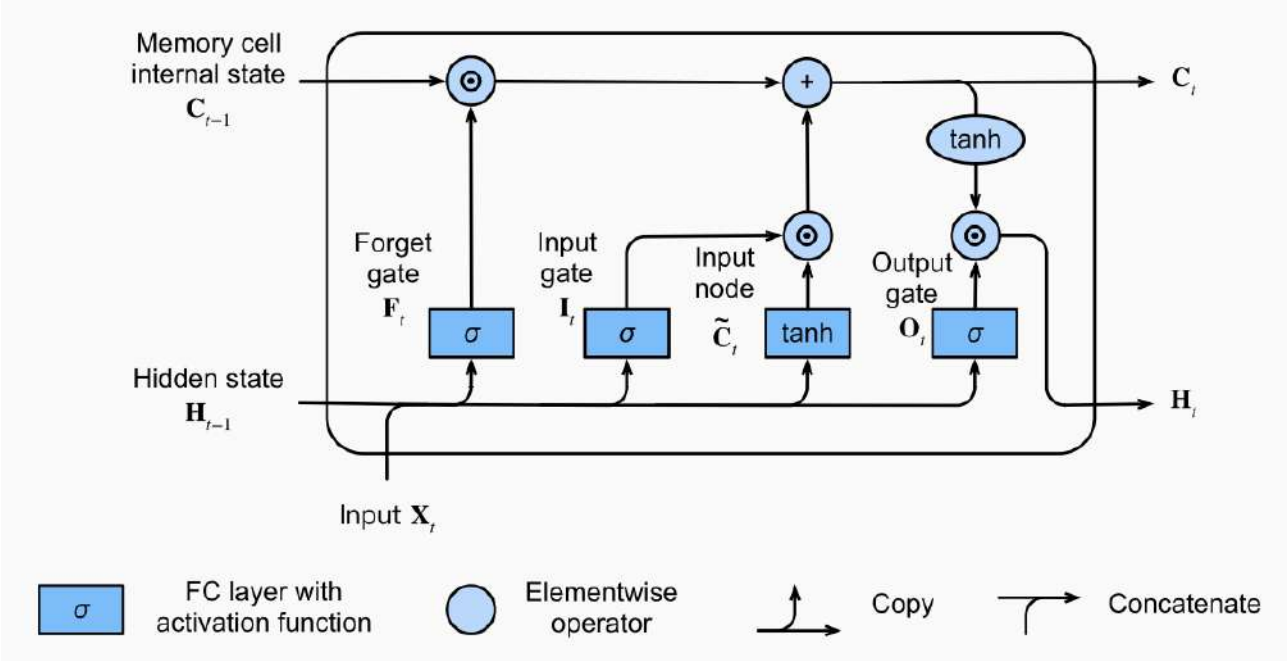


FIGURE 6.1 – Architecture d'une unité LSTM [1]

6.4 Gated Recurrent Unit

Les réseaux GRU, une autre variante des RNN, proposent une architecture simplifiée par rapport aux réseaux LSTM tout en capturant efficacement les dépendances dans les données séquentielles [274]. Les réseaux GRU se composent de deux portes principales : la porte de réinitialisation (r_t) et la porte de mise à jour (z_t). Ces portes régulent le flux d'informations au sein du réseau, permettant de réinitialiser ou de mettre à jour le contenu de la mémoire. Cette architecture permet aux réseaux GRU de capturer les dépendances temporelles en contrôlant la quantité d'informations passées à oublier et la quantité d'informations nouvelles à ajouter. Chaque cellule GRU effectue une série d'opérations pour mettre à jour son état en fonction de l'entrée actuelle et de l'état précédent, comme illustré dans les équations suivantes :

$$z_t = \sigma(W_z \times [h_{t-1}, x_t] + b_z) \tag{6.8}$$

$$r_t = \sigma(W_r \times [h_{t-1}, x_t] + b_r) \tag{6.9}$$

$$\tilde{h}_t = \tanh(W_h \times [r_t \times h_{t-1}, x_t] + b_h) \tag{6.10}$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \tag{6.11}$$

1. Source : `https://d2l.ai/chapter_recurrent-modern/lstm.html`, consultée le 24 novembre 2024.

Ici :

— σ représente la fonction sigmoïde ;

— tanh désigne la fonction tangente hyperbolique ;

— W_z, W_r et W_h sont les matrices de poids associées aux portes et à l'information nouvelle ;

— b_z, b_r et b_h sont les vecteurs de biais correspondants ;

— h_{t-1} est l'état caché précédent ; et

— x_t est l'entrée actuelle.

La porte de mise à jour, décrite dans l'Equation 6.8, détermine la portion de l'état précédent h_{t-1} qui doit être conservée dans l'état actuel. La porte de réinitialisation, expliquée dans l'Equation 6.9, contrôle dans quelle mesure l'état précédent h_{t-1} influence l'activation candidate $\tilde{h}_t$, qui est mise à jour selon l'Equation 6.10. La sortie finale, décrite dans l'Equation 6.11, est une combinaison de l'état précédent et de l'activation candidate, modulée par la porte de mise à jour. Grâce à ces mécanismes, les cellules GRU peuvent efficacement apprendre et mémoriser des informations pertinentes sur des séquences, tout en ayant une structure plus simple comparée aux cellules LSTM. La Figure 6.2 illustre l'architecture d'une unité GRU.

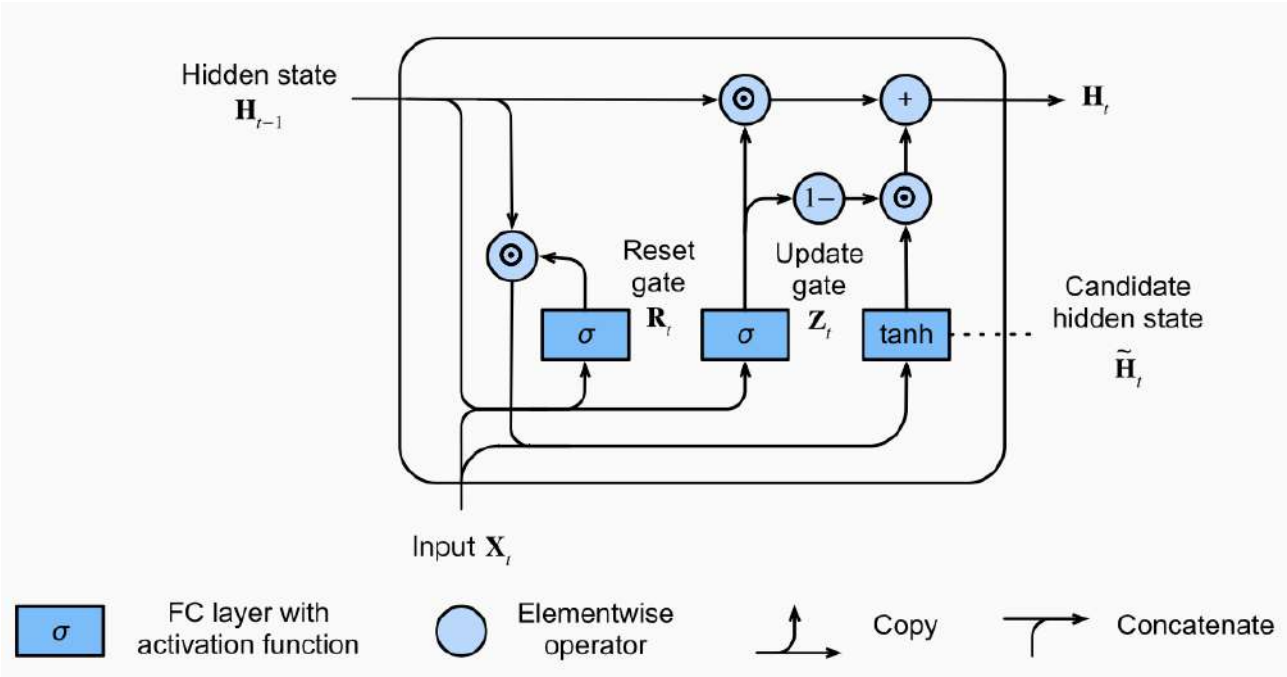


FIGURE 6.2 – Architecture d'une unité GRU [2]

6.5 Configuration du modèle

Le modèle de prédiction a été construit à l'aide du framework Keras, avec une architecture basée sur des couches récurrentes (LSTM ou GRU) pour la modélisation de séquences (Figure 6.3). De nombreuses expérimentations préliminaires ont été réalisées avant la finalisation du modèle

2. Source : `https://d2l.ai/chapter_recurrent-modern/gru.html`, consultée le 24 novembre 2024.

présenté. Les hyperparamètres du modèle ont été optimisés au cours d'une série d'itérations expérimentales, où les paramètres ont été ajustés manuellement et évalués en fonction de leur impact sur les performances. Cette approche, au lieu d'utiliser une méthode automatisée d'optimisation des hyperparamètres telle que le grid search ou l'optimisation bayésienne, a permis des ajustements ciblés basés sur le comportement initial du modèle. Le nombre d'unités récurrentes, la longueur des séquences et le taux de dropout ont été testés dans diverses combinaisons. Le nombre d'unités dans les couches LSTM ou GRU a été ajusté pour équilibrer la capacité du modèle tout en évitant le sur-apprentissage, tandis que le taux de dropout a été choisi dans l'intervalle [0.05, 0.5] afin d'assurer un apprentissage stable. La sélection finale des hyperparamètres a été basée sur des améliorations observées de la perte de validation et de la précision des prédictions. Les modèles conservent le même nombre de caractéristiques en sortie qu'en entrée, un choix soutenu par des résultats expérimentaux montrant de meilleures performances par rapport à des modèles où seule la caractéristique cible était considérée en sortie. Voici une description détaillée du modèle final :

1. **Entrée du modèle :**
   — La forme de l'entrée du modèle est spécifiée comme $(64, 12)$, indiquant que chaque séquence d'entrée a une longueur de 64 pas de temps et que chaque pas de temps comporte 12 caractéristiques.
2. **Couches LSTM ou GRU :**
   — La première couche LSTM ou GRU contient 64 unités et est configurée pour retourner des séquences. Cela signifie qu'elle prend en entrée une séquence de longueur 64 avec 12 caractéristiques à chaque pas de temps et retourne une séquence de sortie pour chaque pas de temps.
   — La deuxième couche LSTM ou GRU contient également 64 unités mais est configurée pour ne pas retourner de séquences. Cela signifie qu'elle prend en entrée la séquence de sortie de la première couche et retourne une seule sortie à la fin de la séquence.
3. **Couche de dropout :**
   — Une couche de régularisation par dropout est ajoutée après la deuxième couche LSTM ou GRU, avec un taux de dropout de 0.1. Cela aide à prévenir le sur-apprentissage en désactivant aléatoirement une fraction des neurones pendant l'entraînement.
4. **Couche de sortie :**
   — La couche de sortie est une couche dense (complètement connectée) contenant 12 unités, ce qui correspond au nombre de caractéristiques dans l'entrée. Cette couche prédit les valeurs des caractéristiques pour chaque pas de temps de la séquence de sortie.
5. **Compilation du modèle :**
   — Le modèle est compilé avec l'optimiseur Adam et la fonction de perte d'erreur quadratique moyenne (MSE). L'erreur absolue moyenne (MAE) est utilisée comme métrique pour évaluer les performances du modèle pendant l'entraînement.

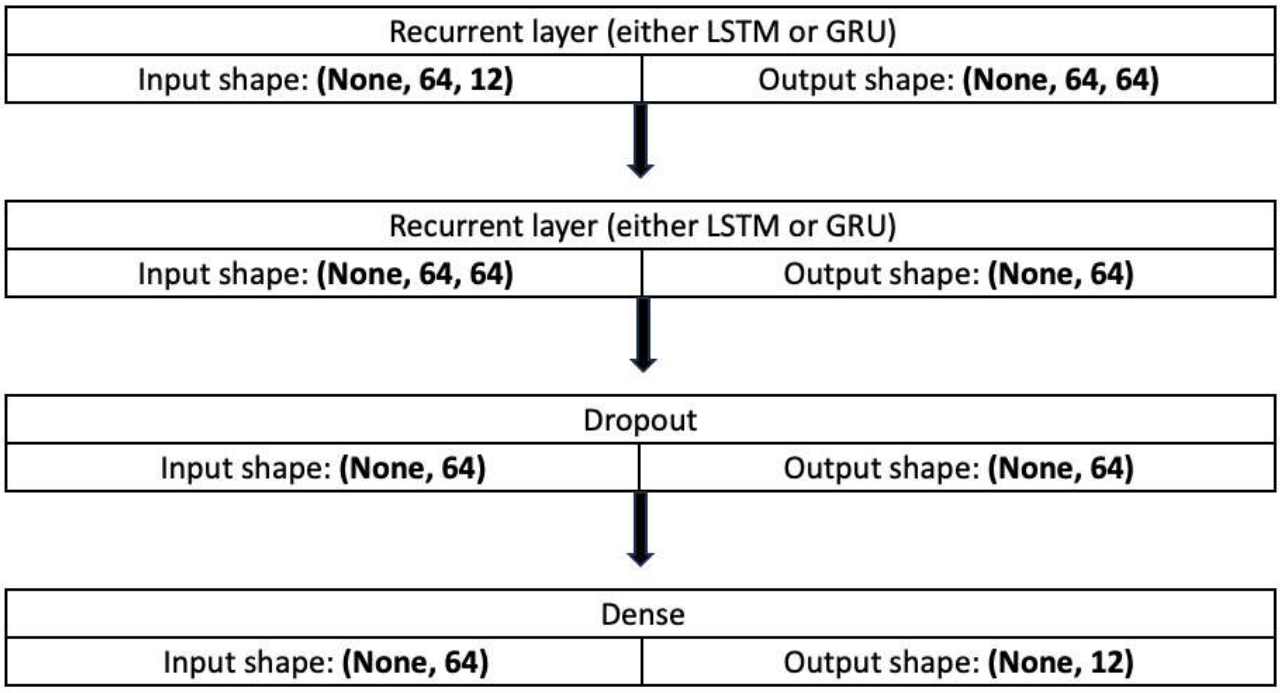


FIGURE 6.3 – Architecture du modèle proposé

6.6 Résultats expérimentaux

Pour réaliser les expérimentations, les données ont été structurées en séries temporelles à partir de données séquentielles en utilisant la classe `TimeseriesGenerator` de la bibliothèque Keras. Les données ont été divisées chronologiquement en 80% pour l'entraînement, 10% pour la validation, et 10% pour les tests, en veillant à préserver la structure temporelle des données et sans appliquer de mélange. De plus, une normalisation des données a été effectuée pour mettre les caractéristiques à l'échelle dans l'intervalle $[0;1]$, en utilisant des échelles variables pour un traitement optimal.

L'entraînement du modèle a été optimisé à l'aide d'un mécanisme de sauvegarde automatique basé sur les performances sur les données de validation. Un callback `ModelCheckpoint` a été configuré pour surveiller la perte de validation (`val_loss`) et sauvegarder uniquement le modèle ayant obtenu la meilleure performance au cours des 10 époques d'entraînement. Les données ont été fournies par des générateurs distincts pour l'entraînement et la validation, permettant une gestion efficace des lots. Cette approche garantit que seul le modèle minimisant la perte de validation est conservé pour des étapes ultérieures.

Les Figures 6.4 et 6.5 illustrent les courbes d'apprentissage, présentant l'évolution de la perte et de l'erreur absolue moyenne pour le modèle basé sur des cellules LSTM ou GRU, au cours des différentes époques d'entraînement et de validation.

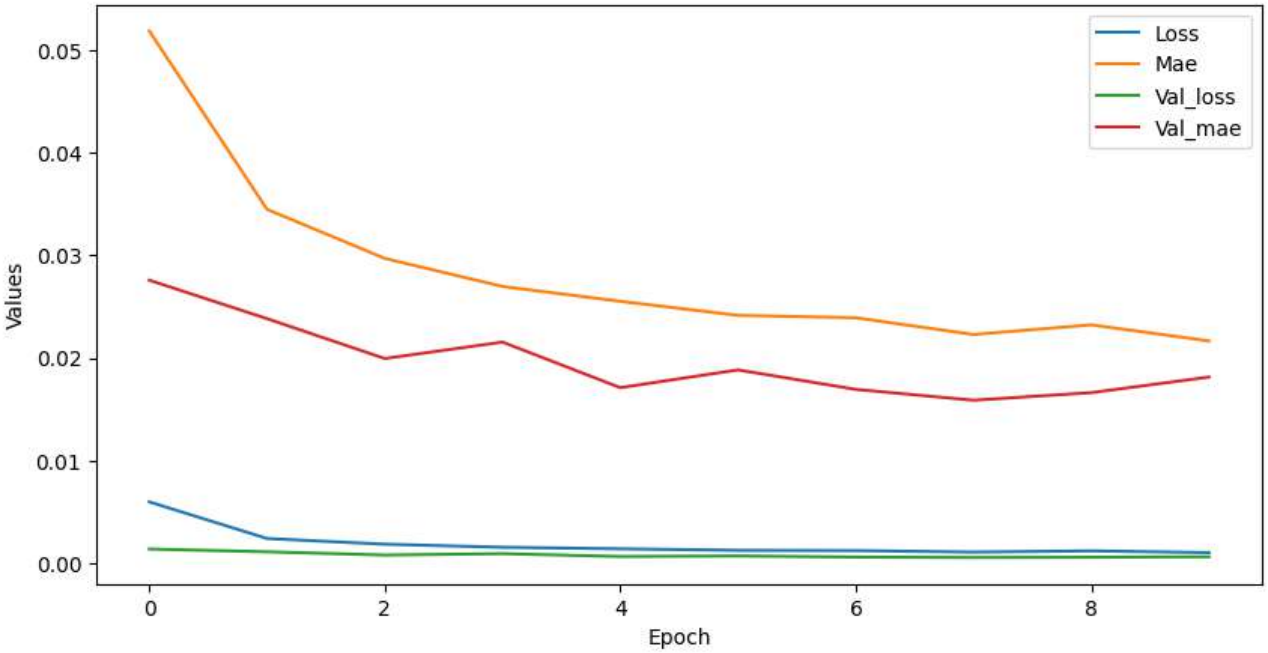


FIGURE 6.4 – Courbes d'apprentissage du modèle basé sur LSTM : MSE et MAE pour l'entraînement et la validation

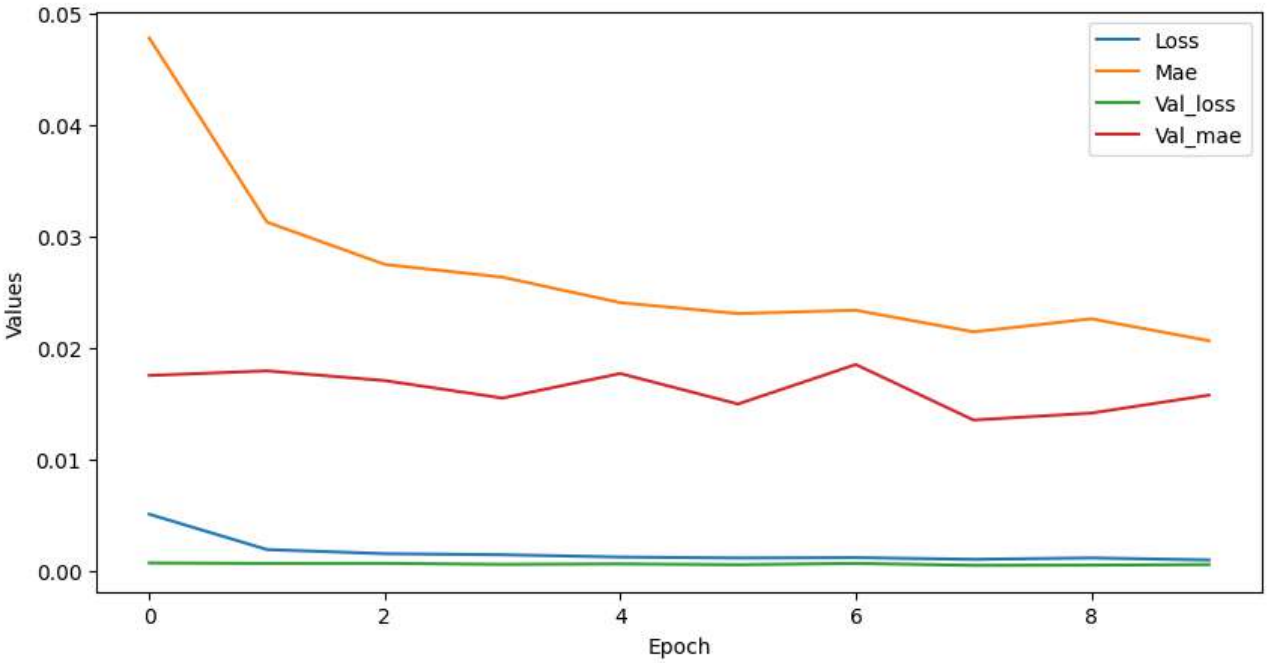


FIGURE 6.5 – Courbes d'apprentissage du modèle basé sur GRU : MSE et MAE pour l'entraînement et la validation

Le Tableau 6.1 présente les résultats de l'évaluation des performances des modèles basés sur LSTM et GRU sur les données réservées pour les tests. Les métriques utilisées pour cette évaluation incluent l'erreur quadratique moyenne (MSE), l'erreur absolue moyenne (MAE), le coefficient de détermination (R^2) global et le coefficient de détermination pour le nombre de jours restants avant

la détérioration. Chaque métrique fournit des informations sur différents aspects de la précision et de l'efficacité, comme détaillé ci-dessous. Elles sont essentielles pour évaluer les performances dans les tâches de régression.

— MSE quantifie la moyenne des carrés des différences entre les valeurs prédites et les valeurs réelles. Elle est calculée comme suit :

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6.12}$$

où n est le nombre d'échantillons, y_i est la valeur réelle et $\hat{y}_i$ est la valeur prédite.

— MAE mesure la moyenne des différences absolues entre les valeurs prédites et les valeurs réelles. Elle est calculée comme suit :

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{6.13}$$

où n est le nombre d'échantillons, y_i est la valeur réelle et $\hat{y}_i$ est la valeur prédite.

— R^2 indique la proportion de la variance dans la variable dépendante qui est prévisible à partir des variables indépendantes. Il est calculé comme suit :

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6.14}$$

où y_i est la valeur réelle, $\hat{y}_i$ est la valeur prédite, et $\bar{y}$ est la moyenne des valeurs réelles.

TABLEAU 6.1 – Évaluation des performances des modèles basés sur LSTM et GRU

	MSE	**MAE**	**R^2 Global**	**R^2 sur les jours restants**
Modèle basé sur LSTM	0.000586	0.01502	0.817	0.9870
Modèle basé sur GRU	0.000505	0.01392	0.832	0.9928

6.7 Discussion

Ces résultats démontrent le potentiel des modèles d'apprentissage profond, en particulier le modèle basé sur GRU, pour prédire la durée de conservation restante des tomates. Ainsi, une approche un peu plus précise et fiable pour la manipulation post-récolte est mise à disposition. Le modèle basé sur GRU a surpassé celui basé sur LSTM, probablement en raison de sa capacité à mieux capturer les dépendances à long terme dans les données séquentielles, ce qui est essentiel pour modéliser le processus de dégradation progressive et souvent non linéaire des tomates au fil du temps. Ce résultat s'aligne sur des recherches antérieures soulignant l'efficacité de GRU dans les tâches de prévision de séries temporelles, où moins de paramètres peuvent conduire à une meilleure généralisation.

L'un des principaux atouts présentés dans ce chapitre réside dans l'application innovante de la technologie de nez électronique pour prédire la détérioration des tomates. Contrairement à la majorité des études existantes, qui se concentrent principalement sur la prévision des rendements agricoles, la détection des maladies ou la classification des cultures à partir de données environnementales, notre approche propose une méthode unique pour suivre le processus de dégradation post-récolte. Prédire le nombre de jours avant que les tomates ne se détériorent constitue un outil précieux pour améliorer la sécurité alimentaire, garantissant aux consommateurs un accès à des produits plus frais et plus sains, impactant ainsi directement la nutrition et la santé publique.

En outre, ce chapitre expose un potentiel significatif pour améliorer à la fois l'efficacité et les normes de qualité dans l'industrie agricole. En intégrant la technologie de nez électronique à la gestion post-récolte, les producteurs peuvent renforcer leurs processus de contrôle qualité. Cette innovation permet une prise de décision un peu plus précise en matière de logistique et de gestion des stocks, promouvant des standards plus élevés d'efficacité et de qualité dans une industrie en rapide évolution.

Conclusion

En résumé, ce chapitre a mis en évidence l'efficacité des modèles basés sur LSTM et GRU pour prédire le nombre de jours restants avant la détérioration des tomates après récolte. Le modèle implémenté se compose de quatre couches : deux couches récurrentes empilées (soit LSTM, soit GRU), une couche de régularisation par dropout et une couche entièrement connectée. Le nombre de caractéristiques en sortie reste identique au nombre de caractéristiques en entrée, en raison des résultats prédictifs obtenus. Ce choix contraste avec les cas où seul le nombre de jours restants était considéré comme la sortie du modèle.
Pour obtenir des résultats encore meilleurs, l'ajout d'images de tomates au jeu de données est une piste prometteuse. Dans ce cas, des recherches futures pourraient envisager l'intégration des CNN afin d'améliorer le modèle proposé. Ces avancées ouvrent des perspectives intéressantes pour optimiser la gestion agricole et promouvoir des pratiques durables dans ce domaine.

CONCLUSION GÉNÉRALE ET PERSPECTIVES

Ce travail de recherche a permis de développer un nez électronique dédié à la prédiction de la qualité des produits locaux, avec un accent particulier sur la tomate. Les objectifs spécifiques de la recherche, allant de la conception du dispositif électronique à la mise en place de modèles d'apprentissage automatique et au test de l'efficacité du nez électronique sur des variétés locales de tomates, ont été atteints. Les hypothèses formulées ont été validées, démontrant que le nez électronique peut détecter des variations subtiles dans les composés volatils émis par les tomates, différencier les différentes étapes de détérioration, et fournir des prédictions un peu plus précises sur la qualité et la date de détérioration des tomates.

Les résultats montrent que les modèles d'apprentissage automatique, tels que les SVM et les forêts aléatoires, ainsi que les réseaux de neurones récurrents (LSTM et GRU), sont efficaces pour analyser les données volatiles collectées. Ces avancées permettent une amélioration significative de la précision et de la fiabilité de la surveillance de la qualité des tomates locales par rapport aux méthodes traditionnelles de contrôle de qualité.

Les perspectives de ce travail de recherche sont nombreuses et prometteuses. Voici quelques axes de développement futurs. Premièrement, l'élargissement du champ d'application, en étendant l'utilisation du nez électronique à d'autres produits alimentaires locaux et non locaux, permettrait de tester sa polyvalence et son efficacité dans divers contextes agroalimentaires. Ensuite, l'optimisation des modèles d'apprentissage pourrait être poursuivie en continuant à affiner les modèles d'apprentissage automatique et en explorant d'autres techniques de machine learning afin d'améliorer encore la précision des prédictions.

En outre, l'intégration de facteurs environnementaux tels que la qualité de l'air, l'exposition aux polluants, et l'exposition à la lumière dans les modèles permettrait de mieux comprendre leur impact sur la détérioration des produits et d'affiner les prédictions. Le développement d'une interface utilisateur conviviale est également essentiel pour faciliter l'utilisation du nez électronique par les producteurs et les contrôleurs de qualité, en proposant des visualisations claires et des alertes précises sur l'état des produits.

La mise en œuvre industrielle du nez électronique pourrait être réalisée en collaboration avec

des entreprises agroalimentaires, en intégrant cette technologie dans les chaînes de production et de distribution afin de tester et valider son efficacité à grande échelle. Enfin, des études économiques et sociales sur l'impact de l'adoption du nez électronique, notamment pour les producteurs locaux, sont nécessaires pour promouvoir une adoption plus large et encourager des politiques de soutien. En somme, ces perspectives ouvrent la voie à de futures applications des technologies de détection et de l'intelligence artificielle dans le domaine agroalimentaire, maximisant ainsi les bénéfices pour les consommateurs et les producteurs.

LISTE DES PUBLICATIONS

Conférences

— **Affognon, L.**, Diallo, A., Diallo, C., & Ezin, E. C. *Supervised Learning Models for Tomato Quality Prediction using Electronic Nose Data*, 2023 International Conference on Electrical, Computer and Energy Technologies (ICECET), Cape Town, South Africa, 2023, pp. 1-6. https://doi.org/10.1109/ICECET58911.2023.10389485

— **Affognon, L.**, Diallo, A., Diallo, C., Ezin, E.C. (2023). Electronic Nose Architecture for Tomato Data Collection. In : Arai, K. (eds) Proceedings of the Future Technologies Conference (FTC) 2023, Volume 3. FTC 2023. Lecture Notes in Networks and Systems, vol 815. Springer, Cham. https://doi.org/10.1007/978-3-031-47457-6_9

— **Affognon, L.**, Diallo, A., Diallo, C., & Ezin, E. C. (2023). *Design of an Experimental Electronic Nose for Data Collection for Food Quality*. IEEE EUROCON 2023 - 20th International Conference on Smart Technologies, Torino, Italy, 2023, pp. 406-411. https://doi.org/10.1109/EUROCON56442.2023.10199067

Journal

— **Affognon, L.**, Ezin, E. C., & Diallo, C. *Forecasting Post-Harvest Tomato Shelf Life with Recurrent Neural Networks and Electronic Nose Data*. Soumis pour examen au journal **SN Computer Science**.

— **Affognon, L.**, Diallo, A., Diallo, C. & Ezin, E. C. *A Survey on Statistical and Machine Learning Algorithms Used in Electronic Noses for Food Quality Assessment*. SN COMPUT. SCI. **4**, 590 (2023). https://doi.org/10.1007/s42979-023-02052-0

Publications dans le cadre du projet IoT-SDC

— Diallo, A., **Affognon, L.**, Diallo, C., & Ezin, E. C. (2023). *A Three-Level Deep Learning Intrusion Detection System for IoT Network*, 2023 International Conference on Electrical, Communication and Computer Engineering (ICECCE), Dubai, United Arab Emirates, 2023, pp. 1-6.

https://doi.org/10.1109/ICECCE61019.2023.10442798

— Diallo, A., **Affognon, L.**, Diallo, C., Ezin, E.C. (2025). Towards the Implementation of a Dynamic IDS for IoT : Anomaly Detection in MQTT Traffic. In : Cheikh M. F. Kebe, K., Gueye, A., Ndiaye, A., Sene, N.A., Maiga, AS. (eds) Innovations and Interdisciplinary Solutions for Underserved Areas. InterSol 2024. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 610. Springer, Cham. https://doi.org/10.1007/978-3-031-86493-3_15

— Diallo, A., **Affognon, L.**, Diallo, C., & Ezin, E. C. (2022, October). *Deep Learning Based Binary and Multi-class Classification Comparison for Anomaly Detection.* In 2022 International Conference on Engineering and Emerging Technologies (ICEET) (pp. 1-6). IEEE. https://doi.org/10.1109/ICEET56468.2022.10007171

BIBLIOGRAPHIE

[1] FOOD et Agriculture Organization (FAO). *The State of Food Security and Nutrition in the World 2021*. Food et Agriculture Organization of the United Nations, 2021.

[2] World Health Organization (WHO). *Food Safety*. https://www.who.int/foodsafety/en/. Accessed on December 07, 2024. 2020.

[3] Alphus D. WILSON et Manuela BAIETTO. « Applications and Advances in Electronic-Nose Technologies ». In : *Sensors* 9.7 (2009), p. 5099-5148. ISSN : 1424-8220. URL : https://doi.org/10.3390/s90705099.

[4] Miguel PERIS et Laura ESCUDER-GILABERT. « Electronic noses and tongues to assess food authenticity and adulteration ». In : *Trends in Food Science & Technology* 58 (2016), p. 40-54. ISSN : 0924-2244. URL : https://doi.org/10.1016/j.tifs.2016.10.014.

[5] Hassan ANWAR, Talha ANWAR et Shamas MURTAZA. « Review on food quality assessment using machine learning and electronic nose system ». In : *Biosensors and Bioelectronics : X* 14 (2023), p. 100365. ISSN : 2590-1370. URL : https://doi.org/10.1016/j.biosx.2023.100365.

[6] Luis GUERRERO et al. « Consumer-driven definition of traditional food products and innovation in traditional foods. A qualitative cross-cultural study ». In : *Appetite* 52.2 (2009), p. 345-354. ISSN : 0195-6663. URL : https://doi.org/10.1016/j.appet.2008.11.008.

[7] A. TRICHOPOULOU et al. « Traditional foods : Why and how to sustain them ». In : *Trends in Food Science & Technology* 17.9 (2006). INTRADFOODS, p. 498-504. ISSN : 0924-2244. URL : https://doi.org/10.1016/j.tifs.2006.03.005.

[8] Mingyang WANG et Yinsheng CHEN. « Electronic nose and its application in the food industry : a review ». In : *European Food Research and Technology* 250.1 (2024), p. 21-67. URL : https://doi.org/10.1007/s00217-023-04381-z.

[9] Harry T LAWLESS et Hildegarde HEYMANN. *Sensory evaluation of food : principles and practices*. Springer Science & Business Media, 2010.

[10] Mian Kamran SHARIF et al. « Sensory evaluation and consumer acceptability ». In : *Handbook of food science and technology* 10 (2017), p. 362-386.

[11] Hubert NAGLE Jr et Susan SCHIFFMAN. « Electronic Taste and Smell : The Case for Performance Standards ». In : *Proceedings of the IEEE* 106 (sept. 2018), p. 1471-1478. URL : https://doi.org/10.1109/JPROC.2018.2859678.

[12] Maimunah MOHD ALI et al. « Principles and recent advances in electronic nose for quality inspection of agricultural and food products ». In : *Trends in Food Science & Technology* 99 (2020), p. 1-10. ISSN : 0924-2244. URL : https://doi.org/10.1016/j.tifs.2020.02.028.

[13] FAO. *Food Wastage Footprint : Impacts on Natural Resources*. Food et Agriculture Organization of the United Nations, 2013. URL : http://www.fao.org/3/i3347e/i3347e.pdf.

[14] Cheng-Jin DU et Da-Wen SUN. « Recent developments in the applications of image processing techniques for food quality evaluation ». In : *Trends in Food Science & Technology* 15.5 (2004), p. 230-249. ISSN : 0924-2244. URL : https://doi.org/10.1016/j.tifs.2003.10.006.

[15] Susannah CRAIG. *This Nose Knows*. Wired, Published December 1, 1999. 1999. URL : https://www.wired.com/1999/12/this-nose-knows/ (visité le 07/04/2025).

[16] Nancy GOHRING. *E-Nose Sniffs Out Nasty Bugs*. Wired, Published October 2, 2001. 2001. URL : https://www.wired.com/2001/10/e-nose-sniffs-out-nasty-bugs/ (visité le 07/04/2025).

[17] Kim PETERSON. *Gee, Your Beer Smells Terrific*. Wired, Published January 28, 2003. 2003. URL : https://www.wired.com/2003/01/gee-your-beer-smells-terrific/ (visité le 07/04/2025).

[18] David REED. *Electronic Noses Sniff Success*. IEEE Spectrum, Published July 27, 2022. 2022. URL : https://spectrum.ieee.org/electronic-noses-sniff-success (visité le 07/04/2025).

[19] Ian GOODFELLOW, Yoshua BENGIO et Aaron COURVILLE. *Deep Learning*. MIT Press, 2016.

[20] Kevin P. MURPHY. *Machine Learning : A Probabilistic Perspective*. MIT Press, 2012.

[21] Christopher M. BISHOP. *Pattern Recognition and Machine Learning*. Springer, 2006.

[22] Trevor HASTIE, Robert TIBSHIRANI et Jerome FRIEDMAN. *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. 2nd. Springer, 2009. URL : https://doi.org/10.1007/978-0-387-84858-7.

[23] Pedro DOMINGOS. « A Few Useful Things to Know About Machine Learning ». In : *Communications of the ACM* 55.10 (2012), p. 78-87. URL : https://doi.org/10.1145/2347736.2347755.

[24] Yann LECUN, Yoshua BENGIO et Geoffrey HINTON. « Deep learning ». In : *nature* 521.7553 (2015), p. 436-444.

[25] Trevor HASTIE, Robert TIBSHIRANI et Jerome FRIEDMAN. *The elements of statistical learning : data mining, inference, and prediction*. 2017. URL : https://doi.org/10.1007/978-0-387-84858-7.

[26] Y. LECUN et al. « Gradient-based learning applied to document recognition ». In : *Proceedings of the IEEE* 86.11 (1998), p. 2278-2324. URL : https://doi.org/10.1109/5.726791.

[27] Aurélien VANNIEUWENHUYZE. « Intelligence artificielle vulgarisée ». In : *Le Machine Learning et le Deep Learning par la pratique* (2019).

[28] Paras LAKHANI et Baskaran SUNDARAM. « Deep Learning at Chest Radiography : Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks ». In : *Radiology* 284.2 (2017). PMID : 28436741, p. 574-582. DOI : 10.1148/radiol.2017162326. URL : https://doi.org/10.1148/radiol.2017162326.

[29] Bjoern KROLLNER, Bruce VANSTONE et Gavin FINNIE. « Financial Time Series Forecasting with Machine Learning Techniques : A Survey ». In : jan. 2010.

[30] Geoffrey E HINTON et Ruslan R SALAKHUTDINOV. « Reducing the dimensionality of data with neural networks ». In : *science* 313.5786 (2006), p. 504-507.

[31] A. K. JAIN, M. N. MURTY et P. J. FLYNN. « Data clustering : a review ». In : *ACM Comput. Surv.* 31.3 (sept. 1999), p. 264-323. ISSN : 0360-0300. URL : https://doi.org/10.1145/331499.331504.

[32] Martin ESTER et al. « Density-Based Clustering in Spatial Databases : The Algorithm GDBSCAN and Its Applications ». In : *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*. AAAI Press, 1996, p. 226-231. URL : https://doi.org/10.1023/A:1009745219419.

[33] Lior ROKACH et Oded MAIMON. « Clustering Methods ». In : *Data Mining and Knowledge Discovery Handbook*. Sous la dir. d'Oded MAIMON et Lior ROKACH. Boston, MA : Springer US, 2005, p. 321-352. ISBN : 978-0-387-25465-4. URL : https://doi.org/10.1007/0-387-25465-X_15.

[34] Ian T JOLLIFFE. *Principal component analysis for special types of data*. Springer, 2002. URL : https://doi.org/10.1007/b98835.

[35] Laurens VAN DER MAATEN et Geoffrey HINTON. « Visualizing data using t-SNE. » In : *Journal of machine learning research* 9.11 (2008).

[36] Badrul SARWAR et al. « Item-based collaborative filtering recommendation algorithms ». In : *Proceedings of the 10th international conference on World Wide Web*. 2001, p. 285-295.

[37] Michel WEDEL et Wagner A. KAMAKURA. *Market Segmentation : Conceptual and Methodological Foundations*. Springer, 2012. URL : https://doi.org/10.1007/978-1-4615-4651-1.

[38] Anton LEUSKI. « Evaluating document clustering for interactive information retrieval ». In : *Proceedings of the tenth international conference on Information and knowledge management*. 2001, p. 33-40.

[39] Richard S SUTTON et Andrew G BARTO. *Reinforcement learning : An introduction*. MIT press, 2018.

[40] Martin L PUTERMAN. *Markov decision processes : discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[41] Ronald J WILLIAMS. « Simple statistical gradient-following algorithms for connectionist reinforcement learning ». In : *Machine learning* 8 (1992), p. 229-256.

[42] Vijay KONDA et John TSITSIKLIS. « Actor-critic algorithms ». In : *Advances in neural information processing systems* 12 (1999).

[43] David SILVER et al. « Mastering the game of Go with deep neural networks and tree search ». In : *nature* 529.7587 (2016), p. 484-489.

[44] Volodymyr MNIH et al. « Human-level control through deep reinforcement learning ». In : *nature* 518.7540 (2015), p. 529-533.

[45] Marcus HUTTER. *Universal artificial intelligence : Sequential decisions based on algorithmic probability*. Springer Science & Business Media, 2005.

[46] Junjie PENG et al. « Machine Learning Techniques for Personalised Medicine Approaches in Immune-Mediated Chronic Inflammatory Diseases : Applications and Challenges ». In : *Frontiers in Pharmacology* 12 (sept. 2021). DOI : https://doi.org/10.3389/fphar.2021.720694.

[47] Zhu XIAOJIN. « Semi-supervised learning literature survey ». In : *Computer Sciences TR* 1530 (2008), p. 60-60.

[48] Augustus ODENA. *Semi-Supervised Learning with Generative Adversarial Networks*. 2016. arXiv : 1606.01583 [stat.ML]. URL : https://doi.org/10.48550/arXiv.1606.01583.

[49] Kihyuk SOHN et al. *FixMatch : Simplifying Semi-Supervised Learning with Consistency and Confidence*. 2020. arXiv : 2001.07685 [cs.LG]. URL : https://doi.org/10.48550/arXiv.2001.07685.

[50] Antti TARVAINEN et Harri VALPOLA. *Mean teachers are better role models : Weight-averaged consistency targets improve semi-supervised deep learning results*. 2018. arXiv : 1703.01780 [cs.NE]. URL : https://doi.org/10.48550/arXiv.1703.01780.

[51] Veronika CHEPLYGINA, Marleen DE BRUIJNE et Josien P.W. PLUIM. « Not-so-supervised : A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis ». In : *Medical Image Analysis* 54 (2019), p. 280-296. ISSN : 1361-8415. URL : https://doi.org/10.1016/j.media.2019.03.009.

[52] Ting CHEN et al. « A Simple Framework for Contrastive Learning of Visual Representations ». In : *Proceedings of the 37th International Conference on Machine Learning*. Sous la dir. d'Hal Daumé III et Aarti SINGH. T. 119. Proceedings of Machine Learning Research. PMLR, 13–18 Jul 2020, p. 1597-1607. URL : https://proceedings.mlr.press/v119/chen20j.html.

[53] Jean-Bastien GRILL et al. « Bootstrap Your Own Latent - A New Approach to Self-Supervised Learning ». In : *Advances in Neural Information Processing Systems*. Sous la dir. de H. LAROCHELLE et al. T. 33. Curran Associates, Inc., 2020, p. 21271-21284. URL : https://proceedings.neurips.cc/paper_files/paper/2020/file/f3ada80d5c4ee70142b17b8192b2958e-Paper.pdf.

[54] Ting CHEN et al. *A Simple Framework for Contrastive Learning of Visual Representations*. 2020. arXiv : 2002.05709 [cs.LG]. URL : https://doi.org/10.48550/arXiv.2002.05709.

[55] Jean-Bastien GRILL et al. *Bootstrap your own latent : A new approach to self-supervised Learning*. 2020. arXiv : 2006.07733 [cs.LG]. URL : https://doi.org/10.48550/arXiv.2006.07733.

[56] Kaiming HE et al. *Momentum Contrast for Unsupervised Visual Representation Learning*. 2020. arXiv : 1911.05722 [cs.CV]. URL : https://doi.org/10.48550/arXiv.1911.05722.

[57] Brendan MCMAHAN et al. « Communication-Efficient Learning of Deep Networks from Decentralized Data ». In : *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*. Sous la dir. d'Aarti SINGH et Jerry ZHU. T. 54. Proceedings of Machine Learning Research. PMLR, 20–22 Apr 2017, p. 1273-1282. URL : https://proceedings.mlr.press/v54/mcmahan17a.html.

[58] H. Brendan MCMAHAN et al. *Communication-Efficient Learning of Deep Networks from Decentralized Data*. 2023. arXiv : 1602.05629 [cs.LG]. URL : https://doi.org/10.48550/arXiv.1602.05629.

[59] Tian LI et al. *Federated Optimization in Heterogeneous Networks*. 2020. arXiv : 1812.06127 [cs.LG]. URL : https://doi.org/10.48550/arXiv.1812.06127.

[60] Praneeth VEPAKOMMA et al. *Split learning for health : Distributed deep learning without sharing raw patient data*. 2018. arXiv : 1812.00564 [cs.LG]. URL : https://doi.org/10.48550/arXiv.1812.00564.

[61] Qiang YANG et al. « Federated Machine Learning : Concept and Applications ». In : *ACM Trans. Intell. Syst. Technol.* 10.2 (jan. 2019). ISSN : 2157-6904. URL : https://doi.org/10.1145/3298981.

[62] Maryam M NAJAFABADI et al. « Deep learning applications and challenges in big data analytics ». In : *Journal of big data* 2 (2015), p. 1-21.

[63] Reem Ibrahim HASAN, Suhaila Mohd YUSUF et Laith ALZUBAIDI. « Review of the State of the Art of Deep Learning for Plant Diseases : A Broad Analysis and Discussion ». In : *Plants* 9.10 (2020). ISSN : 2223-7747. URL : https://doi.org/10.3390/plants9101302.

[64] Connor SHORTEN, Taghi M KHOSHGOFTAAR et Borko FURHT. « Deep Learning applications for COVID-19 ». In : *Journal of big Data* 8.1 (2021), p. 1-54. URL : https://doi.org/10.1186/s40537-020-00392-9.

[65] Alex KRIZHEVSKY, Ilya SUTSKEVER et Geoffrey E. HINTON. « ImageNet classification with deep convolutional neural networks ». In : *Commun. ACM* 60.6 (mai 2017), p. 84-90. ISSN : 0001-0782. URL : https://doi.org/10.1145/3065386.

[66] Sutanu BHOWMICK, Satish NAGARAJAIAH et Ashok VEERARAGHAVAN. « Vision and Deep Learning-Based Algorithms to Detect and Quantify Cracks on Concrete Surfaces from UAV Videos ». In : *Sensors* 20.21 (2020). ISSN : 1424-8220. URL : https://doi.org/10.3390/s20216299.

[67] Garrett B GOH, Nathan O HODAS et Abhinav VISHNU. « Deep learning for computational chemistry ». In : *Journal of computational chemistry* 38.16 (2017), p. 1291-1307. URL : https://doi.org/10.1002/jcc.24764.

[68] Yu LI et al. « Accelerating flash calculation through deep learning methods ». In : *Journal of Computational Physics* 394 (2019), p. 153-165. ISSN : 0021-9991. URL : https://doi.org/10.1016/j.jcp.2019.05.028.

[69] Jigang TANG, Songbin LI et Peng LIU. « A review of lane detection methods based on deep learning ». In : *Pattern Recognition* 111 (2021), p. 107623. ISSN : 0031-3203. URL : https://doi.org/10.1016/j.patcog.2020.107623.

[70] Zhong-Qiu ZHAO et al. « Object Detection With Deep Learning : A Review ». In : *IEEE Transactions on Neural Networks and Learning Systems* 30.11 (2019), p. 3212-3232. URL : https://doi.org/10.1109/TNNLS.2018.2876865.

[71] Kaiming HE et al. « Deep Residual Learning for Image Recognition ». In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Juin 2016.

[72] Laith ALZUBAIDI et al. « Review of deep learning : concepts, CNN architectures, challenges, applications, future directions ». In : *Journal of big Data* 8 (2021), p. 1-74. URL : https://doi.org/10.1186/s40537-021-00444-8.

[73] Hansika HEWAMALAGE, Christoph BERGMEIR et Kasun BANDARA. « Recurrent Neural Networks for Time Series Forecasting : Current status and future directions ». In : *International Journal of Forecasting* 37.1 (2021), p. 388-427. ISSN : 0169-2070. URL : https://doi.org/10.1016/j.ijforecast.2020.06.008.

[74] Yihan JIANG et al. « LEARN Codes : Inventing Low-Latency Codes via Recurrent Neural Networks ». In : *IEEE Journal on Selected Areas in Information Theory* 1.1 (2020), p. 207-216. URL : https://doi.org/10.1109/JSAIT.2020.2988577.

[75] Rohit Abraham JOHN et al. « Optogenetics inspired transition metal dichalcogenide neuristors for in-memory deep recurrent neural networks ». In : *Nature communications* 11.1 (2020), p. 3211. URL : https://doi.org/10.1038/s41467-020-16985-0.

[76] Sepp HOCHREITER et Jürgen SCHMIDHUBER. « Long Short-Term Memory ». In : *Neural Computation* 9.8 (nov. 1997), p. 1735-1780. ISSN : 0899-7667. URL : https://doi.org/10.1162/neco.1997.9.8.1735.

[77] Kyunghyun CHO et al. *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation*. 2014. arXiv : 1406.1078 [cs.CL]. URL : https://doi.org/10.48550/arXiv.1406.1078.

[78] A VASWANI. « Attention is all you need ». In : *Advances in Neural Information Processing Systems* (2017).

[79] Sebastian RUDER. « An overview of gradient descent optimization algorithms ». In : *arXiv preprint arXiv :1609.04747* (2016).

[80] Geoffrey HINTON et al. « Deep Neural Networks for Acoustic Modeling in Speech Recognition : The Shared Views of Four Research Groups ». In : *IEEE Signal Processing Magazine* 29.6 (2012), p. 82-97. URL : https://doi.org/10.1109/MSP.2012.2205597.

[81] Razvan PASCANU, Tomas MIKOLOV et Yoshua BENGIO. « On the difficulty of training recurrent neural networks ». In : *Proceedings of the 30th International Conference on Machine Learning*. Sous la dir. de Sanjoy DASGUPTA et David MCALLESTER. T. 28. Proceedings of Machine Learning Research 3. Atlanta, Georgia, USA : PMLR, 17–19 Jun 2013, p. 1310-1318. URL : https://proceedings.mlr.press/v28/pascanu13.html.

[82] Christian SZEGEDY et al. *Going Deeper with Convolutions*. 2014. arXiv : 1409.4842 [cs.CV]. URL : https://doi.org/10.48550/arXiv.1409.4842.

[83] Karen SIMONYAN et Andrew ZISSERMAN. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. arXiv : 1409.1556 [cs.CV]. URL : https://doi.org/10.48550/arXiv.1409.1556.

[84] Andre ESTEVA et al. « Dermatologist-level classification of skin cancer with deep neural networks ». In : *nature* 542.7639 (2017), p. 115-118.

[85] Ross GIRSHICK et al. « Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation ». In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Juin 2014.

[86] Joseph REDMON et al. « You Only Look Once : Unified, Real-Time Object Detection ». In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Juin 2016.

[87] Dario AMODEI et al. « Deep Speech 2 : End-to-End Speech Recognition in English and Mandarin ». In : *Proceedings of The 33rd International Conference on Machine Learning*. Sous la dir. de Maria Florina BALCAN et Kilian Q. WEINBERGER. T. 48. Proceedings of Machine Learning Research. New York, New York, USA : PMLR, 20–22 Jun 2016, p. 173-182. URL : https://proceedings.mlr.press/v48/amodei16.html.

[88] Alexey DOSOVITSKIY et al. *An Image is Worth 16x16 Words : Transformers for Image Recognition at Scale*. 2021. arXiv : 2010.11929 [cs.CV]. URL : https://doi.org/10.48550/arXiv.2010.11929.

[89] CLUBIC. *Des joueurs professionnels de Dota 2 battus par l'IA OpenAI*. Consulté le 4 décembre 2024. 2019. URL : https://www.clubic.com/technologies-d-avenir/intelligence-artificielle/actualite-854489-joueurs-professionnels-dota-2-battus-plate-couture-ia-openai.html.

[90] CLEVY. *Le machine learning décrypté : 3 types d'apprentissage et limites*. Consulté le 4 décembre 2024. 2019. URL : https://blog.clevy.io/fr/le-machine-learning-decrypte-3-3-types-dapprentissage-et-limites/.

[91] TANGUY. *Le contrôle qualité, définition et intérêts*. Accessed on June 09, 2024. Avr. 2021. URL : https://www.plastiform.info/controle-qualite/controle-qualite-definition-interets/.

[92] *Management, qualité, ISO 9000, la page de Christian Dumortier, http ://www.cdumortier.fr/*. Accessed on June 09, 2024. URL : http://cdumortier.fr/.

[93] Jean BRILMAN. « les meilleurs Pratiques du Management, les éditions d'Organisation ». In : *Paris, France* (2001), p. 412.

[94] Andrew F SMITH. *The Oxford companion to American food and drink*. Oxford University Press, 2007.

[95] J BURNETT. « Plenty and want. A social history of food in England from 1815 to the present day. » In : (1989).

[96] James Harvey YOUNG. *Pure food : securing the Federal Food and Drugs Act of 1906*. T. 1004. Princeton University Press, 2014.

[97] Bee WILSON. *The Dark History of Food Fraud, from Poisoned Candy to Counterfeit Coffee*. Princeton : Princeton University Press, 2009. ISBN : 9780691214085. DOI : doi : 10.1515/9780691214085. URL : https://doi.org/10.1515/9780691214085.

[98] « FAO/WHO Codex Alimentarius Commission ». In : *Food and Cosmetics Toxicology* 4 (1966), p. 601. ISSN : 0015-6264. DOI : https://doi.org/10.1016/S0015-6264(66)80695-8. URL : https://www.sciencedirect.com/science/article/pii/S0015626466806958.

[99] Sara MORTIMORE et Carol WALLACE. *HACCP : A practical approach*. Springer Science & Business Media, 2013.

[100] Singam Suranjoy SINGH et K. V. RAGAVAN. « Food Safety and Quality Testing : Recent Areas of Focus and Research Perspectives ». In : *Engineering Aspects of Food Quality and Safety*. Sous la dir. de H. Umesh HEBBAR et al. Cham : Springer International Publishing, 2023, p. 307-337. ISBN : 978-3-031-30683-9. URL : https://doi.org/10.1007/978-3-031-30683-9_11.

[101] Savoy CONTROL. *Contrôle visuel, méthode de contrôle non destructif (CND)*. Consulté le 11 février 2025. 2025. URL : https://www.savoycontrol.fr/savoir-faire/controle-visuel/.

[102] COFREND. *Contrôle Visuel (VT)*. Consulté le 11 février 2025. 2025. URL : https://www.cofrend.com/jcms/uat_128239/fr/controle-visuel-vt.

[103] PLASTIFORM. *Contrôle non destructif (CND)*. Consulté le 11 février 2025. 2023. URL : https://www.plastiform.info/controle-dimension/controle-non-destructif-cnd/.

[104] COFREND. *Ressuage*. Consulté le 11 février 2025. n.d. URL : https://www.cofrend.com/jcms/uat_128227/fr/ressuage-pt.

[105] COFREND. *Magnétoscopie (MT)*. Consulté le 11 février 2025. n.d. URL : https://www.cofrend.com/jcms/mdc_110679/fr/magnetoscopie-mt.

[106] COFREND. *Ultrasons (UT)*. Consulté le 11 février 2025. 2025. URL : https://www.cofrend.com/jcms/uat_128230/fr/ultrasons-ut.

[107] COFREND. *Radiographie (RT)*. Consulté le 11 février 2025. 2025. URL : https://www.cofrend.com/jcms/mdc_110681/fr/radiographie-rt.

[108] COFREND. *Courants de Foucault (ET)*. Consulté le 11 février 2025. 2025. URL : https://www.cofrend.com/jcms/mdc_110676/fr/courants-de-foucault-et.

[109] MFE ENTERPRISES. *Qu'est-ce que le contrôle destructif? Méthodes, exemples et autres*. Accédé le 11 février 2025. 2025. URL : https://mfe-is.com/fr/essais-destructifs/?utm_source=chatgpt.com.

[110] Knauf AUTOMOTIVE. *Contrôle de la qualité dans la production de pièces pour l'industrie automobile*. Consulté le 21 novembre 2024. 2024. URL : https://knaufautomotive.com/fr/expertises/controle-de-la-qualite/.

[111] Mouhsine ZIANI. *Contrôle de qualité dans l'aéronautique : entre normes et exigences*. Consulté le 21 novembre 2024. Nov. 2024. URL : https://fr.linkedin.com/pulse/contr%C3%B4le-de-qualit%C3%A9-dans-la%C3%A9ronautique-entre-normes-et-mouhsine-ziani-lbple.

[112] HQTS Group LTD. *Contrôle de la qualité des produits électroniques grand public*. Consulté le 21 novembre 2024. Nov. 2024. URL : https://www.hqts.com/fr/industries-que-nous-desservons/produits-de-consommation/electronique/produits-electroniques/.

[113] HQTS Group LTD. *Contrôle qualité dans l'industrie textile : normes et procédures*. Consulté le 21 novembre 2024. Sept. 2024. URL : https://www.hqts.com/fr/controle-qualite-textile-normes-procedures/.

[114] EREA. *Contrôle qualité en milieu pharmaceutique - EREA*. Accessed on June 09, 2024. Avr. 2023. URL : https://www.erea.fr/controle-qualite-milieu-pharmaceutique/#:~:text=Le%5C%20contr%5C%C3%5C%B4le%5C%20qualit%5C%C3%5C%A9%5C%20en%5C%20pharmaceutique,%5C%2C%5C%20immuno%5C%2Dh%5C%C3%5C%A9matologiques%5C%2C%5C%20etc.

[115] Frederick Winslow TAYLOR. *The principles of scientific management*. NuVision Publications, LLC, 1911.

[116] Walter Andrew SHEWHART. « Economic control of quality of manufactured product. » In : (1931).

[117] W Edwards DEMING. « Principles for transformation ». In : *Out of the Crisis* 18 (1986), p. 96.

[118] Joseph M JURAN et Frank M GRYNA. « Juran's quality control ». In : *Handbook. 4ème éd. New York* (1988).

[119] Pete PANDE, R NEUMAN et R CAVANAGH. « The Six Sigma Way McGraw-Hill ». In : *New York, NY* (2000).

[120] James P WOMACK, Daniel T JONES et Daniel ROOS. *The machine that changed the world : The story of lean production–Toyota's secret weapon in the global car wars that is now revolutionizing world industry*. Simon et Schuster, 2007.

[121] a PTC Business ARENA. *Qu'est-ce qu'un système de gestion de la qualité (QMS) | Arena*. Accessed on June 18, 2024. Mai 2024. URL : https://www.arenasolutions.com/fr/what-is-qms/.

[122] Aegis SOFTWARE. *Gestion de la qualité (QMS) | Aegis Software*. Accessed on June 18, 2024. Juin 2021. URL : https://www.aiscorp.com/fr/solutions/gestion-qualite/.

[123] *Contrôle de la qualité dans les aliments - Elex*. Accessed on June 09, 2024. URL : https://www.elex-industrie.com/atteindre-plus-haut-niveau-de-controle-de-qualite-dans-aliments.html#:~:text=Le%5C%20plus%5C%20utilis%5C%C3%5C%A9%5C%20et%5C%20le,la%5C%20cha%5C%C3%5C%AEne%5C%20d'approvisionnement%5C%20alimentaire.

[124] Semih OTLES. *Methods of analysis of food components and additives*. CRC press, 2011.

[125] Leo ML NOLLET et Fidel TOLDRA. « Radionuclide Concentrations in Food ». In : *Handbook of Food Analysis-Two Volume Set*. CRC Press, 2015, p. 1239-1256.

[126] James M JAY, Martin J LOESSNER et David A GOLDEN. *Modern food microbiology*. Springer Science & Business Media, 2006.

[127] TraceX TECHNOLOGIES. *Blockchain for Food Traceability*. Consulté le 20 novembre 2024. 2024. URL : https://tracextech.com/blockchain-for-food-traceability/.

[128] Rizwan Matloob ELLAHI, Lincoln C. WOOD et Alaa El-Din Ahmed BEKHIT. « Blockchain-Based Frameworks for Food Traceability : A Systematic Review ». In : *Foods* 12.16 (2023). ISSN : 2304-8158. URL : https://doi.org/10.3390/foods12163026.

[129] IBM FOOD TRUST. *Enhancing Food Traceability and Transparency with Blockchain Technology*. Consulté le 20 novembre 2024. 2020. URL : https://www.ibm.com/blockchain/solutions/food-trust.

[130] WALMART. *Walmart Implements Blockchain to Improve Food Traceability*. Consulté le 20 novembre 2024. 2018. URL : https://corporate.walmart.com/content/dam/corporate/documents/purpose/environmental-social-and-governance-report-archive/wal-071-2018-grr-full-book1.pdf.

[131] EVERLEDGER. *Using Blockchain for Authenticity and Quality Assurance in the Wine Industry*. Consulté le 20 novembre 2024. 2021. URL : https://everledger.io.

[132] Anton AMANN et al. « The human volatilome : volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva ». In : *Journal of Breath Research* 8.3 (juin 2014), p. 034001. DOI : 10.1088/1752-7155/8/3/034001. URL : https://dx.doi.org/10.1088/1752-7155/8/3/034001.

[133] Julian W GARDNER et Philip N BARTLETT. *Electronic Noses : Principles and Applications*. Oxford University Press, jan. 1999. ISBN : 9780198559559. DOI : 10.1093/oso/9780198559559.001.0001. URL : https://doi.org/10.1093/oso/9780198559559.001.0001.

[134] Antonio RIUL et al. « Wine classification by taste sensors made from ultra-thin films and using neural networks ». In : *Sensors and Actuators B : Chemical* 98.1 (2004), p. 77-82. ISSN : 0925-4005. URL : https://doi.org/10.1016/j.snb.2003.09.025.

[135] Mahdi GHASEMI-VARNAMKHASTI et al. « Meat Quality Assessment by Electronic Nose (Machine Olfaction Technology) ». In : *Sensors* 9.8 (2009), p. 6058-6083. ISSN : 1424-8220. DOI : 10.3390/s90806058. URL : https://www.mdpi.com/1424-8220/9/8/6058.

[136] Yuri VLASOV, Andrey LEGIN et Alisa RUDNITSKAYA. « Electronic tongues and their analytical application ». In : *Analytical and Bioanalytical Chemistry* 373 (2002), p. 136-146. URL : https://doi.org/10.1007/s00216-002-1310-2.

[137] Jersson X. LEON-MEDINA, Maribel ANAYA et Diego A. TIBADUIZA. « New Electronic Tongue Sensor Array System for Accurate Liquor Beverage Classification ». In : *Sensors* 23.13 (2023). ISSN : 1424-8220. URL : https://doi.org/10.3390/s23136178.

[138] Marta PODRAŻKA et al. « Electronic Tongue—A Tool for All Tastes ? » In : *Biosensors* 8.1 (2018). ISSN : 2079-6374. URL : https://doi.org/10.3390/bios8010003.

[139] Carsten STEGER, Markus ULRICH et Christian WIEDEMANN. *Machine vision algorithms and applications*. John Wiley & Sons, 2018.

[140] Cheng-Jin DU et Da-Wen SUN. « Learning techniques used in computer vision for food quality evaluation : a review ». In : *Journal of Food Engineering* 72.1 (2006), p. 39-55. ISSN : 0260-8774. URL : https://doi.org/10.1016/j.jfoodeng.2004.11.017.

[141] Geert LITJENS et al. « A survey on deep learning in medical image analysis ». In : *Medical Image Analysis* 42 (2017), p. 60-88. ISSN : 1361-8415. URL : https://doi.org/10.1016/j.media.2017.07.005.

[142] European Environment AGENCY. *Waste management : Production and Consumption, Towards Zero Pollution*. Consulté le 20 novembre 2024. European Environment Agency, 2021. URL : https://www.eea.europa.eu/publications/zero-pollution/production-consumption/waste-management.

[143] European Environment AGENCY. *The European Environment – State and Outlook 2020 : Knowledge for Transition to a Sustainable Europe*. Consulté le 20 novembre 2024. European Environment Agency, 2020. URL : https://www.eea.europa.eu/soer/publications/soer-2020.

[144] European Environment AGENCY. *Counting the costs of industrial pollution : Assessing the external costs of air pollution from European industrial facilities 2008–2017*. Consulté le 20 novembre 2024. European Environment Agency, 2021. URL : https://www.eea.europa.eu/publications/counting-the-costs-of-industrial-pollution.

[145] William TULLETT. « The Smell of Drugs : Medicines, the Senses, and Efficacy ». In : *Smell in Eighteenth-Century England : A Social Sense*. Oxford University Press, août 2019. ISBN : 9780198844136. eprint : https://academic.oup.com/book/0/chapter/268696652/chapter-pdf/40186437/oso-9780198844136-chapter-4.pdf. URL : https://doi.org/10.1093/oso/9780198844136.003.0004.

[146] Manuela BAIETTO et Alphus D. WILSON. « Electronic-Nose Applications for Fruit Identification, Ripeness and Quality Grading ». In : *Sensors* 15.1 (2015), p. 899-931. ISSN : 1424-8220. URL : https://doi.org/10.3390/s150100899.

[147] Constantin APETREI, Mahdi GHASEMI-VARNAMKHASTI et Irina MIRELA APETREI. « Chapter 27 - Olive Oil and Combined Electronic Nose and Tongue ». In : *Electronic Noses and Tongues in Food Science*. Sous la dir. de María Luz RODRÍGUEZ MÉNDEZ. San Diego : Academic Press, 2016, p. 277-289. ISBN : 978-0-12-800243-8. URL : https://doi.org/10.1016/B978-0-12-800243-8.00027-5.

[148] Susanna BURATTI et Simona BENEDETTI. « Chapter 28 - Alcoholic Fermentation Using Electronic Nose and Electronic Tongue ». In : *Electronic Noses and Tongues in Food Science*. Sous la dir. de María Luz RODRÍGUEZ MÉNDEZ. San Diego : Academic Press, 2016, p. 291-299.

ISBN : 978-0-12-800243-8. URL : https://doi.org/10.1016/B978-0-12-800243-8.00028-7.

[149] Carla Guanais BRANCHINI et al. « Chapter 29 - Wine and Combined Electronic Nose and Tongue ». In : *Electronic Noses and Tongues in Food Science*. Sous la dir. de María Luz RODRÍGUEZ MÉNDEZ. San Diego : Academic Press, 2016, p. 301-307. ISBN : 978-0-12-800243-8. URL : https://doi.org/10.1016/B978-0-12-800243-8.00029-9.

[150] A.H. ABDULLAH et al. « Chapter 11 - Rice and the Electronic Nose ». In : *Electronic Noses and Tongues in Food Science*. Sous la dir. de María Luz RODRÍGUEZ MÉNDEZ. San Diego : Academic Press, 2016, p. 103-113. ISBN : 978-0-12-800243-8. URL : https://doi.org/10.1016/B978-0-12-800243-8.00011-1.

[151] PB BHANDARE, NS PENDBHAJE et AP NARANG. « Electronic nose : A review ». In : *RRJET* 2 (2013), p. 1-8.

[152] Krishna PERSAUD. « Chapter 1 - Electronic Noses and Tongues in the Food Industry ». In : *Electronic Noses and Tongues in Food Science*. Sous la dir. de María Luz RODRÍGUEZ MÉNDEZ. San Diego : Academic Press, 2016, p. 1-12. ISBN : 978-0-12-800243-8. URL : https://doi.org/10.1016/B978-0-12-800243-8.00001-9.

[153] Sundar BALASUBRAMANIAN, Jayendra AMAMCHARLA et Jin-E SHIN. « Chapter 7 - Possible Application of Electronic Nose Systems for Meat Safety : An Overview ». In : *Electronic Noses and Tongues in Food Science*. Sous la dir. de María Luz RODRÍGUEZ MÉNDEZ. San Diego : Academic Press, 2016, p. 59-71. ISBN : 978-0-12-800243-8. URL : https://doi.org/10.1016/B978-0-12-800243-8.00007-X.

[154] Zulfiqur ALI, William O'HARE et B. THEAKER. « Detection of bacterial contaminated milk by means of a quartz crystal microbalance based electronic nose ». In : *Journal of Thermal Analysis and Calorimetry - J THERM ANAL CALORIM* 71 (jan. 2003), p. 155-161. URL : https://doi.org/10.1023/A:1022274419166.

[155] S CAPONE et al. « Analysis of vapours and foods by means of an electronic nose based on a sol–gel metal oxide sensors array ». In : *Sensors and Actuators B : Chemical* 69.3 (2000). Proceedings of the International Symposium on Electronic Noses, p. 230-235. ISSN : 0925-4005. URL : https://doi.org/10.1016/S0925-4005(00)00496-2.

[156] S CAPONE et al. « Monitoring of rancidity of milk by means of an electronic nose and a dynamic PCA analysis ». In : *Sensors and Actuators B : Chemical* 78.1 (2001). Selected Papers from Eurosensors XIV, p. 174-179. ISSN : 0925-4005. URL : https://doi.org/10.1016/S0925-4005(01)00809-7.

[157] Jun FU et al. « Pattern Classification Using an Olfactory Model with PCA Feature Selection in Electronic Noses : Study and Application ». In : *Sensors* 12.3 (2012), p. 2818-2830. ISSN : 1424-8220. URL : https://doi.org/10.3390/s120302818.

[158] Robin C MCKELLAR et al. « The electronic nose as a tool for the classification of fruit and grape wines from different Ontario wineries ». In : *Journal of the Science of Food and Agriculture* 85.14 (2005), p. 2391-2396. URL : https://doi.org/10.1002/jsfa.2240.

[159] T TAN et al. « The electronic nose-a new instrument for sensing vapours ». In : *LC GC INTERNATIONAL* 8 (1995), p. 218-218.

[160] WJ HARPER, S SOHN et K DA JOU. « The role of fatty acids in the aroma profiles of Swiss cheese as determined by an electronic nose ». In : *Olfaction and Electronic Nose, 3rd International Symposium, Toulouse*. 1996.

[161] Kathy ADAMS, Jerry SUCHY, Odean M LUKOW et al. « Detection of wheat variety flour aromas using an electronic nose. » In : *Electronic Journal of Environmental, Agricultural and Food Chemistry* 10.9 (2011), p. 2861-2873.

[162] Jing YE et al. « Differentiation of two types of pu-erh teas by using an electronic nose and ultrasound-assisted extraction-dispersive liquid–liquid microextraction-gas chromatography-mass spectrometry ». In : *Anal. Methods* 8 (3 2016), p. 593-604. DOI : https://doi.org/10.1039/C5AY02652K. URL : http://dx.doi.org/10.1039/C5AY02652K.

[163] H YOUNG et al. « Characterization of Royal Gala Apple Aroma Using Electronic Nose Technology-Potential Maturity Indicator ». In : *Journal of Agricultural and Food Chemistry* 47.12 (1999), p. 5173-5177. URL : https://doi.org/10.1021/jf990276u.

[164] Saïd LABRECHE et al. « Shelf life determination by electronic nose : application to milk ». In : *Sensors and Actuators B : Chemical* 106.1 (2005). ISOEN 2003 - Selected Papers from the 10th International Symposium on Olfaction and Electronic Noses., p. 199-206. ISSN : 0925-4005. URL : https://doi.org/10.1016/j.snb.2004.06.027.

[165] Santina ROMANI et al. « Evaluation of Coffee Roasting Degree by Using Electronic Nose and Artificial Neural Network for Off-line Quality Control ». In : *Journal of Food Science* 77.9 (2012), p. C960-C965. eprint : https://ift.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1750-3841.2012.02851.x. URL : https://doi.org/10.1111/j.1750-3841.2012.02851.x.

[166] Lorenzo SIROLI et al. « Efficacy of natural antimicrobials to prolong the shelf-life of minimally processed apples packaged in modified atmosphere ». In : *Food Control* 46 (2014), p. 403-411. ISSN : 0956-7135. URL : https://doi.org/10.1016/j.foodcont.2014.05.049.

[167] P. LAURIENZO et al. « Microstructure and olfactory quality of apples de-hydrated by innovative technologies ». In : *Journal of Food Engineering* 116.3 (2013), p. 689-694. ISSN : 0260-8774. URL : https://doi.org/10.1016/j.jfoodeng.2013.01.002.

[168] Lakshmi P. PATHANGE et al. « Non-destructive evaluation of apple maturity using an electronic nose system ». In : *Journal of Food Engineering* 77.4 (2006), p. 1018-1023. ISSN : 0260-8774. URL : https://doi.org/10.1016/j.jfoodeng.2005.08.034.

[169] M. FALASCONI et al. « The novel EOS835 electronic nose and data analysis for evaluating coffee ripening ». In : *Sensors and Actuators B : Chemical* 110.1 (2005), p. 73-80. ISSN : 0925-4005. URL : https://doi.org/10.1016/j.snb.2005.01.019.

[170] Changying LI, Paul HEINEMANN et Richard SHERRY. « Neural network and Bayesian network fusion models to fuse electronic nose and surface acoustic wave sensor data for apple defect detection ». In : *Sensors and Actuators B : Chemical* 125.1 (2007), p. 301-310. ISSN : 0925-4005. URL : https://doi.org/10.1016/j.snb.2007.02.027.

[171] Francesco MASULLI et al. « Boosting and Classification of Electronic Nose Data ». In : *Multiple Classifier Systems*. Sous la dir. de Fabio ROLI et Josef KITTLER. Berlin, Heidelberg : Springer Berlin Heidelberg, 2002, p. 262-271. ISBN : 978-3-540-45428-1. URL : https://doi.org/10.1007/3-540-45428-4%5C_26.

[172] Guohua HUI. « Chapter 2 - Apple Analysis and the Electronic Nose ». In : *Electronic Noses and Tongues in Food Science*. Sous la dir. de María Luz RODRÍGUEZ MÉNDEZ. San Diego : Academic Press, 2016, p. 15-19. ISBN : 978-0-12-800243-8. URL : https://doi.org/10.1016/B978-0-12-800243-8.00002-0.

[173] Tomomi MICHISHITA et al. « Gas Chromatography/Olfactometry and Electronic Nose Analyses of Retronasal Aroma of Espresso and Correlation with Sensory Evaluation by an Artificial Neural Network ». In : *Journal of Food Science* 75.9 (2010), S477-S489. URL : https://doi.org/10.1111/j.1750-3841.2010.01828.x.

[174] Emmanuelle SCHALLER, Jacques O. BOSSET et Felix ESCHER. « 'Electronic Noses' and Their Application to Food ». In : *LWT - Food Science and Technology* 31.4 (1998), p. 305-316. ISSN : 0023-6438. URL : https://doi.org/10.1006/fstl.1998.0376.

[175] Evor L. HINES et al. « Pattern Analysis for Electronic Noses ». In : *Handbook of Machine Olfaction*. John Wiley & Sons, Ltd, 2002. Chap. 6, p. 133-160. ISBN : 9783527601592. eprint : https://onlinelibrary.wiley.com/doi/pdf/10.1002/3527601597.ch6. URL : https://doi.org/10.1002/3527601597.ch6.

[176] Evandro BONA et Rui Sérgio DOS SANTOS FERREIRA DA SILVA. « Chapter 4 - Coffee and the Electronic Nose ». In : *Electronic Noses and Tongues in Food Science*. Sous la dir. de María Luz RODRÍGUEZ MÉNDEZ. San Diego : Academic Press, 2016, p. 31-38. ISBN : 978-0-12-800243-8. URL : https://doi.org/10.1016/B978-0-12-800243-8.00004-4.

[177] Ferath KHERIF et Adeliya LATYPOVA. « Chapter 12 - Principal component analysis ». In : *Machine Learning*. Sous la dir. d'Andrea MECHELLI et Sandra VIEIRA. Academic Press, 2020, p. 209-225. ISBN : 978-0-12-815739-8. URL : https://doi.org/10.1016/B978-0-12-815739-8.00012-2.

[178] Markus RINGNÉR. « What is principal component analysis ? » In : *Nature biotechnology* 26.3 (2008), p. 303-304. URL : https://doi.org/10.1038/nbt0308-303.

[179] The NOBLES. *Understanding Principle Component Analysis (PCA) step by step.* https://medium.com/analytics-vidhya/understanding-principle-component-analysis-pca-step-by-step-e7a4bb4031d9. Accessed October 31, 2022. Jan. 2020.

[180] Alaa THARWAT et al. « Linear discriminant analysis : A detailed tutorial ». In : *AI communications* 30.2 (2017), p. 169-190. URL : https://doi.org/10.3233/AIC-170729.

[181] Suresh BALAKRISHNAMA et Aravind GANAPATHIRAJU. « Linear discriminant analysis-a brief tutorial ». In : *Institute for Signal and information Processing* 18.1998 (1998), p. 1-8.

[182] Magdalena ŚLIWIŃSKA et al. « Food Analysis Using Artificial Senses ». In : *Journal of Agricultural and Food Chemistry* 62.7 (2014). PMID : 24506450, p. 1423-1448. URL : https://doi.org/10.1021/jf403215y.

[183] Juan Carlos Gonzalez AVELLA. *Using Linear Discriminant Analysis (LDA) for data Explore : Step by Step.* https://apsl.tech/en/blog/using-linear-discriminant-analysis-lda-data-explore-step-step. Accessed October 31, 2022. Juill. 2017.

[184] Kee Siong NG. « A simple explanation of partial least squares ». In : *The Australian National University, Canberra* (2013), p. 1-10.

[185] Svante WOLD, Michael SJÖSTRÖM et Lennart ERIKSSON. « PLS-regression : a basic tool of chemometrics ». In : *Chemometrics and Intelligent Laboratory Systems* 58.2 (2001). PLS Methods, p. 109-130. ISSN : 0169-7439. URL : https://doi.org/10.1016/S0169-7439(01)00155-1.

[186] Richard G. BRERETON et Gavin R. LLOYD. « Partial least squares discriminant analysis : taking the magic away ». In : *Journal of Chemometrics* 28.4 (2014), p. 213-225. URL : https://doi.org/10.1002/cem.2609.

[187] Gaston SANCHEZ et Ethan MARZBAN. *All Models Are Wrong : Concepts of Statistical Learning.* https://allmodelsarewrong.github.io. Accessed November 2, 2022. 2020.

[188] Itziar FRADES et Rune MATTHIESEN. « Overview on techniques in cluster analysis ». In : *Bioinformatics methods in clinical research* (2010), p. 81-107. URL : https://doi.org/10.1007/978-1-60327-194-3%5C_5.

[189] Vikramaditya JAKKULA. « Tutorial on support vector machine (SVM) ». In : *School of EECS, Washington State University* 37.2.5 (2006), p. 3.

[190] ANALYTIXLABS. *Introduction to SVM – Support Vector Machine Algorithm of Machine Learning.* https://www.analytixlabs.co.in/blog/introduction-support-vector-machine-algorithm. Published On July 7, 2021 and Accessed November 3, 2022. 2021.

[191] Gongde GUO et al. « KNN Model-Based Approach in Classification ». In : *On The Move to Meaningful Internet Systems 2003 : CoopIS, DOA, and ODBASE*. Sous la dir. de Robert MEERSMAN, Zahir TARI et Douglas C. SCHMIDT. Berlin, Heidelberg : Springer Berlin Heidelberg, 2003, p. 986-996. ISBN : 978-3-540-39964-3. URL : https://doi.org/10.1007/978-3-540-39964-3%5C_62.

[192] Onel HARRISON. *Machine Learning Basics with the K-Nearest Neighbors Algorithm*. https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761. Published On September 10, 2018 and Accessed November 3, 2022. 2018.

[193] L.A. ZADEH. « Fuzzy sets ». In : *Information and Control* 8.3 (1965), p. 338-353. ISSN : 0019-9958. URL : https://doi.org/10.1016/S0019-9958(65)90241-X.

[194] Tim C PEARCE et al. *Handbook of machine olfaction : electronic nose technology*. John Wiley & Sons, 2006.

[195] Jayeeta SHARMA, Amlan Jyoti BHAGAWATI et Riku CHUTIA. « Fuzzy Logic based Odour Classification System in Electronic Nose ». In : *International Journal of Computer Applications* 78.15 (2013).

[196] Runu Banerjee (ROY) et al. « Fusion of Electronic Nose and Tongue Response Using Fuzzy based Approach for Black Tea Classification ». In : *Procedia Technology* 10 (2013). First International Conference on Computational Intelligence : Modeling Techniques and Applications (CIMTA) 2013, p. 615-622. ISSN : 2212-0173. URL : https://doi.org/10.1016/j.protcy.2013.12.402.

[197] Ritaban DUTTA et al. « Electronic nose based tea quality standardization ». In : *Neural Networks* 16.5 (2003). Advances in Neural Networks Research : IJCNN '03, p. 847-853. ISSN : 0893-6080. URL : https://doi.org/10.1016/S0893-6080(03)00092-3.

[198] Cosimo DISTANTE, Pietro SICILIANO et Lorenzo VASANELLI. « Odor discrimination using adaptive resonance theory ». In : *Sensors and Actuators B : Chemical* 69.3 (2000). Proceedings of the International Symposium on Electronic Noses, p. 248-252. ISSN : 0925-4005. URL : https://doi.org/10.1016/S0925-4005(00)00502-5.

[199] Abbas GORJI-CHAKESPARI et al. « Performance Comparison of Fuzzy ARTMAP and LDA in Qualitative Classification of Iranian Rosa damascena Essential Oils by an Electronic Nose ». In : *Sensors* 16.5 (2016). ISSN : 1424-8220. URL : https://doi.org/10.3390/s16050636.

[200] TUTORIALSPOINT. *Artificial Intelligence - Fuzzy Logic Systems*. https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_fuzzy_logic_systems.htm. Accessed November 3, 2022.

[201] S AGATONOVIC-KUSTRIN et R BERESFORD. « Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research ». In : *Journal of Pharmaceutical and Biomedical Analysis* 22.5 (2000), p. 717-727. ISSN : 0731-7085. URL : https://doi.org/10.1016/S0731-7085(99)00272-1.

[202] DEEPANSHI. *Beginners Guide to Artificial Neural Network*. https://www.analyticsvidhya.com/blog/2021/05/beginners-guide-to-artificial-neural-network. Published On May 25, 2021, Last Modified On May 31, 2021 and Accessed November 3, 2022. 2021.

[203] Chaohua YAN et An LU. « A deep learning method combined with electronic nose to identify the rice origin ». In : *Journal of Instrumentation* 17.08 (août 2022), P08016. URL : https://dx.doi.org/10.1088/1748-0221/17/08/P08016.

[204] Juan C. RODRIGUEZ GAMBOA et al. « Validation of the rapid detection approach for enhancing the electronic nose systems performance, using different deep learning models and support vector machines ». In : *Sensors and Actuators B : Chemical* 327 (2021), p. 128921. ISSN : 0925-4005. URL : https://doi.org/10.1016/j.snb.2020.128921.

[205] Massimo ARIA et Corrado CUCCURULLO. « bibliometrix : An R-tool for comprehensive science mapping analysis ». In : *Journal of Informetrics* 11.4 (2017), p. 959-975. ISSN : 1751-1577. URL : https://doi.org/10.1016/j.joi.2017.08.007.

[206] FAO et al. « L'État de la sécurité alimentaire et de la nutrition dans le monde 2022 ». In : *Réorienter les politiques alimentaires et agricoles pour rendre l'alimentation saine plus abordable* (Rome 2022). URL : https://doi.org/10.4060/cc0639fr.

[207] M.T.S.R. GOMES. « Chapter 3 - Electronic Nose in Dairy Products ». In : *Electronic Noses and Tongues in Food Science*. Sous la dir. de María Luz RODRÍGUEZ MÉNDEZ. San Diego : Academic Press, 2016, p. 21-30. ISBN : 978-0-12-800243-8. URL : https://doi.org/10.1016/B978-0-12-800243-8.00003-2.

[208] José de JESÚS RUBIO. « A method with neural networks for the classification of fruits and vegetables ». In : *Soft Computing* 21.23 (2017), p. 7207-7220. URL : https://doi.org/10.1007/s00500-016-2263-2.

[209] Wenjiang DONG et al. « Differentiation of Chinese robusta coffees according to species, using a combined electronic nose and tongue, with the aid of chemometrics ». In : *Food Chemistry* 229 (2017), p. 743-751. ISSN : 0308-8146. URL : https://doi.org/10.1016/j.foodchem.2017.02.149.

[210] Angiras MODAK et al. « A novel fuzzy based signal analysis technique in electronic nose and electronic tongue for black tea quality analysis ». In : *2016 IEEE First International Conference on Control, Measurement and Instrumentation (CMI)*. 2016, p. 279-283. URL : https://doi.org/10.1109/CMI.2016.7413755.

[211] Wandee AUNSA-ARD et al. « Prediction of Rice Odor by Using an Electronic Nose and Artificial Neural Network ». In : *2019 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)*. 2019, p. 1-2. URL : https://doi.org/10.1109/ICCE-TW46550.2019.8992028.

[212] Santina ROMANI et María Teresa RODRIGUEZ-ESTRADA. « Chapter 5 - Bakery Products and Electronic Nose ». In : *Electronic Noses and Tongues in Food Science*. Sous la dir. de María Luz RODRÍGUEZ MÉNDEZ. San Diego : Academic Press, 2016, p. 39-47. ISBN : 978-0-12-800243-8. URL : https://doi.org/10.1016/B978-0-12-800243-8.00005-6.

[213] Thomas HÜBERT, Carlo TIEBE et Ulrich BANACH. « Chapter 12 - Electronic Noses for the Quality Control of Spices ». In : *Electronic Noses and Tongues in Food Science*. Sous la dir. de María Luz RODRÍGUEZ MÉNDEZ. San Diego : Academic Press, 2016, p. 115-124. ISBN : 978-0-12-800243-8. URL : https://doi.org/10.1016/B978-0-12-800243-8.00012-3.

[214] Yan SHI et al. « A deep feature mining method of electronic nose sensor data for identifying beer olfactory information ». In : *Journal of Food Engineering* 263 (2019), p. 437-445. ISSN : 0260-8774. URL : https://doi.org/10.1016/j.jfoodeng.2019.07.023.

[215] Xiaojing TIAN et al. « Combination of an E-nose and an E-tongue for adulteration detection of minced mutton mixed with pork ». In : *Journal of Food Quality* 2019 (2019). DOI : 10.1155/2019/4342509. URL : https://doi.org/10.1155/2019/4342509.

[216] Xing-yi HUANG et al. « Evaluating quality of tomato during storage using fusion information of computer vision and electronic nose ». In : *Journal of Food Process Engineering* 41.6 (2018), e12832. eprint : https://onlinelibrary.wiley.com/doi/pdf/10.1111/jfpe.12832. URL : https://doi.org/10.1111/jfpe.12832.

[217] Z XIAOBO, Z JIEWEN et L YANXIAO. « Objective quality assessment of apples using machine vision, NIR spectrophotometer, and electronic nose ». In : *Transactions of the ASABE* 53.4 (2010), p. 1351-1358. DOI : 10.13031/2013.32572. URL : https://doi.org/10.13031/2013.32572.

[218] Stijn SAEVELS et al. « An electronic nose and a mass spectrometry-based electronic nose for assessing apple quality during shelf life ». In : *Postharvest Biology and Technology* 31.1 (2004), p. 9-19. ISSN : 0925-5214. URL : https://doi.org/10.1016/S0925-5214(03)00129-7.

[219] Gemma ECHEVERRÍA et al. « Characterization of Fuji apples from different harvest dates and storage conditions from measurements of volatiles by gas chromatography and electronic nose ». en. In : *J Agric Food Chem* 52.10 (mai 2004), p. 3069-3076. URL : https://doi.org/10.1021/jf035271i.

[220] Fiorella SINESIO et al. « Use of electronic nose and trained sensory panel in the evaluation of tomato quality ». In : *Journal of the Science of Food and Agriculture* 80.1 (2000), p. 63-71. URL : https://doi.org/10.1002/(SICI)1097-0010(20000101)80:1%3C63::AID-JSFA479%3E3.0.CO;2-8.

[221] Changying LI, Paul Heinz HEINEMANN et Joseph IRUDAYARAJ. « Detection of apple deterioration using an electronic nose and zNose^{TM} ». In : *Transactions of the Asabe* 50.4 (2007), p. 1417-1425. URL : https://doi.org/10.13031/2013.23614.

[222] Antihus Hernández GÓMEZ et al. « Evaluation of tomato maturity by electronic nose ». In : *Computers and Electronics in Agriculture* 54.1 (2006), p. 44-52. ISSN : 0168-1699. URL : https://doi.org/10.1016/j.compag.2006.07.002.

[223] Antihus Hernández GÓMEZ et al. « Monitoring storage shelf life of tomato using electronic nose technique ». In : *Journal of Food Engineering* 85.4 (2008), p. 625-631. ISSN : 0260-8774. URL : https://doi.org/10.1016/j.jfoodeng.2007.06.039.

[224] Amalia Z. BERNA et al. « Electronic nose systems to study shelf life and cultivar effect on tomato aroma profile ». In : *Sensors and Actuators B : Chemical* 97.2 (2004), p. 324-333. ISSN : 0925-4005. URL : https://doi.org/10.1016/j.snb.2003.09.020.

[225] Paola PANI et al. « Influence of an osmotic pre-treatment on structure-property relationships of air-dehydrated tomato slices ». In : *Journal of Food Engineering* 86.1 (2008), p. 105-112. ISSN : 0260-8774. URL : https://doi.org/10.1016/j.jfoodeng.2007.09.017.

[226] I. CONCINA et al. « Early detection of microbial contamination in processed tomatoes by electronic nose ». In : *Food Control* 20.10 (2009), p. 873-880. ISSN : 0956-7135. URL : https://doi.org/10.1016/j.foodcont.2008.11.006.

[227] Kazimierz BRUDZEWSKI, Stanislaw OSOWSKI et Anna DWULIT. « Recognition of Coffee Using Differential Electronic Nose ». In : *IEEE Transactions on Instrumentation and Measurement* 61.6 (2012), p. 1803-1810. URL : https://doi.org/10.1109/TIM.2012.2184011.

[228] M.L. RODRIGUEZ-MENDEZ et al. « Evaluation of oxygen exposure levels and polyphenolic content of red wines using an electronic panel formed by an electronic nose and an electronic tongue ». In : *Food Chemistry* 155 (2014), p. 91-97. ISSN : 0308-8146. URL : https://doi.org/10.1016/j.foodchem.2014.01.021.

[229] Yang Ji WEI et al. « Application of Electronic Nose for Detection of Wine-Aging Methods ». In : *Material Research and Applications*. T. 875. Advanced Materials Research. Trans Tech Publications Ltd, mars 2014, p. 2206-2213. URL : https://doi.org/10.4028/www.scientific.net/AMR.875-877.2206.

[230] Jesús LOZANO et al. « Automatic Sensor System for the Continuous Analysis of the Evolution of Wine ». In : *American Journal of Enology and Viticulture* 66.2 (2015), p. 148-155. ISSN : 0002-9254. eprint : https://www.ajevonline.org/content/66/2/148.full.pdf. URL : https://doi.org/10.5344/ajev.2014.14103.

[231] Lili WU et al. « Identification of Early Moldy Rice Samples by PCA and PNN ». In : *Communications and Information Processing*. Sous la dir. de Maotai ZHAO et Junpin SHA. Berlin, Heidelberg : Springer Berlin Heidelberg, 2012, p. 506-514. ISBN : 978-3-642-31965-5. URL : https://doi.org/10.1007/978-3-642-31965-5_59.

[232] Bo ZHOU et Jun WANG. « Use of electronic nose technology for identifying rice infestation by Nilaparvata lugens ». In : *Sensors and Actuators B : Chemical* 160.1 (2011), p. 15-21. ISSN : 0925-4005. URL : https://doi.org/10.1016/j.snb.2011.07.002.

[233] Arun JANA et al. « Regression model on electronic nose data from aromatic rice samples ». In : *2012 Sixth International Conference on Sensing Technology (ICST)*. 2012, p. 418-421. URL : https://doi.org/10.1109/ICSensT.2012.6461712.

[234] Danfeng WANG et al. « Prediction of total viable counts on chilled pork using an electronic nose combined with support vector machine ». In : *Meat Science* 90.2 (2012), p. 373-377. ISSN : 0309-1740. URL : https://doi.org/10.1016/j.meatsci.2011.07.025.

[235] Xiaojing TIAN, Jun WANG et Shaoqing CUI. « Analysis of pork adulteration in minced mutton using electronic nose of metal oxide sensors ». In : *Journal of Food Engineering* 119.4 (2013), p. 744-749. ISSN : 0260-8774. URL : https://doi.org/10.1016/j.jfoodeng.2013.07.004.

[236] Shankara NAIKA et al. « La culture de la tomate : production, transformation et commercialisation ». In : *Agrodok* (2005).

[237] FAOSTAT. *Tomato production data*. https://www.fao.org/faostat/en/#data/QCL/visualize. Accessed on March 20, 2023.

[238] Louis Albert de BROGLIE, Dominique GUÉROULT et Sonia BUCHARD. *Tomates d'hier et d'aujourd'hui*. Hoëbeke, 2005.

[239] J DOSSOU, I SOULÉ et Marcelline MONTCHO. « Evaluation des caractéristiques physico-chimiques et sensorielles de la purée de tomate locale produite à petite échelle au Bénin ». In : *Tropicultura* 25.2 (2007), p. 119-125.

[240] Olga FAGBOHOUN et Denis G KIKI. « Aperçu sur les principales variétés de tomate locale cultivées dans le sud du Bénin ». In : *Bulletin de la Recherche Agronomique Numéro* 24 (1999), p. 10-21.

[241] Paul Ayihadji Ferdinand HOUSSOU et al. « Technologie de production simultanée de purée et du jus de tomate ». In : *International Journal of Biological and Chemical Sciences* 9.5 (2015), p. 2468-2476.

[242] MD MONTCHO. « La transformation de la tomate locale en purée : une approche de solution pour réduire les pertes de production en période de saturation du marché ». In : *InRural. Universe Network SPIPM Interactive* (2002).

[243] Joseph HOUNDEHIN et Yves AGNOUN. *Capitalisation des méthodes de conservation/transformation de la tomate*. AGRI-HUB Bénin Communauté d'Innovation Recherche-Développement. Accessed on November 23, 2022. Juill. 2014.

[244] F ASSOGBA KOMLAN, R SIKIROU et J AZAGBA. « Comment réussir la culture de la tomate en toute saison : Cas des régions urbaines et périurbaines du sud-Bénin ». In : *RTE. 2ème édition. INRAB* (2013).

[245] Armel C. G. MENSAH et al. *Mieux produire la tomate en toute période au Bénin*. 56p. Référentiel Technico-Economique (RTE). MAEP/INRAB/FIDA/ProCar/PADMAR/World Vegetable Center/Bénin. Dépôt légal N° 11553, du 26/08/2019, Bibliothèque Nationale (BN) du Bénin, 3ème trimestre. ISBN : 978-99982-53-13-1., 2019. ISBN : 978-99982-53-13-1.

[246] Delphine M POTT, Sonia OSORIO et José G VALLARINO. « From central to specialized metabolism : An overview of some secondary compounds derived from the primary metabolism for their role in conferring nutritional and organoleptic characteristics to fruit ». In : *Frontiers in plant science* 10 (2019), p. 835. URL : https://doi.org/10.3389/fpls.2019.00835.

[247] Harry J KLEE et Denise M TIEMAN. « The genetics of fruit flavour preferences ». In : *Nature Reviews Genetics* 19.6 (2018), p. 347-356. URL : https://doi.org/10.1038/s41576-018-0002-5.

[248] Diane M. BECKLES. « Factors affecting the postharvest soluble solids and sugar content of tomato (Solanum lycopersicum L.) fruit ». In : *Postharvest Biology and Technology* 63.1 (2012), p. 129-140. ISSN : 0925-5214. URL : https://doi.org/10.1016/j.postharvbio.2011.05.016.

[249] Reza GHAFFARI et al. « Early detection of diseases in tomato crops : An Electronic Nose and intelligent systems approach ». In : *The 2010 International Joint Conference on Neural Networks (IJCNN)*. 2010, p. 1-6. URL : https://doi.org/10.1109/IJCNN.2010.5596535.

[250] Ronnie S. CONCEPCION et al. « Implementation of Multilayer Perceptron Neural Network on Quality Assessment of Tomato Puree in Aerobic Storage using Electronic Nose ». In : *2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*. 2019, p. 65-70. DOI : 10.1109/CIS-RAM47153.2019.9095783. URL : https://doi.org/10.1109/CIS-RAM47153.2019.9095783.

[251] Leo BREIMAN. « Random forests ». In : *Machine learning* 45 (2001), p. 5-32. URL : https://doi.org/10.1023/A:1010933404324.

[252] T. COVER et P. HART. « Nearest neighbor pattern classification ». In : *IEEE Transactions on Information Theory* 13.1 (1967), p. 21-27. URL : https://doi.org/10.1109/TIT.1967.1053964.

[253] Kurt HORNIK, Maxwell STINCHCOMBE et Halbert WHITE. « Multilayer feedforward networks are universal approximators ». In : *Neural Networks* 2.5 (1989), p. 359-366. ISSN : 0893-6080. URL : https://doi.org/10.1016/0893-6080(89)90020-8.

[254] Corinna CORTES. « Support-Vector Networks ». In : *Machine Learning* (1995). URL : https://doi.org/10.1007/BF00994018.

[255] Anjaneyulu Babu SHAIK et Sujatha SRINIVASAN. « A Brief Survey on Random Forest Ensembles in Classification Model ». In : *International Conference on Innovative Computing and Communications*. Sous la dir. de Siddhartha BHATTACHARYYA et al. Singapore : Springer Singapore, 2019, p. 253-260. ISBN : 978-981-13-2354-6. URL : https://doi.org/10.1007/978-981-13-2354-6%5C_27.

[256] H. TAUD et J.F. MAS. « Multilayer Perceptron (MLP) ». In : *Geomatic Approaches for Modeling Land Change Scenarios*. Sous la dir. de María Teresa CAMACHO OLMEDO et al. Cham : Springer International Publishing, 2018, p. 451-455. ISBN : 978-3-319-60801-3. URL : https://doi.org/10.1007/978-3-319-60801-3%5C_27.

[257] David E RUMELHART, Geoffrey E HINTON et Ronald J WILLIAMS. « Learning representations by back-propagating errors ». In : *nature* 323.6088 (1986), p. 533-536. URL : https://doi.org/10.1038/323533a0..

[258] Anusha BHARAT, N POOJA et R Anishka REDDY. « Using Machine Learning algorithms for breast cancer risk prediction and diagnosis ». In : *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*. 2018, p. 1-4. URL : https://doi.org/10.1109/CIMCA.2018.8739696.

[259] Abdoulaye DIALLO et al. « Deep Learning Based Binary and Multi-class Classification Comparison for Anomaly Detection ». In : *2022 International Conference on Engineering and Emerging Technologies (ICEET)*. 2022, p. 1-6. URL : https://doi.org/10.1109/ICEET56468.2022.10007171.

[260] MISSOURI BOTANICAL GARDEN. *Environmental (Abiotic) Problems of Tomatoes*. Consulté le 21 novembre 2024. 2023. URL : https://www.missouribotanicalgarden.org/gardens-gardening/your-garden/help-for-the-home-gardener/advice-tips-resources/insects-pests-and-problems/environmental/environmental-abiotic-problems-of-tomatoes.

[261] Elnaz AMIRAHMADI et al. « Impacts of Environmental Factors and Nutrients Management on Tomato Grown under Controlled and Open Field Conditions ». In : *Agronomy* 13.3 (2023). ISSN : 2073-4395. URL : https://doi.org/10.3390/agronomy13030916.

[262] Edward J. COLLINS et al. « Tomatoes : An Extensive Review of the Associated Health Impacts of Tomatoes and Factors That Can Affect Their Cultivation ». In : *Biology* 11.2 (2022). ISSN : 2079-7737. URL : https://doi.org/10.3390/biology11020239.

[263] Preeti SAINI et Bharti NAGPAL. « Deep-LSTM Model for Wheat Crop Yield Prediction in India ». In : *2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT)*. 2022, p. 73-78. URL : https://doi.org/10.1109/CCiCT56684.2022.00025.

[264] Sonal JAIN et Dharavath RAMESH. « AI based hybrid CNN-LSTM model for crop disease prediction : An ML advent for rice crop ». In : *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. Juill. 2021, p. 1-7. URL : https://doi.org/10.1109/ICCCNT51525.2021.9579587.

[265] Maysa Malfiza Garcia de MACEDO, Andrea Britto MATTOS et Dário Augusto Borges OLIVEIRA. « Generalization of Convolutional LSTM Models for Crop Area Estimation ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020), p. 1134-1142. URL : https://doi.org/10.1109/JSTARS.2020.2973602.

[266] Seyed Mahdi Mirhoseini NEJAD et al. « Multispectral Crop Yield Prediction Using 3D-Convolutional Neural Networks and Attention Convolutional LSTM Approaches ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 (2023), p. 254-266. URL : https://doi.org/10.1109/JSTARS.2022.3223423.

[267] V KIRAN KUMAR, KV RAMESH et V RAKESH. « Optimizing LSTM and Bi-LSTM models for crop yield prediction and comparison of their performance with traditional machine learning techniques ». In : *Applied Intelligence* 53.23 (2023), p. 28291-28309. DOI : https://doi.org/10.1007/s10489-023-05005-5.

[268] Sony M KURIAKOSE et Tripty SINGH. « Indian Crop Yield Prediction using LSTM Deep Learning Networks ». In : *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. 2022, p. 1-5. URL : https://doi.org/10.1109/ICCCNT54827.2022.9984407.

[269] Khadijeh ALIBABAEI, Pedro D. GASPAR et Tânia M. LIMA. « Crop Yield Estimation Using Deep Learning Based on Climate Big Data and Irrigation Scheduling ». In : *Energies* 14.11 (2021). ISSN : 1996-1073. URL : https://doi.org/10.3390/en14113004.

[270] Saeed KHAKI, Hieu PHAM et Lizhi WANG. « Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning ». In : *Scientific Reports* 11.1 (2021), p. 11132. URL : https://doi.org/10.1038/s41598-021-89779-z.

[271] Liyun GONG et al. « Deep Learning Based Prediction on Greenhouse Crop Yield Combined TCN and RNN ». In : *Sensors* 21.13 (2021). ISSN : 1424-8220. URL : https://doi.org/10.3390/s21134537.

[272] Colin LEA et al. *Temporal Convolutional Networks for Action Segmentation and Detection*. 2016. arXiv : 1611.05267 [cs.CV]. URL : https://doi.org/10.48550/arXiv.1611.05267.

[273] Usharani BHIMAVARAPU, Gopi BATTINENI et Nalini CHINTALAPUDI. « Improved Optimization Algorithm in LSTM to Predict Crop Yield ». In : *Computers* 12.1 (2023). ISSN : 2073-431X. URL : https://doi.org/10.3390/computers12010010.

[274] Tien Vu HUU et al. « QoE Aware Video Streaming Scheme Utilizing GRU-Based Bandwidth Prediction and Adaptive Bitrate Selection for Heterogeneous Mobile Networks ». In : *IEEE Access* 12 (2024), p. 45785-45795. URL : https://doi.org/10.1109/ACCESS.2024.3382155.

TABLE DES MATIÈRES

Dédicace **ii**

Remerciements **iii**

Sommaire **v**

Résumé **viii**

Abstract **x**

Sigles et abréviations **xi**

Glossaire **xviii**

Liste des figures **xix**

Liste des tableaux **xxii**

Introduction générale **1**

1 Généralités sur l'apprentissage automatique **6**

Introduction 6

1.1 Fondements de l'apprentissage automatique 6

1.1.1 Les bases mathématiques et statistiques 6

1.1.2 Les bases informatiques 7

1.1.3 Les paradigmes de modélisation 7

1.2 Catégories de l'apprentissage automatique 7

1.2.1 Apprentissage supervisé 7

1.2.2 Apprentissage non supervisé 10

1.2.3 Apprentissage par renforcement 12

1.2.4 Apprentissage semi-supervisé 16

1.2.5 Apprentissage auto-supervisé . . . 16
1.2.6 Apprentissage fédéré . . . 17
1.2.7 Apprentissage profond . . . 17
1.3 Limites de l'apprentissage automatique . . . 21
Conclusion . . . 22

2 Approches et outils du contrôle de qualité 23
Introduction . . . 23
2.1 Historique et évolution du contrôle de qualité . . . 23
2.2 Typologie des méthodes de contrôle qualité . . . 26
2.2.1 Contrôle Non Destructif (CND) . . . 26
2.2.2 Contrôle Destructif (CD) . . . 28
2.3 Contrôle de qualité dans quelques secteurs . . . 29
2.3.1 Industrie automobile . . . 29
2.3.2 Industrie aéronautique . . . 29
2.3.3 Industrie électronique . . . 30
2.3.4 Industrie textile . . . 30
2.3.5 Industrie pharmaceutique . . . 31
2.3.6 Industrie manufacturière . . . 32
2.3.7 Industrie alimentaire . . . 33
2.4 Technologies modernes de contrôle de qualité des aliments . . . 34
2.4.1 Blockchain . . . 35
2.4.2 Nez électronique . . . 35
2.4.3 Langue électronique . . . 37
2.4.4 Imagerie et vision par ordinateur . . . 39
2.5 Impact environnemental et économique . . . 41
Conclusion . . . 41

3 Nez électronique et traitement de données 42
Introduction . . . 42
3.1 Technologies du nez électronique . . . 43
3.2 Analyse des données du nez électronique . . . 44
3.2.1 Techniques statistiques . . . 50
3.2.1.1 Analyse en composantes principales . . . 50
3.2.1.2 Analyse discriminante linéaire . . . 51
3.2.1.3 Moindres carrés partiels . . . 51
3.2.1.4 Analyse discriminante fonctionnelle . . . 52
3.2.1.5 Analyse de clusters . . . 52
3.2.2 Techniques intelligentes . . . 53
3.2.2.1 Machine à vecteurs de support . . . 54

3.2.2.2 k plus proches voisins 56
3.2.2.3 Logique floue 57
3.2.2.4 Réseau de neurones artificiels 58
3.3 Discussion 60
Conclusion 62

4 Nez électronique expérimental 64
Introduction 64
4.1 Étude des applications du nez électronique 65
4.2 Conception et spécifications techniques du nez électronique 68
4.2.1 Réseau de capteurs 68
4.2.2 Microcontrôleur 69
4.2.3 Description 70
4.3 Discussion sur l'optimisation et la validation du dispositif 71
4.4 Protocole expérimentale de collecte des données sensorielles de la tomate 75
4.4.1 Tomate en lumière 75
4.4.2 Espèces de tomates 76
4.4.3 Collecte de données 77
4.4.4 Dataset 79
Conclusion 84

5 État des tomates : approches supervisées 86
Introduction 86
5.1 Approches d'apprentissage automatique pour les tomates 86
5.2 Random Forest 88
5.3 k-Nearest Neighbors 89
5.4 Multi-Layer Perceptron 89
5.5 Support Vector Machine 90
5.6 Métriques d'évaluation 91
5.7 Matrice de confusion 92
5.8 Pré-traitement et nettoyage des données 92
5.9 Evaluation des performances des modèles 93
5.10 Analyse des résultats et comparaison avec les méthodes existantes 99
Conclusion 99

6 Prédiction de la durée de vie des tomates 101
Introduction 101
6.1 Application des séries temporelles dans l'agriculture 101
6.2 Pré-traitement et nettoyage des données 103
6.3 Long Short-Term Memory 103

6.4 Gated Recurrent Unit 105
6.5 Configuration du modèle 106
6.6 Résultats expérimentaux 108
6.7 Discussion 110
Conclusion 111

Conclusion générale et perspectives **112**

Liste des publications **114**

Bibliographie **140**

Table des matières **141**