THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : I-MEP² - Ingénierie - Matériaux, Mécanique, Environnement, Energétique, Procédés, Production
Spécialité : MEP - Mécanique des fluides Energétique, Procédés
Unité de recherche : CEA/LITEN - Laboratoire d'Innovation pour les Technologies des Energies Nouvelles et les Nanomatériaux

Diagnostic robuste au vieillissement des piles à combustibles par des méthodes d'apprentissage profond basées physiques

Fault diagnosis of fuel cells using Physics-Informed Deep Learning: towards aging-aware methods

Présentée par :

Chiara PETTOROSSI

Direction de thèse :

Mathias GÉRARD DIRECTEUR DE RECHERCHE, CEA Centre de Grenoble	Directeur de thèse
Vincent HEIRIES INGENIEUR DOCTEUR, CEA-Leti	Co-encadrant de thèse
Sébastien ROSINI INGENIEUR CHERCHEUR, CEA	Co-encadrant de thèse

Rapporteurs :

Fei GAO
PROFESSEUR DES UNIVERSITES, Université de Technologie de Belfort Montbéliard
Rachid OUTBIB
PROFESSEUR DES UNIVERSITES, Aix-Marseille Université

Thèse soutenue publiquement le **3 décembre 2025**, devant le jury composé de :

Nadia YOUSFI STEINER, PROFESSEURE DES UNIVERSITES, Université Marie et Louis Pasteur	Présidente
Mathias GERARD, DIRECTEUR DE RECHERCHE, CEA centre de Grenoble	Directeur de thèse
Fei GAO, PROFESSEUR DES UNIVERSITES, Université de Technologie de Belfort Montbéliard	Rapporteur
Rachid OUTBIB, PROFESSEUR DES UNIVERSITES, Aix-Marseille Université	Rapporteur
Christophe BERENGUER, PROFESSEUR DES UNIVERSITES, Grenoble INP - UGA	Examinateur
Julien D'ARBIGNY, DOCTEUR EN SCIENCES, ALSTOM	Examinateur
Dani JURICIC, FULL PROFESSOR, Univerza v Ljubljani	Examinateur

Invités :

Marian FAUR
INGENIEUR SYSTEME, Genvia

Acknowledgments

A PhD is a long journey, and writing the thesis is the last demanding chapter. Yet here I am, at the end of it, feeling both exhausted and fulfilled, aware of the effort it required, and grateful for the experiences, the lessons, and the people who shaped this path.

First, I would like to thank the DuraSyS-PAC project (ANR-22-PEHY-0002) for funding my research and giving me the opportunity to exchange ideas and collaborate with researchers across France.
My deepest gratitude goes to my thesis director, Mathias Gerard, for his constant guidance and support throughout these years, and to my supervisors Vincent Heiries, for his motivation and efforts in creating a positive environment, and Sébastien Rosini, for his valuable explanations and patience.

I am sincerely grateful to the *Laboratorie Architecture Électrique et Hybridation* in CEA LITEN for hosting me, for the convivial moments and for helping me to improve my French, to the point where I can finally formulate few coherent sentences. Special thanks to my office mates, Adrien and Baptiste, for the psychological support, occasional arguments, and philosophical debates about absolutely everything.
My thanks also go to all the people who supported my research with their expertise. To Guillaume Serre, for welcoming me and introducing me to the world of MePHYSTO. And to Maxime Piffard, for the dedication and patience in accompanying my work.
A special thank goes to the *Laboratoire Signaux et Systèmes de Capteurs* in CEA LETI, which adopted me during these three years. A sincere thank-you to Raphaël, both a colleague and a friend, always available with his kindness to provide sincere opinions and advices. I am also grateful to all the members of the *Physics-Informed Learning Machine – Predictive Maintenance (Tiger) team*: you were a constant presence, a source of ideas, feedback, and discussions. In particular, a big thank you to Youssof, Misael, Quentin, Félix and Pascal.

Thank you to all the essential people in my life, who turned these challenging years into some of the happiest and most surprising ones. To my friends in Grenoble, my second family: Paolo, Marco, Fede, Nic and Euge. Thank you for sharing daily and night-life, and thesis struggles. Ilaria, my mountain buddy and half flatmate, thank you for bringing me back to reality whenever I got lost in overthinking. To Lucia, my sunshine in the greyness of this city. Thank you for being there for a beer, a workout, or a long talk, always listening without judgment. To Antonio, for always spreading joy and love around you, and for introducing me to pizzica.
To my friends far away, who are a constant support and reason of happiness in my life. To Anna Lisa and Irene, thank you for being an endless source of funny moments and for encouraging me to see things differently with your inspiring perspectives. To my lifelong friends, Carlotta and Lucia, for your daily reminder that there is a world outside research. To Anna, for being like a sister to me since our university days, and for continuing to be no matter the distance between us. Thank you Mery for always finding ways to cheer me up, for sharing incredible (and exaggerated) stories which can make me cry and laugh at the same time. Your creative and artistic soul will always be an inspiration.

Finally, to my mum and dad, and my brother Matteo: thank you for listening to my complaints, for encouraging me, and for always being the place where I can recover energies when I feel overwhelmed by life.

Chiara
Grenoble, October 2025

Abstract

Fuel cells are electrochemical converters with a pivotal role in decarbonizing critical systems and transportation through hydrogen, with particular relevance to heavy-duty applications. The combination of chemical and physical complex processes within fuel cells inevitably leads, over the system lifetime, to faults and degradation, jeopardizing durability, reliability, and ultimately large-scale deployment of the technology. While research in materials science has delivered advancements in the past, the next major technological breakthroughs are expected from improved operation and maintenance, driving growing interest in the integration of prognosis and diagnosis strategies into fuel cell management systems. In this context, the present PhD thesis focuses on developing fault diagnosis approaches aiming at: i) accurately detecting and identifying faults in fuel cells for automotive applications, under dynamic loads, where sensors availability is limited and fast inference is required; ii) ensuring algorithm robustness with respect to system degradation and aging throughout the fuel cell lifetime; iii) exploiting all the available sources of information, including detailed physics-based models and experimental data. The main novelty of this work lies in leveraging deep learning for automatic feature extraction and for the analysis of dynamic monitoring data, while informing the fault diagnosis algorithm with prior physics knowledge. This integration not only enables a more accurate approach to fault diagnosis but also reduces reliance on the availability of large datasets and highly detailed physics-based models. The effectiveness of the proposed methods is tested on a real experimental dataset representing over 1,000 hours of fuel-cell operation, capturing realistic fault scenarios under aging conditions. The PhD work demonstrates that embedding physics into deep learning algorithms allows relying solely on existing onboard sensors, decreases the need for costly experimental campaigns to collect datasets, and enhances the robustness of fault diagnosis to aging. Hence, the proposed framework paves the way for fault-tolerant and aging-resilient control in fuel cells.

Résumé

Les piles à combustible à hydrogène sont des convertisseurs électrochimiques jouant un rôle clé dans la décarbonation des systèmes énergétiques et en particulier dans le transport, avec une pertinence toute particulière pour les applications lourdes. La combinaison des processus physiques et chimiques au sein des piles à combustible entraîne inévitablement, au cours de la durée de vie du système, à des défauts et des dégradations. Ces dernières menacent leur durabilité, leur fiabilité et, in fine, le déploiement à grande échelle de cette technologie. La recherche sur les matériaux a réalisé des progrès significatifs ces dernières années, permettant d'améliorer la durabilité. Les axes de progrès sur le contrôle et la maintenance des systèmes piles permettront de compléter l'atteinte des objectifs de durabilité et fiabilité attendus. Il existe donc un intérêt grandissant pour l'intégration de stratégies de diagnostic et de pronostic dans les systèmes de gestion des piles à combustible. Dans ce contexte, cette thèse se concentre sur le développement d'approches de diagnostic des défauts visant à : i) détecter et identifier avec précision les défauts dans les piles à combustible pour applications automobiles, sous charges dynamiques, où la disponibilité des capteurs est limitée et la prise de décision en temps réel requise ; ii) assurer la robustesse des algorithmes face à la dégradation et au vieillissement du système tout au long de la durée de vie des piles ; iii) exploiter toutes les sources d'information disponibles, y compris les modèles physiques détaillés et les données expérimentales. L'originalité de ce travail réside dans l'utilisation d'algorithmes d'apprentissage profond pour automatiquement analyser et extraire des informations au sein de données de surveillance dynamique, en intégrant des connaissances physiques du système. Cette intégration permet non seulement une approche plus précise du diagnostic, mais réduit également le besoin d'un grand nombre de données expérimentales et de modèles physiques très détaillés. L'efficacité des méthodes proposées est démontrée sur un jeu de données expérimental réel représentant plus de 1000 heures de fonctionnement d'une pile, ce qui permet de capturer des scénarios de défauts dans des conditions de vieillissement réalistes. Les travaux de cette thèse démontrent que l'intégration de connaissances physiques du système au sein d'algorithmes d'apprentissage profond permet de s'appuyer uniquement sur les capteurs embarqués existants, de réduire le besoin de campagnes expérimentales coûteuses pour collecter des données et d'améliorer la robustesse du diagnostic des défauts face au vieillissement. Ainsi, l'approche proposée ouvre la voie à un contrôle tolérant aux défauts et résilient au vieillissement des piles à combustible.

Résumé Étendu

Contexte

La demande énergétique mondiale continuera d'augmenter au cours des prochaines décennies, avec une augmentation correspondante des émissions de carbone si les combustibles fossiles restent la source d'énergie dominante. En particulier, le secteur des transports est à lui seul responsable de plus d'un tiers des émissions totales de CO_2, et la réduction de ces émissions est difficile. L'hydrogène est apparu comme une alternative propre et prometteuse aux carburants conventionnels. Dans ce contexte, les piles à combustible sont considérées comme un levier clé de la décarbonation. Elles représentent une solution à faible émission de carbone pour répondre à la forte demande énergétique, à la haute densité d'énergie gravimétrique et au ravitaillement très rapide du transport routier lourd et du transport maritime. Cependant, la pénétration à grande échelle des technologies de piles à combustible dans le marché est entravée par deux barrières persistantes : la limitée durée de vie des stacks et le coût élevé. En pratique, la durabilité des piles à combustible, bien qu'étant cruciale pour le transport, est actuellement inférieure aux objectifs requis, particulièrement dans les applications de transport lourd. La combinaison de processus chimiques et physiques complexes conduit inévitablement, au cours de la durée de vie du système, à des défauts et à la dégradation des piles à combustible, compromettant la durabilité, la fiabilité et le déploiement à grande échelle de la technologie. Cela indique un besoin d'améliorations considérables. Afin de prolonger la durabilité des piles à combustible, une solution viable et largement soutenue est le développement d'un fonctionnement avancé, accompagné de stratégies de maintenance prédictive et intelligente. En effet, le pronostic et la gestion de l'état de santé (en anglais *Prognostic and Health Management* (PHM)), qui permettent de diagnostiquer les défauts et d'adapter le fonctionnement du système au vieillissement, sont l'un des piliers de l'industrie moderne. Ils peuvent aider à préserver des performances optimales et à atténuer la dégradation avec un coût limité, constituent une étape vers des piles à combustible plus durables. Cette thèse contribue à ce domaine de recherche fondamental en développant des solutions de diagnostic des défauts robustes et fiables pour les piles à combustible. En détectant et en isolant rapidement les défauts dans le système, le diagnostic permet des actions correctives rapides et soutient des décisions de commande éclairées pour prolonger la durabilité du système.

Enjeux majeurs

Le diagnostic de piles à combustible à membrane échangeuse de protons (PEMFC) a traditionnellement été réalisé à l'aide de mesures de haute précision et de capteurs dédiés, notamment la spectroscopie d'impédance électrochimique, qui garantit une compréhension approfondie des processus électrochimiques internes au stack. Bien qu'ils fournissent des informations pertinentes, le déploiement continu de ces outils pour le diagnostic dans les applications automobiles embarquées est difficilement disponible. À l'inverse, il serait très avantageux de s'appuyer principalement sur des capteurs déjà intégrés dans le stack. Cela pourrait également entraîner une réduction des coûts associés au déploiement du diagnostic et atténuer les impacts négatifs sur la fiabilité causés par les défaillances des capteurs. Cependant, l'interprétation des informations fournies par ces capteurs n'est pas simple. Plusieurs effets se chevauchent dans les signaux de surveillance conventionnels, en particulier dans des conditions de charge dynamique, ce qui rend extrêmement difficile la détection d'un défaut et l'isolation de son origine. Pour qu'un diagnostic fiable soit réalisé en utilisant uniquement des signaux embarqués, il est essentiel de développer des méthodes capables d'extraire et de capturer avec précision les informations latentes contenues dans les données de surveillance. Pour distinguer correctement les différentes signatures des défauts, les dépendances non linéaires et les tendances entre les combinaisons de signaux doivent être analysées. Cela permettrait de reduire au maximum l'instrumentation du système.
Au cours des dernières années, un effort substantiel s'est concentré sur le développement d'algorithmes pour le

diagnostic des piles à combustible, avec une nette orientation vers les méthodes basées données, motivé par le grand succès de l'intelligence artificielle. En partant des méthodes conventionnelles d'apprentissage automatique (machine learning), les méthodes d'apprentissage profond (deep learning) ont progressivement été largement adoptées en raison de leurs fortes capacités à capturer les dynamiques non linéaires et les dépendances dans les données. Néanmoins, en raison de la complexité de ces algorithmes et du grand nombre de paramètres sur lesquels ils sont construits, les méthodes d'apprentissage profond sont extrêmement exigeant en données. Dans le secteur des piles à combustible, où la technologie n'est pas encore totalement mature pour un déploiement sur le marché, l'acquisition de grands ensembles de données est particulièrement difficile, et les industriels sont réticents à partager publiquement leurs données. Cela limite la disponibilité de vastes et représentatifs ensembles de données nécessaires pour construire des solutions d'apprentissage profond. Dans les scénarios d'application réelle, les données disponibles sont de petite taille et ne contiennent qu'un petit nombre d'échantillons des défauts représentatifs. Pour exploiter les solutions d'apprentissage profond lorsque de grands ensembles de données ne sont pas disponibles, une solution prometteuse consiste à recourir aux connaissances physiques. En effet, l'intégration de modèles basés sur la physique dans l'apprentissage profond peut être bénéfique pour réduire les besoins en données, tout en maintenant une haute généralisation des algorithmes et en renforçant leur robustesse. Cependant, aucune méthodologie universelle n'existe pour cela, et chaque domaine technologique et contexte d'application nécessite une investigation dédiée. Ainsi, différentes méthodes doivent être explorées pour combiner efficacement les connaissances intégrées dans un modèle physique de pile à combustible avec l'apprentissage profond, afin de fournir des solutions de diagnostic robustes.
Enfin, les piles à combustible subissent intrinsèquement une dégradation, provoquant une dérive de leur comportement de fonctionnement et de leurs performances au fil du temps. La détérioration de la pile à combustible est étroitement liée à la perte de fiabilité du diagnostic du système. Ce problème est remarquablement courant, mais il est souvent négligé dans les stratégies de surveillance et nécessite une prise en compte directe. Le vieillissement du système entraîne une dérive dans les signaux mesurés, ce qui peut compromettre la validité à long terme des algorithmes. Cela implique que, lorsque la pile à combustible se dégrade, les défauts au sein du stack deviennent de plus en plus difficiles à détecter. Garantir qu'un algorithme de diagnostic reste fiable tout au long de la durée de vie du système est crucial pour éviter un diagnostic des défauts incorrect, ce qui pourrait réduire davantage la durabilité du système. Actuellement, le domain des piles à combustible manque de méthodologies capables de réaliser un diagnostic précis tout en limitant les coûts associés et en compensant la faible disponibilité des données. Dans cette thèse, nous abordons directement ces défis en proposant de tirer parti d'une approche hybride qui intègre un modèle physique avec l'apprentissage profond, fournissant une base pour des stratégies de surveillance plus fiables et robustes au vieillissement.

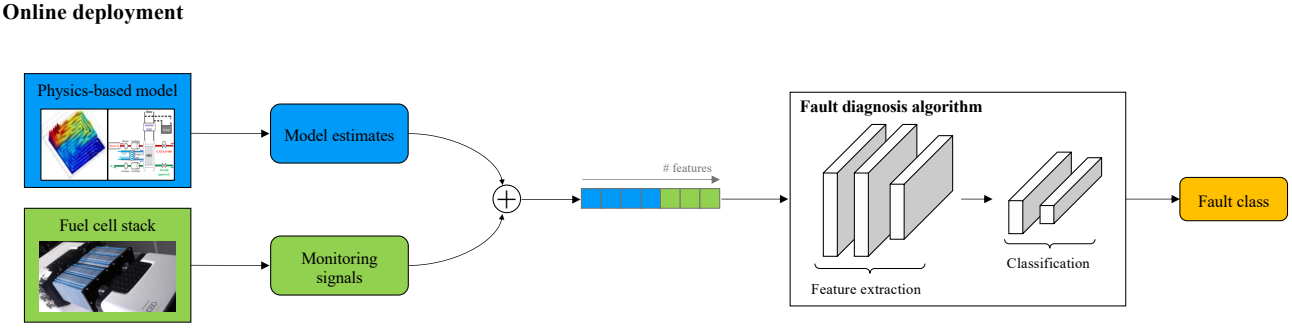
Méthodologie

Il existe trois catégories de méthodes de diagnostic : basées modèle, basées données, ou une combinaison des deux (hybride). Chaque approche présente ses propres forces et limites. Les méthodes basées sur un modèle bénéficient de l'interprétabilité physique et de la généralisation, mais peinent souvent à capturer le comportement réel du système en raison des limites liées à le calage du modèle et des hypothèses simplificatrices. Les approches basées données, en revanche, peuvent apprendre les characteristiques des défauts directement à partir de données expérimentales, mais nécessitent des ensembles de données vastes et de haute qualité pour être entraînées. Dans ce travail, nous adoptons un approche hybride qui tire parti à la fois de la modélisation basée sur la physique et des données expérimentales, afin de bénéficier des avantages des deux tout en compensant leurs limites. Le modèle utilisé dans ce travail s'appelle MePHSYTO-FC. Il s'agit d'un modèle multiphysique des PEMFC basé sur les lois physiques du transport des réactifs et des produits de réaction, du transfert de masse et de chaleur, et sur les lois électrochimiques. L'ensemble de données expérimentales des défauts disponible pour cette étude est particulièrement précieux, car il comprend une large gamme des défauts, également dans des conditions de vieillissement et sous charge dynamique, eprésentant plus de 1000 heures de fonctionnement d'une pile. Ainsi, il permet une compréhension plus approfondie de la manière dont les défauts influencent les performances des

PEMFC, de la manière dont le vieillissement affecte le comportement du système, et de la manière dont ces deux aspects se combinent. L'objectif de cette recherche est d'explorer des méthodes capables d'intégrer efficacement les connaissances physiques avec l'apprentissage profond pour répondre aux limites de l'état de l'art dans le diagnostic des défauts des piles à combustible. L'incorporation de connaissances physiques dans les algorithmes offre plusieurs avantages, notamment une meilleure généralisation aux conditions inédites du système. Bien qu'un modèle physique puisse être utilisé pour effectuer directement le diagnostic des défauts, son hybridation avec des algorithmes basés sur les données est souvent privilégiée lorsque l'erreur de modélisation est importante. À l'inverse, un modèle physique peut garantir que les algorithmes respectent la physique sous-jacente du système, car les lois physiques peuvent limiter la région de solutions réalisables. Dans ce travail, nous visons à démontrer les avantages de la combinaison des connaissances intégrées dans un modèle physique avec un algorithme d'apprentissage automatique pour le diagnostic des défauts des PEMFC. Le modèle physique peut produire des estimations qui ne sont pas forcément représentatives avec précision du système PEMFC réel, mais qui peuvent bien se généraliser à différentes conditions de fonctionnement, et les données de surveillance en temps réel peuvent aider à compenser la précision limitée des estimations du modèle.

Diagnostic des défauts guidé par la physique

Bien que plusieurs stratégies puissent être utilisées pour hybrider physique et données, nous proposons d'abord d'améliorer les informations d'entrée de l'algorithme de diagnostic en ajoutant aux signaux mesurés, collectés via des capteurs sur le stack, des estimations observable et inobservable de la dynamique du système fournies par le modèle physique. Le modèle physique est utilisé en parallèle du système PEMFC réel. Cependant, au lieu d'effectuer un diagnostic des défauts basé sur l'analyse des résidus comme dans les méthodes basées modèle traditionnelles, ce qui nécessite un modèle très précis, les estimations du modèle et les signaux mesurés dans le stack sont agrégés ensemble. Ce faisant, le modèle physique offre un accès à de nouvelles variables de processus non mesurées qui fournissent une meilleure description de l'état interne des systèmes. En conséquence, la capacité de l'algorithme à distinguer les différents défauts dans le stack peut être améliorée. Entre autres, le modèle physique peut fournir des estimations de la résistance de la membrane, de la teneur en eau de la membrane et de la densité de courant. Ces connaissances supplémentaires peuvent aider à démêler l'effet des défauts de celui du vieillissement, qui se cumule sur les signaux de surveillance. Les signaux mesurés sont souvent limités en nombre et affectés par le bruit, ce qui rend difficile pour les algorithmes uniquement basés données de distinguer de manière fiable les défauts des effets du vieillissement. En combinant les estimations de variables de processus non mesurées avec les signaux de surveillance directement mesurés sur le stack, l'approche permet un diagnostic des défauts plus précis sans nécessiter de capteurs physiques supplémentaires.

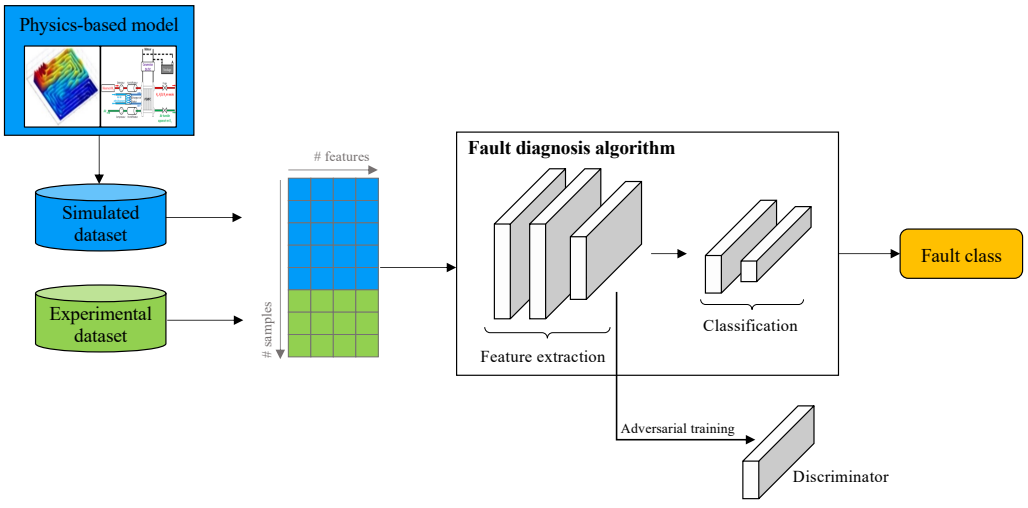


Physics-guided fault diagnosis (chapitre 4).

Transfert de connaissances issues de modèles physiques par adaptation de domaine

Une stratégie différente qui ne nécessite pas le déploiement en ligne d'un modèle physique est ensuite proposée. En effet, dans les applications automobiles embarquées, cela peut représenter un défi, car les ressources de calcul sont limitées et une inférence rapide est requise. Par conséquent, nous explorons comment les connaissances détaillées intégrées dans un modèle physique peuvent être utilisées hors ligne pour soutenir le développement d'un diagnostic des défauts, et comment elles peuvent soulager les exigences en données typiquement associées à l'entraînement des algorithmes d'apprentissage profond. La taille de données est souvent insuffisante, et il est difficile d'apprendre des signatures caractéristiques du défaut et de développer des algorithmes efficaces avec un bon niveau de généralisation. Un manque de représentativité dans l'ensemble de données d'entraînement peut affecter considérablement les performances des algorithmes sur les systèmes en fonctionnement réel. En apprentissage automatique, l'intégration de modèles physique dans les algorithmes de machine learning est apparue comme une solution prometteuse à ce problème. Plus précisément, un modèle de simulation peut efficacement fournir des quantités illimitées de données labelisées, reproduisant de variations des conditions de fonctionnement. Néanmoins, les ensembles de données simulées ne peuvent pas être directement utilisés pour entraîner des algorithmes destinés à un déploiement dans le monde réel, car les modèles sont affectés par des erreurs. Ces erreurs provoquent un décalage de distribution entre les données simulées et les données mesureées. Une solution prometteuse à ce problème est offerte par les méthodes d'adaptation de domaine. En utilisant l'adaptation de domaine, un transfert efficace de connaissances du modèle physique vers un réseau de neurones profonds est possible. Cela est fait pendant l'entraînement hors ligne de l'algorithme. Ainsi, nous proposons une méthodologie qui fusionne profondément les connaissances du modèle physique lors de l'entraînement hors ligne de l'algorithme d'apprentissage profond, visant à surmonter la rareté des données dans le diagnostic des défauts des PEMFC. Plus précisément, nous développons un réseau antagoniste supervisé (domani adversarial neural network) pour réduire correctement l'écart entre les données simulées et réelles, tout en parvenant à une correcte identification des défauts dans la PEMFC. Cette stratégie vise à réduire la dépendance des grands ensembles de données expérimentales et améliore la généralisation de l'algorithme.

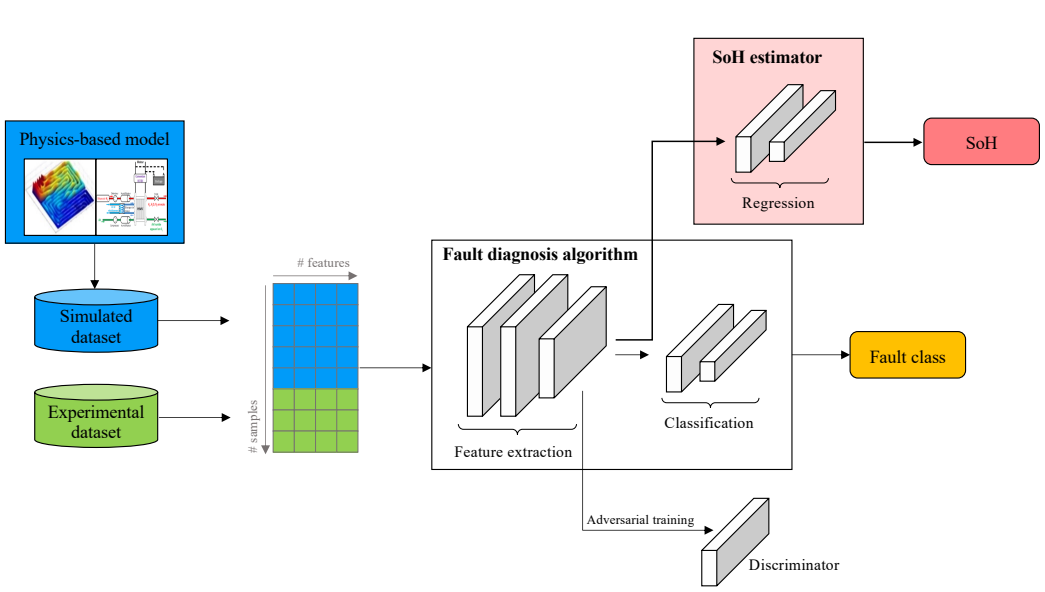
Offline training



Domain adversarial neural network (chapitre 5).

Approche intégrée pour le diagnostic et l'estimation de l'état de santé

Dans de nombreuses applications de surveillance du monde réel, les mécanismes de défauts et la dégradation du système ne sont pas des phénomènes indépendants, mais strictement liés. Cela affecte également les systèmes PEMFC, dans lesquels le vieillissement impacte non seulement les performances du stack mais aussi la manière dont les défauts se manifestent et évoluent. Le diagnostic se concentre sur la détection et la classification des conditions de fonctionnement anormales, tandis que l'estimation de l'état de santé fournit une mesure de la dégradation du système. Ces deux tâches ont souvent été abordées indépendamment, négligeant le couplage inhérent entre elles. En effet, la dégradation physique du stack altère les performances du système et, par conséquent, les signaux utilisés pour surveiller le système.
Sur cette base, nous proposons une nouvelle méthodologie visant à intégrer le diagnostic des défauts et l'estimation de l'état de santé au sein d'une architecture d'apprentissage profond. Contrairement au pronostic, notre objectif n'est pas de prévoir la durée de vie utile restante ou de prédire la perte de tension future du stack, mais plutôt de fournir une estimation précise de l'état de dégradation actuel. Cette information est précieuse en soi, car elle soutient une prise de décision plus éclairée pour l'interprétation des défauts, la planification de la maintenance et les stratégies potentielles de contrôle correctif. L'approche proposée consiste en un réseau de neurones profonds multitâche qui exploite des objectifs complémentaires pour extraire des informations plus pertinentes à partir des signaux de surveillance disponibles. L'apprentissage multitâche est un paradigme d'apprentissage automatique conçu pour résoudre simultanément plusieurs tâches. Le réseau multitâche que nous présentons est capable d'aborder conjointement le diagnostic des défauts et l'estimation de l'état de santè, en imposant l'extraction de caractéristiques pertinentes pour les défauts ainsi que pour la dégradation à partir des signaux de surveillance.



Multitask domain adversarial neural network (chapitre 6).

Résultats principaux

Cette thèse traite de l'utilisation de l'apprentissage profond guidé par la physique (physics-informed deep learning) pour faire progresser le diagnostic des défauts des piles à combustible et répondre aux contraintes majeures du domaine. Le travail démontre que guider un algorithme d'apprentissage profond avec des connaissances physiques améliore la discrimination des défauts dans la pile à combustible. L'augmentation de la robustesse et de la capacité de généralisation de la méthode guidée par la physique se traduit par une augmentation de la précision

de 55,9 % lors d'un test sur une pile à combustible ayant subi une dégradation, par rapport à une méthode purement basée données. De plus, en utilisant le modèle physique pour obtenir un ensemble de données simulées qui peut assister l'entraînement des algorithmes d'apprentissage profond, nous permettons une disponibilité accrue des données. En recourant à l'adaptation de domaine, nous réduisons considérablement les exigences en données expérimentales pour l'entraînement des algorithmes, tout en maintenant une excellente précision de diagnostic, même quand la dégradation est observée dans la pile à combustible. C'est une réussite majeure car elle réduit la charge pratique de la collecte de données et le coût associé aux longues campagnes expérimentales. Enfin, partant du principe que le diagnostic des défauts et l'estimation de la santé sont des tâches complémentaires, nous proposons une nouvelle architecture impliquant un réseau multitâche. Nous démontrons que cette approche peut atteindre des performances élevées sur la détection des défauts ainsi que sur l'estimation de l'état de santé, sans rencontrer de transfert négatif. Cela représente une étape vers une approche de diagnostic des défauts et de surveillance de la santé tenant compte du vieillissement pour les systèmes piles à combustible.
Ces conclusions majeures permettent de prouver qu'en utilisant des algorithmes d'apprentissage profond dédiés à l'analyse de données temporelles, il est possible de réaliser une extraction automatique de caractéristiques à partir des signaux de surveillance embarqués traditionnels, facilement disponibles dans les applications automobiles, sans recourir à des techniques de mesures dédiées ou à des procédures d'extraction de caractéristiques ad hoc. Néanmoins, la faible disponibilité de données pour les technologies de piles à combustible, particulièrement dans des conditions de défauts et de dégradation du stack, peut être atténuée par l'utilisation d'un modèle physique. En conclusion, ce travail démontre qu'un diagnostic très précis peut être obtenu grâce à des algorithmes d'apprentissage profond performants, même avec une quantité très limitée de données réelles labellisées. Cela réduit considérablement la charge pratique liée à la collecte de données.

Contents

Acknowledgments **i**

Abstract **iii**

Résumé Étendu **v**

List of Figures **xvii**

List of Tables **xx**

1 Introduction **1**
1.1 Context and motivation . . . 2
1.2 Scientific challenges . . . 2
1.3 Contributions of the thesis . . . 3

2 Background on PEMFC and Fault Diagnosis **5**
2.1 Introduction . . . 6
2.2 PEMFC technology . . . 7
2.2.1 Principle . . . 7
2.2.2 PEMFC durability . . . 8
2.2.3 Degradation mechanisms . . . 9
2.2.4 Faults in PEMFC . . . 11
2.3 Health monitoring and fault diagnosis: state of the art in fuel cells . . . 12
2.3.1 Model-based methods . . . 13
2.3.2 Non-model-based methods . . . 15
2.3.3 Hybrid methods . . . 18
2.3.4 Summary . . . 20
2.4 Signals to diagnose the stack . . . 20
2.5 Impact of aging in PEMFC fault diagnosis . . . 21

3 Dataset and Physics-based Model **23**
3.1 Introduction . . . 24
3.2 The experimental faulty dataset . . . 24
3.2.1 Analytical overview of the dataset . . . 28
3.3 Physics-based model . . . 36
3.3.1 Description . . . 36
3.3.2 Calibration . . . 38
3.3.3 Faults representation using the model . . . 39
3.3.4 Insights on modeling error . . . 42
3.4 Data-preprocessing: instructions and details . . . 43
3.5 Summary . . . 45

4 Guiding Machine Learning Algorithms with Physics Knowledge for PEMFC fault diagnosis 47
4.1 Introduction . . . 49
4.2 Background on time series analysis . . . 50
4.2.1 Time series . . . 50
4.2.2 Time series classification . . . 50
4.2.3 Recurrent Neural Network . . . 51
4.3 The methodology . . . 53
4.3.1 Description . . . 53
4.3.2 Diagnosis variables . . . 55
4.4 Proposed algorithm architectures and implementation details . . . 56
4.4.1 Physics-based model deployment . . . 57
4.4.2 Algorithms architecture . . . 57
4.4.3 Data preparation . . . 58
4.4.4 Performance evaluation metrics . . . 58
4.5 Results and performance analysis . . . 58
4.5.1 Validation on the initial phase of stack operation . . . 59
4.5.2 Validation on the entire dataset, with the influence of aging . . . 60
4.5.3 Assessment of the methodology robustness with unseen aging data . . . 61
4.5.4 Discussion . . . 63
4.6 Summary . . . 64

5 Transfer Information from Physics Models via Domain Adaptation 67
5.1 Introduction . . . 69
5.2 Background on domain adaptation . . . 70
5.2.1 Domain adaptation . . . 70
5.2.2 Domain adversarial neural network . . . 72
5.2.3 Open issues . . . 72
5.3 The methodology . . . 73
5.3.1 Problem description . . . 74
5.3.2 Supervised DANN . . . 74
5.3.3 Comprehensive framework . . . 76
5.4 Implementation details and results . . . 77
5.4.1 Physics-based simulated dataset . . . 77
5.4.2 Insights on the sim-to-real transfer . . . 78
5.4.3 Details on the source and target domain datasets . . . 81
5.4.4 Results for the supervised DANN . . . 82
5.5 Summary . . . 86

6 An Integrated Framework for Fault Diagnosis and SoH estimation 89
6.1 Introduction . . . 91
6.2 Background on multitask learning . . . 91
6.3 SoH identification in PEMFC . . . 92
6.3.1 SoH estimation: reference definition and faulty dataset annotation . . . 93
6.4 The multitask methodology . . . 95
6.4.1 Problem statement . . . 95
6.4.2 The multitask DANN network . . . 96
6.5 Implementation details and results . . . 97
6.5.1 Datasets . . . 97
6.5.2 Hyperparameter settings and multitask DANN training . . . 97

6.5.3 Evaluation metrics 98
6.5.4 Performance evaluation 99
6.6 Summary 101

7 Conclusions **103**
7.1 Scientific contributions 104
7.2 Key findings 105
7.3 Limitations and future perspectives 105

List of Publications **107**

Acronyms **110**

Bibliography **111**

List of Figures

2.1 Fuel cell electric vehicles stock by type, 2019-2025. 2025* Includes data from January to July [13]. 6
2.2 Cost of fuel cell systems [15]. 7
2.3 Schematic of the PEMFC. 8
2.4 Polarization curve. 8
2.5 Performance level criteria: voltage loss, maximum power, efficiency loss. In blue are illustrated the performances that meet the requirements, at BoL; in red the performances at EoL. 9
2.6 PEMFC degradation and consequences (adapted from [27]). 10
2.7 Schematic of the impact of faults and their effect on PEMFC degradation (adapted from [33]). 11
2.8 PHM framework [8]. 12
2.9 Categorization of fault diagnosis methods. 13
2.10 Schematic of model-based fault diagnosis. 14
2.11 Data-driven fault diagnosis [44]. 16

3.1 Pacific test bench at CEA [165]. 24
3.2 FC-DLC automotive load cycle profile. 25
3.3 Current load and stack voltage recorded during the entire experimental campaign on the PEMFC stack. 26
3.4 Load current during the conditioning and characterization phase of the PEMFC stack. 26
3.5 Distribution of the faults across the entire experimental campaign. 28
3.6 Number of occurrences for the different fault classes: nominal (0), flooding (1), drying (2), air starvation (3), hydrogen starvation (4). 28
3.7 Distribution of flooding fault class, polarization curve (top) and FC-DLC (bottom). The injected fault is shown in blue. 29
3.8 Distribution of drying fault class, polarization curve (top) and FC-DLC (bottom). The injected fault is shown in blue. 29
3.9 Distribution of air starvation fault class, polarization curve (top) and FC-DLC (bottom). The injected fault is shown in blue. 30
3.10 Distribution of H2 starvation fault class, polarization curve (top) and FC-DLC (bottom). The injected fault is shown in blue. 30
3.11 Polarization curve at stack beginning of life, for nominal condition (green) and faults: flooding (blue), drying (yellow), air starvation (red), hydrogen starvation (purple). 31
3.12 Nyquist diagram of the EIS measurements, at 0.5 A/cm2 current density, for nominal condition (green) and faults: flooding (blue), drying (yellow), air starvation (red), hydrogen starvation (purple). 31
3.13 Kernel density estimation (KDE) plots for voltage, anode and cathode pressure drop, for the different fault classes. The following current values, with a tolerance of ± 5 A, are considered: 55 A, 110 A, 140 A, 185 A. 33
3.14 Polarization curves acquired at different operating time (nominal conditions). 34
3.15 Stack power vs stack current, acquired at different operation times (nominal conditions). 34
3.16 Kernel density estimation (KDE) plots for voltage, anode and cathode pressure drop, for the different fault classes available in the dataset, at beginning-of-life and end-of-life. The following current values, with a tolerance of ± 5 A, are considered: 55 A, 110 A, 140 A, 185 A. 35
3.17 Schematic of the physics-based PEMFC model (MePHYSTO). 37

3.18 Flooding injection: coolant temperature varies from 80°C to 60°C. Cell voltage (a), liquid saturation at electrodes (b), membrane water content (c). 40
3.19 Drying injection: inlet gases relative humidity drops to 0. Cell voltage (a), membrane water content (b), membrane resistivity (c). 40
3.20 Air starvation injection: air stoichiometry drops from 2 to 1.5. Cell voltage (a), liquid saturation (b), current density distribution along the channel, before and after fault injection (c). 41
3.21 Hydrogen starvation injection: hydrogen stoichiometry drops from 1.5 to 1.1. Cell voltage (a), membrane resistivity (b), current density distribution along the channel, before and after fault injection (c). 41
3.22 PEMFC response for a segment of the FC-DLC cycle: voltage, anode and cathode pressure drop. 42
3.23 Train-validation-test split (0.7-0.1-0.2) of a polarization curve, without dedicated procedure. . 44
3.24 Dedicated train-validation-test split procedure. 44
3.25 Train-validation-test split (0.7-0.1-0.2) of a polarization curve, with the dedicated procedure. . 45
3.26 Illustration of the sliding window technique. 45

4.1 **Physics-guided fault diagnosis.** The physics-based model provides estimates of observable and non-observable variables of the fuel cell stack. This knowledge is combined with the monitoring signals and guides a deep learning algorithm to correctly classify faults within the fuel cell. . . 47
4.2 Schematic of feed-forward network and recurrent neural network. 51
4.3 Recurrent neuron (left), unfolded through time (right) [192]. 52
4.4 LSTM cell structure. 52
4.5 Flowchart of the proposed physics-guided LSTM method. 54
4.6 Simplified illustration of the dataset splitting (see train-validation-test split in section 3.4) for the validation on the initial phase of stack operation. 59
4.7 True (unfilled) vs predicted (filled) fault state of the PEMFC stack, showing an air starvation fault (left) and a drying fault (right). The physics-guided LSTM is used in the first 250 hours of the dataset. 60
4.8 Simplified illustration for the dataset splitting (see train-validation-test split in section 3.4) for the validation on the entire dataset. 60
4.9 True (unfilled) vs predicted (filled) fault state of the PEMFC stack, showing drying fault (left) and a flooding fault (right). The physics-guided LSTM is used on the entire dataset. 61
4.10 Illustration of the dataset splitting for assessing the robustness with unseen aging data. 61
4.11 Confusion matrix for the data-driven LSTM (left) and the physics-guided LSTM (right) method. The train and test datasets are illustrated in Figure 4.10. 62
4.12 Two-dimensional representation by means of t-SNE of the features extracted by the LSTM neural network, (left) data-driven LSTM, (right) physics-guided LSTM. The dataset is shown in Figure 4.10 63

5.1 **Domain adversarial neural network.** The physics-based model is used offline to generate a physics-relevant, simulated dataset. The information embedded in this dataset is then used to build a deep learning algorithm via adversarial training, together with the experimental data available. At testing time, only the monitoring signals measured on the fuel cell stack are used to diagnose the system. 67
5.2 Domain adaptation: illustration of the distribution shift between source and target domain. . . 71
5.3 Scheme of unsupervised DANN (adapted from [221]). 72
5.4 Scheme of supervised DANN: feature extractor G_f, fault classifier G_y and domain discriminator G_d. A RNN network is used in G_f to effectively process the multivariate time series. 75
5.5 Flowchart of the proposed methodology. 77

5.6 Kernel density estimation (KDE) plots of voltage, anode and cathode pressure drop, for a simulated average current load of 0.5 A/cm^2, with added noise. The nominal class and the four fault classes are represented. 78
5.7 t-SNE visualization of the input space for source (simulated, light-blue) and target (experimental, orange) domain. For the target domain, we use only the first 100 hours of stack operation from the experimental dataset. 79
5.8 Confusion matrices in the case of no adaptation i) for the source (a) and target (b) domain. . . 79
5.9 t-SNE visualization of the features extracted from source $G_f(X_s)$ and target $G_f(X_t)$ dataset with the unsupervised DANN, case ii). Left: source and target data are shown in light blue and orange, respectively. Right: features are colored according to the fault classes: nominal (green), flooding (blue), drying samples (red), air starvation (cyan), hydrogen starvation (magenta). . . 80
5.10 Confusion matrices in the case of unsupervised DANN ii) for the source (a) and target (b) domain. 80
5.11 t-SNE visualization of the features extracted from source $G_f(X_s)$ and target $G_f(X_t)$ dataset with the supervised DANN, case iii). Left: source and target data are shown in light blue and orange, respectively. Right: features are colored according to the fault classes: nominal (green), flooding (blue), drying samples (red), air starvation (cyan), hydrogen starvation (magenta). . . 81
5.12 Confusion matrices in the case of supervised DANN iii) for the source (a) and target (b) domain. 81
5.13 Schematic representation of the three sub-datasets considered as the target domain. 82
5.14 Accuracy vs target dataset (sub-dataset 1) fraction for data-driven and supervised DANN algorithms. Different feature extractor networks are tested: GRU, LSTM, BiLSTM. 84
5.15 Accuracy of supervised DANN and data-driven algorithm vs number of samples for sub-dataset 1, sub-dataset 2, sub-dataset 3. The LSTM architecture is used. 85

6.1 **Multitask domain adversarial neural network.** Starting from the methodology proposed in chapter 5, we develop a multitask network to: *i*) perform fault diagnosis and *ii*) estimate the state of health of the fuel cell stack. The physics-based model is used to generate a simulated dataset that supports the training of the deep learning algorithm; at testing time, only the monitoring signals measured on the fuel cell stack are used to serve the twofold objective. 89
6.2 General architecture of multitask learning. 92
6.3 Relative voltage loss rate (RVLR) at $1A/cm^2$. 93
6.4 Derived empirical SoH model for the faulty dataset. 95
6.5 Multitask supervised DANN. 96
6.6 Training history of the multitask DANN. From top to bottom: $\mathcal{L}_y$ (task 1), $\mathcal{L}_h$ (task 2), $\mathcal{L}_d$ (discriminator). 98
6.7 Comparison of the accuracy for the data-driven, the supervised DANN, and the proposed multitask DANN algorithms. The fraction of the target dataset used for training is progressively reduced: 0.8, 0.5, 0.1, 0.05. 99
6.8 Illustration of the dataset partitioning. In dark gray, the interval of operating time used for training and validation; in green, the intervals used for evaluating algorithms' performance in test 1, test 2, and test 3. 100
6.9 Fault classification accuracy for t = 100 h (left) and t = 250 h (right), for tests illustrated in Figure 6.8. 101

List of Tables

2.1 Possible integration strategies for physics-based model (Phys. model) and machine learning (ML). The physics-based model is represented in light blue, the machine learning algorithm is represented dark blue. Adapted from [137]. . . . 19
2.2 Comparison of the advantages and disadvantages of model-based and data-driven fault diagnosis methods. . . . 20

3.1 Nominal operating conditions for the PEMFC stack used in the experimental campaign. . . . 25
3.2 Faults injection modes. . . . 26
3.3 Monitoring variables during the experimental campaign. . . . 27
3.4 Kullback-Leibler divergence between nominal and faults distribution, for each variable. . . . 32
3.5 Kullback-Leibler divergence between BoL and EoL, for each variable, for each fault. . . . 36
3.6 Set of β_k calibrated. . . . 39
3.7 PEMFC model error: stack voltage. . . . 42
3.8 PEMFC model error: anode pressure drop. . . . 43
3.9 PEMFC model error: cathode pressure drop. . . . 43

4.1 Model estimates from the physics-based PEMFC model. . . . 55
4.2 Measured signals from PEMFC stack. . . . 56
4.3 Actuators inferred signals. . . . 56
4.4 Comparison of the data-driven and physics-guided fault diagnosis methods, for the different algorithms, over the initial phase (250 h) of stack operation. . . . 59
4.5 Comparison of the data-driven and physics-guided fault diagnosis methods, for the different algorithms, over the 1000 h dataset. . . . 61
4.6 Comparison of the data-driven and physics-guided fault diagnosis methods, for the different algorithms, on the dataset shown in Figure 4.10. . . . 62

5.1 Number of samples per fault class, for the source dataset and the sub-datasets of the target domain. 82
5.2 Optimize algorithms architectures. Optuna is used to optimize the feature extractor, while the fault classifier and domain discriminator are empirically optimized. . . . 83
5.3 Accuracy results for sub-dataset 1, averaged over 10 independent training runs with different random seeds. We compare the supervised DANN method with the data-driven algorithms, for GRU, LSTM and BiLSTM as feature extractor networks. For each experiment, the best between DANN and data-driven is in bold. . . . 83
5.4 Accuracy results for sub-dataset 2, averaged over 10 independent training runs with different random seeds. We compare the supervised DANN method with the data-driven algorithms, for GRU, LSTM and BiLSTM as feature extractor networks. For each experiment, the best between DANN and data-driven is in bold. . . . 84
5.5 Accuracy results for sub-dataset 3, averaged over 10 independent training runs with different random seeds. We compare the supervised DANN method with the data-driven algorithms, for GRU, LSTM and BiLSTM as feature extractor networks. For each experiment, the best between DANN and data-driven is in bold. . . . 84
5.6 Accuracy results, averaged over 10 independent training runs with different random seeds. We compare the data-driven and supervised DANN BiLSTM with the optimized architecture, and the supervised DANN BiLSTM with the same G_f as the data driven. . . . 86

6.1 RVLR at 1 A/cm^2 for the polarization curves in nominal condition. The reference value is highlighted in blue, and in bold the polarization curves with monotonic trend. 94
6.2 Architectures of the two comparative algorithms and of the multitask DANN network. 97
6.3 Comparison of fault diagnosis performance in terms of accuracy and prediction confidence for the data-driven, the supervised DANN, and the proposed multitask DANN, for which the MAE for SoH estimation is also reported. A BiLSTM-based feature extractor is used. H_n is the normalized entropy. 99
6.4 Fault classification accuracy and SoH estimation performance for different dataset partitions, for the experimental testing protocol illustrated. Comparison between data-driven, supervised DANN and multitask DANN. H_n is the normalized entropy. 101

1

Introduction

Contents

1.1 Context and motivation **2**

1.2 Scientific challenges **2**

1.3 Contributions of the thesis **3**

1.1 Context and motivation

Global energy demand is projected to continue rising in the coming decades, with a corresponding increase in carbon emissions if fossil fuels remain the dominant energy source. In particular, the transportation sector alone is responsible for more than one-third of total CO_2 emissions, and decreasing these corresponding emissions is difficult. Hydrogen has emerged as a promising clean alternative to conventional fuels. In this context, fuel cells are seen as a key enabler of decarbonization, since they represent a low carbon solution to the high energy demand, high gravimetric energy density and very fast refueling of heavy-duty road transport and maritime shipping. However, the large-scale market penetration of fuel cell technologies is obstructed by two persistent barriers: limited lifetime of fuel cell stacks and high cost [1]. In practice, the durability of fuel cell systems, while being crucial for transportation, currently falls short of the required targets, particularly in the case of heavy-duty applications. This indicates a need for considerable improvements [2]. In the interest of extending fuel cells durability, a widely supported viable solution is the development of advanced operation, accompanied by smart and predictive maintenance strategies. Indeed, prognosis and health management (PHM) [3], which allows diagnosing incipient faults and adapting operation according to system degradation, is one of the pillar of modern fuel cell industry. It can help preserve optimal performance and mitigate degradation with limited cost, representing a step towards more durable fuel cell systems [4]. This PhD thesis aims at contributing to this fundamental research field by developing robust and reliable fuel cell fault diagnosis. By promptly detecting and isolating emerging faults in the system, fault diagnosis enables rapid correcting actions and supports informed control decisions to limit degradation.

1.2 Scientific challenges

Diagnosis of fuel cell systems has traditionally been assessed with high-fidelity measurements and dedicated sensors, including electrochemical impedance spectroscopy (EIS), which guarantees deep understanding of internal electrochemical processes in the stack [5]. Despite providing insightful information, the continuous deployment of such tools for onboard fault diagnosis in automotive applications is hardly available. Conversely, it would be highly advantageous to rely primarily on the few sensors already integrated in fuel cell stacks. This could entail also a reduction of the costs associated with the deployment of fault diagnosis, and mitigate negative impacts on reliability caused by sensors failures and degradation. However, interpreting the information provided by these sensors is not straightforward. Multiple effects often overlap in the conventional monitoring signals, especially under dynamic load conditions, making it extremely challenging to detect the occurrence of a fault and isolate its origin. Hence, the following question has to be addressed:

1. *Can accurate and reliable fault diagnosis be achieved using only on-board monitoring signals while keeping instrumentation minimal?*

To achieve this, it is essential to develop methods capable of accurately extracting and capturing the latent information contained in monitoring data. To correctly discriminate between different fault signatures, nonlinear dependencies and trends across combinations of signals have to be analyzed. In recent years substantial effort has focused on developing algorithms for fault diagnosis in fuel cells, with a pronounced shift toward data-driven methods motivated by the large success of artificial intelligence. Starting from conventional machine learning methods, deep learning methods have progressively gained widespread adoption due to strong capabilities in capturing non-linear dynamics and dependencies in the data. Thus, they are effective in analyzing multivariate time series for fuel cell fault diagnosis. However, due to the complexity of these algorithms, and the large number of parameters they are built on, deep learning methods are extremely data intensive. In the fuel cell sector, where the technology has not yet fully matured for market deployment, obtaining large datasets is particularly challenging, and manufacturers are reluctant to publicly share data. This limits the availability of large and representative datasets necessary to build deep learning solutions. In real application scenarios, the available datasets are small, and contain only a small number of representative fault samples. Hence, the following question arises:

2. *How can deep learning solutions be exploited when large faulty datasets are unavailable for fuel cell technologies?*

A promising answer is to resort to physics knowledge. Indeed, the integration of physics-based models into deep learning can be beneficial for reducing data requirements, while maintaining high generalization of the algorithms and boosting their robustness [6]. While pursuing this direction appears promising, no universal methodological framework exists, and each technological field and application context require dedicated investigation. Accordingly, a major question is raised:

3. *How can the knowledge embedded in a fuel cell physics-based model be effectively combined with deep learning to provide robust fault diagnosis solutions?*

Eventually, fuel cells intrinsically undergo degradation, causing a drift in their operating behavior and performance over time. Fuel cell deterioration is closely linked to the loss of reliability of the fault diagnosis methodology implemented on the system [7]. Although being remarkably common, this is often disregarded in health monitoring strategies, and requires direct consideration. Variations in the system operating conditions result in corresponding changes in the sensor signals, potentially undermining the long-term validity of the algorithms. This implies that, as the fuel cell degrades, faults within the stack become increasingly difficult to detect. Consequently, incorrect fault diagnosis can accelerate the degradation process, further reducing the system durability. This raises a fourth, extremely important question:

4. *How can we ensure that a fault diagnosis algorithm remains reliable throughout the lifetime of the fuel cell system?*

The preceding discussion highlights the deficiencies in the current state-of-the-art of fuel cell fault diagnosis, namely the lack of methodologies capable of achieving accurate fault diagnosis while limiting the associated cost and compensating for scarce data availability. We directly address these challenges by proposing to leverage a hybrid framework that integrates a physics-based model with deep learning. This thesis work advances the frontiers of fuel cell fault diagnosis, providing a foundation for more reliable and aging-robust monitoring strategies.

1.3 Contributions of the thesis

The broad objective of this work is to advance health-monitoring solutions for fuel cell systems, enabling cost-effective and reliable operation, and supporting a larger deployment of the technology in transportation sector. The main contributions will be presented in the following chapters:

- **Chapter 4** proposes a hybrid method for fuel cell fault diagnosis that integrates a physics-based model operating online with a deep neural network. By combining model-based estimates of unmeasured process variables with monitoring signals directly measured on the stack, the approach enables accurate fault diagnosis without requiring additional physical sensors.

- **Chapter 5** presents a deep learning-based method for fuel cell fault diagnosis that leverages physics-based simulations as prior knowledge. Within a domain adaptation framework, effective knowledge transfer from physics-based model to a deep neural network is enabled during the offline training of the algorithm. This strategy reduces the dependence on large experimental datasets and enhances the generalization of the fault diagnosis algorithm.

- **Chapter 6** introduces a novel methodology that aims at integrating fault diagnosis and state of health estimation to enforce aging awareness in fuel cell system monitoring. It consists of a multitask deep neural network that leverages complementary objectives to extract more relevant and robust information from the available monitoring signals.

2

Background on PEMFC and Fault Diagnosis

Contents

2.1 Introduction 6
2.2 PEMFC technology 7
2.2.1 Principle 7
2.2.2 PEMFC durability 8
2.2.3 Degradation mechanisms 9
2.2.4 Faults in PEMFC 11
2.3 Health monitoring and fault diagnosis: state of the art in fuel cells 12
2.3.1 Model-based methods 13
2.3.2 Non-model-based methods 15
2.3.3 Hybrid methods 18
2.3.4 Summary 20
2.4 Signals to diagnose the stack 20
2.5 Impact of aging in PEMFC fault diagnosis 21

2.1 Introduction

On the way to a decarbonized world, sustainable mobility has, and will continue to have, a significant impact. It leverages vehicle powertrains based on electrochemical storage and conversion systems, especially lithium-ion batteries and hydrogen fuel cell systems, for their high efficiency and zero emissions [8]. Proton exchange membrane fuel cells (PEMFCs) are electrochemical converters that convert hydrogen into electricity. PEMFCs constitute one of the promising blocks for cutting down carbon emissions, and they are increasingly becoming part of the contemporary energy landscape, both for mobility applications, i.e., automotive, rail, trucks, ships, airplanes, and for stationary applications, e.g., emergency systems and power generators [9]. Specifically, the main reason for the increased deployment of fuel cell electric vehicles (FCEVs) in the automotive sector is that the power supply is given by an external hydrogen tank. This offers advantages in terms of higher driving ranges, fast refueling, and no degradation during down-time of the system, which occurs in lithium-ion batteries [10]. PEMFCs are currently the leading technology for hydrogen mobility solutions due to their high specific power, low temperature enabled fast start-up and high efficiency [11]. The 2025 Global Hydrogen Review of the International Energy Agency (IEA) highlights, indeed, that the number of FCEVs has been increasing in the past years. Major FCEVs manufacturers, including Toyota (Mirai 2016), Honda (Clarity, 2017), Hyundai (Nexo, 2025), are maintaining their strategic commitments [12]. Although the outlook for light-duty vehicles appears less favorable recently, the demand for heavy-duty vehicles presents a promising trend, particularly for trucks, which are expected to dominate the hydrogen transport market [13], and maritime ships.

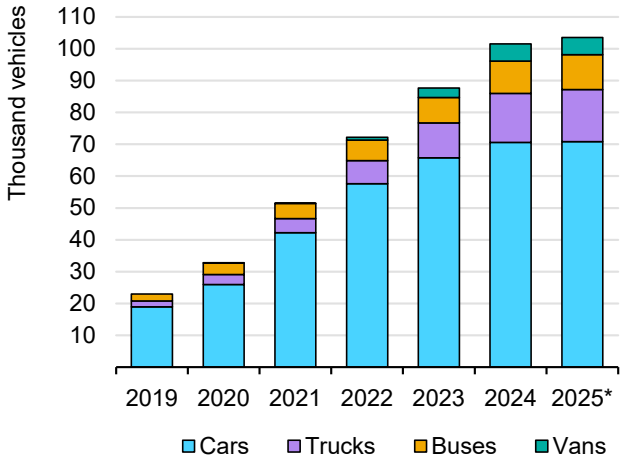


Figure 2.1: Fuel cell electric vehicles stock by type, 2019-2025. 2025* Includes data from January to July [13].

Nevertheless, two major obstacles are hindering a large-scale deployment of the technology: high cost and low durability [14]. A PEMFC system for heavy-duty transportation is expected to cost $80 per kilowatt by 2030. The U.S. Department of Energy (DoE) estimated that fuel cell cost has been reducing in the past years owing to technological breakthroughs. Nonetheless, for the forthcoming years, the envisaged cost reduction is expected to come mostly from optimized stack operation [15], as illustrated in Figure 2.2. For what concerns durability, the ultimate targets for fuel cell system lifetime under realistic operating conditions is set at 8,000 hours for light-duty vehicles, 30,000 hours for heavy-duty trucks, and 80,000 hours for stationary power systems [16]. Research has been focusing on identifying and understanding the fuel cell degradation mechanisms and developing materials and strategies to mitigate their effects. Efforts towards these objectives have to be made from several perspectives: working on the materials, reducing the causes of degradation, improving the structural design, implementing new supervision and management designs [17]. Optimizing the operating conditions of the fuel cell is considered as one of the main actions to push the development of the technology.

Non-optimized operation could be a critical reason for unexpected systems shutdowns and further degradation of the components [14]. Indeed, PEMFC systems are demanding applications for which reliability and robustness is required under dynamic and harsh operating conditions. Realistic operating conditions include starting and

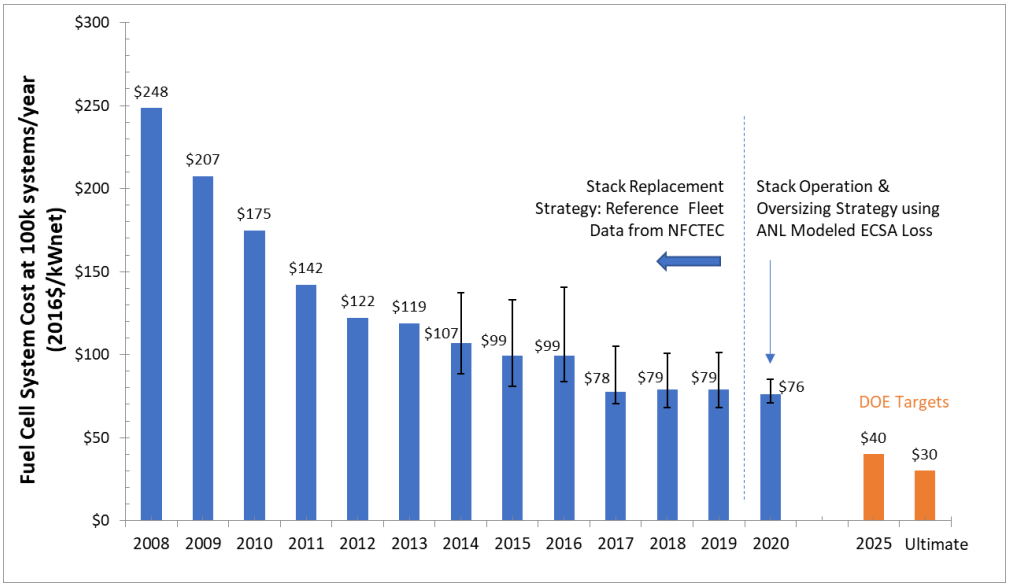


Figure 2.2: Cost of fuel cell systems [15].

stopping, freezing, presence of impurities in the fuel and air, high humidity levels and dynamic load cycles. Degradations of the materials, suboptimal operating procedures, disturbances from the external environment and load, have a severe impact on the durability performance of PEMFC. Such conditions induce stress on the chemical and mechanical stability of the fuel cell system materials and components. Consequently, developing innovative system designs and operation strategies to optimize performance while mitigating system degradation, is considered crucial for enabling large commercialization of PEMFC based technologies.
PHM solutions are promising to cope with the short lifetime and high-cost bottlenecks of PEMFCs. PHM is of great interest since it allows facing the problems of health assessment and life prediction [18], enabling adequate actions to extend the system durability, without the need for new designs and materials research [19].

2.2 PEMFC technology

2.2.1 Principle

A PEMFC is an electrochemical converter that transforms air and hydrogen gases into electrical energy generating zero carbon emissions. Electricity is generated as long as the reactants are supplied, and the reaction products are removed [20]. The reacting material as well as the reaction products are liquids or gases. This permits a continuous supply of reactants to the cell and a continuous removal of the reaction products. Hydrogen is fed to the cell at anode side, protons are produced with the oxidation reaction and transferred to the cathode electrode through the polymer membrane. At the same time, electrons are delivered from the anode to the cathode electrode through an external electrical circuit. Then, the reduction reaction takes place at the cathode electrode by electrons, oxygen and protons. The two semi-reactions taking place at the two electrodes are:

$$H_2 \rightleftharpoons 2H^+ + 2e^- \tag{2.1}$$

$$\frac{1}{2}O_2 + 2H^+ + 2e^- \rightleftharpoons H_2O \tag{2.2}$$

The core of the cell is the membrane electrode assembly (MEA) [21], which is an assembled stack of:

- proton exchange membrane (PEM), which transports protons from anode to cathode, acts as electrical insulator and separates reactants (i.e., hydrogen and oxygen). The material of choice for the membrane is Nafion, a polymer with high chemical and mechanical resistance, which allows the passage of protons

when hydrated;

- catalyst layers (CLs), which provide the reactant sites where electrochemical reactions occur. CLs are tightly attached on both sides of the membrane. The most common catalyst used is platinum, which is supported by carbon particles to allow a more efficient distribution of the material;
- Gas diffusion layers (GDLs), which are porous layers that favor the diffusion of reactant gases and the evacuation of the reaction product, i.e., water. GDLs allow uniform gas distribution and serve as support and protection for CLs.

Thereafter, bipolar plates (BPs) are placed on both sides of the MEA. They are used to collect electrons, to circulate coolant, to insulate reactants between different cell units, and to homogeneously distribute reactants on the overall surface of the cell. A schematic of a single PEMFC is given in Figure 2.3.

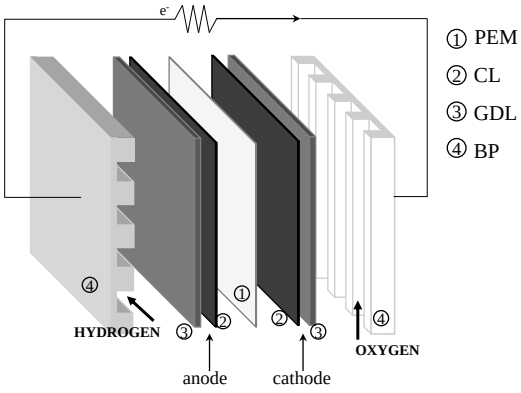


Figure 2.3: Schematic of the PEMFC.

Thermodynamically, the voltage of a single electrochemical cell is 1.23 V when pure oxygen is used. Nevertheless, several factors induce cell voltage drop: activation losses, ohmic losses and transportation losses. The resulting voltage-current curve is the polarization curve, illustrated in Figure 2.4.

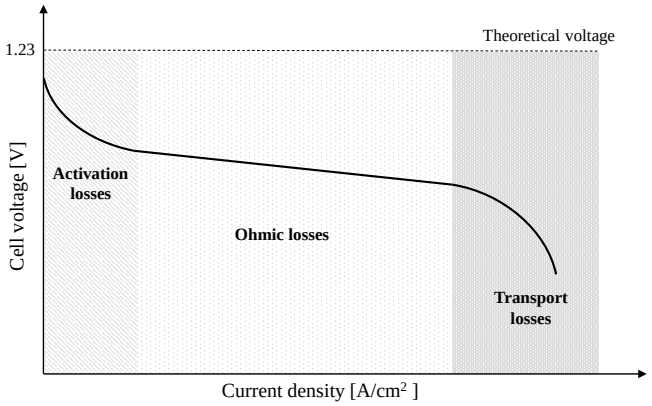


Figure 2.4: Polarization curve.

To provide an output voltage that is sufficient to power the system of interest, several cells have to be assembled in series and pressed together. This constitutes the PEMFC stack, which eventually can be used as power source in mobility or stationary applications.

2.2.2 PEMFC durability

Durability is defined as the time interval during which a component is able to operate at its required performance level, from the beginning of life (BoL) to the end of life (EoL) [22]. The definition of the required performance level for a PEMFC is not straightforward, and different criteria can be deployed:

- Voltage loss: the voltage loss, at a reference current density i_{ref}, has to be less than 10 % of its nominal value at BoL. As the voltage deterioration is larger, the fuel cell efficiency decreases faster, leading to poor fuel economy [23]. Thus, $V_{\text{EoL}} = 0.9 \cdot V_{\text{BoL}}$;
- Maximum power: the maxim power P_{max} generated by the fuel cell has to be above a certain threshold value, $P_{\text{max,EoL}}$;
- Efficiency loss: to generate a reference power density P_{ref}, the current density of the fuel cell cannot exceed a certain threshold i_{EoL}. As the required current density becomes larger, a significant degradation of the system efficiency is expected. Accordingly, the fuel cell system can no longer be used.

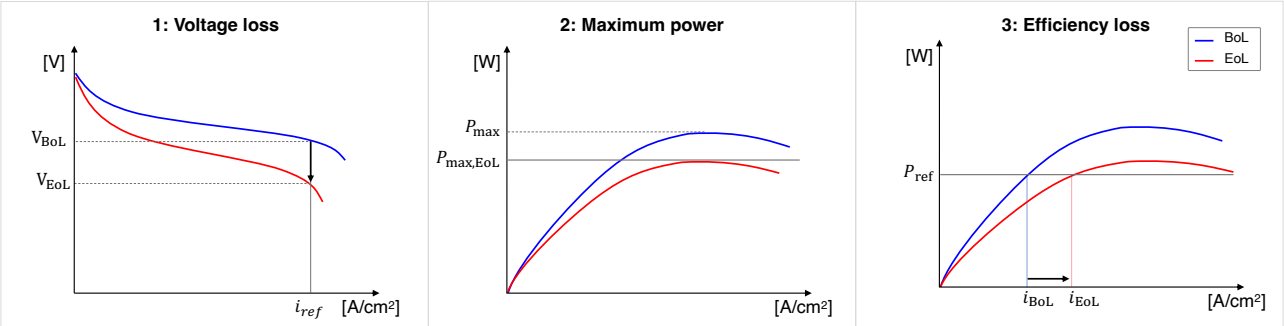


Figure 2.5: Performance level criteria: voltage loss, maximum power, efficiency loss. In blue are illustrated the performances that meet the requirements, at BoL; in red the performances at EoL.

Therefore, improving PEMFC durability involves extending the time the system meets the required performance level, eventually minimizing voltage degradation over time and reducing the fuel consumption efficiency. In addition to performance degradation, catastrophic component failures can occur originating from manufacturing defects or continuous degradation, such as membrane rupture leading to reactant gas mixing [24].
By 2023 the durability of PEMFC systems for heavy-duty transportation was around 10,000 hours. The goal for 2030 is 25,000 hours, and the ultimate target is set to 30,000 hours, with progressively decreasing costs [17]. Despite the progress made, the relatively short lifespan of fuel cells remains a major obstacle for accelerating the introduction of these technologies in industrial and consumer applications. In recent years, the fuel cell community has significantly focused on enhancing durability through advancements in materials, such as improving catalysts and membranes, as well as optimizing GDLs and BP designs. The DoE highlighted that the technological maturity of PEMFC is now high [25], and that the most promising strategies for boosting the technology rely on optimizing the operational conditions through adequate fuel cell management and control [26].

2.2.3 Degradation mechanisms

As introduced in section 2.2.1, the combination of electrical, fluidic and thermal phenomena that take place inside the PEMFC, make it a complex multi-physics system. The complexity is increased by the irreversible modifications of the physical and chemical properties of the materials composing the stack. Consequently, it is hard to have a clear understanding of all the degradation mechanisms taking place in the system: they are multi cause and multi effect, and their impact on the subcomponents of the fuel cell is different. Hence, our intention is not to exhaustively address all the degradation mechanisms related to PEMFC, but rather to provide the reader with an overview of the impact of operation on degradation.
The performance of a PEMFC stack gradually deteriorates throughout its lifetime, due to usage, start and stop cycles, environmental conditions, and non-optimal operating conditions, particularly when used for vehicular applications [27]. PEMFCs degradation is linked to two major phenomena:

- Loss of electrochemically active surface area (ECSA), which refers to the surface of catalyst participating in chemical reactions, i.e., platinum [28]. Catalyst deterioration represents a continuous trend in time, and it is the major cause of efficiency loss;

- Membrane degradation, which refers to the ionomer which favor the transport of protons from anode to cathode [29]. It includes membrane thinning, and, in time, pinholes and cracks. When a membrane rapture occurs, the fuel cell is no more able to generate power;

Loss of ECSA is induced by several phenomena. First, platinum dissolution occurs. If the water content is too high, the platinum particles are more mobile and the Ostwald ripening mechanism is enhanced [30]. Particle dissolution is favored in high voltage, or fuel starvation, or high temperature conditions. Moreover, very low or very high current densities, high partial pressure of the gases and low membrane humidity induce the formation of radicals, which, in turn, induce the corrosion of the carbon support.
For what concerns the membrane, the most severe degradation is provoked by radicals, which cause ionomer decay and pinholes formation. Radicals encourage membrane thinning, as well as reduction of proton conductivity and water intake capability of the membrane [31]. Very high and very low current densities, high temperature and aggressive operating conditions, favor the presence of radicals. Furthermore, impurities intake, coming from air or coolant pollutants, has a negative impact on the proton conductivity. Membrane degradation, in time, induces an increase in gas crossover rate, which makes the hydrogen use less efficient. At the same time, the mechanical strength of the membrane is reduced.
In addition to catalyst and membrane degradation, mechanical degradation of the GDLs, as well as of the BPs, occurs. The effects include an increase of the contact resistance, delamination of the components of the MEA, and loss of hydrophobicity. Mechanical degradations are enhanced by pressure and humidity cycling, vibrations, freeze starts and low stoichiometric ratios [32]. Overall, prolonged use exposes PEMFC to several degradation mechanisms, that reversibly and/or irreversibly deteriorate chemical and mechanical stability of several subcomponents of the cell. The damaging action of these phenomena is accelerated by operating the cell under certain faulty conditions [33]. This includes, too low or too high temperature, inappropriate humidity levels, bad management of stoichiometry levels, high concentration of pollutants in the inlet gases. Furthermore, load cycling, starts and stops, idling and high power operation boost degradation [34]. Therefore, it is clearly necessary to optimize the use and control of PEMFC systems, so as to boost performance and durability [35].

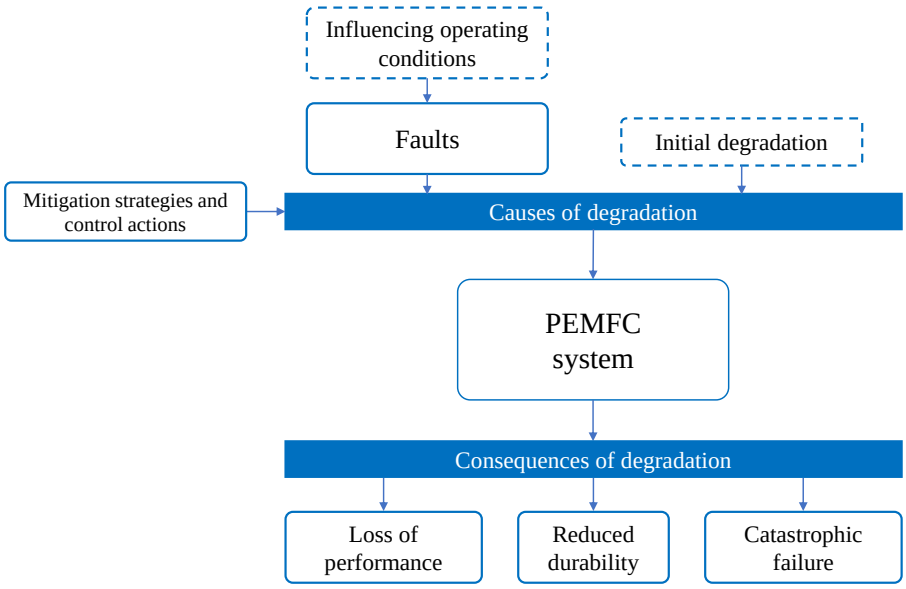


Figure 2.6: PEMFC degradation and consequences (adapted from [27]).

We highlight that the degradation mechanisms described occur at cell level. All the cells of a fuel cell stack will be affected differently, depending on their position inside the stack assembly. Accordingly, the durability of the stack depends on the durability of the individual cells [36]. In this work we will refer to stack degradation, keeping in mind that the degradation occurs at a cell level, and the loss of efficiency that is undermining the durability is due to the cumulative degradation of all the cells constituting the stack.

2.2.4 Faults in PEMFC

A fault is defined as a set of operating conditions that leads to loss of performances and causes accelerated degradation of the PEMFC stack, and that can be reversed or recovered by means of an appropriate control action [37]. At first, the effects of a fault are to reduce the efficiency and the power output of the fuel cell. However, a prolonged permanence of the stack in a fault state can favor the setting of those conditions leading to permanent degradation, e.g., radicals formation [38]. A fault can be caused by the incorrect operation of the auxiliary systems to the stack, as well as by the imposed load demand [39]. Faults in PEMFC have been widely investigated in the literature. Massive water droplets formation in the flow channels or inside the electrodes leads to a flooding fault, during which the excess of water blocks the reactants pathways [40]. Flooding can occur at both the cathode and the anode side. Although water is generated at the cathode during the electrochemical reaction, water diffuses through the membrane to the anode side, where its removal is limited by the low gas flow rate [41]. Pressure drop increases inside the stack, eventually leading to oscillations in the partial pressures of the active sites and, thus, to oscillations of the stack voltage. In the opposite case, drying indicates a poor hydration of the membrane. Given that membrane ionic conductivity is highly dependent on its hydration, a drying fault will cause an increase in the ohmic resistance and therefore a drop in the voltage [42]. Therefore, water management in PEMFCs is crucial to ensure stable operation, high efficiency and membrane durability [11]. Poor gas control can lead to a starvation fault of air and/or hydrogen. Indeed, when the gas flow is insufficient to sustain the imposed current load, local starvation can occur in the fuel cell inducing a variation in the current density distribution and a voltage drop [43]. Specifically, low hydrogen stoichiometry causes localized starvation at the outlet, with reverse current setting up and inducing carbon support corrosion. Eventually, if the stack is operated in open cathode mode, carbon monoxide (CO) poisoning is typically considered a major fault, since high CO concentrations are found in the air [44]. A schematic of the impacts of the faults on PEMFC degradation is illustrated in Figure 2.7.

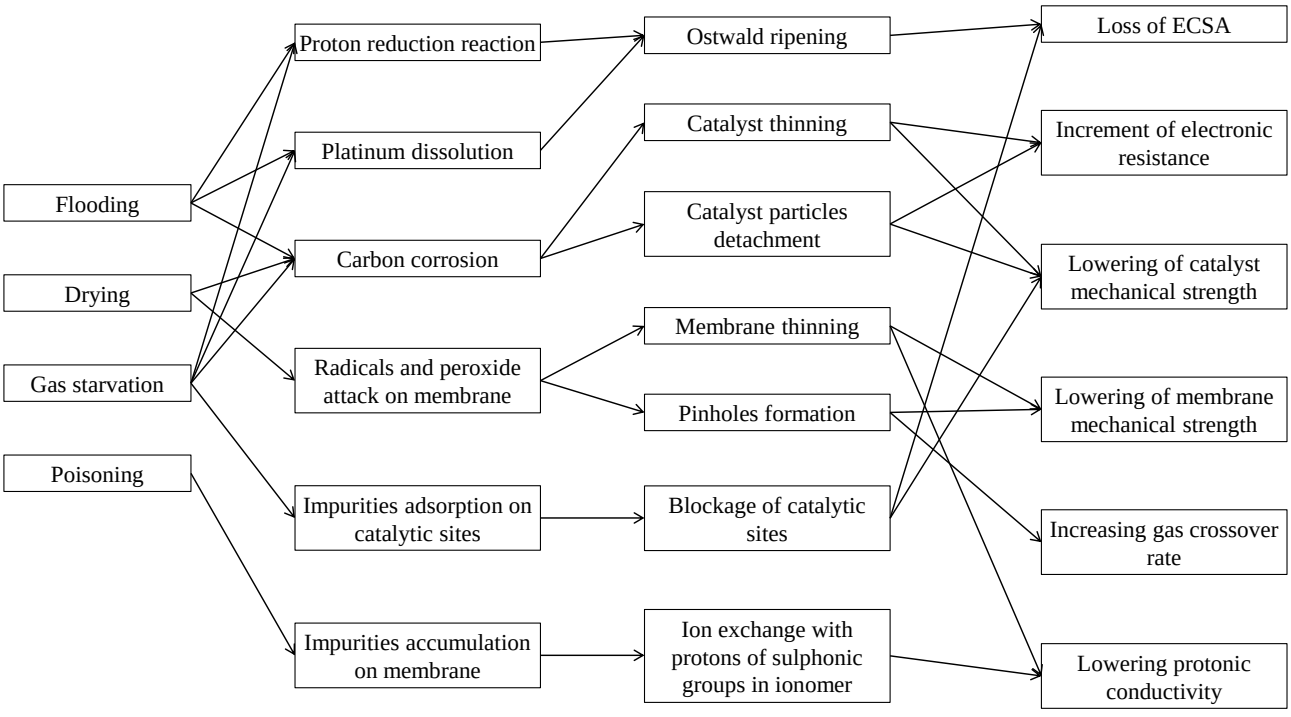


Figure 2.7: Schematic of the impact of faults and their effect on PEMFC degradation (adapted from [33]).

The faults have a reversible impact on stack efficiency, reducing the output voltage within seconds or minutes. Over time, however, these degrading conditions also cause irreversible damages to the PEMFC, which become visible over months of operation [45]. The extent of the irreversible degradation is enhanced with prolonged exposure to such conditions. In order to immediately remedy the faults and prevent the stack from being irreversibly damaged, it is essential to accurately monitor, in real-time, the state of the stack and correctly

diagnose a fault when it occurs. Therefore, the main goal of this study is to develop a method that can accurately estimate the state of the PEMFC stack under dynamic conditions, using the signals monitored on the stack, with a particular focus on water management faults, i.e., flooding and drying, and on starvation faults.

2.3 Health monitoring and fault diagnosis: state of the art in fuel cells

PHM has been playing a crucial role for the successful adoption and widespread deployment of critical and complex industrial systems [46]. PHM allows ensuring performance, safety, availability and reliability of a system throughout its lifetime by detecting and preventing failures and degradations. It defines a set of activities in which the main objective is to achieve high availability while reducing operational costs, by identifying optimal operational strategies and predictive maintenance policies [6]. Indeed, the dominant trend in maintenance is to avoid costly reactive maintenance, and shift towards predictive maintenance. By predicting degradations and preserving the integrity of complex systems, unanticipated operational problems leading to performance deficiencies, and adverse effects to mission safety can be avoided [18].
As illustrated in Figure 2.8, the PHM framework integrates the detection of an incipient fault (fault detection), its isolation, the identification of its origin and the specific fault type (fault diagnosis), the prediction of the remaining useful life (prognosis), and eventually the maintenance decision-making and fault corrective actions. Its practical implementation includes system observation through data acquisition and preprocessing to enable detection, diagnosis and prognosis tasks. Consequently, health monitoring is a fundamental component of PHM. It provides real-time data essential for assessing system condition and predicting future failures. Typically, sensors are employed to monitor operating conditions, performance, and degradation levels, which allows for a timely detection of anomalies and faults. Minor faults within a system component can escalate into major issues, leading to failure of the entire system. A proactive approach helps to avoid unforeseen incidents, downtime, and fault propagation.

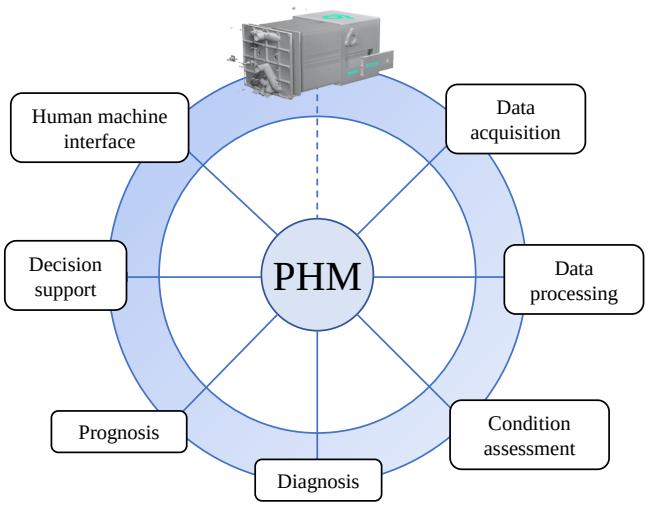


Figure 2.8: PHM framework [8].

PHM is being increasingly studied for PEMFC in order to deal with system assessment, diagnosis, prognosis, and decision-making support. While advancements in material design and new assembly configurations are essential for mitigating degradation, ensuring that the PEMFC stack operates under optimal conditions can significantly contribute to enhance durability and has a positive impact on the performance. For these reasons, combined with relatively low cost and ease of implementation, health management strategies are being extensively adopted. In this context, fuel cell management system (FCMS) is a comprehensive set of functions used to operate properly and safely the whole PEMFC system [47]. Its primary objective is to ensure the delivery of the required power at every moment of operation while maximizing the overall efficiency of the system. This includes optimizing the performance of the fuel cell stack and the associated auxiliary components, i.e., air compressor, hydrogen supply system, cooling circuit, and humidifier. Also, optimizing the performance allows decreasing the fuel

consumption, thus operational costs. Within the health management framework, FCMS embeds prognosis and diagnosis solutions. Prognosis solutions deal with the prediction of fuel cell future degradation trends, and they are required to adapt the system operation to the severity of its degradation. Diagnosis is the process of identifying the system failures and determining the root causes [48]. Diagnosis is targeted to raise alarms in the case of faulty operation modes and analyze the main factors for fuel cell fault operations. The control strategy is subsequently adjusted in order to maintain PEMFC performance and reliability targets, as well as to prevent further damage. Early fault diagnosis enables predicting the fault development trends, which is vital in the prevention of faults propagation and system failure [49]. In order to avoid the occurrence of subsequent major accidents caused by minor faults, in several mission-critical engineering systems is granted great importance to the intelligent fault diagnosis. This is the case also for PEMFC. Particularly in non-stationary applications, PEMFC fault diagnosis is strongly required for detecting and isolating the faults, providing insightful information which allows to run the system and to maintain optimal operating conditions [21]. Nevertheless, performing PEMFC fault diagnosis presents significant challenges due to the stringent requirements to be met. When integrating an algorithm in the FCMS, the use of high-performance computing units is impractical or cost-prohibitive. Also, the use of physical sensors to assess the internal states of the systems has to be reduced as much as possible. Additionally, PEMFCs exhibit highly dynamic behavior, which is intricately linked to the dynamic profile of driving cycles, and it induces rapid changes in load demand and in environmental conditions. The complex electrochemical and thermodynamic processes occurring in the fuel cell make the monitoring signals highly nonlinear. As a result, the fault signatures are obscured. Moreover, as the fuel cell ages, its performance characteristics change, altering the robustness of fault detection in time.

The high complexity and demanding nature of PEMFC fault diagnosis have attracted significant research activities and efforts in recent years, pushing the development of numerous methodology and advanced algorithms [50]. In this chapter we aim to provide an overview of available fault diagnosis strategies, with a particular focus on fuel cell fault diagnosis. Various frameworks have been proposed for categorizing fault diagnosis methods. We choose to adhere to the most common categorization in PEMFC fault diagnosis, which distinguishes between model-based, non-model-based, and hybrid methods, as illustrated in Figure 2.9. The choice of which method to use is typically driven by the available data and/or the performance models, the sensors that can be integrated into the system, cost benefit consideration and the safety criticality related to the implementation of the method [51].

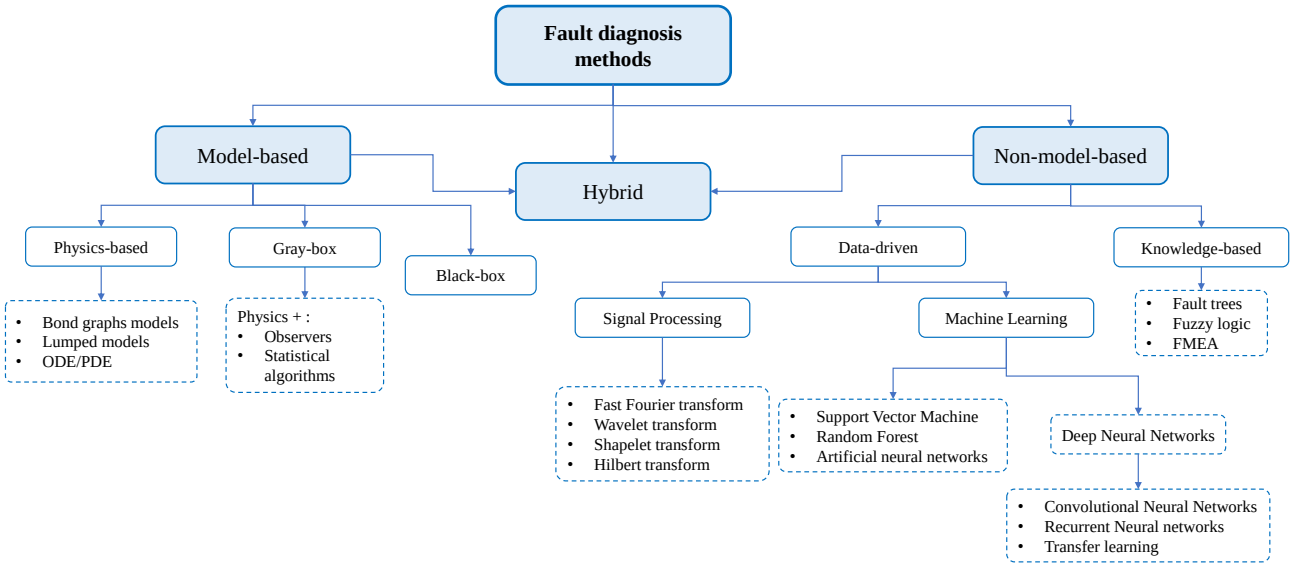


Figure 2.9: Categorization of fault diagnosis methods.

2.3.1 Model-based methods

Model-based defines techniques that employ a detailed model of the system to simulate how the system behaves under various external and internal conditions [52]. A model of the system is required to reproduce the behavior

of the system in nominal conditions, i.e., in a faultless state. The consistency between the real system outputs and those obtained by the model is monitored, and any deviation is used to derive considerations on the possible faults in the system [53]. Indeed, model-based diagnosis can be divided in two subtasks: fault detection and fault isolation. Fault detection checks the consistency of observed signals while fault isolation assesses the fault type. The differences between system output and model output are called residuals, which are used to detect the occurrence of a fault [54]. The model of a complex system can either be white box or gray box. A white box model is based on first principles or physical laws, and can be developed through bond graphs modeling or parameter estimation methods [55]. A gray-box model can be based on physical laws, with some complex mathematical equations replaced by empirical formulae, map tables or artificial intelligence tools like Principal Components Analysis (PCA) and artificial neural networks (ANN) [56], [57]. Alternatively, observer-based techniques including Kalman Filter (KF), and its nonlinear variation, e.g., Unscented KF (UKF), can serve to estimate the internal state of the system using a dynamic mathematical model, recursively corrected by noisy measurements. Ultimately, entirely black-box models can be obtained by using machine learning tools, like ANN, to replicate the behavior of the system in optimal conditions.
Once a model of the system is available, by comparing model estimates with real measured signals, residuals can be obtained. The residuals have to be analyzed to detect the occurrence of the fault, and eventually isolate it. If the fault diagnosis algorithm is dedicated to only one fault, a simple threshold-based approach is sufficient to isolate the fault [58]. To make the method robust to noise in the measurements, adaptive thresholds over a time moving window [59] or statistical tests on the residuals can be used [60]. However, when the faults considered are more than one, with different levels of severity, using threshold-based methods can be unsatisfactory, and signature matrices or classification algorithms are preferred to identify the origin of the fault [61]. A schematic of the model-based method is presented in Figure 2.10.

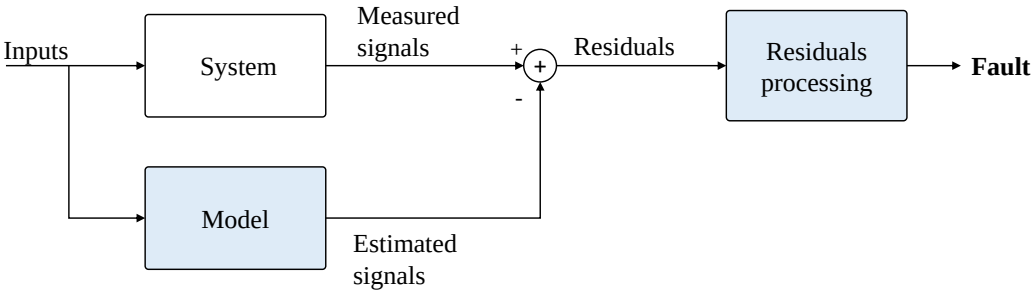


Figure 2.10: Schematic of model-based fault diagnosis.

Regardless of the modeling technique, accurate and complete models are hardly available for engineering systems. The major reason is that modeling large complex systems requires expert knowledge and lots of human labor. Notwithstanding, the fault diagnosis performance is strictly dependent on the model accuracy. Thus, the major disadvantage for model-based methods is the limited fidelity that several models exhibit with respect to the reality that they are expected to represent. Also, employing physics-based approaches, could be challenging due to computational times required in solving physics-based governing equations and predicting system performances, and their complexity can further hinder implementation in certain embedded applications.

Model-based methods in PEMFC

The main difficulties in developing model-based fault diagnosis methods for PEMFC are linked to the physical behavior of the fuel cell stack by itself. Indeed, a fuel cell stack is a highly multi-physics and multi-scale system, where the estimation or measure of internal parameters is difficult, if not impossible. In literature, many modeling approaches based on the description of the physical and chemical processes in a fuel cell stack, either as systems of (partial) differential equations or as electrical circuits have been proposed. In PEMFC modeling, the Nernst–Planck equation is typically employed to describe charge transport, the Fick's law to characterize ionic diffusion, and the Butler–Volmer equation to represent electrochemical charge-transfer kinetics [62]. Accordingly, a PEMFC physical model can be highly complex particularly if the modeling of stack subcomponents, i.e., anode,

cathode, GDL, catalyst and membrane is included. Such models can be useful for stationary applications in which the dynamic is much slower, like PEM electrolyzers [63], but the computational complexity makes it hard to deploy them in vehicular applications. Equivalent circuit models represent an alternative, since they offer a means to estimate internal states of the fuel cell, such as vapor saturation or gas partial pressures, which can support fault detection, but they require detailed measurements to be calibrated [64]. Randle's models offer a commonly used option, since they can be easily calibrated using impedance measurements. However, they often yield only a limited and coarse description of the system, which may introduce systematic modeling errors that propagate to the fault diagnosis of the system. In this regard, a combination of analytic equations and observers has proven to be a good strategy to limit errors [60]. Also, approaches comprising KF, or its variations, are easier to be implemented for online applications [65]. To overcome issues linked to missing knowledge or physics, calibration and computational complexity, black-box models of the system in nominal condition can be built using historical data. Pioneering works have employed support vector machines (SVM) algorithms to build a model of the air flow system [66]. Model based on ANN have proven useful to estimate voltage and pressure drop in nominal conditions, giving in input current, temperature and coolant flux signals [67]. Recurrent neural networks (RNN), like Elman Neural Networks, have been proposed to deal with the dynamic nature of the system [59, 68].
Eventually, fuzzy logic rules can be established to classify faults [68]. An alternative is represented by fault signature matrices, which are the preferred strategy nowadays to go beyond binary matrix and single fault identification [69–72].

2.3.2 Non-model-based methods

Non-model-based fault diagnosis refers to techniques and methods used to detect and isolate faults, without relying on a physical or mathematical model of the system optimal behavior. Indeed, unlike model-based approaches which use system models, non-model-based methods rely on direct measurements, historical data, and patterns recognition to detect deviations indicative of faults. Non-model-based methods are commonly categorized in two different approaches: knowledge-based and data-driven.
Knowledge-based methods are based on knowledge sources available in the specific engineering and industrial domain. They comprise qualitative approaches, e.g., fault trees, fuzzy logic, failure mode and effect analysis (FMEA) databases, historical maintenance logs, physical cause and effect relationships, reliability data, as well as heuristic or empirical expert knowledge [73]. To achieve fault diagnosis, monitored data are evaluated in terms of sets of rules that are known from past experience of human experts [74]. On the contrary, data-driven fault diagnosis relies only on data for the recognition of anomalous patterns attributable to faults, without the need for extensive domain experts' knowledge. Current advancements in sensor technologies, data analysis, and machine learning make it possible to develop data-driven methods, in which no elaborate performance model of the system is needed [75], making these methods increasingly popular. For complex industrial systems, historical data, rather than a model, are available. Then, instead of utilizing the experience and knowledge of engineers, data-driven methods apply an empirical database that represents the system in healthy and faulty conditions.
Two main strategies can be used to extrapolate the information from the data: signal processing techniques or machine learning. Signal processing methods consist in extracting relevant features from the frequency content of the monitoring signals [76]. The signals obtained from the system are transferred from the time domain to the frequency domain to extrapolate significant information which cannot be conceived in the time domain. The most common methods are Fast Fourier Transform (FFT) and Wavelet Transform (WT), but Hilbert Transform (HT) can also be used in certain contexts. These frequency-based features are then input to a classification algorithm, such as SVM, k-Nearest Neighbors (k-NN), or decision trees, which performs the fault classification. Machine learning based techniques, on the other hand, rely on artificial intelligence to compensate for the deficiencies of traditional approaches. Machine learning algorithms are used to adaptively learn the fault patterns from the historical data using relevant features extracted from the signals [77]. In this case, the extraction of the information (feature extraction), and the subsequent classification, can be supervised or unsupervised [78]. Supervised methods learn to distinguish faulty patterns by training on historical labeled data, i.e., for every

input signals, the corresponding fault label has to be known [61]. Commonly implemented algorithms are SVM, Random Forest (RF) and ANN. Unsupervised methods do not need fault labels to be trained, instead they require a larger amount of data to learn fault patterns. They are based on clustering technique, the most common of which is k-Means clustering.
Eventually, data-driven fault diagnosis is carried out by checking the consistency of the historical knowledge, used offline to train an algorithm, to that extracted in real-time from the online monitored data [74]. The framework of data-driven fault diagnosis is illustrated in Figure 2.11.

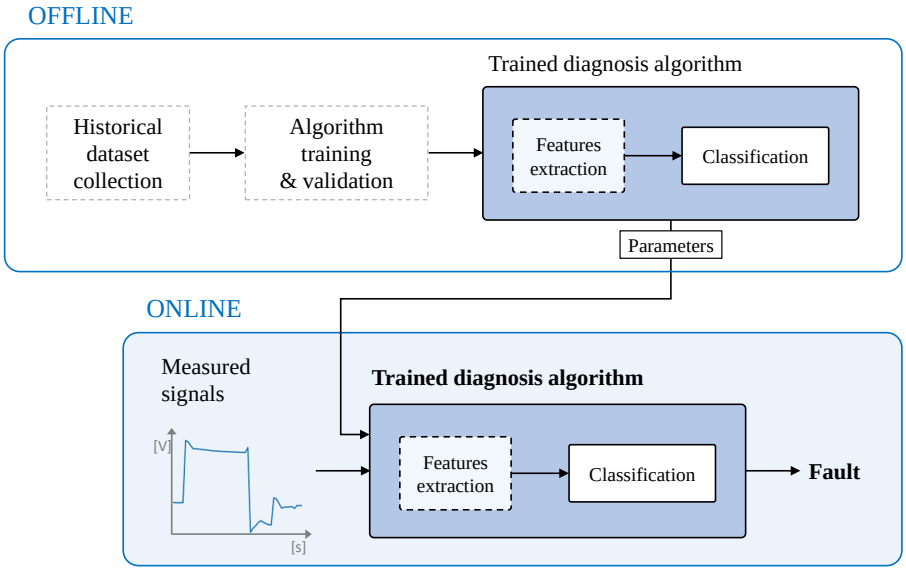


Figure 2.11: Data-driven fault diagnosis [44].

The main advantage of data-driven methods is the reduced expert knowledge and computational effort compared to model-based approaches, as it does not operate simulation models, but rather relies on the recognition of anomalous patterns in the signals. Additionally, it does not require a structured and deep expert knowledge, and, thus, it reduces the period of development [79]. However, traditional machine learning approaches present some limitations. They rely on carefully engineered features extracted from the raw data, making performance highly dependent on the quality of preprocessing. Furthermore, their ability to capture complex, nonlinear dependencies is also limited, which can hinder accuracy when dealing high-dimensional datasets.

The rise of Deep Learning Nowadays, the condition of complex systems is typically monitored by a large number and different types of sensors, which results in very heterogeneous condition monitoring data to be analyzed. In addition, different fault types are affecting different signals and the correspondence is generally not one-to-one. This implies that not all the signals are relevant for the fault, but more signals could carry relevant information to diagnose a given fault. Consequently, the fault diagnosis task, and specifically the extraction of relevant features, has become more difficult [80]. This, together with increased data availability and advancements in computing resources, has been the driver for the great success and development of deep learning in fault diagnosis [81]. Deep learning has transformed how categorization, pattern recognition, and regression tasks are performed across various application domains. Indeed, deep learning is particularly well positioned as it offers advantages in terms of ability to: process large volumes of condition monitoring data; automatically extract relevant features from high-dimensional and heterogeneous data sources, bypassing the need for dedicated feature extraction procedures; learn both functional and temporal relationships within and between time series signals [82]. It opens the paths to novel approaches, which could not be envisaged with machine learning. For instance, transfer learning enables the transfer of knowledge between different operating conditions and different machines. Lastly, it empowers and facilitates synthetic data generation, probabilistic learning and reinforcement learning [83], which could not be easily envisioned just few years ago.

Data-driven methods in PEMFC

Data-driven fault diagnosis methods entail a good solution for PEMFC since they are less computationally demanding than model-based approaches, and thus better suited for online applications. Considering the temporal characteristics of the PEMFC monitoring data, approaches relying on signal processing have been widely applied because of their ability to extract valuable features from non-stationary signals [84]. WT can extrapolate wavelet packet coefficients from the stack voltage, which are relevant to identify fault patterns [85]. WT allows separating the low and high frequency contributions in the monitoring signals, i.e., voltage and pressures, which is extremely useful in the case of dynamic load applied to the PEMFC [86, 87]. Similarly, techniques like HT, FFT or shapelet transform, can be useful to capture oscillations in signals, extract relevant features and use them to diagnose the stack [25, 88]. The extraction of meaningful information by means of signal processing is often combined with machine learning algorithms. SVM is often favored owning to its high flexibility and robustness [89]. Also, reservoir computing (RC), a specific RNN that can model temporal signals requiring small data to be trained, has proven particularly suited [90, 91]. A different approach to extrapolate information in the frequency domain is given by total harmonic distortion (THD) technique, which consists in sending low frequency signals to the stack and, by so doing, identify characteristics frequencies for the faults [92, 93]. However, the online applicability of this method has yet to be proven [94, 95]. Nonetheless, signal processing methods are highly sensitive to noise. Parameters selection is tricky, but highly impactful on their performance, leading to no guarantees on their robustness. Also, the extracted features might be non-exhaustive, not sufficient to capture the overall trend and the non-linearities between signals. To increase the robustness of these methods one could augment the number of electrochemical signals acquired [96], but this clearly entails an increase in complexity and cost. Recently, the distribution of relaxation times (DRT) has emerged as a technique for analyzing the frequency response of electrochemical devices via EIS, as it separates electrochemical processes with similar time constants, revealing them as distinct peaks in the frequency domain [97]. Nonetheless, the assumption that EIS is available, while being valid for stationary applications, it is unlikely at the moment for vehicular applications [98].
Methods based on multivariate statistical analysis offer an alternative to the frequency domain analysis, since they can reduce effectively the dimensionality of monitoring signals while retaining the most valuable information for fault diagnosis [62]. For instance, if individual cell voltages are used as monitoring signals, the resulting data will be high-dimensional and rather demanding to process. Dimensionality reduction methods, like PCA, can extract features that retain most of the variance in the data [99, 100]. Alternatively, Fisher discrimnant analysis (FDA) serves the same scope by maximizing class separability, thereby facilitating fault identification from monitoring data [101]. PCA and FDA are linear algorithms, clearly not suited when dynamic loads are applied to the PEMFC stack. Thus, their performances are often enhanced by combining them with supervised classification tools like SVM [102, 103] or RF [104], or with unsupervised ones like fuzzy C-means [105]. Eventually, it has been demonstrated that if the measured signals have high quality, machine learning algorithms can achieve good diagnosing performance even without a prior feature reduction step [106]. If the application is stationary and individual cell voltages are available, SVM can be successfully applied [107]. When polarization curves and/or EIS spectra are available, a relatively simple k-NN can be sufficient [43, 108, 109]. Other tested algorithms include k-NN, RF, Gaussian Naïve Bayes (GNB) [9].
With the raise of deep learning, it quickly became clear that these techniques were well suited to the complexity and high non-linearity of PEMFC systems [110]. By handling high-dimensional input data and extracting intricate features from diverse information sources and signals, they can effectively offer a broader perspective of the state of the PEMFC system [111]. Specifically, given the temporal nature of the signals measured from fuel cell stacks, RNNs have often been selected. Indeed, they can properly manage high-dimensional time series data, providing a comprehensive perspective and automatic feature extraction from sequential signals. In general, the number of monitoring signals used to diagnose the stack with these algorithms is much larger, going beyond stack voltage and EIS measurements. A long short-term memory (LSTM) network has been used to diagnose flooding with 12 signals [112], a bidirectional LSTM (BiLSTM) has been proposed to diagnose flooding and drying with 14 sensored signals [113]. The signals often include auxiliary components' information,

e.g., compressor speed. Also, Convolutional Neural Network (CNN), with the promise of being highly versatile, have been tested [114–116]. Nonetheless, deep learning algorithms comes with a drawback: overfitting is a common issue in deep learning, particularly when training data is limited. To address this, ensemble strategies, for example combing CNN and LSTM networks, can be employed [117]. Separately, transfer learning offers a way to leverage knowledge from related tasks or datasets to enhance performance when labeled data is scarce to train a supervised algorithm. For example, transfer learning can be useful to transfer knowledge from single-cell to stack data, using adversarial learning [118], or to transfer knowledge between different PEMFC datasets, using instance reweighting techniques [119].

2.3.3 Hybrid methods

Fault diagnosis is inherently a complex task, requiring a deep understanding of the system, whether derived from physical modeling or empirical data. When multiple knowledge sources are available, diagnosis performance can be significantly enhanced by integrating this heterogeneous information in a coherent framework [120]. Hybrid approaches aim to combine all available knowledge, i.e., models, data, and expert insights, to leverage their complementary strengths and improve the accuracy, robustness, and reliability of fault detection and isolation. Data and physics-based model are the two main source of knowledge available for most of the industrial engineering systems: in the past century, physical models have been created, while in the new century machine learning has emerged [121]. To implement fault diagnosis, the choice between these two depends on the complexity of the underlying system [122]. Physics-based models are generally hindered by their limited ability to properly tune the parameters and by the modeling error that originates from model simplification and omission [123], but they do not require large amounts of data, retain the interpretability of a model, and provide the opportunity to generate synthetic data. In contrast, data-driven approaches are limited by the representativeness of the training datasets but are simple to implement, and one can use data-driven models to discover complex patterns from large volumes of data [124]. However, the volume of useful experimental data available for complex physical systems is limited. This motivates the adoption of strategies that fuse the two approaches in order to address specific industrial requirements and needs [125]. The underlying assumption of data-driven methods is that the relevant information concerning the faults can be learned from past data. This holds if the data are representative of all potential conditions that the system could experience [126]. However, the collection of such a representative dataset is extremely difficult, or impossible, because faults, on average, are rare, the environmental conditions and the mission profile of the system can vary. On the other hand, model-based methods rely on the model capability of accurately representing the real system.
Hybrid methods have the objective of integrating the robustness and interpretability of model-based methods with the specificity and accuracy of data-driven methods [127]. A specific category of hybrid methods is characterized by the approaches that integrate a physical model within machine learning or deep learning algorithms. Nowadays, the literature does not have a clear nomenclature for integrating previous knowledge of a physical phenomenon with machine learning. 'Physics-informed,' 'physics-based,' 'physics-guided,' and 'theory-guided' are often some used terms [128]. The main obstacle to face when developing hybrid approaches, combining machine learning and physics-based models, is the lack of a generic framework for integration [129]. Indeed, several strategies can be adopted, as illustrated in Table 2.1. Physics can be used to constrain or guide the construction a deep learning algorithms [130, 131], or to enforce physics-based generative networks [132, 133]. In Physics-Informed Neural Networks (PINNs), the physics can be assimilated during the learning process by means of specialized loss functions [134]. Implementing a physics-informed learning algorithm mainly involves incorporating relevant observational, inductive, or learning biases to guide the learning process towards physically consistent solutions [135]. Concisely, when prior physics knowledge is integrated in the algorithms, improvements can be achieved in terms of required data as well as interpretability [136].

Hybrid methods in PEMFC

First attempts of developing hybrid fault diagnosis for PEMFC were using machine learning algorithms, like k-NN, SVM, ANN or Naïve Bayesian Networks, for residual classification instead of fault matrices [54, 138, 139].

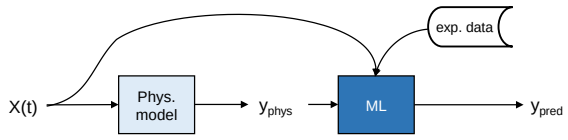
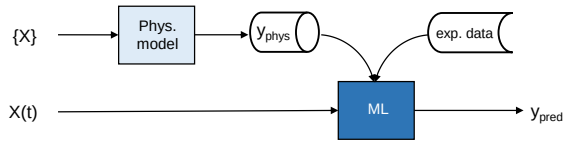
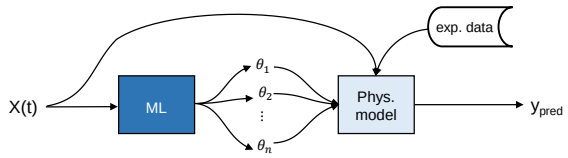
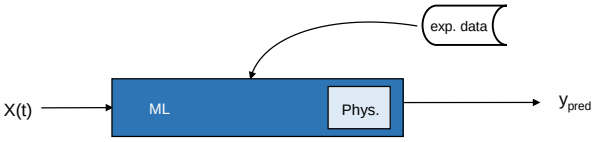
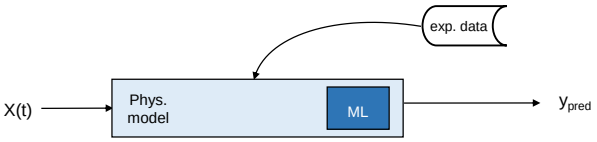
Type	Architecture	Description
Residual learning		• ML learns difference between physics model predictions and mesurements • Physics model is part of the prediction pipeline
Transfer learning		• ML trains on experimental data augmented with physics model generated data • Physics model is not part of the prediction pipeline
Parameters learning		• ML learns parameters of physics model that best explain experimental data • Output is strictly constrained by physics • Interpretability
Physics-constrained ML algorithm		• ML architecture contains physics-based elements by design, e.g., loss function
ML accelerated physics model		• ML assists solution of differential equations • ML guide data acquisition for physics model

Table 2.1: Possible integration strategies for physics-based model (Phys. model) and machine learning (ML). The physics-based model is represented in light blue, the machine learning algorithm is represented dark blue. Adapted from [137].

It has been demonstrated that if one has a mapping model for temperature, voltage and current density of a PEMFC, than it could combine it with ANN to effectively classify and localize the faults [140]. A physical model can also help in finding the best set of sensors needed to diagnose the stack, so as to provide prior knowledge that is fundamental to develop a data-driven approach [141]. On the other hand, a neural network can be used to predict the parameters of an electrochemical impedance model, which in turn serve to diagnose the stack [142–145].

Lately, PINNs have emerged as neural networks that force the output to satisfy physical constraints, by using additional loss terms that enforce physics laws during training [146]. By so doing, the algorithm is simultaneously constrained to fit the observed data and respect the physics. For example, an equivalent circuit model has been enforced into the loss function to guide the training of a deep learning network [147]. Beyond enforcing physical consistency, physics-based models can significantly reduce the amount of experimental data required to train diagnosis algorithms. Also, the physics models can guide the discovery and the generation of degradation trajectories of PEMFC [148]. By providing synthetic data or guiding the learning process through physically meaningful constraints, these physics models enhance generalization and reduce dependence on costly or time-consuming measurements. In this context, digital twins (high-fidelity, physics-informed virtual replicas of electrochemical systems) are particularly powerful tools [149]. They enable the simulation of a wide range of operating conditions and fault scenarios, which are afterward used to train pattern recognition algorithm, supporting the development of more efficient and reliable fault diagnosis strategies [150]. Through this approach, a high fidelity model of solid oxide fuel cells has been combined with machine learning algorithms [151]. Also, knowledge has been transferred from a PEMFC simulator to a real system with CNN algorithm [152]. However, recent works have highlighted that the limited accuracy of physics models hinders this direct transfer, suggesting that adaptation strategies should be investigated to achieve more robust transfer from physics-based to deep

learning algorithms [153, 154].

2.3.4 Summary

In this section, we have reviewed the most significant research works on PEMFC fault diagnosis, with a predominant focus on model-based and data-driven methods. These methodologies represent the foundation of current fault diagnosis strategies, each offering unique advantages in interpreting the complex behaviors of fuel cell systems. However, each method also comes with inherent limitations.
Model-based methods rely on representations of the system in normal behavior, often based on physical laws, which is valued for the interpretability, together with no need for faulty data. Nonetheless, in real-world application they encounter numerous challenges, primarily resulting from the complex modeling process and the limited scalability. The associated computational cost often hinders their online capabilities. Moreover, attaining high accuracy in physics-based models demands well-characterized data to calibrate their detailed representations of the system. Data-driven methods, on the other hand, are more flexible and simpler to implement for nonlinear systems, as no model is required. This comes at the cost of little interpretability, together with the dependency on high quality and large datasets. A summary of the advantages and disadvantages of model-based and non-model-based methods is given in Table 2.2.

	Model-Based	**Data-Driven**
Advantages	Interpretability	Limited expert knowledge
	Physics constraints	Adaptability
	Few data	Easy implementation
		Good non-linear capabilities
		Fast inference
Disadvantages	Expert knowledge required	Large datasets
	Limited scalability	Limited generalization
	Computational cost	No interpretability
	Dependency on modeling error	

Table 2.2: Comparison of the advantages and disadvantages of model-based and data-driven fault diagnosis methods.

Hybrid methods have emerged as a natural evolution to overcome the limitations of pure model-based or pure data-driven paradigms [155]. As the field moves forward, the fusion of high-fidelity physics-based models with deep learning algorithms represents a promising direction in addressing persistent challenges in PEMFC fault diagnosis. These challenges include:

- need for robust time series analysis when dynamic load is imposed to the stack;
- scarcity of real-world data in faulty conditions;
- ability to cope with system aging and progressive signals deterioration;
- necessity for online implementation;
- accurate fault isolation in presence of multiple faults.

2.4 Signals to diagnose the stack

Fault diagnosis, as part of the health monitoring, requires monitoring sensors and devices to acquire signals that provide relevant information to characterize the operational faults in the system.
While numerous techniques exist for monitoring electrochemical systems, not all are compliant with real-time, high-frequency monitoring, which is crucial for effective fault diagnosis. For instance, methods including polarization curves or cyclic voltammetry (CV) require time-consuming acquisitions. The former provides valuable insights into the performance degradation of PEMFCs, which is tightly linked to the occurrence of

a fault, while the latter is of particular interest for determining the electrochemical surface area, and thus the degradation level [156, 157]. EIS is well known to be a powerful tool to monitor the PEMFC stack conditions, as it allows separating the contribution of different physics and chemical processes. It consists in exciting an electrochemical system by a sinusoidal electric input at different frequency, while measuring the complementary electric output, thus it represents a non-destructive solution to monitor the stack [158]. However, despite several advancements have been made towards the viability of online EIS, the methodology is not mature yet [106]. Alternatively, external magnetic field measurements can serve as noninvasive tools for analyzing abnormalities in the internal current distribution of PEMFCs [159]. However, their implementation on real systems is not yet envisaged due to the high sensitivity to noise [160].
In practice, to enable fast and cost-effective online fault diagnosis, the preferred strategy is to rely on signals that are already monitored within the system, typically those that are low-cost, readily accessible, and do not require additional instrumentation. Among these, stack voltage is most often used as a primary indicator of system performance, as deviations from expected voltage levels can indicate the presence of internal faults. However, voltage measurements alone often lack the specificity required to discriminate among different types of faults. The effects of several phenomena cumulate on the voltage measurements, i.e., dynamic load, degradation, faults [161]. To gain a more detailed view of the system condition and enhance diagnosis resolution, individual cell voltage monitoring can be used [162]. Nonetheless, its implementation requires high sensors' density, which ignores practical and economic constraints, particularly in compact stack configurations. Alternatively, current density distribution has strong correlations with faults, as higher current induces greater reactant gas velocities, influencing water removal dynamics, but it is difficult to measure in an embedded system [37, 85, 163]. Non-electric signals can also be considered. When targeting water management-related faults, coolant temperature has proven to be useful, as it is sensitive to thermal imbalances caused by issues such as flooding or dehydration [9, 164]. Furthermore, pressure drop, at cathode as well as anode, is closely associated with the presence of liquid water and evolving flow patterns, offering indirect but valuable insights into two-phase flow behavior. Used separately, these signals do not provide direct information about the health state of the stack. Yet, they embed diagnostic-relevant information. Accordingly, with appropriate processing, complementary information can be extracted and leveraged to improve the overall accuracy and reliability of the fault diagnosis framework.

2.5 Impact of aging in PEMFC fault diagnosis

PEMFC systems are inherently exposed to degradation over time, resulting in changes to their chemical, mechanical, and electrical properties. This temporal evolution directly affects the monitoring signals, leading to drifts, biases, or increased variability, which pose a significant challenge for reliable fault diagnosis. In essence, the PEMFC stack is a non-stationary system, and any diagnosis methodology should account for its time-varying behavior. In model-based approaches, degradation necessitates recalibration or adaptation of the physical model, so that it continues to accurately capture the system's dynamics. In data-driven methods, degradation can alter the features extracted from the signals, potentially shifting them outside the data distribution the algorithm was trained on. Thus, to avoid false negatives or misclassifications, the algorithm should be capable of distinguishing between normal aging effects and true fault conditions [107].
This remains a major open challenge in PEMFC fault diagnosis, which only a limited number of studies have thoroughly investigated. An ideal solution would involve integrating knowledge of the system's degradation behavior into the fault diagnosis process, either through accurate degradation models or historical data representative of the stack aging. Some approaches treat degradation as an additional fault class [60, 91] or even define multiple fault classes corresponding to different degradation stages [89]. While conceptually useful, these methods lack the ability to detect faults during degradation, accelerating further the degradation. A different strategy involves developing multiple diagnosis models tailored to different degradation levels [165]. This is a promising direction as it ensures diagnosis throughout the overall lifetime of the stack, but it requires knowledge of the current degradation state, and it involves considerable effort in model training and validation.
Eventually, the development of aging-robust fault diagnosis methods requires access to experimental data from

systems operating under fault and degradaded conditions, a scenario that is both costly and extremely challenging, especially given the current maturity level of PEMFC technology and deployment. From a broader perspective, we argue that fault diagnosis and prognosis should not be treated as independent tasks. Instead, prognosis insights should inform and enhance diagnosis decisions, particularly in non-stationary systems. While fully addressing this challenge requires knowledge and data that are either incomplete or not publicly accessible, the methodologies developed in this work are explicitly designed with the objective of enhancing the aging robustness of fault diagnosis algorithms.

3

Dataset and Physics-based Model

Contents

3.1 Introduction 24
3.2 The experimental faulty dataset 24
3.2.1 Analytical overview of the dataset 28
3.3 Physics-based model 36
3.3.1 Description 36
3.3.2 Calibration 38
3.3.3 Faults representation using the model 39
3.3.4 Insights on modeling error 42
3.4 Data-preprocessing: instructions and details 43
3.5 Summary 45

3.1 Introduction

As discussed in chapter 2, there exist three major categories for fault diagnosis methods: model-based, data-driven, or a hybrid combination of both. Each approach carries its own strengths and limitations. Model-based methods benefit from physical interpretability and generalization, but often struggle to capture the real-world behavior of the system due to limitations in the calibration procedure and of simplifying assumptions. Data-driven approaches, by contrast, can learn faulty patterns directly from experimental data but require large and high-quality datasets to be trained. In this work, we adopt a hybrid framework that leverages both physics-based modeling and experimental data, in order to benefit from the advantages of the two while compensating for their limitations. The purpose of this chapter is twofold. First, we introduce the experimental faulty dataset available for this study. This dataset is particularly valuable, as it includes a broad range of faults, also under aging conditions and dynamic load cycling. Thus, it enables a deeper understanding of how faults influence PEMFC performances, how aging affects the system's behavior, and how these two aspects combine. Second, we present the physics-based model used in this thesis, providing an overview of the underlying physical principles, of its capabilities, as well as of its limitations. While an exhaustive description of the model can be found in the literature [28], here we highlight how it can contribute in simulating PEMFC behavior under both normal and faulty conditions, eventually demonstrating its applicability. Based on the information that we present here, the subsequent chapters will explore how these two elements can be integrated and employed to develop hybrid, physics-informed, fault diagnosis methodologies for PEMFC systems.

3.2 The experimental faulty dataset

The experimental campaign was conducted at CEA in the framework of a past project [165]. The low-temperature PEMFC stack used has been assembled at CEA, and it is based on the design F developed in-house. It is a 1.6 kW water-cooled stack, composed of 20 cells with an active area of 220 cm^2. The MEA has been manufactured by SolviCore, and it has a platinum loading of 0.1 mg/cm^2 at anode and 0.4 mg/cm^2 at cathode. The membrane consists of Nafion and its thickness is 30 μm. BPs have been manufactured by shaping metal sheets. The anode plate is traversed by 12 parallel channels that serpentine through 5 passes. The cathode plate, on the other hand, is traversed by 20 parallel channels that serpentine through 3 passes. It is based on the closed-cathode technology. The gases used to sustain the electrochemical reactions are air and hydrogen. To run the experiment, the Pacific test bench, developed at CEA, was used. It is illustrated in Figure 3.1.



Figure 3.1: Pacific test bench at CEA [165].

This bench can accommodate PEMFC stacks with maximum power of 4 kW, and it can be used either with O2 or with air. Pressure can vary from 1×10^4 to 5×10^4 Pa, and the maximum tolerated pressure drop is 3×10^4 Pa. Electrically, the system is designed to operate under maximum allowable conditions of 100 V, 1000 A and

a total power output of 10 kW. The air flow rate at the cathode side must be regulated to maintain an oxygen stoichiometry of 2, and at anode side the flow rate has to be regulated to have a hydrogen stoichiometry of 1.5. The coolant flux is set at 9.8 L/min. A programmable logic controller of the ControlLogix type ensured the control and data acquisition from the sensors integrated into the fuel cell stack. Prior to initiating the experimental campaign for faulty dataset generation, the fuel cell stack underwent a conditioning phase. This process involved operating the stack under nominal conditions for a duration of 20 hours with a constant current density of 0.5 A/cm^2. During this period, the inlet gas relative humidity was set to 80 %, deviating from the standard 50 %. Eventually, the nominal operating conditions for the stack are given in Table 3.1

Variable	**Unit**	**Value**
Temperature	°C	80
Anode relative humidity	%	0
Cathode relative humidity	%	0
Anode relative humidity	%	0
O_2 stoichiometry	-	2
H_2 stoichiometry	-	1.5
O_2 inlet pressure	Pa	1.5e5
H_2 inlet pressure	Pa	1.5e5

Table 3.1: Nominal operating conditions for the PEMFC stack used in the experimental campaign.

The fuel cell stack was ultimately operated for a total of 1057 hours, during which the Fuel Cell Dynamic Load Cycle (FC-DLC) was applied to emulate real-world automotive driving conditions. One cycle, as shown in Figure 3.2, has a duration of 1180 s. The first 800 s represent urban usage, the last 380 s represent extra-urban usage.

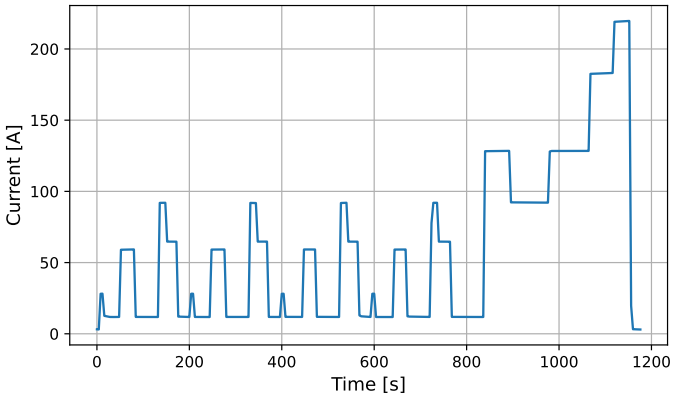


Figure 3.2: FC-DLC automotive load cycle profile.

Figure 3.3 presents the complete dataset, detailing the applied current load profile and the resulting stack voltage over a duration of more than 1000 hours. It is important to highlight that, prior to the initiation of the cycling phase with the FC-DLC, the stack underwent an extended characterization period lasting several hours. This is evident in Figure 3.4, where the typical step-wise profile of the current load during the acquisition of polarization curves can be observed. Throughout the test campaign, periodic characterization phases, including the acquisition of polarization curves, have been conducted to monitor the performance evolution and assess the degradation state of the PEMFC over time.

Fault injection

Throughout the operation of the stack, by adjusting the coolant temperature, the gases relative humidity and the gases' stoichiometry, the following faults have been injected: flooding, drying, air and hydrogen starvation. CO poisoning has not been considered in this experimental campaign. In detail, flooding has been triggered either

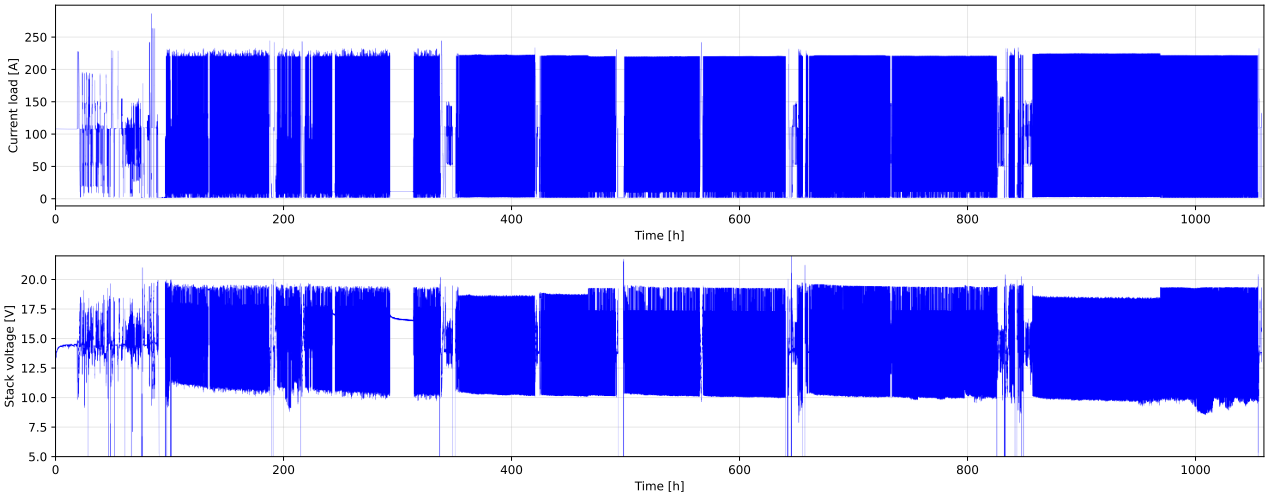


Figure 3.3: Current load and stack voltage recorded during the entire experimental campaign on the PEMFC stack.

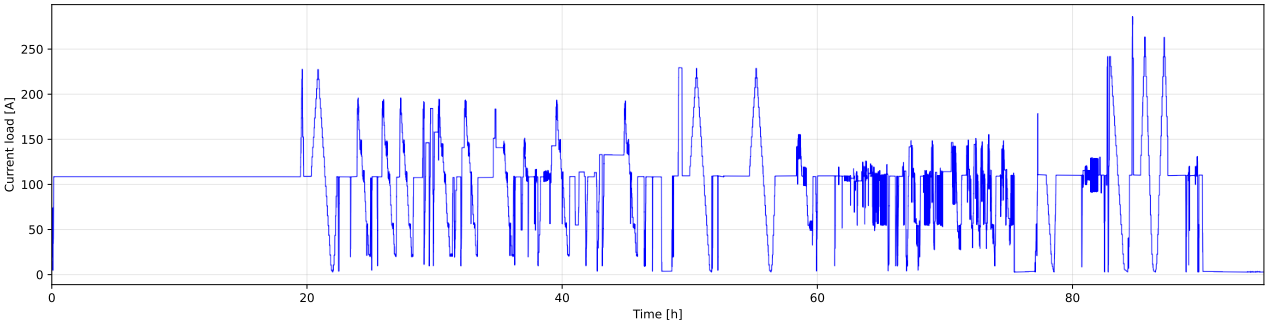


Figure 3.4: Load current during the conditioning and characterization phase of the PEMFC stack.

by lowering the coolant temperature from 80°C to 60°C, or by increasing the relative humidity, both at anode and at cathode side, from 50 % to 75 %. In this way, conditions necessary for liquid water formation inside the fuel cell stack were established. For what concerns drying, it has been triggered by increasing the coolant temperature up to 85 °C, or by lowering the anode and cathode relative humidity levels to 25 %. Subsequently, air starvation has been injected by lowering the air stoichiometry to 1.5, while hydrogen starvation by lowering the hydrogen stoichiometry to 1.2. These faults have been intentionally injected throughout the entire duration of the experimental campaign, at various operating times, with a variable duration. It is important to note that, although in practice the manifestation of faults can vary significantly depending on the magnitude, i.e., deviation from nominal conditions, or on the duration of the fault injection, this variability has not been explicitly considered when designing the experimental campaign. Thus, all the possible realizations of a fault were included in a single fault scenario.

Fault	**Injection mode**
Flooding	i) Anode RH = 75% & Cathode RH = 75 % ii) Coolant temperature = 60 °
Drying	i) Anode RH = 25% & Cathode RH = 25 % ii) Coolant temperature = 85 °
Air starvation	O_2 stoichiometry = 1.5
H_2 starvation	H_2 stoichiometry = 1.2

Table 3.2: Faults injection modes.

Monitoring variables

Several sensors have been installed on the test bench to monitor the PEMFC stack during the experimental campaign. The complete list of monitoring variables, with the corresponding measurement unit as recorded from sensors, is provided in Table 3.3. All variables have been continuously recorded for the entire duration of the experiment. Although the sensors operated at different sampling rates, a preprocessing step was applied to synchronize and resample the monitoring variables, yielding to a dataset with consistent time resolution of one sample every 2 s across all variables.

Variable	**Unit**
Anode pressure drop	mbar
Cathode pressure drop	mbar
Coolant flowrate main loop	L/h
Coolant flowrate stack output	L/h
Gas flowrate anode	NL/h
Gas flowrate cathode	NL/h
Current	A
Inlet pressure anode	bar
Inlet pressure cathode	bar
Outlet pressure anode	bar
Outlet pressure cathode	bar
Inlet coolant pressure	bar
Inlet H_2 relative humidity	%
Inlet air relative humidity	%
Outlet H_2 relative humidity	%
Outlet air relative humidity	%
H_2 stoichiometry	-
Air stoichiometry	-
Temperature anode inlet	°C
Temperature cathode inlet	°C
Temperature anode outlet	°C
Temperature cathode outlet	°C
Temperature coolant inlet	°C
Temperature coolant outlet	°C
Stack voltage	V
$i-th$ cell voltage, $i = 1, \ldots, 20$	V

Table 3.3: Monitoring variables during the experimental campaign.

Labeling procedure

Eventually, we performed a procedure to annotate the dataset with fault labels corresponding to the different faults injected in the PEMFC stack. We assigned a fault label to each data point based on the timing and the operating conditions, for each time step. Further classification according to the severity of the fault has not been possible, as a direct quantification of the faults magnitude is not explicitly available in the dataset. Thus, for the nominal condition and the faults injected, five classes are identified (each with a discrete label): nominal (0), flooding (1), drying (2), air starvation (3), hydrogen starvation (4). The distribution of the five classes is presented in Figure 3.5, which illustrates the occurrence of each fault across the entire duration of the experimental campaign. The resulting labeled dataset will serve as the ground truth for the development of the fault diagnosis algorithms, in a supervised framework, and for the subsequent evaluation of their performance.

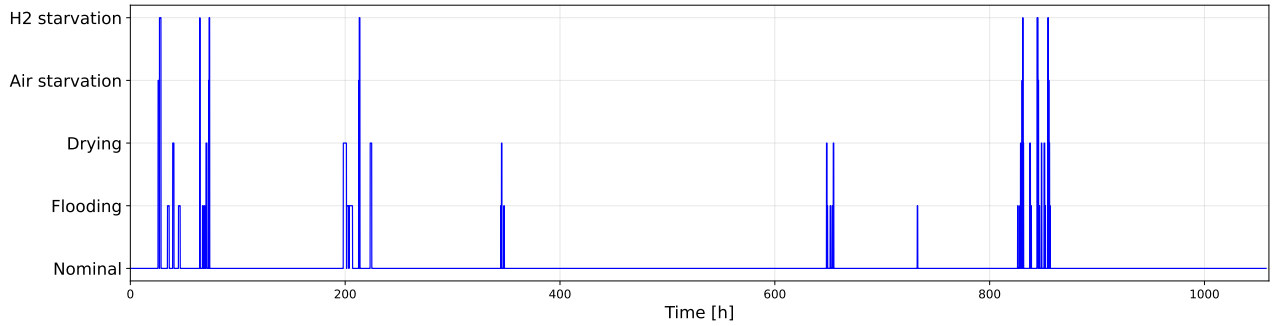


Figure 3.5: Distribution of the faults across the entire experimental campaign.

3.2.1 Analytical overview of the dataset

Dataset imbalance

One of the key challenges associated with the dataset is the highly imbalanced distribution of fault classes. The majority of the dataset consists of cycling the stack with FC-DLC cycles in nominal conditions, with a limited number of time intervals in which faults have been injected. This uneven distribution of the faults leads to an imbalanced dataset. In particular, the nominal class is heavily overrepresented, while the other fault classes account for a comparatively small number of samples. The extent of the imbalance is visually illustrated in the histogram in Figure 3.6, which shows the number of data samples associated with each fault class.

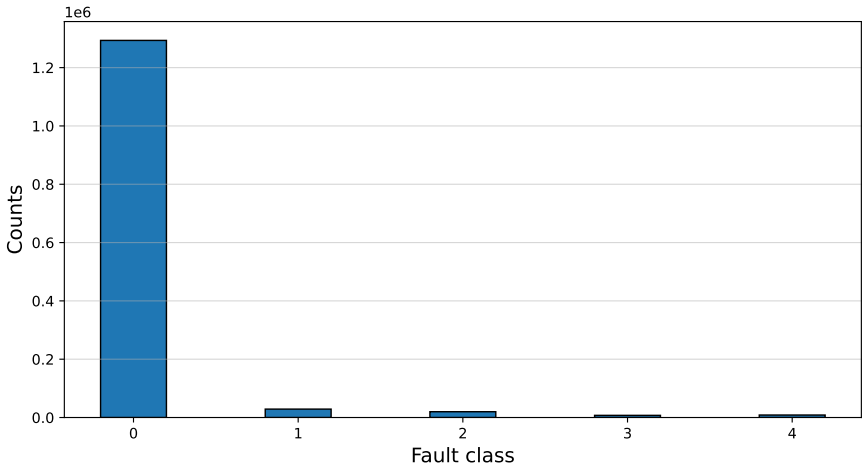


Figure 3.6: Number of occurrences for the different fault classes: nominal (0), flooding (1), drying (2), air starvation (3), hydrogen starvation (4).

Furthermore, the dataset structure is heterogeneous: in some cases, only FC-DLC cycles were applied; in others, polarization curves were acquired; occasionally, both were included. This results in a non-uniform fault class distribution, both in terms of their occurrence over time and the variety of fault types represented, which is clearly illustrated in Figure 3.7, Figure 3.8, Figure 3.9, Figure 3.10. For instance, if we consider the starvation faults, we can observe a huge time interval that goes from 250 hours to 800 hours, in which these faults are not represented. Additionally, no faults have been injected during dynamic cycles in the first 200 hours of stack operation.

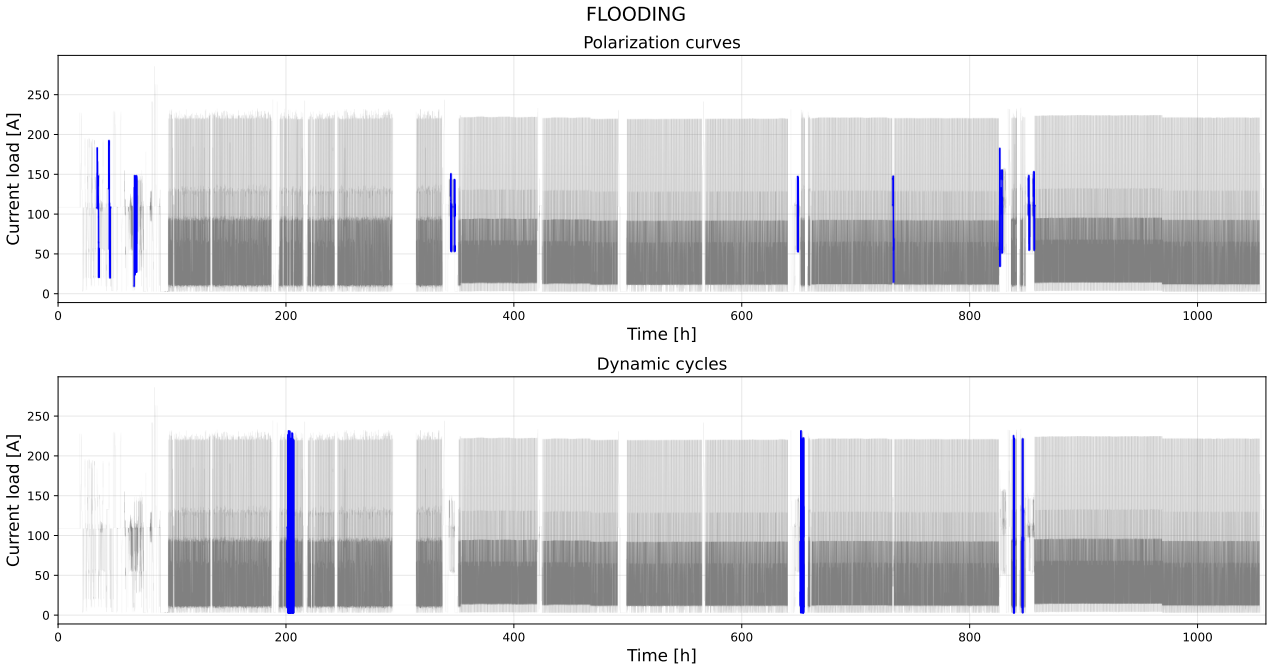


Figure 3.7: Distribution of flooding fault class, polarization curve (top) and FC-DLC (bottom). The injected fault is shown in blue.

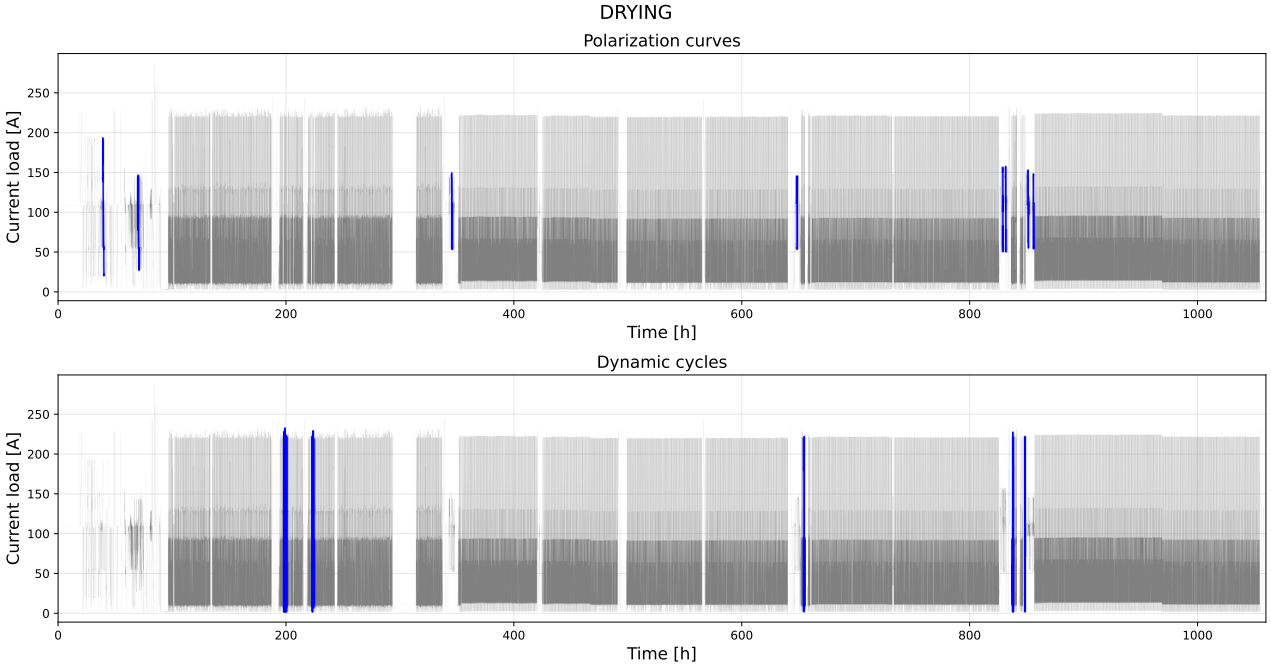


Figure 3.8: Distribution of drying fault class, polarization curve (top) and FC-DLC (bottom). The injected fault is shown in blue.

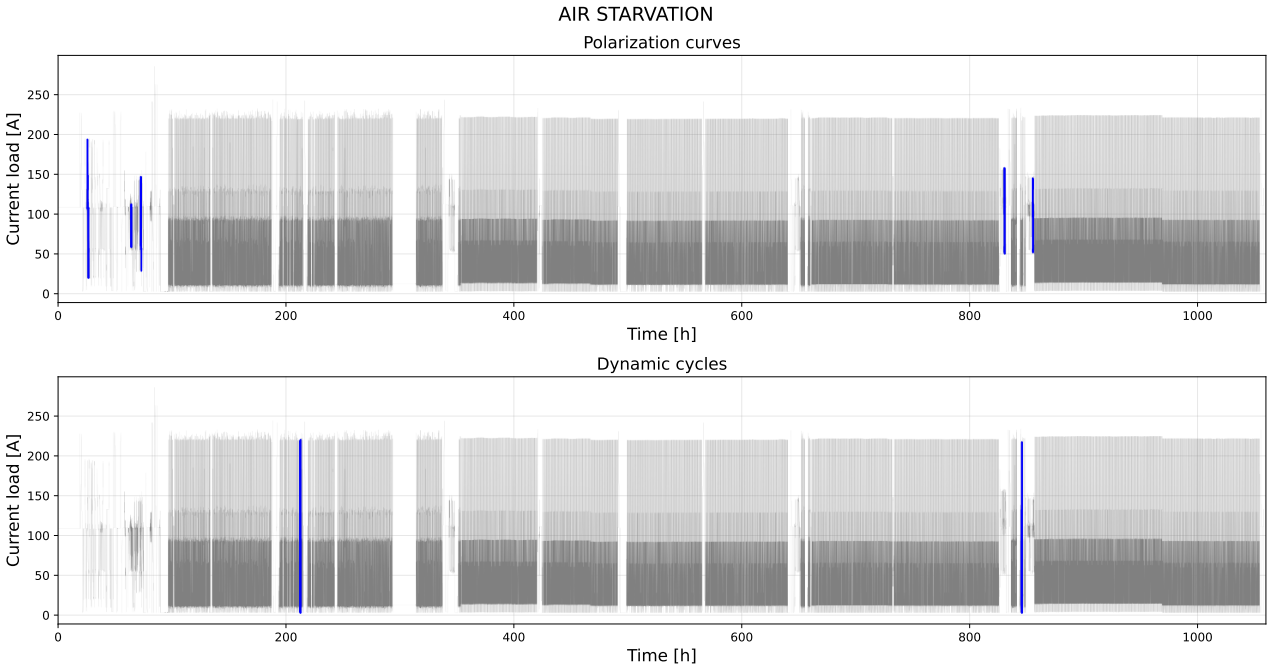


Figure 3.9: Distribution of air starvation fault class, polarization curve (top) and FC-DLC (bottom). The injected fault is shown in blue.

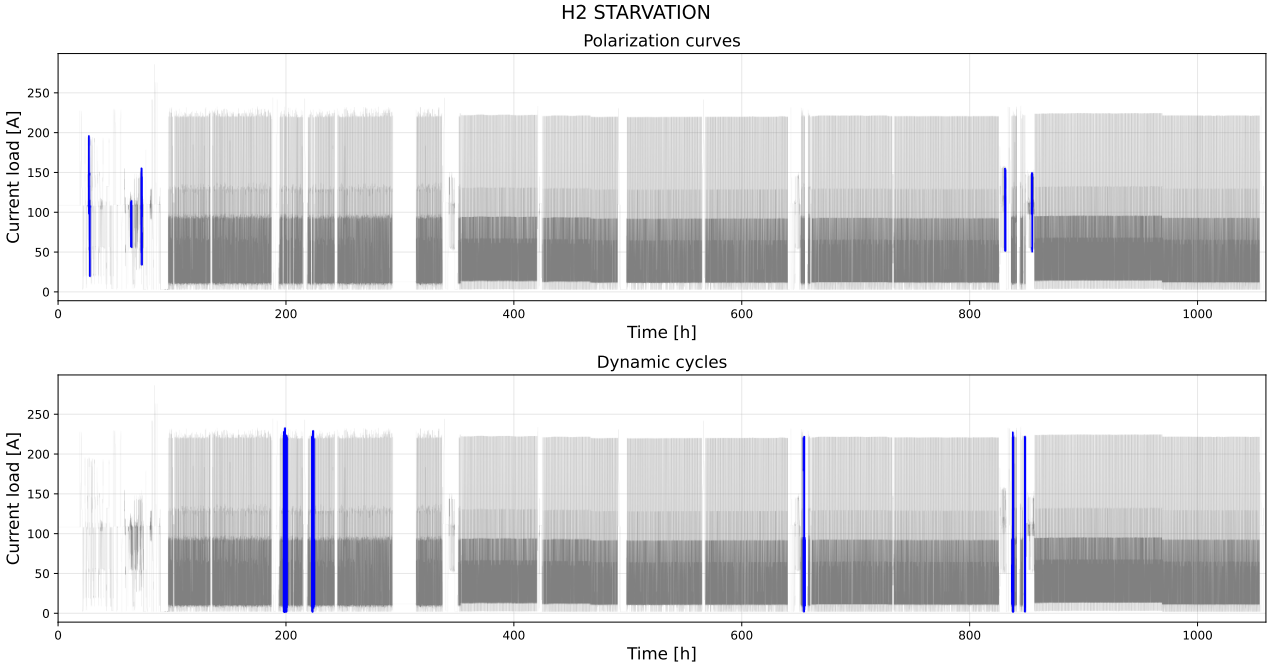


Figure 3.10: Distribution of H2 starvation fault class, polarization curve (top) and FC-DLC (bottom). The injected fault is shown in blue.

Inspecting the effect of faults on variable distribution

This subsection presents the behavior of the PEMFC under the different fault conditions. Polarization curves enable the quantification of the PEMFC performance. In Figure 3.11 we provide the polarization curves acquired during the characterization phase in the first 100 hours of stack operation, for the nominal condition and the faults. The best performance, i.e., highest voltage, is guaranteed when the stack is operated in nominal condition. As faults are injected, the stack voltage output decreases, reflecting the detrimental impact of these faults on the PEMFC performance.

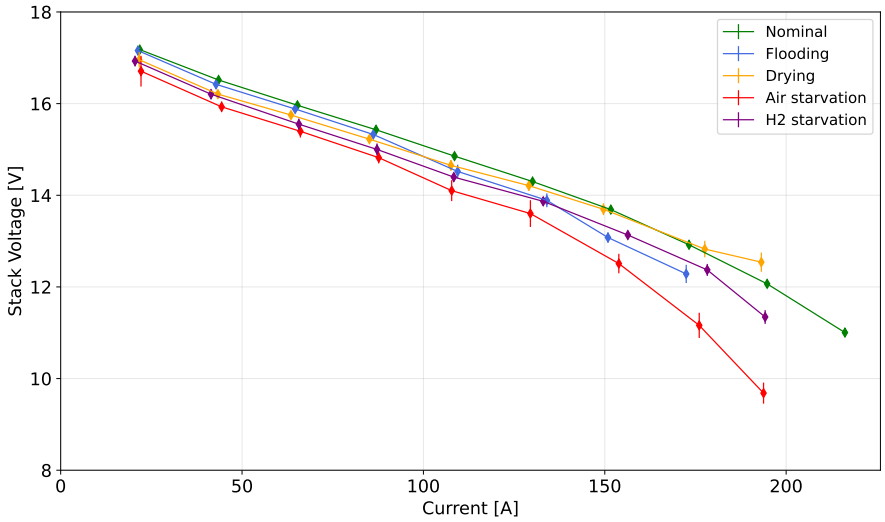


Figure 3.11: Polarization curve at stack beginning of life, for nominal condition (green) and faults: flooding (blue), drying (yellow), air starvation (red), hydrogen starvation (purple).

Additionally, EIS was occasionally performed to assess the health state and degradation of the PEMFC. Impedance measurements are widely acknowledged as powerful diagnostic tools for evaluating the state of PEMFC, and of electrochemical systems in general.

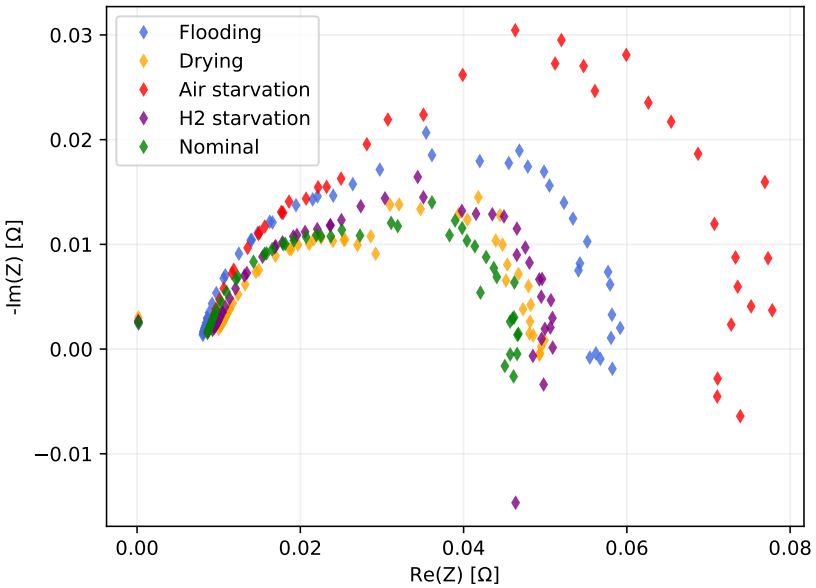


Figure 3.12: Nyquist diagram of the EIS measurements, at 0.5 A/cm2 current density, for nominal condition (green) and faults: flooding (blue), drying (yellow), air starvation (red), hydrogen starvation (purple).

Figure 3.12 presents the Nyquist diagram obtained for each fault during the characterization phase, in the first 60

hours of stack operation, at a current density of 0.5 A/cm^2. Indeed, different patterns can be observed for the different faults. Nonetheless, given the infrequent acquisition of these measurements, the temporal resolution of the data obtained through EIS is limited. Despite the valuable insights EIS can offer, its infrequent acquisition and lack of continuous temporal coverage led us to exclude these measurements from the development of the fault diagnosis strategy.

For the purpose of fault diagnosis, it is interesting to investigate the distributional behavior of key variables, i.e., stack voltage, anode and cathode pressure drop, across different fault classes. For each variable, we compare its distribution across fault classes using Kernel Density Estimation (KDE) plots. Unlike histograms, KDE plots provide a smoother, continuous representation of the underlying probability distribution of the data. KDE is a non-parametric method that estimates the probability density function of a variable by applying kernel functions, typically Gaussian, to smooth the empirical data distribution [166]. In Figure 3.13 we visualize the KDE plot for data of the first 100 hours of stack operation. To enhance clarity, the visualization is restricted to the following current values: 55 A, 110 A, 140 A, and 185 A, with a tolerance of ± 5 A. The plots illustrate how the physical behavior of the PEMFC evolves under faulty conditions, revealing distinct shifts in the distributions compared to nominal operation. To quantify the distance between distributions, we resort to the Kullback-Leibler (KL) divergence [167]. In Table 3.4 we provide the KL as a quantitative measure of the distance between the distribution of each fault and the distribution in nominal condition, emphasizing the effect of the faults on the different diagnosis signals.

KL divergence (Nominal - Fault)				
	Flooding	Drying	Air starvation	H2 starvation
Voltage	1.475	2.506	0.608	1.496
Pressure drop anode	1.743	2.346	1.381	2.707
Pressure drop cathode	1.932	4.632	3.789	2.088

Table 3.4: Kullback-Leibler divergence between nominal and faults distribution, for each variable.

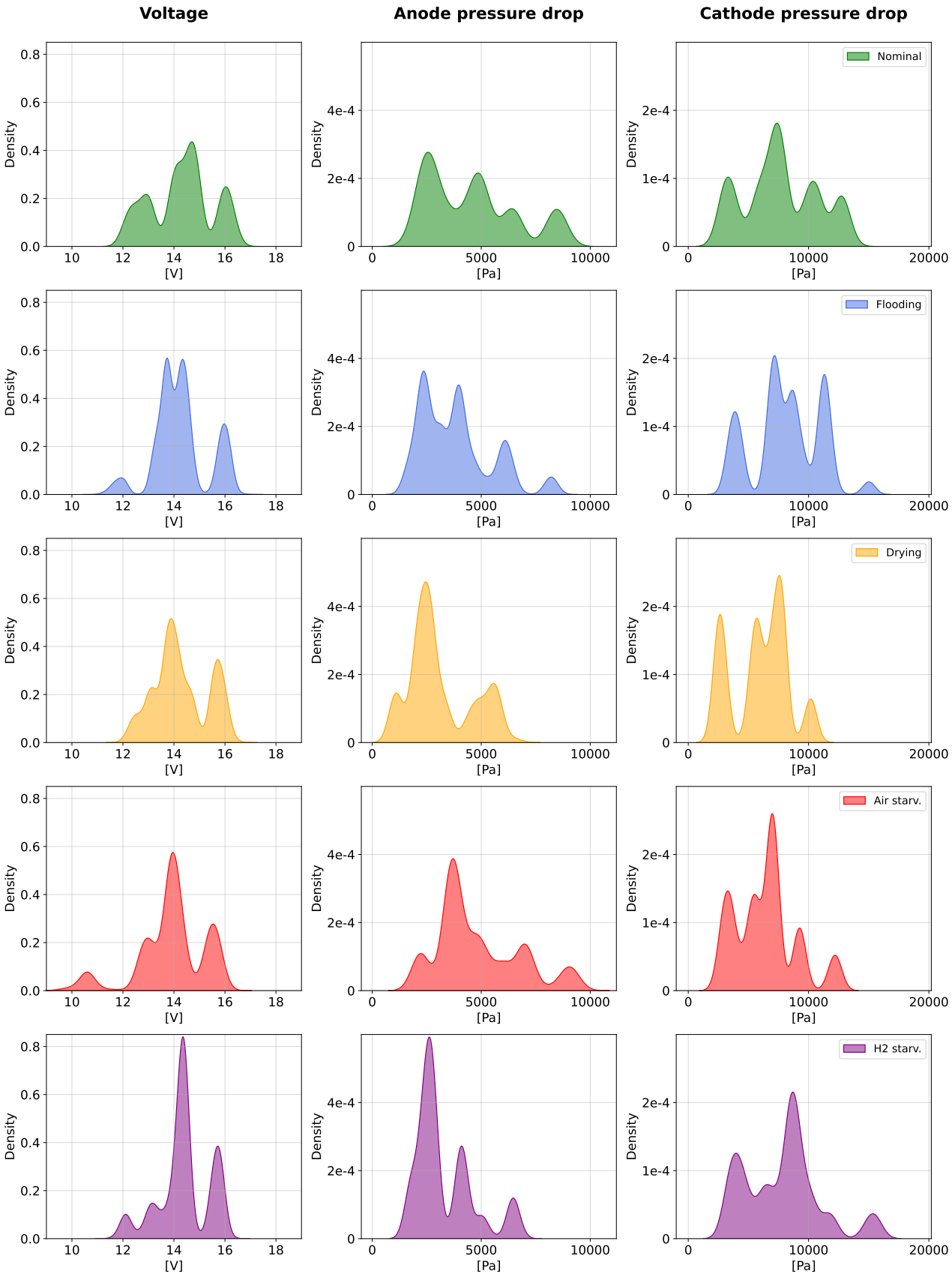


Figure 3.13: Kernel density estimation (KDE) plots for voltage, anode and cathode pressure drop, for the different fault classes. The following current values, with a tolerance of ± 5 A, are considered: 55 A, 110 A, 140 A, 185 A.

Inspecting the effect of aging on the faults

Eventually, the polarization curves acquired in nominal conditions at the different operating times are illustrated Figure 3.14, and the corresponding evolution of the stack power is given in Figure 3.15. The aging degrades the performance of the PEMFC in time, which implies reduced stack voltage, and, thus, a reduction in the power generated by the stack. If we consider for example a current value of 110 A, corresponding to a current density of 0.5 A/cm^2, we see that between the first polarization curve acquired at 21 hours and the last polarization curve recorded at 1055 hours, the cell voltage loss is 0.04 V while the overall stack voltage loss is 0.8 V, and the power loss is 54 W. This degradation is the major cause for the limited durability of PEMFCs.

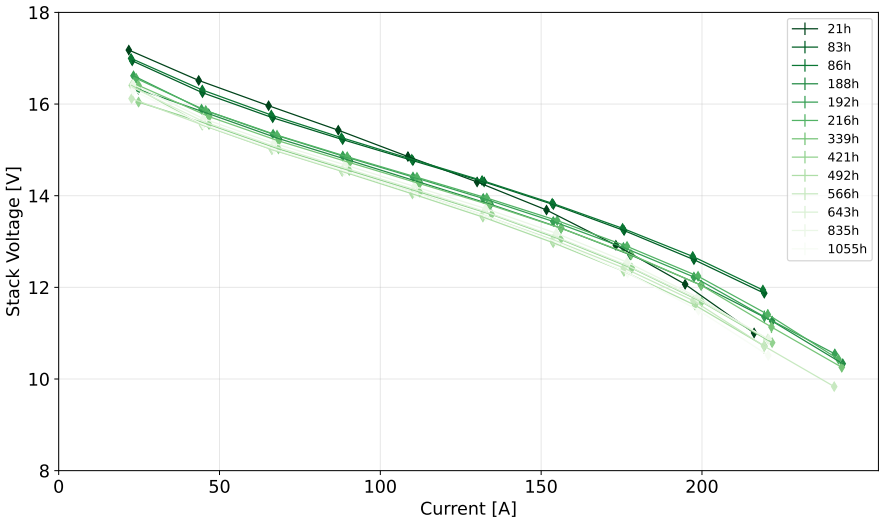


Figure 3.14: Polarization curves acquired at different operating time (nominal conditions).

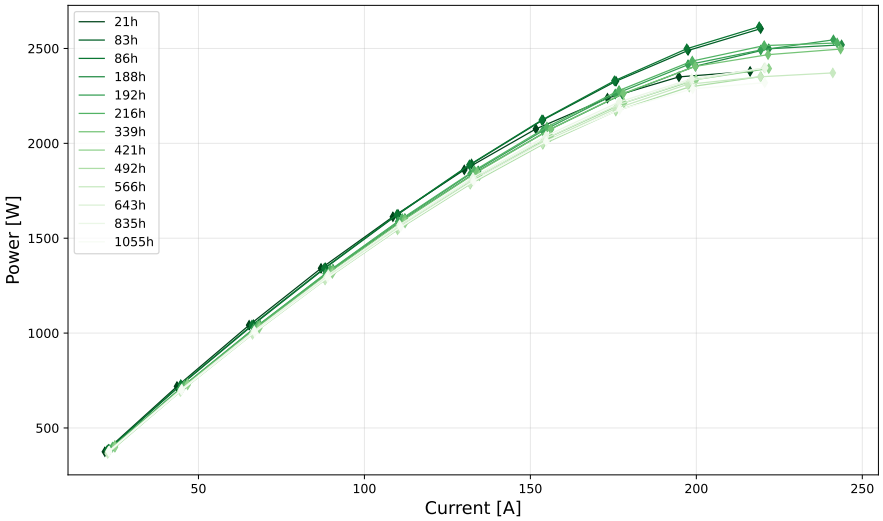


Figure 3.15: Stack power vs stack current, acquired at different operation times (nominal conditions).

However, as previously discussed, aging affects the performance of the stack in nominal conditions, as well as the impact of the faults in the system. To understand this, we consider the first 100 hours of operation of the stack, which we refer to as BoL, and the last recordings of stack operation, from 800 hours to the end, which we refer to as EoL. In Figure 3.16 we present KDE density plots comparing the distribution at BoL and EoL, for the following current values: 55 A, 110 A, 140 A, and 185 A, with a tolerance of ± 5 A.
A shift in the distributions is observed between BoL and EoL, confirming that the aging alters the behavior of the stack in nominal condition and, also, under faulty conditions. We observe that the voltage and the pressure drops, particularly at anode, decreases with aging. This is counterintuitive: the PEMFC is expected to lose

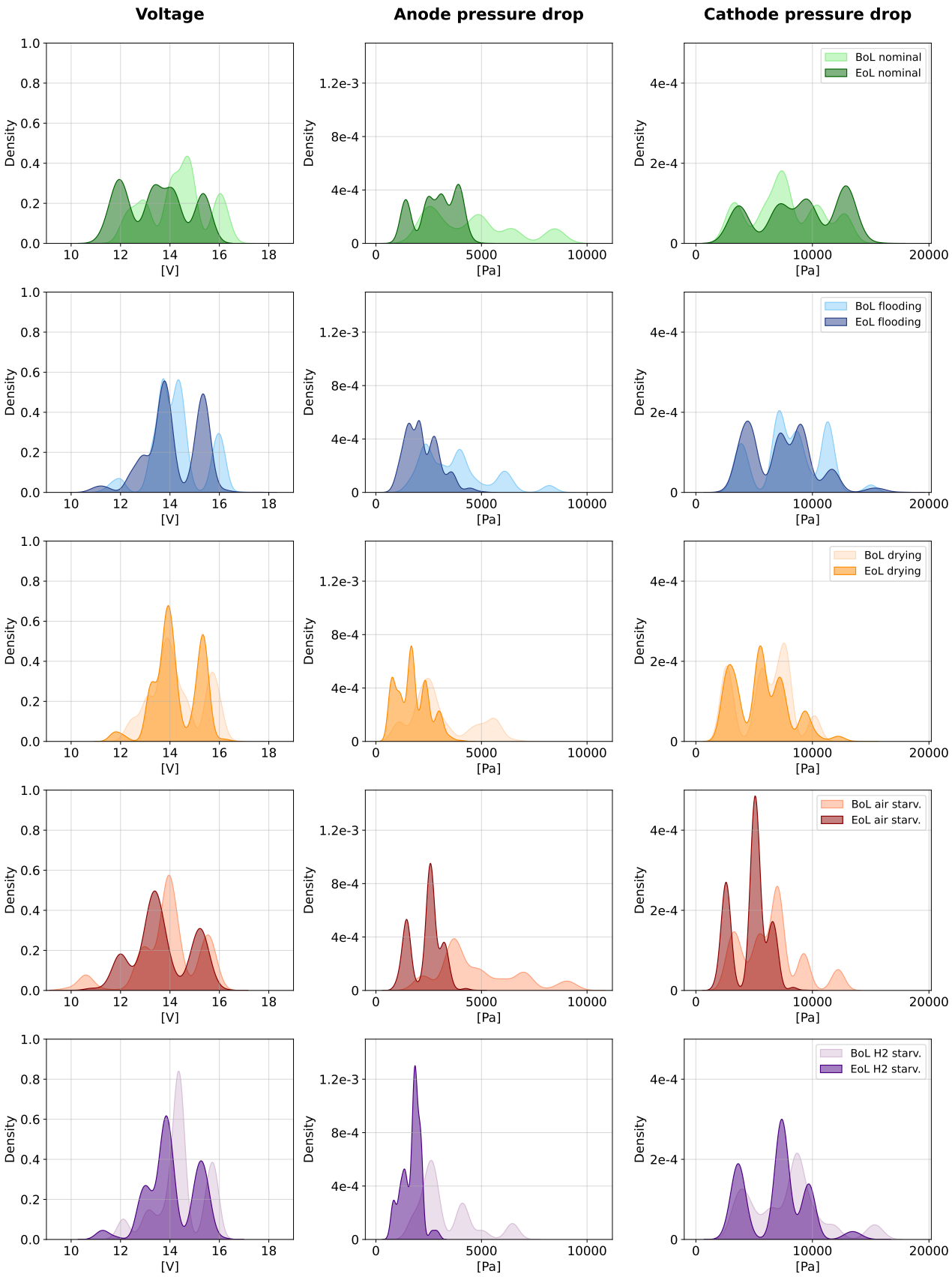


Figure 3.16: Kernel density estimation (KDE) plots for voltage, anode and cathode pressure drop, for the different fault classes available in the dataset, at beginning-of-life and end-of-life. The following current values, with a tolerance of ± 5 A, are considered: 55 A, 110 A, 140 A, 185 A.

hydrophobicity, leading to increased liquid water accumulation within the stack and, consequently, a higher pressure drop. The observed decrease in pressure drop suggests that flooding issues may have occurred during the early stages of the experimental campaign, potentially masking the expected trend. To quantify these distribution shifts, we calculate the KL divergence between the density distributions at BoL and EoL, as shown in Table 3.5.

	KL divergence (BoL- EoL)				
	Nominal	Flooding	Drying	Air starvation	H2 starvation
Voltage	0.429	0.749	0.962	0.733	1.272
Pressure drop anode	3.724	1.627	3.349	4.875	4.604
Pressure drop cathode	0.984	0.366	0.611	3.333	1.761

Table 3.5: Kullback-Leibler divergence between BoL and EoL, for each variable, for each fault.

The results indicate that the anode pressure drop exhibits the largest distribution shift, with KL divergence values consistently above 1. This supports the hypothesis of water management issues within the stack at BoL. The large KL values suggests that distinguishing faults will become increasingly challenging as the stack ages.

3.3 Physics-based model

3.3.1 Description

The PEMFC model used in this work is called MePHSYTO-FC. It has been developed in the past years in CEA. It is a multi-physics model based on the physical laws of transport of reactants and reaction products, mass and heat transfer, and on electrochemical laws. It is based on ODE equations solved with a lumped model approach and pseudo bond-graph construction. The main assumption consists in modeling the stack as a single equivalent cell, embodying the same characteristics of the individual cells constituting the stack. Therefore, the individual cells are assumed to behave identically, and temperature boundary conditions within the cells are neglected, with the model considering only thermal exchange between the terminal plate and the bipolar plate. This is a 0D+2D model, as it is illustrated in Figure 3.17. The in-plane flux is modelled in 2D, i.e., the cell surface is divided in meshes, while the through-plane flux, i.e., through the membrane, is modelled in 0D. This implies that the transport phenomena in the GDL and CL are modelled as linear. The model is modular, built with balance blocks (blocks C, which account for energy and mass balances) and transport blocks (blocks R, which account for electrical, heat, gas and water transport).
Specifically, in the BP the physics-based model accounts for:

- Gas and water transport, with Darcy's law, and mass balances along the channels;
- Electric transport and heat transfer;
- Transport of the coolant in the cooling circuit, and heat balance.

In the membrane, the following phenomena are considered:

- Water and heat transfer, as well as mass balance and gas permeation through the membrane;
- Transport by convection and diffusion, with mass and energy balance in the GDL;
- Electrochemical response of the system.

Eventually, for the overall stack, the heat transfer in the cooling circuit and in the terminal plates is modelled. To estimate the PEMFC performance, i.e., the output voltage, the electrochemical response of the system is computed with a semi-empirical equation, based on the Butler-Volmer law [168]:

$$U_{\text{rev}} = E_{\text{rev}} + \eta_{\text{act}} - R_{\text{ohm}} \cdot i_d \tag{3.1}$$

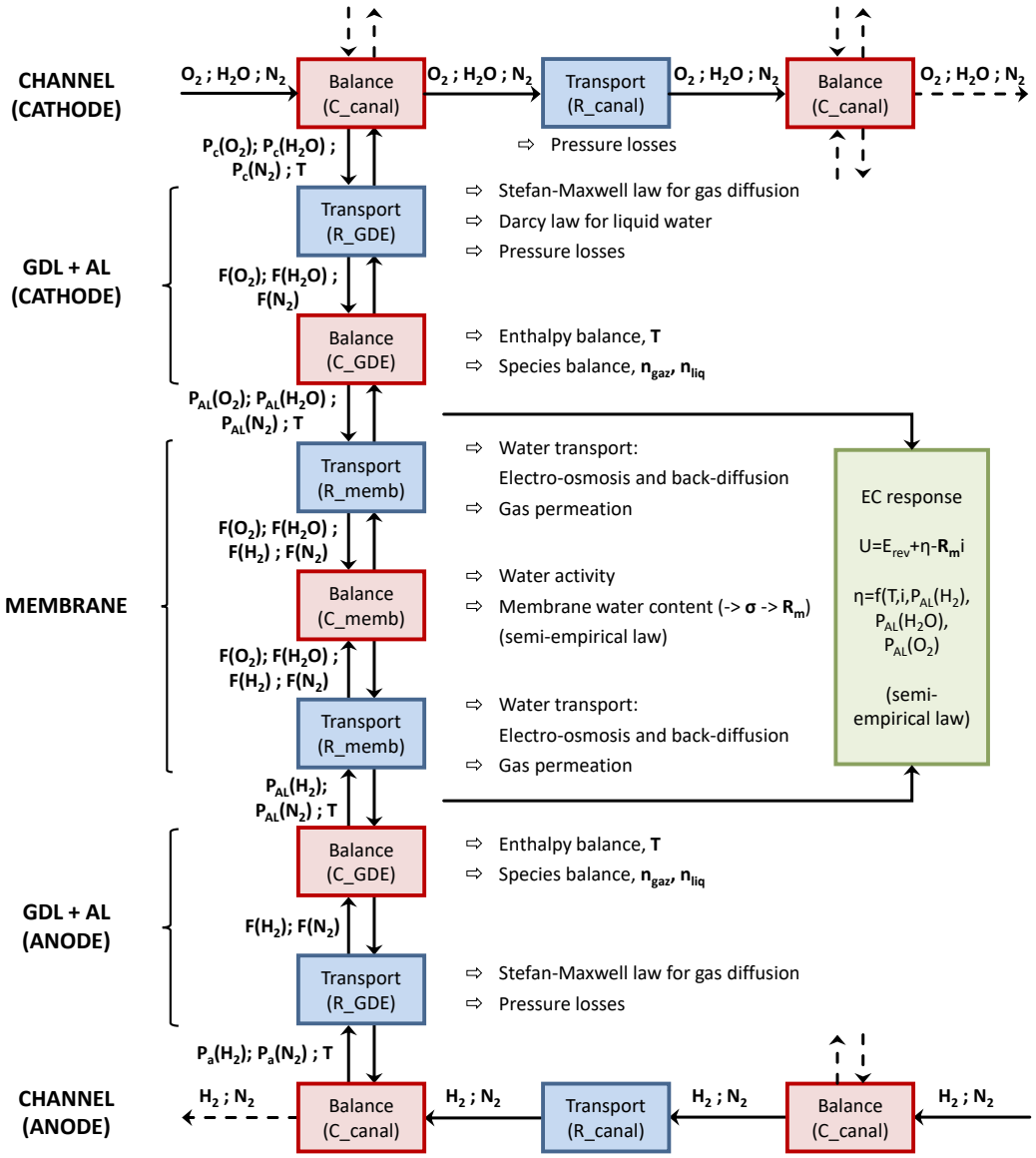


Figure 3.17: Schematic of the physics-based PEMFC model (MePHYSTO).

where E_{rev} (V) is the reversible thermodynamic voltage, η_{act} (V) represents the activation losses, R_{ohm} ($\Omega \cdot m^2$) is the ohmic resistance, i_d (A/m^2) is the local current density. The current density i_d corresponds to:

$$i_d = \frac{I}{S_{\text{act}}} \tag{3.2}$$

where I (A) is the current and S_{act} (m^2) is the active surface of the cell. The reversible voltage is given by:

$$E_{\text{rev}} = E^0_{\text{rev}} + \frac{\Delta S_0}{nF}(T - T_0) + \frac{RT}{nF} \cdot \ln\left(\frac{P^*_{H_2}}{P_0}\left(\frac{P^*_{O_2}}{P_0}\right)^{1/2}\right) \tag{3.3}$$

where E^0_{rev} is the reversible potential in standard conditions, i.e., temperature $T_0 = 25$ °C and pressure $T_0 = 1$ atm, ΔS_0 is the entropy in standard conditions, T is the temperature, $P^*_{H_2}$ and $P^*_{O_2}$ are the partial pressures of hydrogen and oxygen, respectively. Then, the activation losses η_{act} are computed by means of the following semi-empirical law:

$$\eta_{\text{act}} = \beta_1 + \beta_2 \cdot T + \beta_3 \cdot T \cdot \ln(i_d) + \beta_4 \cdot T \cdot \ln\left(\frac{P^*_{O_2}}{P^0}\right) + \beta_5 \cdot T \cdot \ln\left(\frac{P^*_{H_2}}{P^0}\right) + \beta_6 \cdot T \cdot \ln\left(\frac{P^*_{H_2O}}{P^0}\right) \tag{3.4}$$

where $P^*_{H_2O}$ is the partial pressure of water steam and β_k, $k = 1, ..., 6$ are semiempirical coefficients describing the electrochemical response at the electrodes, as function of the local operating conditions (local partial pressures of the gases, water steam, temperature. . .). The ohmic resistance is given by the contribution of the membrane resistance R_m and contact resistance R_c:

$$R_{\text{ohm}} = R_m + R_c \tag{3.5}$$

A simplified description of two-phase flow phenomena is adopted in the model. The presence of liquid water in the channels can prevent the gases to reach the active sites for the reactions, thus it influences the electrochemical response. Accordingly, the impact of the liquid water on the PEMFC performance is described as a temporary loss of S_{act}. The following empirical model describes the variation of the cell active surface:

$$S'_{\text{act}} = \frac{S_{\text{act}}}{(s_{\max} - s)^{\gamma}} \tag{3.6}$$

where S'_{act} is the corrected active surface, s is the liquid saturation, $s_{\max}$ and γ are two empirical parameters. This correction entails a variation of the current density i_d and, as a consequence, of the total output voltage. Eventually, an aging model is established to account for the degradation of the PEMFC in time. Several studies highlighted that the major contributions to fuel cell aging are given by catalyst and membrane degradation [34], both of which are described in MePHYSTO. However, since membrane degradation has a secondary impact with respect to catalyst degradation on the system performance [169], in this work we focus only on the latter. Specifically, only platinum dissolution is considered. While physical models, e.g., Ostwald ripening models [170], are available to describe the platinum dissolution, these are computationally expensive. Accordingly, we resort to a light-weight empirical model that captures the macroscopic effect of platinum dissolution: the loss of ECSA. Hence, S_{act} is adequately modified: $S'_{\text{act,deg}} = S'_{\text{act}} \cdot \tau_{\text{deg}}$, where τ_{deg} is the degradation rate. Since platinum dissolution is exponentially increasing in time [171], it is straightforward to describe the degradation rate as $\tau_{\text{deg}} = a \cdot t^{-b}$ where t is the operation time, and a and b are two empirical parameters. Then, equation (3.4) becomes:

$$\eta_{\text{act}} = \beta_1 + \beta_2 \cdot T + \beta_3 \cdot T \cdot \ln\left(\frac{i_d}{\tau_{\text{deg}}}\right) + \beta_4 \cdot T \cdot \ln\left(\frac{P^*_{O_2}}{P^0}\right) + \beta_5 \cdot T \cdot \ln\left(\frac{P^*_{H_2}}{P^0}\right) + \beta_6 \cdot T \cdot \ln\left(\frac{P^*_{H_2O}}{P^0}\right) \tag{3.7}$$

3.3.2 Calibration

While MePHYSTO provides a general description for a PEMFC system, to employ it effectively it must be adapted for the desired fuel cell technology. The model requires calibration against empirical data and adjustment of key parameters to reflect the operational behavior of the real PEMFC stack. Among these parameters, there are physical properties of the membrane, of the CLs, of the BPs and of the GDL, available from manufacturer specifications. The number of meshes for the 2D cell plates description are critical parameters to be adjusted. While a high number of meshes implies high precision in the local conditions, this comes with a drawback on the computational cost. Relying on prior studies [61], we set one mesh to represent the coolant circuit, and 10 meshes to represent the reactants flow in the channels. As a result, temperature is considered uniform along the circuit, while a gradient is modelled for the partial pressures of the reactant gases, allowing the estimation of the current density distribution.

The parameters β_k, $k = 1, ..., 6$ used in equation (3.4) have a significant influence on the model prediction of the PEMFC electrochemical response. Due to their critical importance, a dedicated calibration procedure has been developed [165]. This procedure employs an iterative optimization approach utilizing experimental polarization curves acquired under different operating conditions, including varying temperature, relative humidity and gases stoichiometry. Incorporating polarization curves measured under a large range of operating conditions ensures that the model captures the system's behavior comprehensively, across various scenarios. The calibration process

is initiated by assigning initial estimates to the β_k parameters. Based on these provisional values, the local conditions, i.e., temperature and pressures, within the fuel cell are estimated, and the corresponding stack voltage is computed. The optimization consists in finding the set of β_k that minimizes the discrepancy between simulated and experimental polarization curves. The iterative procedure is repeated until the modeling error satisfies the convergence criterion. Although the procedure can be computationally demanding and may require a substantial number of iterations, it is performed only once prior to model deployment. The set of β_k identified with the calibration procedure are given in Table 3.6.

β_1	β_2	β_3	β_4	β_5	β_6
0.1605	$-2.405e-03$	$-1.320e-04$	$1.025e-04$	$1.909e-04$	$-9.070e-05$

Table 3.6: Set of β_k calibrated.

The average root mean squared error (RMSE) between the simulated and the experimental stack voltage is 0.416 V, corresponding to 20.8 mV for the single cell voltage. This error is relatively low and was considered acceptable for the purposes of this study. Thus, the calibration procedure has been validated.

3.3.3 Faults representation using the model

If we want to use the physics-based PEMFC model to support the development of the hybrid fault diagnosis algorithm, we have to be sure that it is capable of simulating the fuel cell behavior in case of a fault. To assess the model capabilities, we started by injecting faults in different conditions and analyzing the model response. The objective is to evaluate whether the model can capture some variations, or shifts, in the estimated signals when faulty conditions are injected. This is intended to be a qualitative evaluation of the model capabilities, without focusing extensively on the modeling error. Indeed, some of the model estimates that we are going to consider, e.g., membrane water content or liquid saturation, are non-observable states for the PEMFC, so that the experimental equivalents are not available for comparison. However, they are critical to understand the effect and impact of the fault injection. This evaluation is performed for flooding, drying, air starvation and hydrogen starvation. For each of these faults, we simulate the PEMFC behavior for 500 s, and the fault is injected at 250 s. A static current density load of 0.5 A/cm^2 is imposed.

Flooding

Flooding in PEMFC refers to the accumulation of liquid water within the gas flow channels or at the electrodes. One of the primary factors that can induce flooding is a reduction in system temperature, which promotes water condensation within the stack. To investigate the impact of flooding, we simulated a scenario in which the coolant temperature undergoes a step change from its nominal operating value of 80°C down to 60°C. The dynamic response of the PEMFC model to this temperature perturbation is illustrated in Figure 3.18. As shown in Figure 3.18a, the cell voltage begins to decline shortly after t = 250 s. This behavior is due to the onset of liquid water accumulation within the electrodes, which impedes the transport of reactant gases to the CLs, thereby reducing the number of available sites for the electrochemical reaction. Figure 3.18b shows that this effect is more prominent at the cathode, where water is generated as a by-product of the oxygen reduction reaction. Also, in Figure 3.18c we observe a simultaneous increase in the membrane water content, confirming the presence of excess liquid water within the stack as a result of flooding injection. Overall, the simulation demonstrates that the model can capture how liquid accumulation hinders reactant transport to the catalyst sites, thereby limiting the electrochemical reaction during PEMFC flooding.

Drying

Drying arises when the hydration level of the electrolyte membrane is insufficient to sustain effective proton conduction. One of the most common causes of membrane dehydration is inadequate humidification of the inlet reactant gases. In this scenario, the membrane gradually loses its water content. This impacts the proton conductivity, leading to a significant increase in ohmic resistance, and to a decrease in cell voltage. Eventually, if the membrane becomes too dry, irreversible damages, like pinholes formation, may occur. To analyze the

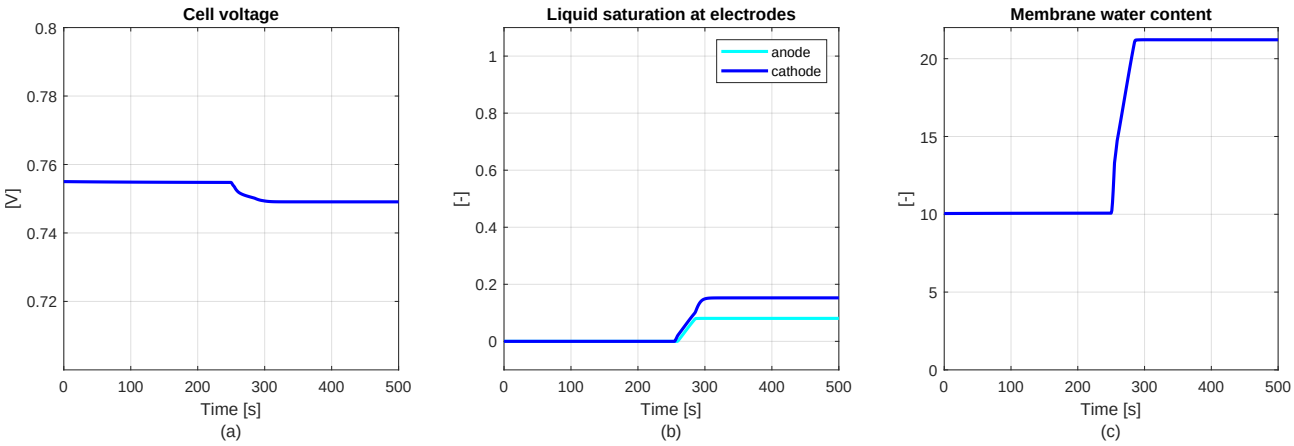


Figure 3.18: Flooding injection: coolant temperature varies from 80°C to 60°C. Cell voltage (a), liquid saturation at electrodes (b), membrane water content (c).

effects of drying, we conducted a simulation in which the relative humidity of hydrogen and air at the inlets is reduced from 50% to 0. In Figure 3.19 the model response is presented. Following the removal of humidification, a progressive decline of the cell voltage is observed, which has been induced by the decrease in membrane water content. The drop in the membrane water content corresponds to an increase in membrane resistivity. Interestingly, the voltage response exhibits a relatively slow transient behavior despite the rapid change in membrane hydration. This can be attributed to competing electrochemical and thermal processes occurring when drying is injected. In general, the simulation demonstrated that the model can capture some of the phenomena related to drying of PEMFC stack.

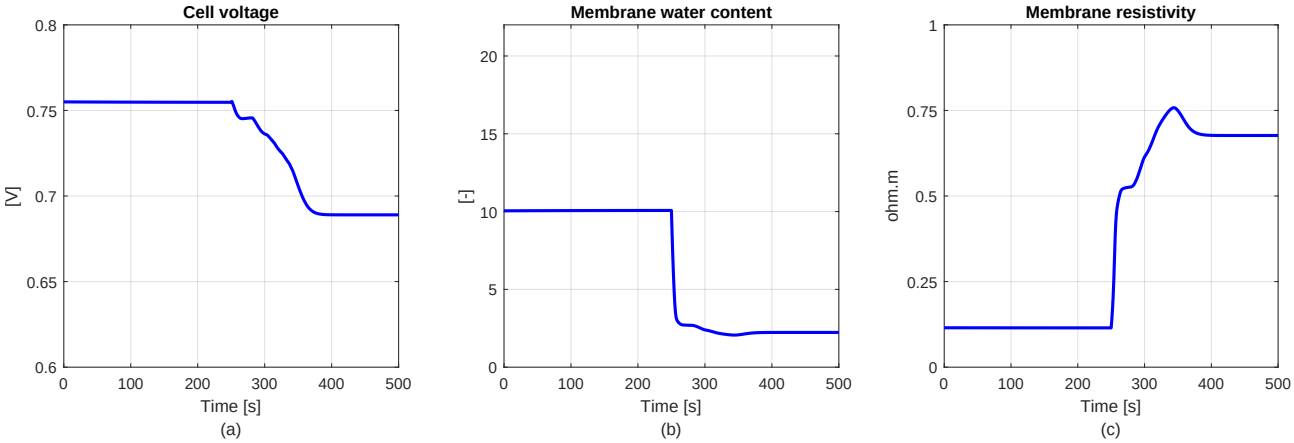


Figure 3.19: Drying injection: inlet gases relative humidity drops to 0. Cell voltage (a), membrane water content (b), membrane resistivity (c).

Air starvation

Air starvation arises when the supply of oxygen to the cathode is insufficient to sustain the electrochemical reaction at the desired load demand. This induces strong heterogeneity in the current density distribution, exhibiting higher values at the inlet compared to the outlet. Then, the rate of the electrochemical reactions decreases, which leads to a drop in stack voltage. In addition, issues related to water management can emerge. Indeed, sufficient airflow is necessary to remove the water produced in the oxygen reduction reaction. When the airflow rate is too low, liquid water tends to accumulate and may lead to localized flooding. To assess the model capability to represent this behavior, we reduce the air stoichiometry from 2.0 to 1.5. The resulting PEMFC response is illustrated in Figure 3.20: the cell voltage experiences a sharp decline when the fault is injected, accompanied by an increase in the liquid water saturation at the cathode side. Furthermore, in Figure 3.20c we observe the current density distribution along the channel. The current density experiences a gradient,

being higher near the inlet and significantly lower towards the outlet, and, when air starvation is injected, this distribution varies.

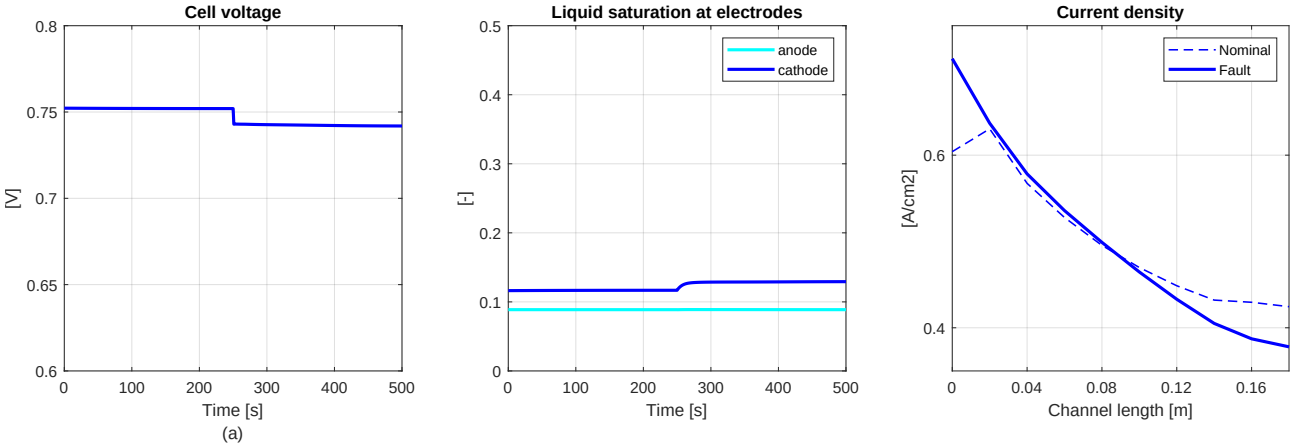


Figure 3.20: Air starvation injection: air stoichiometry drops from 2 to 1.5. Cell voltage (a), liquid saturation (b), current density distribution along the channel, before and after fault injection (c).

Hydrogen starvation

Analogously to air starvation, hydrogen starvation occurs when the hydrogen supply at the anode is insufficient to sustain the electrochemical reaction at the desired load demand. Theoretically, a hydrogen stoichiometry of 1.0 should be adequate, since it corresponds to the exact molar requirement for the reaction. However, a stoichiometry of 1.5 is employed in PEMFC to ensure consistent reactant availability and mitigate transport losses. To explore how the model captures hydrogen starvation, we simulate a scenario in which the hydrogen stoichiometry is reduced from 1.5 to 1.1. As illustrated in Figure 3.20a, the cell voltage exhibits a transient response immediately following the fault injection; however, the system quickly stabilizes, and the steady-state voltage remains only slightly below the nominal value. One possible explanation lies in the fact that the membrane resistivity shows only a marginal increase. Moreover, the current density distribution along the flow channel do not show significant spatial variation. The limited impact observed suggests that the model struggles to capture pronounced hydrogen starvation effects when the hydrogen stoichiometry remains above the theoretical threshold of 1.0. Indeed, hydrogen starvation involves complex, multi-physical interactions, such as local gas transport limitation and reverse current mechanism at the outlet, that are not fully described by the model. As a result, no significant impact on the PEMFC performance predicted by the model is observed as long as the stoichiometry is maintained above 1.0.

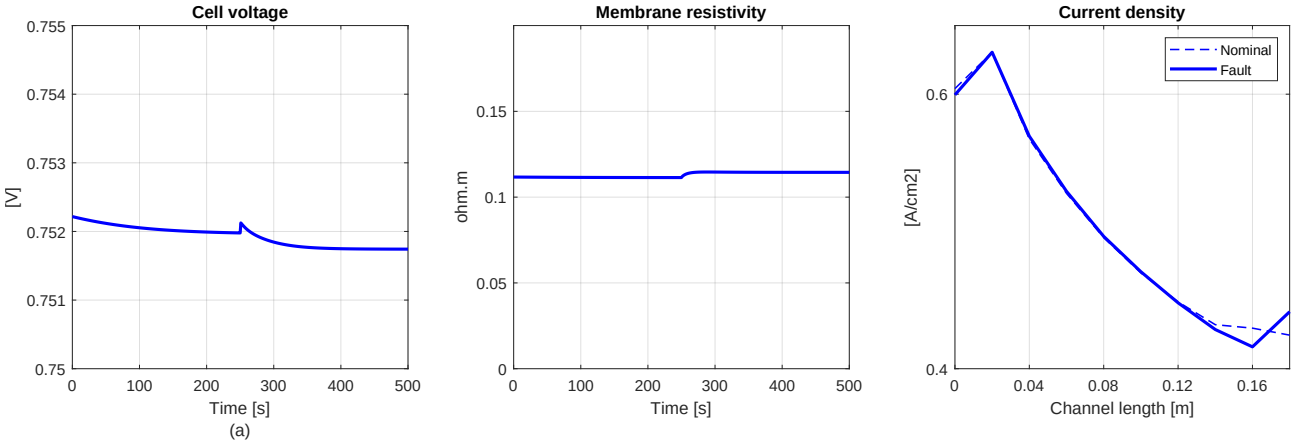


Figure 3.21: Hydrogen starvation injection: hydrogen stoichiometry drops from 1.5 to 1.1. Cell voltage (a), membrane resistivity (b), current density distribution along the channel, before and after fault injection (c).

3.3.4 Insights on modeling error

Following the calibration and validation of the physics-based PEMFC model, we proceed to quantitatively evaluate its accuracy under dynamic load conditions. This assessment is conducted by comparing the model outputs to the experimental measurements collected from the PEMFC stack during operation. It is important to note that this comparison can only be performed for state variables and outputs that are directly measurable from the physical system. Indeed, while the model is capable of estimating internal and unobservable variables, such as membrane water content or internal liquid saturation levels, these quantities are not available from experimental measurements and, thus, cannot be validated directly. Therefore, the validation focuses solely on the subset of variables that were acquired from the experimental campaign. Specifically, for the purposes of fault diagnosis, the most relevant variables include the stack voltage and the pressure drops across the anode and cathode channels. Figure 3.22 presents the time-domain comparison between the experimental measurements and the corresponding variables predicted by the model for a dynamic current load profile (segment of FC-DLC) in nominal conditions.

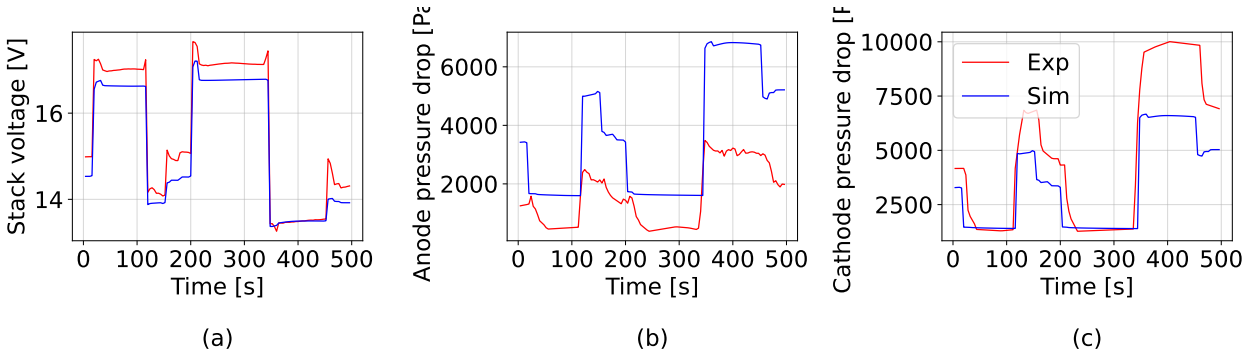


Figure 3.22: PEMFC response for a segment of the FC-DLC cycle: voltage, anode and cathode pressure drop.

The model estimates demonstrate good agreement with the experimental measurements, capturing the dynamic trends of the voltage and pressure signals, with minor discrepancies during rapid transients. Nonetheless, positive or negative effects are observed in the steady-state intervals, which are due to limited accuracy. To quantitatively assess the modeling error, we adopted three standard performance metrics:

- Mean Absolute Error (MAE), which captures the average magnitude of the absolute errors;
- RMSE, which gives greater weight to larger deviations and is sensitive to transient mismatches;
- Mean Absolute Percentage Error (MAPE), which expresses the error relative to the magnitude of the measured values and is useful for comparing across variables with different scales.

These metrics were computed for each fault class as well as for nominal class, and the results were averaged over the entire duration of the dynamic test. The results are reported for the stack voltage, anode pressure drop and cathode pressure drop, in Table 3.7, Table 3.8, Table 3.9, respectively.

	MAE [V]	**RMSE [V]**	**MAPE [%]**
Nominal	0.667	0.763	4.59
Flooding	0.991	1.286	10.85
Drying	0.535	0.654	3.63
Air starvation	0.926	1.129	6.53
H2 starvation	0.633	0.766	4.26
average	0.750	0.920	5.97

Table 3.7: PEMFC model error: stack voltage.

	MAE [Pa]	RMSE [Pa]	MAPE [%]
Nominal	1180.8	1389.2	59.3
Flooding	1052.6	1282.4	71.4
Drying	1519.5	1788.7	136.7
Air starvation	1325.1	1547.0	47.7
H2 starvation	947.5	1190.0	52.1
average	1205.1	1439.5	73.5

Table 3.8: PEMFC model error: anode pressure drop.

	MAE [Pa]	RMSE [Pa]	MAPE [%]
Nominal	1681.9	2167.6	27.4
Flooding	2702.9	3214.7	54.1
Drying	1113.3	1548.2	27.2
Air starvation	1669.7	1986.4	32.0
H2 starvation	2046.4	2536.9	30.0
average	1842.8	2290.8	34.2

Table 3.9: PEMFC model error: cathode pressure drop.

The modeling errors associated with the stack voltage are relatively low across all fault scenarios, with the lowest MAPE of 3.63% for drying and the highest of 10.65% for flooding, and an average MAPE of 5.97%. This suggests that the model effectively describes the electrochemical response of the PEMFC stack. This is critical for fault diagnosis, as the stack voltage is the primary indicator of the overall system health and performance. However, the errors observed for pressure drop predictions are notably higher, especially under certain operating conditions or faults. The average MAPE is 73.5% for anode pressure drop, and it becomes 136.5% if we consider drying. The pressure drop at cathode is better estimated, as it has an average MAPE of 34.2%. This discrepancy may have two main explanations. First, it may be attributed to simplifications or omissions in the physical modeling the multiphase flow behavior of PEMFC. In particular, the liquid water transport model in the channels is overly simplified. Second, the increased error may also be a consequence of calibration uncertainties encountered during the parameter identification process. Specifically, the calibration has been centered on the electrochemical response of the system, and it has been performed minimizing the error on the voltage, while pressure drop has been neglected.

3.4 Data-preprocessing: instructions and details

Whether we are working with the experimental dataset or with a simulated dataset generated through the physics-based PEMFC model, we are considering temporal signals, i.e., time series, that capture the evolution of monitoring variables for the PEMFC in time. Before these datasets can be effectively used, they must undergo a careful preprocessing phase to ensure their suitability for developing fault diagnosis algorithms. This is crucial for aligning the statistical properties of both experimental and simulated data, correctly manage the inherent temporal structure, and addressing disparities such as noise, scale differences, and imbalance. Only after a dedicated preprocessing procedure the datasets can provide a reliable foundation for training robust and generalizable algorithms.
The preprocessing procedure for time series dataset consists in the following steps:

1. **Data cleaning**. Remove segments for which not all the monitoring variables are available, or interpolate if necessary.

2. **Normalization**. Scale all the signals to ensure uniformity and ease convergence when training machine learning or deep learning algorithms. For example, voltage and pressure present great disparity in the numerical range: the voltage ranges between 0 V and 20 V, while the pressure can go up to 2e05 Pa. The

approach we use consists in normalizing the data with min-max values, in order to scale all the signals between 0 and 1.

3. **Train-validation-test split**. Divide the preprocessed dataset into training, validation, and test sets. Performing this split at an early stage is critical to avoid data leakage, which could artificially inflate the performance of the algorithms.
However, for our particular case, this procedure presents additional challenges due to the nature of the experimental data. Specifically, for certain fault conditions, measurements are available only within limited time intervals, as we have seen in section 3.2.1. Furthermore, when considering polarization curves, it is not appropriate to arbitrarily split the sequence into training, validation, and test sets. Doing so could result in testing the algorithm on current density values for which it has not been trained on. As it is illustrated in Figure 3.23, we could end up using only high current measurements (blue) for training, and testing only on low current values (yellow).

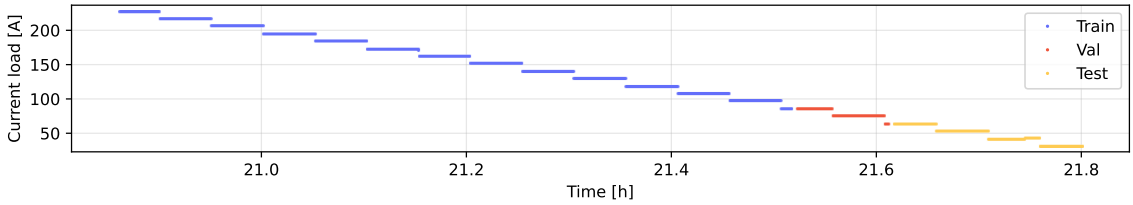


Figure 3.23: Train-validation-test split (0.7-0.1-0.2) of a polarization curve, without dedicated procedure.

To address these issues, we adopted the following strategy:

(a) Segment the dataset into fixed-length temporal frames. The dataset is first divided into non-overlapping frames of 6 minutes each, in order to account for the slow dynamics of water management faults.

(b) Split each frame into training, validation, and test subsets. Within each frame, the data were further partitioned into training, validation, and test sets.

4. This segmentation preserves the temporal coherence, ensures that each frame reflects a consistent fault class and guarantees a more uniform distribution of the different operating conditions (e.g., varying current loads) across the train, validation and test sets. Moreover, it avoids contamination between training and test data across different time intervals, as shown in Figure 3.25.

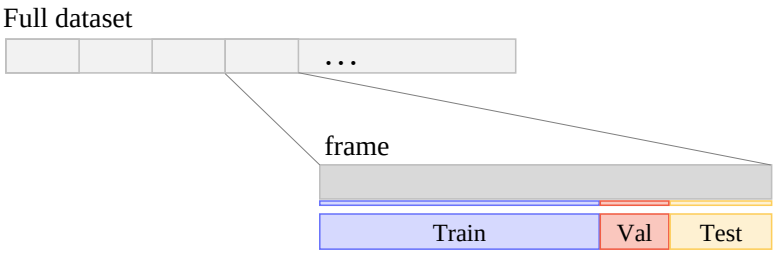


Figure 3.24: Dedicated train-validation-test split procedure.

5. **Windowing**. Divide the continuous signals into fixed-length windows. This is necessary for the algorithm constraints, which require inputs of uniform dimensions, as well as for their real-time implementation, where data are collected and analyzed in discrete intervals. This process is usually done with a sliding window (Figure 3.26): the window moves over the signal with a specified step, or stride, creating overlapping segments. The shorter is the stride, the larger the number of samples available. We set a stride of 2 s.

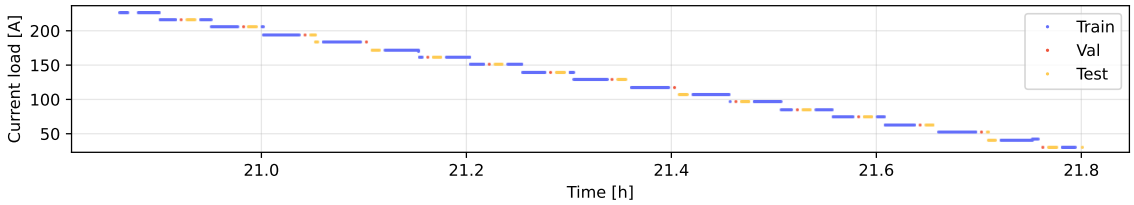


Figure 3.25: Train-validation-test split (0.7-0.1-0.2) of a polarization curve, with the dedicated procedure.

6. **Balancing**. Equalize number of samples across fault classes. A faulty dataset is often imbalanced, as we could observe for the experimental dataset described in this chapter, and this induces biases when training machine learning algorithms. To avoid this situation, we balance the dataset by using the second most populated class, i.e., flooding, as a reference, oversampling the minority classes and undersampling the majority class, i.e., nominal. The minority classes are increased by factors ranging from 2 to 5.

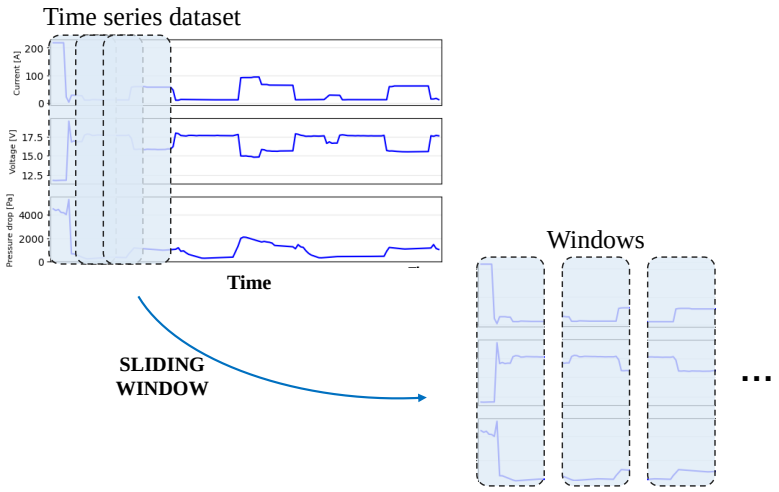


Figure 3.26: Illustration of the sliding window technique.

3.5 Summary

This chapter presented the experimental faulty dataset and the physics-based PEMFC model employed for the development of fault diagnosis methodologies.

The experimental dataset was analyzed in detail, with particular emphasis on fault types, their distributions, and their impact on PEMFC performance. The analysis of the faulty PEMFC dataset reveals key challenges that must be addressed for effective fault diagnosis. The data is characterized by strong imbalance, with nominal operating condition dominating, and fault events occurring much less frequently. While aging-related degradation trends are clearly visible, they interact with faults in complex ways, shifting the behavior of key variables over time, entangling the effects of faults with aging, and complicating their interpretation. Although such distribution shifts between fault classes can be intuitively observed with KDE plots and quantified with KL divergence measures, these techniques alone do not enable disentanglement of the impacts of aging and faults. Although constructing a detailed mapping of these effects for each operating condition would be theoretically beneficial, it is practically infeasible. Consequently, straightforward discrimination of fault classes is hindered. Instead, the nature of the data calls for dedicated machine learning or deep learning algorithms, inherently well-suited to extract nonlinear, informative features that can separate faults classes, even in the presence of aging and operating conditions variability.

The physics-based PEMFC model has been introduced as a complementary tool to the experimental data. It incorporates detailed electrochemical and transport equations to simulate the fuel cell system. It has been

calibrated against experimental polarization curves and validated across various fault scenarios, demonstrating its capability in capturing the physical phenomena related to faults. While the model accuracy on the electrochemical response, i.e., output voltage, is good, the accuracy on the pressure drop estimation is worse. Further refinements, including enhanced parameters identification techniques and improved two-phase flow modeling, are expected to increase model accuracy. Nonetheless, despite the presence of systematic biases in certain estimates, the model remains highly valuable for the intended purposes, as it provides a detailed representation of the underlying physical processes, and it correctly captures the dynamics of several physical phenomena happening in the PEMFC.

Building upon the insights gained in this chapter, the following chapters will present hybrid, physics-informed methodologies that leverage both the rich physical information provided by the model and the flexibility of deep learning algorithms, with the intent of achieving robust, versatile, and accurate PEMFC fault diagnosis.

4

Guiding Machine Learning Algorithms with Physics Knowledge for PEMFC fault diagnosis

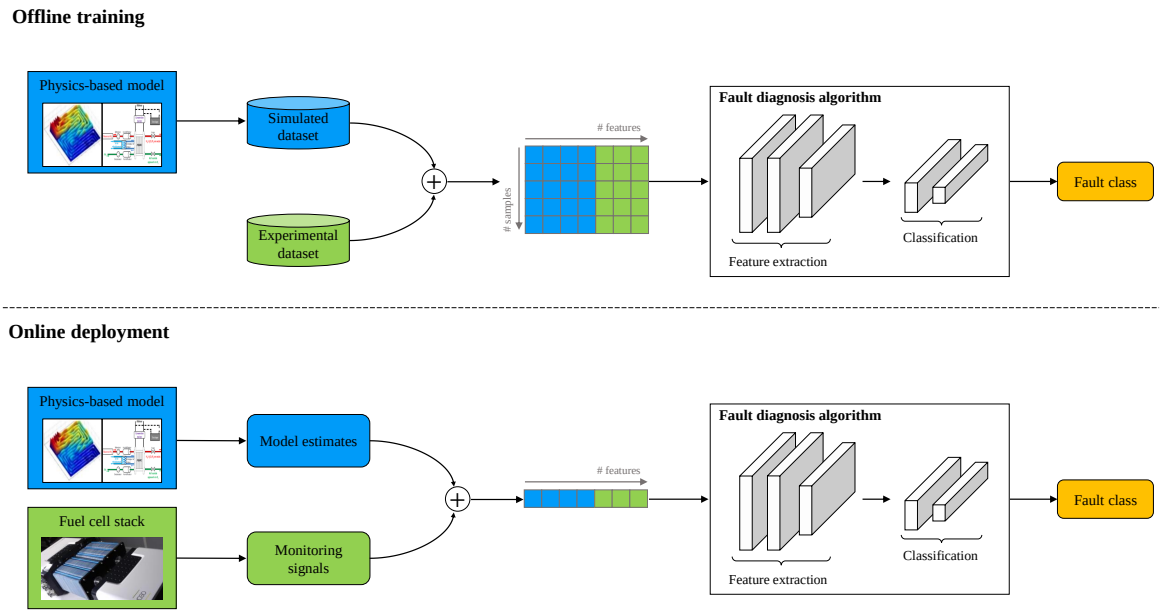


Figure 4.1: **Physics-guided fault diagnosis.** The physics-based model provides estimates of observable and non-observable variables of the fuel cell stack. This knowledge is combined with the monitoring signals and guides a deep learning algorithm to correctly classify faults within the fuel cell.

Contents

4.1 Introduction 49
4.2 Background on time series analysis 50
4.2.1 Time series 50
4.2.2 Time series classification 50
4.2.3 Recurrent Neural Network 51
4.3 The methodology 53
4.3.1 Description 53
4.3.2 Diagnosis variables 55
4.4 Proposed algorithm architectures and implementation details . . 56
4.4.1 Physics-based model deployment 57
4.4.2 Algorithms architecture 57
4.4.3 Data preparation 58
4.4.4 Performance evaluation metrics 58
4.5 Results and performance analysis 58
4.5.1 Validation on the initial phase of stack operation 59
4.5.2 Validation on the entire dataset, with the influence of aging . 60
4.5.3 Assessment of the methodology robustness with unseen aging data 61
4.5.4 Discussion 63
4.6 Summary 64

4.1 Introduction

Incorporating physical knowledge into data-driven machine learning algorithms offers several attractive benefits, including better generalization to unseen conditions for the system. While a high confidence physics-based model could be used to directly perform fault diagnosis, its hybridization with data-driven algorithms is often privileged when modeling error is relevant [172]. As a reciprocal measure, a physics-based model can ensure that the algorithms comply with the underlying physics of the system, as physical laws can limit the feasible region of solutions [173]. In this work, we aim to demonstrate the benefits of combining the knowledge embedded in a PEMFC physics-based model with a machine learning algorithm for fault diagnosis. The physics-based model can produce estimates that may not be accurately representative of the real PEMFC system but can generalize well to different operating conditions, and the real time monitoring data can help to compensate for limited accuracy of the model estimates [174].
Although several strategies can be used to hybridize the two, we propose to enhance the input space of the classification algorithm by compounding the measured signals, collected through sensors on the PEMFC stack [175], with estimates of observable and unobservable system dynamics provided by the physics-based PEMFC model described in chapter 3. The physics-based model is used in parallel to the real PEMFC stack. However, instead of performing fault diagnosis based on residual analysis as in traditional model-based methods, which requires a highly accurate model, the model estimates and the measured signals from the stack are aggregated together. By so doing, the physics-based model, which is calibrated offline, offers access to new, unmeasured process variables of the PEMFC which provide a better description of the internal state of the systems and are sensitive to fault signatures [176]. Consequently, the ability of the algorithm to discriminate the different faults in the stack can be improved. Among others, the physics-based model can provide estimates of the membrane resistance, the membrane water content and the current density distribution. This additional knowledge can help to disentangle the effect of the faults from that of aging, which cumulates on the monitoring signals, hindering a robust fault detection in real PEMFC systems (see section 3.3.2). The measured signals are often limited in numbers and affected by noise, which makes it challenging for data-driven algorithms to reliably discriminate between faults and aging effects. This chapter serves two objectives. First, it proposes to illustrate the problem of time series analysis. Second, it describes the hybrid fault diagnosis framework. We start by providing an exhaustive description of the specific characteristics of time series, giving insights on time series analysis and classification. Indeed, extracting valid and useful information from time series data is not a trivial task. We describe how conventional machine learning algorithms can be used to classify time series, and how RNN, which are better adapted for the scope, can be used. Specifically, we focus on LSTM networks, since they have proven to be particularly well-suited to capture temporal information in data and have shown an excellent ability to reveal hidden complex functional mappings between input and target fault labels [177]. Indeed, using LSTM networks, we can take into consideration that the state of dynamical systems is not only correlated to the value of the diagnosis signals at the time of measurement, but also to their values at previous time steps. By testing the physics-guided framework both with conventional machine learning algorithms and a LSTM network, we demonstrate the flexibility of the method as well as the superior performance achieved by the LSTM. Furthermore, the results highlight that the physics-based model brings outstanding contributions to PEMFC fault diagnosis.
This chapter is partly based on a work published in an international conference and on an article published in a scientific journal:

- C. Pettorossi, M. Gerard, V. Heiries, S. Rosini. “A novel hybrid method for PEMFC systems fault diagnosis”. European Fuel Cells and Hydrogen Pietro Lunghi Conference (EFC23), Capri, Naples, Italy, 2023.

- C. Pettorossi, R. Morvillier, V. Heiries, S. Rosini, and M. Gerard, “Physics-guided fault diagnosis method for proton exchange membrane fuel cells based on LSTM neural network,” J. Power Sources, vol. 626, p. 235696, Jan. 2025, doi: 10.1016/J.JPOWSOUR.2024.235696.

4.2 Background on time series analysis

Industrial systems are continuously monitored using sensors, which generate streams of measurements sampled at regular time intervals [178]. These streams naturally form time series, where each observation is not independent but instead reflects the system's evolving state over time. Effective fault diagnosis in dynamic systems requires the investigation of temporal dependencies inherent in these signals. Since faults typically develop as processes rather than instantaneous events, each measurement is embedded within a temporal context whose dynamics can contain essential diagnosis information. Disregarding this temporal dimension, the risk is to not capture the patterns that distinguish normal operation from fault behavior [179]. To preserve the temporal characteristics of the signals, adequate strategies for extracting relevant information and classify it in the correct fault class have to be used.

4.2.1 Time series

A time series is an ordered sequence of values of a variable, typically observed at uniform time intervals. Unlike cross-sectional data, where each observation is assumed independent and identically distributed (i.i.d.), time series frequently exhibit autocorrelation, trends, periodicities, or non-stationary behaviors [180]. For example, in many physical processes, the current state evolves smoothly from the past states, so that each observation depends on the previous one. A time series is generally defined by a sequence of ordered points taken at successive timestamps. Namely, it is a sequence $X_T = (x_1, \dots, x_T) \in \mathbb{R}^T$ where x_t is the variable at time t, and $t = 1, \dots, T$ are the sampling times.
In many applications, a time series consists not of a single measurement but of multiple variables recorded simultaneously, like voltage, temperature, and pressure in PEMFC stacks. Such multivariate time series contain both temporal dependencies within each variable and relationships between variables, and carry crucial information about the underlying process dynamics. If N variables are considered, one can define $\underline{x}_t = [x_t^j]_{j=1}^N \in \mathbb{R}^N$ and the multivariate time series is given as $\underline{X}_T = (\underline{x}_1, \dots, \underline{x}_T) \in \mathbb{R}^{N \times T}$, which can be unfolded as:

$$\underline{X}_T = \begin{bmatrix} x_1^1 & \cdots & x_T^1 \\ \vdots & \ddots & \vdots \\ x_1^N & \cdots & x_T^N \end{bmatrix} \tag{4.1}$$

The time dependency, alongside the high dimensionality and the noisy nature of the data, makes the analysis of time series a real challenge. Specifically, this temporal structure is often key to detecting gradual degradation, identifying precursors to failure, or distinguishing normal behavior from faults in the system. In the context of fault diagnosis, analyzing time series data to detect and distinguish between different fault conditions can be framed as a classification task, where the goal is to assign each sequence to a predefined fault category or to normal operation.

4.2.2 Time series classification

Time series classification is a supervised learning problem aimed at assigning to a specific class,temporally structured sequences of variable length. This means that to every multivariate time series $\underline{X}_T$ is associated a class label $y_T \in \mathbb{R}^K$, where $y_T = \{0, \dots, K\}$ and K is the total number of classes. A data sample is given by the pair $(\underline{X}_T, y_T)$ and the corresponding time series dataset is $D = \left\{ \left(\underline{X}_T, y_T\right)^i \right\}_{i=1}^n$, which comprises a collection of n pairs. Then, the scope of a classification algorithm trained on the time series dataset D, is to map the input space of the multivariate time series $\underline{X}_T$ over the space of the fault classes y_T [181].
Several time series classification algorithms have been developed over the years [182]. Among these, distance-based methods address the problem by finding a suitable measure of similarity between pairs of time series [183]. This solution is intuitive, but it becomes increasingly complex when multivariate time series are considered. An alternative solution that can better deal with high dimensionality is given by features-based classification algorithms, in which new representations are found to represent the data instead of raw time series. Traditional machine learning algorithms, like SVM or RF have been widely proposed for this scope. In this approach,

time series data $\underline{X}_T = (\underline{x}_1, \ldots, \underline{x}_T)$ is segmented into discrete observations $\underline{x}_t$, each of which is treated as an independent sample by the classification algorithm, thus neglecting the critical temporal dependencies. Indeed, these algorithms do not have the ability to process the temporal correlation of the data appropriately [184]. The implicit assumption used when deploying traditional machine learning classification algorithms is that each observation contains sufficient information to discriminate between normal and faulty states without resorting to historical context. This implies neglecting the order of observations and ignoring that adjacent time points are typically highly correlated [185]. As an improvement, one could try to extract informative summary features from the time series data, and then provide this information to the classification algorithm. Yet, extracting relevant features remains challenging, and the most common approach is to compute the mean over the temporal window of the signals. The intrinsic characteristics of time series data are often difficult to capture with handcrafted descriptors alone. More advanced methods rely on frequency domain features extraction, derived via the FFT or WT, which can help highlighting periodicities and transient patterns, or on dimensionality reduction techniques like PCA to retain the most information from the data. Nevertheless, all these approaches typically require substantial effort in data preprocessing and feature engineering and might also reduce the discriminative power of the algorithm. Recently, some effort has been spent to exploit deep neural networks, which can extract automatically the features from time series data [186]. Specifically, recurrent networks, i.e., RNNs, have been designed to record and preserve correlation information from sequential data [187, 188]. Variants include Gated Recurrent Unit (GRU) and LSTM networks, which address the vanishing gradient issue, as well as RC methods that leverage fixed recurrent structures [189]. Although in this work we focus primarily on RNNs, it is worth noting that other architectures are also commonly used for time series classification. CNNs can be effective for handling very long sequences, as they are able to extract local temporal patterns efficiently. In contrast, fully connected deep networks are often applied to shorter sequences where temporal dependencies are less pronounced [190]. More recently, transformer architectures, which rely on self-attention mechanisms, have emerged as powerful tools for handling complex tasks, e.g., natural language process.

4.2.3 Recurrent Neural Network

The basic neural network is a FNN, which has a feedforward architecture in which the information flows only in one direction. A RNN looks much like a FNN, but the activations are directed also backward, as illustrated in Figure 4.2. Broadly speaking, RNNs are networks learning from sequences, where each neuron feeds the next one with a memory parameter so that prediction at time t depends on the whole sequence. RNN maintains a hidden state that evolves across the sequence, which is useful when the temporal information has to be preserved, and it is very effective if we want to capture early signs from the data [191]. The simplest RNN consists of a neuron that, at each time step, receives in input the current input and its past output. The unfolded structure of the recurrent neuron is given in Figure 4.3.

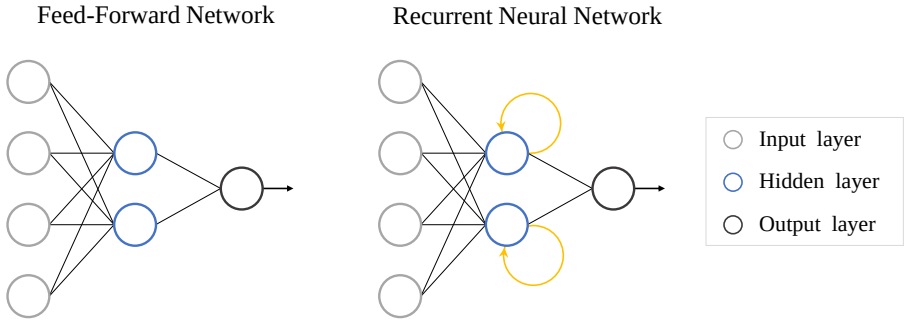


Figure 4.2: Schematic of feed-forward network and recurrent neural network.

Training a RNN is often challenging. To train it, the network must first be unrolled through time, and then traditional backpropagation is performed [193]. Since the gradients must be propagated backward through all the time steps, the unrolled RNN effectively becomes a very deep network and, consequently, suffers from unstable

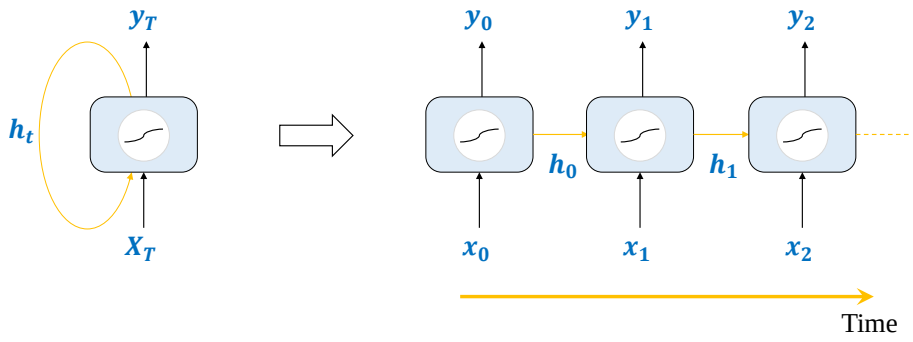


Figure 4.3: Recurrent neuron (left), unfolded through time (right) [192].

gradients, such as vanishing or exploding gradients. Additionally, when sequences are long, the network may struggle to retain information from earlier time steps. As data passes through the network, some information is inevitably lost at each stage. After a while the network contains no trace of the first input.
To tackle this problem, various types of cells have been proposed [194]. Among those, LSTM and GRU are the architectures that contributed the most to the large success of RNNs. The LSTM incorporates specialized gates and a memory cell that selectively retain, update, or forget information, while the GRU network introduced gating mechanisms that effectively mitigate vanishing gradients and enable learning of longer-term dependencies with fewer parameters and improved training stability.

Long Short-Term Memory Network

The core of the LSTM neural network is the memory cell, that replaces the conventional neurons in the hidden layers, and it is shown in Figure 4.4.

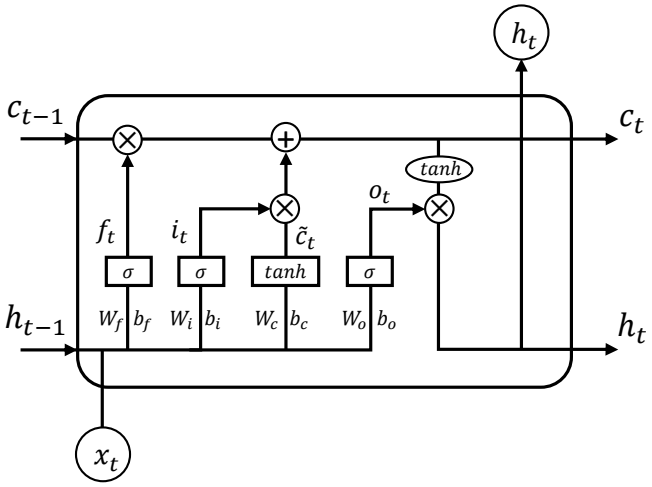


Figure 4.4: LSTM cell structure.

Let us consider a time series X_T, an input signal containing the information from an initial time $t = 1$ until the time $t = T$. In each memory cell, a cell state C_t controls the network information using the input gate, output gate and forget gate. At time t, when a new observation x_t is fed to the network, the forget gate determines whether to keep or remove the information of the preceding memory block output h_{t-1}. The output of the forget gate is:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{4.2}$$

where W_f and b_f are the input weights and bias of the forget gate, respectively. The input gate $i_t$determines whether x_t is stored in the cell state C_t:

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{4.3}$$

where W_i and b_i are the input weights and bias of the input gate, respectively. A tanh layer function is used to generate a new information vector $\tilde{C}_t$ to be added to C_t:

$$\tilde{C}_t = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{4.4}$$

where W_c and b_c are the input weights and bias of the input gate, respectively. Then, the cell state C_t is updated as follow:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{4.5}$$

Eventually, the output of the memory cell h_t is generated by using the output gate and a tanh layer:

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{4.6}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{4.7}$$

where W_o and b_o are the input weights and bias of the input gate, respectively. Eventually, a layer of the LSTM neural network consists of N_c memory cells, which determines the expressive capacity of the network. Given N signals of the multivariate time series, the input of the LSTM neural network is the multivariate time series $\underline{X}_T \in \mathbb{R}^{N \times T}$. Then, each LSTM layer maps the input matrix to the hidden state matrix $\underline{H}_T \in \mathbb{R}^{N_c \times T}$, and the output layer translates the last hidden state into a probability vector $\underline{\hat{y}}_t$, from which the predicted class label $\hat{y}_t$ is typically obtained using argmax function.
In short, an LSTM cell can learn to recognize an important input (input gate), store it in the long-term state, preserve it as long as needed (forget gate), and extract it whenever is needed. This capability is especially relevant in complex systems where faults do not induce immediate and unambiguous signal changes, but rather manifest as subtle temporal patterns. By processing a time series dataset, LSTM networks can aggregate contextual information from different variables, learn temporal dynamics and thus provide richer representations to discriminate between the system dynamic and the fault progression. A common extension is the BiLSTM network [195], which processes sequences in both forward and backward direction. This enables the network to capture both past and future context at each time step, often leading to improved performance in tasks where complete temporal information enhances prediction accuracy.

4.3 The methodology

The hybrid method that we propose aims at effectively combining a physics-based PEMFC model and a machine learning classification algorithm for PEMFC fault diagnosis. The idea is to enhance the information available on the stack conditions by means of prior physical knowledge to achieve better and more robust fault diagnosis. We will introduce the methodology using a LSTM network as classification algorithm, and traditional machine learning methods will be later introduced for comparison. The crux of the proposed fault diagnosis method is to leverage the physics knowledge embedded in the physics-based model to access estimates of the internal states of the PEMFC stack. At the same time, it exploits the LSTM ability of processing time series and extracting features to identify fault patterns in complex dynamic signals. By integrating a physics-based model, which provides estimates of internal system states grounded in physical laws, the proposed approach guides the classification algorithm with more relevant input information, resulting in a significantly more reliable fault diagnosis method than purely data-driven approaches. This demonstrates that the physics-based model can effectively contribute to the learning process of classification algorithms, ultimately improving their ability to generalize across a wider range of conditions and increasing the overall robustness in real-world scenarios.

4.3.1 Description

The physics-guided fault diagnosis method is illustrated in Figure 4.5. It involves using the physics-based PEMFC model in parallel to the PEMFC stack and an LSTM neural network as classification algorithm, that can adequately process the joint information collected through sensors and provided by the model estimates.

Indeed, variables like membrane resistance are well known to be good indicators of faults. However, measuring the membrane resistance requires impedance measurements, which is not a practical option. Nonetheless, an estimate of the membrane resistance could be obtained through the model. Thus, the description of the faults in the real PEMFC comes from the real time recorded sensors measurements as well as from the model estimates. However, unlike residual-based methods, where the model is used to compute the difference between measured and predicted outputs, here the model itself is not employed to generate residuals. Relying solely on model outputs can be problematic, as modeling errors may strongly affect the accuracy of the fault diagnosis and amplify the impact of discrepancies between predictions and reality. Instead, in the proposed framework, the model estimates are directly integrated as inputs to the classification algorithm. This results in a physics-guided algorithm which has an enhanced input space and gains representation power for achieving fault diagnosis without increasing the number of sensors required on the stack.

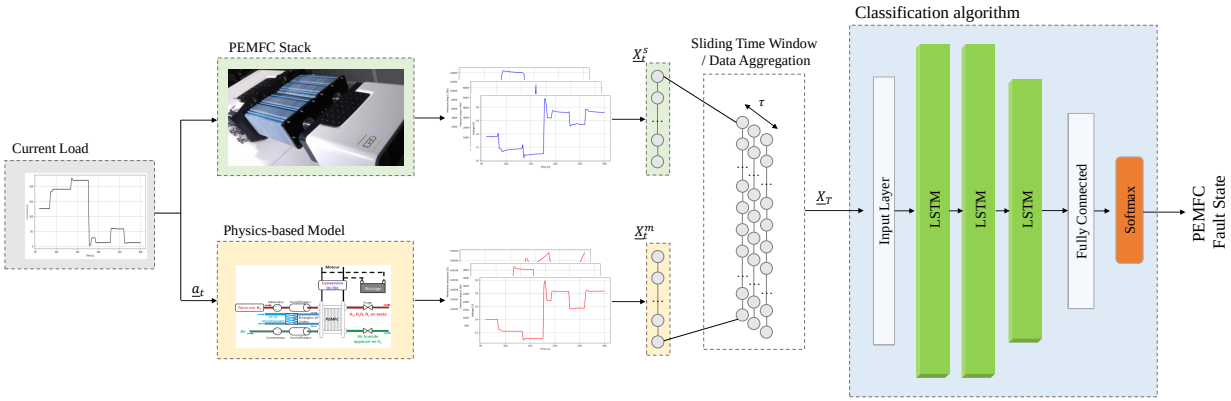


Figure 4.5: Flowchart of the proposed physics-guided LSTM method.

Once it has been calibrated, the physics-based model takes in input the actuator signals, i.e., command variables inferred from the system, at time t, $\underline{a}_t = [a_t^i]_{i=1}^{N_a}$, where a_t^i is the i-th actuator inferred signal and N_a is the total number of actuator signals required to run the model. Then, the physics-based model outputs the estimates of several process variables, i.e., model estimates of the process variables at time t, $\underline{x}_t^m = [x_t^{(m,j)}]_{j=1}^{N_m}$, where $x_t^{(m,j)}$ is the j-th model estimate at time t and N_m are the number of process variables estimated by the model. In addition, we have access to measured signals at time t, $\underline{x}_t^s = [x_t^{(s,l)}]_{l=1}^{N_s}$, where $x_t^{(s,l)}$ is the l-th signal measured from the stack and N_s are the number of measured signals. Eventually, the model estimates and the measured signals are aggregated, i.e., $\underline{x}_t = [\underline{x}_t^m, \underline{x}_t^s] \in \mathbb{R}^{N_m+N_s}$ to enhance the information available for the LSTM neural network. Then, the continuous stream of model estimates and measured signals during a time window of length T is the multivariate time series $\underline{X}_T = (\underline{x}_{t-T}, \ldots, \underline{x}_t) = [\underline{X}_T^m, \underline{X}_T^s] \in \mathbb{R}^{T\times(N_m+N_s)}$. This serves as input of the LSTM neural network which, once trained, outputs the fault state prediction $\hat{y}_T$. The training procedure for the LSTM network is described in Algorithm 1.

Algorithm 1 Training procedure for the physics-guided LSTM network

Require: Time series dataset $D = \{[\underline{X}_T^m, \underline{X}_T^s], y_T\}$; batch size h; learning rate μ; max epochs $n_{\text{epochs,max}}$
Ensure: Optimal LSTM network parameters $\hat{\theta}$
1: Initialize θ parameters for the LSTM network
2: **while** $n_{\text{epochs}} < n_{\text{epochs,max}}$ **do**
3: Sample batches $\{[\underline{X}_T^m, \underline{X}_T^s], y_T\}_{i=1}^h$ from D
4: Update the parameters θ with backpropagation
5: **end while**

At training time, we have access to a dataset of measured signals, collected through sensors from the PEMFC stack, and to model estimates of the process variables provided by the physics-based PEMFC model. We also have access to the corresponding labels for the fault states of the stack. We are referring to a situation in which

labeled data are available, therefore the standard supervised approach can be implemented to build a direct mapping from the input, i.e., $\underline{X}_t = [\underline{X}_t^m, \underline{X}_t^s]$, and the fault state label, i.e.,y_T. Then, the validity of the proposed physics-guided LSTM method is assessed on a separate test set. As we emphasized, among the model estimates $\underline{x}_t^m$ there are estimates of process variables and internal states of the PEMFC that are not directly measurable [61, 196]. This can be related to the unavailability of dedicated sensors for real-time applications, to the high cost of sensors or to the inherent non-observability of the variables. The model estimates are listed in Table 4.1.

# Symbol	Description	Unit
1	Stack voltage	[V]
2	Anode pressure drop	[Pa]
3	Cathode pressure drop	[Pa]
4	Membrane resistance	[Ω]
5	Membrane water content	[-]
6	Membrane temperature	[K]
7	Relative humidity at cathode outlet	[%]
8	Relative humidity at anode outlet	[%]
9	Current density distribution	[A/m^2]
10	Thermal power	[W]

Table 4.1: Model estimates from the physics-based PEMFC model.

The assumption we make here is that the additional information given by these variables allows the algorithm to be more discriminative with respect to the PEMFC stack fault state. Thus, integrating them into the multivariate time series $\underline{X}_T$ as input of the LSTM neural network could be advantageous for a number of reasons. First, it allows for an increase in the representation power of the available information. In addition, the extended knowledge provided by the physics-based PEMFC model could increase the ability of the LSTM classification algorithm to better isolate the root causes of the faults, since different variables can capture different aspects of the underlying phenomena. For instance, the model estimates could enforce the disentanglement of similar patterns generated by different faults in the measured signals. Furthermore, the enhanced input space enforces the coupling between sensored signals and model estimates, guiding the LSTM neural network to be more generalizable and, thus, making it robust to noise fluctuations. In fact, while the measurements from sensors may be affected by noise and may deteriorate in time, there is no such an impact on the model estimates, which are rather affected by modeling errors, thus not depending on the environmental conditions. At the same time, having a classification algorithm that learned from historical faulty data can limit the influence of modeling errors and/or biases on the model estimates. Conversely, the physics-based model has a positive effect in limiting the consequences related to varying environmental condtions and to the aging of the PEMFC stack on the performances of the diagnosing algorithm. In a pure data-driven LSTM the complex degradation mechanisms of PEMFC cannot be discerned from the faults due to the limited representativeness of the available data. By complementing the measured signals with informative and realistic model estimates, the classification algorithm has a larger set of input that allows it to capture a richer view of the data, making it less sensitive to shifts caused by aging, and more robust. Lastly, the interpretability of the algorithm, and more widely of the fault diagnosis method, is improved.

4.3.2 Diagnosis variables

The input signals provided to the classification algorithm, i.e., the LSTM network, must be carefully selected. These signals constitute the sole source of information upon which the algorithm depends to determine the fault state of the PEMFC stack. As they directly influence the performance of the fault diagnosis, their quality and relevance are fundamental.

We have widely highlighted the role of the stack voltage as a relevant indicator of a deviation in the state of the PEMFC stack. This, together with its easy online attainability, make it the favored diagnosing signal in most of the PEMFC fault diagnosis algorithms [103]. However, the stack voltage alone falls short of effectively discerning

the inherent dynamic behavior of the system from the dynamics induced by the faults. Thus, the current load and the coolant flow have also been included as monitoring signals to decouple the effect of load variation from the effect of state variation on the fault indicators. It is crucial to separate the two effects to proactively detect the occurrence of a fault. Moreover, to successfully isolate the different fault states, pressure drops measured at cathode and anode are included [197]. The measured signals $\underline{x}_t^s \in \mathbb{R}^T$ used in the present study are listed in Table 4.2. These signals are conventionally acquired by embedded sensors in real world PEMFC systems.

# Symbol	Description	Unit
1	Current	[A]
2	Stack voltage	[V]
3	Anode pressure drop	[Pa]
4	Cathode pressure drop	[Pa]
5	Inlet coolant flow	[m^3/s]

Table 4.2: Measured signals from PEMFC stack.

To provide estimates of the process variables, the physics-based PEMFC model requires in input the operating signals $\underline{a}_t \in \mathbb{R}^{N_a}$ which can be inferred from the actuators on the PEMFC system. It is worth noting that, even though these signals are attainable in real PEMFC systems applications, they are usually affected by a certain level of uncertainty or noise, and, in practice, only a rough estimate is available. A detailed overview of the actuator inferred signals is provided in Table 4.3.

# Symbol	Description	Unit
1	Coolant temperature	[K]
2	H_2 inlet relative humidity	[%]
3	Air inlet relative humidity	[%]
4	H_2 stoichiometry	[-]
5	Air stoichiometry	[-]
6	H_2 inlet pressure	[Pa]
7	Air inlet pressure	[Pa]

Table 4.3: Actuators inferred signals.

Eventually, the process variables estimated with the calibrated physics-based PEMFC model $\underline{x}_t^m$, which can be observable or unobservable variables, are those listed in Table 4.1. As previously mentioned, variables 1-3 are critical for diagnosing the PEMFC stack and are the model estimate equivalents of measured signals. Variables 4-8 are internal states of the PEMFC stack that are especially well adapted to describe phenomena related to water management faults in the stack [61], while variables 9-10 have been selected due to their correlation with starvation faults [140, 161]. Obtaining some of these process variables through sensors is usually unfeasible. Dedicated sensors can be overly costly, or cannot be adapted for online applications, or some signals could not directly be measurable. Thus, these model estimates enable a more exhaustive and detailed representation of the PEMFC, in order to guide the fault diagnosis method with physics relevant knowledge.

4.4 Proposed algorithm architectures and implementation details

We give here the implementation details for the development and training of the physics-guided fault diagnosis method, as well as for the pure-data-driven method used for comparison, for the LSTM network and four machine learning algorithms considered: SVM, RF, Gradient Boosting (GB), FNN. This includes defining how to deploy the physics-based model in a realistic framework, defining the algorithms architectures and the dedicated preprocessing steps. Eventually we describe the metrics used for validating the performance of the algorithms.

4.4.1 Physics-based model deployment

To evaluate whether the additional physics-knowledge provided by the physics-based PEMFC model can improve the performance and the robustness of the fault diagnosis algorithm, we tested the proposed framework in an offline setting. Indeed, the aim of our study is to assess the benefits of integrating prior knowledge to the machine learning algorithms, and specifically to the LSTM network. Despite testing the methodology in an offline setting, the model simulations have to be conducted to represent the conditions of the actual deployment. In realistic scenarios the measurements of the inputs required to run the model can be affected by low sensors' accuracy. Moreover, direct measures might be unavailable. Consequently, the variables used as input of the physics-based model cannot be directly observed but must be indirectly inferred based on actuator signals or highly noisy sensor measurements. Thus, to reflect these practical limitations and ensure that the offline experiments approximate the conditions encountered during real-time operation, we perturbed the input provided to the physics-based model with synthetic uncertainty. Concretely, the signals $\underline{a}_t$ were perturbed by both a proportional bias β and additive stochastic noise η, according to the following formulation:

$$\tilde{\underline{a}}_t = \underline{a}_t + \beta + \eta \tag{4.8}$$

where the bias term is defined as:

$$\beta = 0.1 \cdot \underline{a}_t \tag{4.9}$$

and the noise is drawn from a Gaussian distribution:

$$\eta \sim \mathcal{N}(0, \sigma^2), \quad \text{where } \sigma = 0.1 \cdot \underline{a}_t \tag{4.10}$$

This allows to approximate the combined effects of sensor inaccuracy, estimation error, and environmental disturbances that would be present in a real application. After introducing this controlled perturbation, the model is executed to generate the estimates of the variables listed in Algorithm 2.

Algorithm 2 Deployment of the physics-guided algorithm

Require: Model inputs: $\underline{a}_t$; bias scaling factor 0.1; noise scaling factor 0.1
Ensure: Model estimates $\underline{X}_T^m$
1: Perturb the N_a input signals: $\tilde{\underline{a}}_t \leftarrow \underline{a}_t + \beta + \mu$
2: Run a simulation
3: Retrieve model estimates $\underline{X}_T^m$

4.4.2 Algorithms architecture

In this section we provide an overview of the classification algorithms architecture employed in this study, both for the physics-guided and the pure data-driven methods. From the exploration of the parameters space, we selected the hyperparameters to maximize the accuracy on the test set and to ensure the best achievable performance. The architecture of the LSTM neural network used comprises 4 layers. The network has three initial LSTM layers, with number of hidden neurons equal to 20, 20 and 15, respectively, and one dense layer with 5 neurons and softmax activation function. We utilize the Adam algorithm with adaptive learning rate to optimize the training of the neural network through backpropagation. The initial learning rate is set to 0.02. The batch size is set to 512 and the number of epochs is set to 100. The implementation is done through the TensorFlow package in Python [198]. To evaluate the effectiveness of the proposed physics-guided LSTM method, a comparative analysis against traditional machine learning algorithms widely used in the literature for PEMFC fault diagnosis is also included in the study. The machine learning algorithms implemented are: SVM, with regularization parameter $C = 2$ and Radial Basis Function (*rbf*) kernel; RF, with number of estimators equal to 10; GB, with number of estimators equal to 50; FNN, with 2 hidden layers, 50 neurons per layer. These algorithms were implemented using Scikit-learn library in Python [199].

To thoroughly validate the performance of the algorithms, ensure a robust assessment and avoid overfitting, stratified k-fold cross validation, with $k = 5$, is implemented.

4.4.3 Data preparation

The dimension of the input data $\underline{X}_T$ varies depending on the method selected. The pure data-driven LSTM takes in input solely the measured signals $\underline{X}_T^s$. The proposed physics-guided LSTM takes in input also the model estimates $\underline{X}_T^m$, thus take in input the multivariate time series $\underline{X}_T = [\underline{X}_T^s, \underline{X}_T^m]$. The input space to the LSTM neural network is normalized to a range [0,1] by a min/max normalization given by the available dataset in the first 50 hours of operation. Subsequently, the signals will be processed with a sliding time window of size 20 s. The sliding time window means that the first sample comprises the signals from time 1-20 s, the second from 21-40 s, and so on. The time series data are subsequently used as input for the classification algorithms. In this process, the LSTM neural network analyzes the complete time window $\underline{X}_T = (\underline{x}_{t-T}, \cdot, \underline{x}_t)$, capturing temporal dependencies, whereas the traditional machine learning algorithms used in the comparative analysis receive in input the average of the time series along the time window T. The objective here is to maintain methodological simplicity and therefore we deliberately chose the simplest statistical summary, avoiding the extraction of more complex features.

4.4.4 Performance evaluation metrics

To quantitatively compare the performance of the fault diagnosis methods, we apply the following metrics, which are commonly used in classification algorithms: accuracy (*acc*), precision (*prec*), recall (*rec*) and F1 score (*F1*). They are defined as follows:

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.11}$$

$$prec = \frac{TP}{TP + FP} \tag{4.12}$$

$$rec = \frac{TP}{TP + FN} \tag{4.13}$$

$$F1 = 2\frac{prec \cdot rec}{prec + rec} \tag{4.14}$$

where TP is the number of correctly detected faults, FP that of incorrectly detected ones, TN is the number of correctly assigned fault-free cases and FN that of incorrectly assigned fault-free. Each metric captures a different aspect of model performance. Accuracy reflects overall correctness, but can be misleading in imbalanced datasets. Precision measures how many predicted positives are actually correct, while recall indicates how well the model identifies actual positives. F1 score balances precision and recall, making it useful when both false positives and false negatives are important. Eventually, the confusion matrix is also used to provide the reader with a more detailed description of the performance across the fault states.
In addition to the above metrics, for a more precise characterization of the quality and robustness of the fault diagnosis methods, we propose the time to detect metric. This is the time between the moment the PEMFC stack changes state and the moment the algorithm detects the change, classifying it as the correct fault. The motivation behind this metric is that by detecting a fault as soon as possible, we can promptly restore nominal operating conditions, thereby limiting the detrimental impact of the fault on the PEMFC stack. Therefore, it provides useful information about algorithms' performance.

4.5 Results and performance analysis

To exhaustively validate the proposed physics-guided method, we conducted three experiments on the dataset presented in chapter 3. Initially, to illustrate the benefits of using the physics-guided LSTM method over the

pure data-driven counterpart, we test their performances on the first 250 hours of operation of the PEMFC stack. Then, tests are performed over the entire dataset available, comprising more 1000 hours of stack operation. For each experiment, the train-validation-test split described in section 3.4 is used, in order to ensure avoiding data leakage issues and to reduce the risk of overfitting. Nonetheless, differently from the data preprocessing procedure described in section 3.4, for these experiments we chose not to balance the dataset for training the algorithms. The decision was made to better reflect the realistic scenario we would face when using faulty datasets of PEMFC systems, in which faults are rare events, while being aware that it could induce biases in the classification algorithms.
In addition, to demonstrate the generalization and the robustness to stack aging of the physics-guided LSTM method, we use the first 250 hours of operation to train the classification algorithms and, then, we test their performance on the data collected after 600 hours of operation.

4.5.1 Validation on the initial phase of stack operation

In this section, the performance of the proposed physics-guided LSTM method is evaluated and compared to a pure data-driven LSTM neural network on the data collected at initial phase of stack operation. Specifically, we keep the data in the first 250 hours of operation only, as shown in Figure 4.6. This allows to evaluate the performance of the algorithms with a reduced influence of stack aging, thereby enabling a more focused assessment of the algorithm capabilities. The dataset, which has not been balanced, consists of 8437 samples for the nominal state (0), 3846 samples for the flooding fault (1), 1692 samples for the drying fault (2), 477 samples for the air starvation fault (3) and 485 samples for the H2 starvation fault (4).

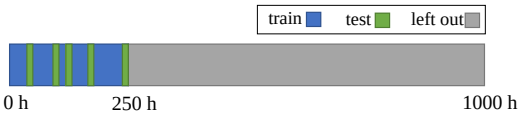


Figure 4.6: Simplified illustration of the dataset splitting (see train-validation-test split in section 3.4) for the validation on the initial phase of stack operation.

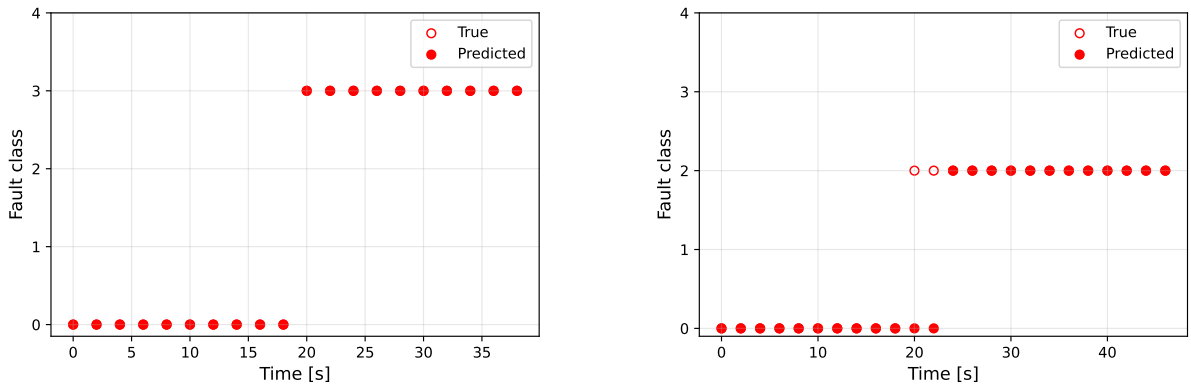
The performance metrics for the physics-guided LSTM method and the data-driven LSTM, are given in Table 4.4, validated with cross validation. Also, the performances of the SVM, RF, GB and FNN are given, for the data-driven as well as the physics-guided approach.

	acc	*prec*	*rec*	*F1*
Data-driven				
SVM	0.776	0.862	0.534	0.604
RF	0.939	0.945	0.908	0.925
GB	0.877	0.900	0.828	0.857
FNN	0.795	0.807	0.559	0.612
LSTM	0.966	0.964	0.954	0.958
Physics-guided				
SVM	0.947	0.968	0.871	0.913
RF	0.980	0.985	0.952	0.968
GB	0.959	0.975	0.927	0.949
FNN	0.954	0.959	0.919	0.938
LSTM	**0.995**	**0.995**	**0.994**	**0.995**

Table 4.4: Comparison of the data-driven and physics-guided fault diagnosis methods, for the different algorithms, over the initial phase (250 h) of stack operation.

The results highlight that the physics-guided LSTM method surpasses the data-driven LSTM neural network, in all the metrics, with an increase of 3.00 % in the accuracy and of 3.86 % in the F1 score, as well as all the

other machine learning algorithms considered. We evaluate the time to detect for the physics-guided LSTM neural network, which results in 10.1 s, in average, against 14.2 s for the data-driven LSTM, proving that the physics-guided method facilitates the diagnosis of the stack by reducing the time to detect of 28.9 %. To clarify the significance of the time to detect metrics, in Figure 4.7 we show the true fault state vs the fault state predicted by the physics-guided LSTM for two arbitrarily selected time intervals, representing the detection of air starvation (3) on the left, and of drying (2) on the right.



(a) (b)

Figure 4.7: True (unfilled) vs predicted (filled) fault state of the PEMFC stack, showing an air starvation fault (left) and a drying fault (right). The physics-guided LSTM is used in the first 250 hours of the dataset.

4.5.2 Validation on the entire dataset, with the influence of aging

Considering that the PEMFC stack has a durability of thousand of hours, it is not enough to prove that the fault diagnosis algorithm is valid in the first 250 hours of operation. In fact, under the effect of degradation phenomena the monitoring signals tend to deteriorate. The effects of degradation cumulate with the effect of faults on the measured signals, e.g., on the voltage. This induces a shift in the distribution of the fault classes and, eventually, hinders fault discriminability. A robust fault diagnosis method must be capable of addressing this challenge. Accordingly, we extend the study of the previous paragraph to the entire dataset, which consists of more than 1000 hours of operation of the stack. The data samples available for each system state are the following: 12726 for the nominal state (0), 5383 for the flooding fault (1), 2682 for the drying fault (2), 658 for the air starvation fault (3) and 785 for the H2 starvation fault (4). A simplified representation of the dataset after the train-validation-test splitting procedure described in section 3.4 is shown in Figure 4.8. The splitting procedure ensures that there is no overlap between the train and the test sets, and, consequently, guarantees a solid assessment of the methods performances.

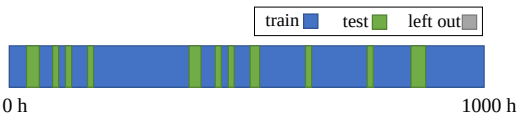


Figure 4.8: Simplified illustration for the dataset splitting (see train-validation-test split in section 3.4) for the validation on the entire dataset.

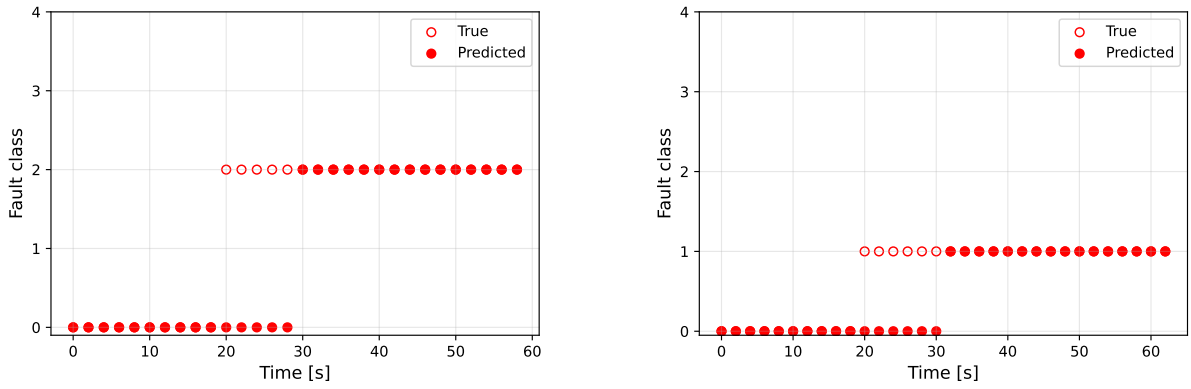
Table 4.5 shows the results, validated through cross validation, for the physics-guided LSTM and the data-driven LSTM, as well as for the machine learning algorithms introduced for comparison. The proposed physics-guided LSTM method outperforms the data-driven LSTM in terms of fault diagnosis performances, with an improvement is 5.89 % for the accuracy and 15.0 % for the F1 metrics. Although the machine learning classification algorithms considered show commendable performances with the physics-guided approach, they fall short of the physics-guided LSTM effectiveness.

For the physics-guided LSTM method, the time to detect has an average value of 17.9 s, while for the data-driven LSTM it is 43.2 s, corresponding to a reduction of 58.6 %. Figure 4.9 shows the true fault state and the predicted

	acc	*prec*	*rec*	*F1*
Data-driven				
SVM	0.748	0.848	0.505	0.586
RF	0.904	0.916	0.859	0.885
GB	0.799	0.859	0.677	0.745
FNN	0.743	.755	0.512	0.561
LSTM	0.933	0.913	0.881	0.859
Physics-guided				
SVM	0.944	0.968	0.871	0.913
RF	0.973	0.980	0.937	0.957
GB	0.939	0.972	0.866	0.912
FNN	0.944	0.951	0.907	0.928
LSTM	**0.988**	**0.990**	**0.987**	**0.988**

Table 4.5: Comparison of the data-driven and physics-guided fault diagnosis methods, for the different algorithms, over the 1000 h dataset.

one for the physics-guided algorithm, in arbitrary time intervals: detection of drying (2) on the left, and detection of flooding (1) on the right.



(a) (b)

Figure 4.9: True (unfilled) vs predicted (filled) fault state of the PEMFC stack, showing drying fault (left) and a flooding fault (right). The physics-guided LSTM is used on the entire dataset.

4.5.3 Assessment of the methodology robustness with unseen aging data

To further investigate the robustness of the physics-guided LSTM method to PEMFC stack aging, we split the entire dataset as shown in Figure 4.10.

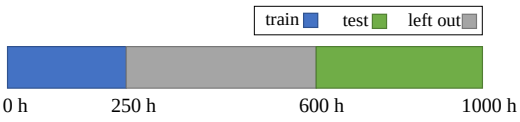


Figure 4.10: Illustration of the dataset splitting for assessing the robustness with unseen aging data.

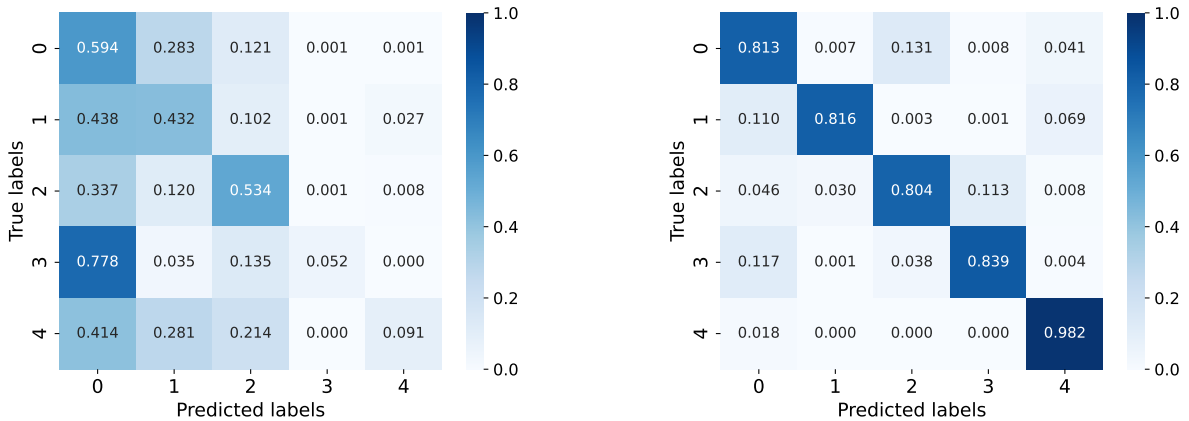
The first 250 hours are used to train and validate the algorithms, then the performances are tested on the last part of the dataset, from 600 hours to 1000 hours. This enables the validation of the algorithms' ability to discriminate faults when PEMFC aging is established and distribution shift becomes an issue in the measured signals. Since in real life it is hard to obtain a faulty dataset comprehending the whole operation time of the PEMFC stack to

train the fault diagnosis algorithm, this study enables the validation of the robustness of fault diagnosis method in a more rigorous and realistic scenario. This scenario is extraordinarily challenging, and this work stands out as one of the pioneering studies to address this critical issue in the literature, providing a valuable benchmark for investigating fault diagnosis under non-stationary, degrading conditions in PEMFC systems. For this study, the results for the physics-guided LSTM and the data-driven LSTM, and for the machine learning algorithms included for comparison, are shown in Table 4.6.

	acc	*prec*	*rec*	*F1*
Data-driven				
SVM	0.495	0.389	0.296	0.283
RF	0.448	0.378	0.302	0.296
GB	0.434	0.219	0.230	0.222
FNN	0.486	0.356	0.304	0.283
LSTM	0.524	0.584	0.439	0.459
Physics-guided				
SVM	0.721	0.833	0.592	0.669
RF	0.731	0.731	0.541	0.579
GB	0.698	0.641	0.433	0.467
FNN	0.662	0.668	0.514	0.539
LSTM	**0.817**	**0.807**	**0.853**	**0.812**

Table 4.6: Comparison of the data-driven and physics-guided fault diagnosis methods, for the different algorithms, on the dataset shown in Figure 4.10.

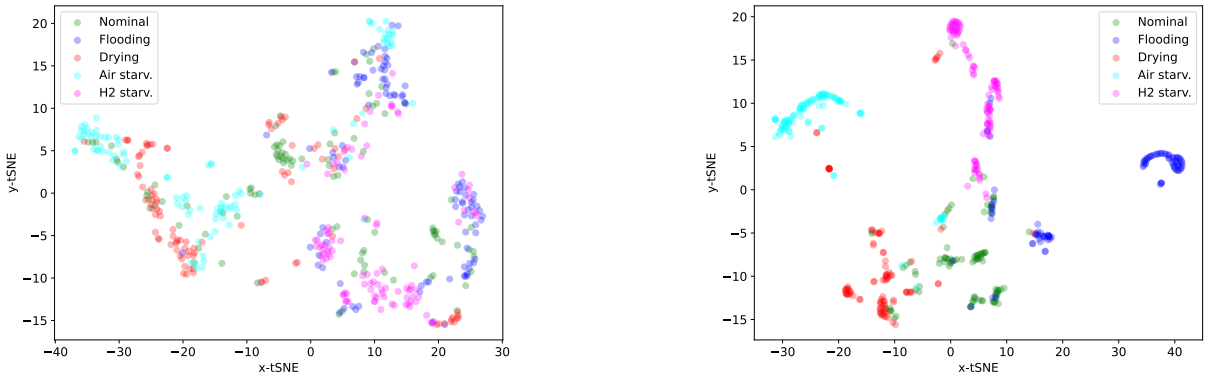
The findings presented in Table 4.6 illustrate that the sole method that can achieve satisfactory performances in diagnosing the PEMFC stack under aging condition is the physics-guided LSTM. The accuracy score and the F1 score are 0.817 and 0.812, respectively. All the machine learning algorithms included for comparison, both data-driven and physics-guided, fail in accurately diagnosing the stack, underscoring the importance of adopting a LSTM neural network for correctly analyzing time series data. Compared to the data-driven LSTM neural network, the proposed physics-guided LSTM method shows a notable improvement, 55.9 % for the accuracy and 76.9 % for the F1 score. In order to illustrate the diagnosis performances more effectively, in Figure 4.11 we provide the confusion matrices for the data-driven LSTM neural network (left) and the physics-guided LSTM method (right).



(a) (b)

Figure 4.11: Confusion matrix for the data-driven LSTM (left) and the physics-guided LSTM (right) method. The train and test datasets are illustrated in Figure 4.10.

The physics-guided LSTM diagnoses well the faults in the test dataset, remarkably showing a score of 0.982 for the hydrogen starvation fault (4), while the data-driven LSTM is not capable of diagnosing starvation faults (3,4) and achieves inadequate scores on the other fault classes. This suggests that the additional information about the internal states of the PEMFC, provided by the physics-based model, is remarkably useful to make the fault diagnosis more robust. Furthermore, in Figure 4.12 we show the 2D representation of the features extracted by the LSTM neural network in the two methods, after the last hidden layer, through t-distributed Stochastic Neighbor Embedding (t-SNE) visualization [200]. This is a nonlinear dimensionality reduction technique used to project high-dimensional data into a low-dimensional space, typically 2D or 3D. It is widely used to visualize the distribution of learned features, as it preserves local structure and reveals clusters or separability in the data. Figure 4.12a shows the features extracted by the data-driven LSTM neural network, while Figure 4.12b shows the distribution of the features extracted by the physics-guided LSTM neural network. Only in the latter 5 different clusters, one for each fault state, are clearly distinguishable, highlighting the motivation behind the notable performance improvement.



(a) (b)

Figure 4.12: Two-dimensional representation by means of t-SNE of the features extracted by the LSTM neural network, (left) data-driven LSTM, (right) physics-guided LSTM. The dataset is shown in Figure 4.10

Eventually, the time to detect metric confirms these results. The physics-guided LSTM requires 66.9 s to diagnose the faults at test time, against 298.1 s of the data-driven LSTM, with a reduction of the detection time of 77.6 %.

4.5.4 Discussion

As stated in the introduction and demonstrated in the three elucidative cases study, the proposed physics-guided LSTM method is designed to increase the discriminative power and improve the generalization ability of the fault diagnosis algorithm. The estimates of the fault related variables provided by the physics-based PEMFC model are informative of the state of the PEMFC stack and, as a result, increase the accuracy and decrease the time to detect of the physics-guided LSTM compared to the data-driven LSTM. Indeed, by integrating the model estimates and the measured signals, it is possible to significantly enhance the input space of the LSTM neural network without increasing the number of sensors installed on the stack. Concurrently, the measurements acquired through sensors help mitigate the impact of errors and biases in the physical model estimates. By continuously incorporating measured signals from the PEMFC stack, the methodology can compensate for deviations introduced by modeling inaccuracies. The last case study, which tested the methods with unseen data, evidences that the physics-guided LSTM is capable of effectively adapting to degraded signals, and thus, it is robust to PEMFC aging. This proves a greater generalization ability, which is a clear advantage with respect to the data-driven LSTM neural network, as well as traditional machine learning algorithms, i.e., SVM, RF, GB, FNN. Our findings yield several key insights:

1. The LSTM neural network is confirmed to be the most suited algorithm to deal with time series signals, collected through sensors on the PEMFC stack. It is capable of capturing complex functional patterns in

the input data, bringing a clear advantage in diagnosing the faults within the stack.

2. The model estimates of process variables relevant for the fault description compensate for the limited sensors that can be embedded on a PEMFC in real vehicular applications. Accordingly, the physics-based model can provide meaningful prior knowledge on the PEMFC fault states, allowing to infer informative process variables not available from the measured signals, which demonstrate to be beneficial for diagnosing the stack.

3. Moreover, as proven in the last case study, the process variables provided by the model serve as additional information for the algorithms to better discern the faults, limiting the confusion given by the data distribution shift due to aging. Indeed, instead of relying only on voltage and pressure drop measurements, we can additionally include membrane temperature, membrane water content, relative humidity in the channels and current density. Increasing the amount of information that the LSTM neural network has access to, the algorithm has greater flexibility in representing complex fault phenomena and, thus, it can better capture the underlying relationship in the data.

However, to some extent, certain limitations in the presented method persist. First, the feasibility of the online implementation of the physics-based PEMFC model is not investigated here. Even if we reproduce the assessment of the inputs for the model through the actuators signals, the realistic use comes with more challenges. Limited computational capability to embed the model on the FCMSs, and in general on electronic boards, is a well-known issue of physics-based models. This highlights the need for a dedicated model order reduction technique to adapt the proposed method for online implementation. Second, the sensitivity of the hybrid fault diagnosis framework to varying levels of modeling error has not been assessed. This would clarify whether precise model calibration is required for optimal performance or if the approach remains robust with reduced calibration effort. Moreover, the accuracy of the physics-based model could degrade over time due to changes in system behavior or environmental conditions. In such scenarios, improved offline calibration or online parameters' adaptation could help to preserve the relevance and reliability of the model estimates. Finally, although our primary focus was to investigate the contribution of embedded physics knowledge to fault diagnosis, further improvements of the classification algorithms could be achieved through automated optimization of LSTM network architecture. While we selected the architecture to maximize classification accuracy, we acknowledge that this heuristic process may not lead to a globally optimal solution. Utilizing tools for systematic hyperparameter tuning could enhance the overall performance and robustness of the physics-guided fault diagnosis method.

4.6 Summary

In this chapter, we presented and discussed three major contributions of this PhD work to the field of PEMFC fault diagnosis.

First, we emphasized the importance of using dedicated time series processing techniques. We demonstrated that traditional machine learning algorithms are insufficient for handling the dynamic and complex nature of real-time signals measured on the fuel cell stack, worsening the fault diagnosis performance. By leveraging algorithms designed to explicitly account for the temporal structure of the data, specifically LSTM networks, we could better capture the fault progression in the signals, and thus improve the diagnosis accuracy and reduce the time to detect the occurrence of a fault.

Second, we proposed and detailed a novel hybrid framework for integrating a physics-based PEMFC model with a LSTM network for fault diagnosis, outlining the practical implementation of the model. The key idea is to enhance the input space of the classification algorithm with informative estimates of the PEMFC internal state provided by the physics-based model. This physics-guided method increases the knowledge on the internal conditions of the system, which is otherwise limited when data come only from sensors, thus significantly improving the capability of the LSTM network to discriminate fault states.

Third, we demonstrated that the proposed framework is beneficial also for the robustness of the fault diagnosis

algorithm, particularly in the presence of PEMFC aging and signal degradation. The physics-guided method maintains higher performance even as the stack undergoes gradual performance loss. These findings were validated through experiments conducted on a real PEMFC stack over more than 1000 hours of operation, offering an unprecedented validation in a real-world dataset. The physics-guided LSTM method represents a notable improvement for PEMFC fault diagnosis. Its ability to accurately diagnose the faulty states of the stack under dynamic load profiles and in a wide range of operation times is essential for making the PEMFC technology more reliable and reduce the effects of degradation on the diagnosis performances.

Eventually, we discussed the limitations of the method and outlined future directions. While the diagnosis results are promising, further work is needed to adapt the physics-based model for online implementation deployment. The limited computational resources of embedded systems call for the use of model order reduction techniques and potentially surrogate modeling to retain accuracy while ensuring feasibility. Future investigations will also focus on determining whether the physics-guided method can be beneficial in reducing the need for large experimental datasets for algorithms training.

In conclusion, the methodology developed in this chapter represents a significant step toward realistic and practical PEMFC fault diagnosis. By bridging physics-based modeling with data-driven learning, it lays the foundation for generalizable and scalable fault diagnosis algorithms. This work lays the ground for more innovations in the field, with the ultimate goal of extending PEMFC durability and supporting their envisaged adoption in automotive applications.

5

Transfer Information from Physics Models via Domain Adaptation

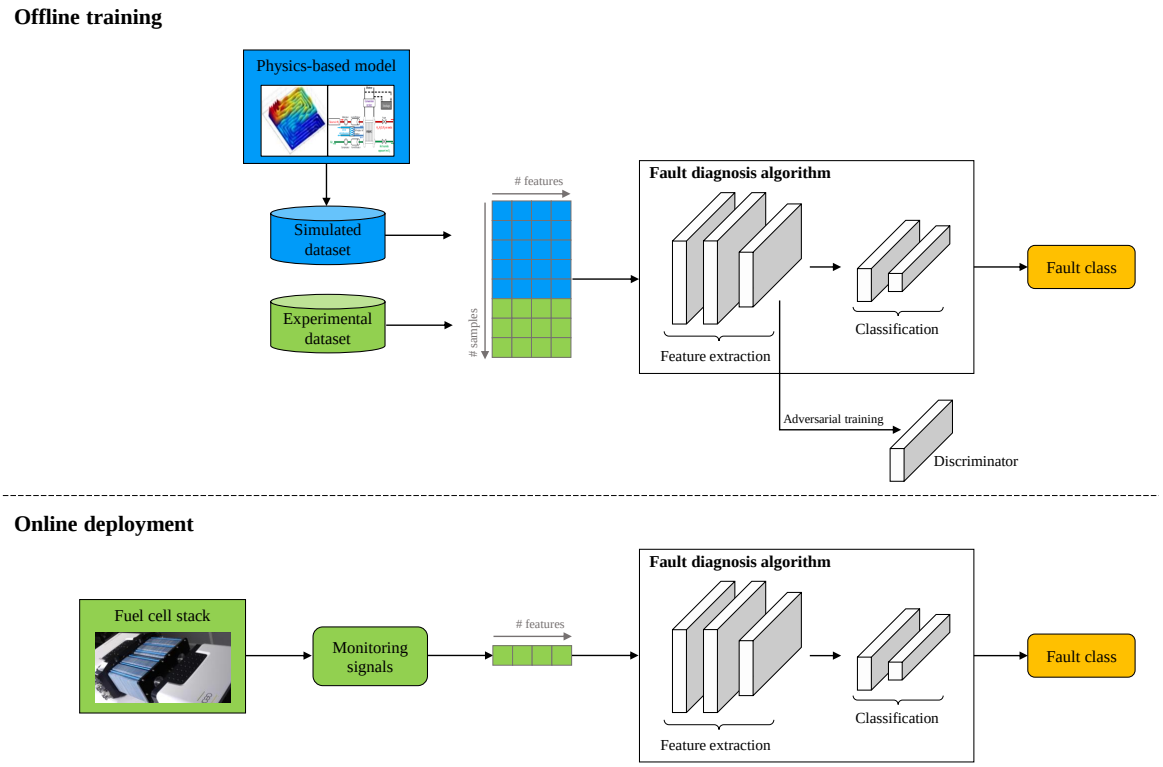


Figure 5.1: **Domain adversarial neural network.** The physics-based model is used offline to generate a physics-relevant, simulated dataset. The information embedded in this dataset is then used to build a deep learning algorithm via adversarial training, together with the experimental data available. At testing time, only the monitoring signals measured on the fuel cell stack are used to diagnose the system.

Contents

5.1 Introduction 69
5.2 Background on domain adaptation 70
5.2.1 Domain adaptation 70
5.2.2 Domain adversarial neural network 72
5.2.3 Open issues 72
5.3 The methodology 73
5.3.1 Problem description 74
5.3.2 Supervised DANN 74
5.3.3 Comprehensive framework 76
5.4 Implementation details and results 77
5.4.1 Physics-based simulated dataset 77
5.4.2 Insights on the sim-to-real transfer 78
5.4.3 Details on the source and target domain datasets 81
5.4.4 Results for the supervised DANN 82
5.5 Summary 86

5.1 Introduction

In the previous chapter we explored how a physics-based model can serve to expand our understanding of the PEMFC internal states, offering valuable information that enhances the robustness of the fault diagnosis. However, the proposed approach requires the online deployment of a physics-based model. In embedded automotive applications, this represents a challenge, since the computational resources are limited and fast inference is required. Consequently, in this chapter we explore how the detailed knowledge embedded in a physics-based model can be used offline to support the development of robust fuel cell fault diagnosis, and how it can relieve the heavy data requirements typically associated with the training of deep learning algorithms.
We have seen that, due to their superior capabilities in addressing nonlinearities, deep learning algorithms, including LSTM networks, are powerful tools for capturing fault patterns from the complex signals measured on PEMFC systems, and they generally outperform traditional algorithms. Nonetheless, they pose specific challenges, mainly the amount of data needed to train them correctly. While being a critical issue, data scarcity has not been investigated much in the field of PEMFC fault diagnosis yet. It is generally assumed that sufficiently high-quality datasets are available for estimating the underlying statistical distribution in nominal and faulty conditions. However, collecting large datasets, with an adequate number of representative samples, is difficult and expensive for PEMFC stacks. When addressing PEMFC stack fault diagnosis, we often deal with datasets of insufficient size, from which it is challenging to learn fault patterns and develop efficient fault diagnosis algorithms with a correct level of generalization capability. To build a highly generalizable, pure data-driven, deep learning algorithm, the training data must cover all the possible faulty states in a wide range of operating conditions, with usage profiles close to real world application scenarios. A lack of representativeness in the training dataset can dramatically affect the algorithms' performance on real operating systems, thus making them less competitive. If the algorithm fails in generalizing to different operating and environmental conditions, it will not be robust enough to be embedded in the FCMSs of the vehicles.
As previously discussed, in machine learning, the integration of physics-based models into data-driven algorithms has emerged as a promising solution to address deficiencies in the available dataset and enlarge their representativeness [201, 202]. Specifically, a simulation model can effectively provide unlimited amounts of labeled data, reproducing large variations in the operating conditions. Nonetheless, simulated datasets cannot be directly used to train algorithms intended for real-world deployment [203], since even accurate physics-based models are affected by errors. Moreover, in PEMFC models, only the main component is usually modeled, while the surrounding is disregarded, leading to imprecise estimations [204]. These errors cause a distributional shift between the simulated data and the measurements collected on real systems. While this shift could be reduced by making the simulation model even more accurate, this is time-consuming [205]. A promising solution to this problem is offered by domain adaptation methods, a subfamily of transfer learning [206]. Several works have demonstrated that by correctly implementing domain adaptation strategies to incorporate prior knowledge from physics-based models into deep learning algorithms, the dependence on data can be reduced, and the fault diagnosis algorithms could benefit of some interpretability [207, 208]. Hence, we focus on developing a PEMFC fault diagnosis methodology by means of domain adaptation techniques, with the aim of transferring information from a physics-based model to a deep learning algorithm.
The major contributions of this study are as follows:

1. We propose a methodology, in the framework of domain adaptation, that profoundly fuses physics-based modeling knowledge within the offline training of the deep learning algorithm, aiming to overcome data scarcity in PEMFC fault diagnosis;

2. We develop a supervised adversarial network to correctly reduce the discrepancy between simulated and real data, while achieving fault discriminability for PEMFC;

3. We demonstrate that the methodology that we propose, by correctly leveraging physics-relevant simulated data, requires much less data than the pure data-driven algorithms.

This Chapter is partly based on two works published in two international conferences:

- C. Pettorossi, V. Heiries, M. Gerard and S. Rosini, "Addressing Data Scarcity in PEMFC Fault Diagnosis Using Adversarial Learning," 2024 IEEE International Conference on Prognostics and Health Management (ICPHM), Spokane, WA, USA, 2024, pp. 393-398, doi: 10.1109/ICPHM61352.2024.10627464.

- C. Pettorossi, V. Heiries, R. Morvillier, S. Rosini and M. Gerard, "PEM Fuel Cell Fault Diagnosis in Automotive Applications Under Aging Conditions: A Physics-Informed Deep Learning Method," 2025 IEEE/AIAA Transportation Electrification Conference and Electric Aircraft Technologies Symposium (ITEC+EATS), Anaheim, CA, USA, 2025, pp. 1-6, doi: 10.1109/ITEC63604.2025.11098144.

5.2 Background on domain adaptation

The effectiveness of machine learning algorithms relies on the assumption that the training and test data are drawn from the same underlying distribution. Consequently, deep learning algorithms require to be trained with large, diverse and representative datasets in order to perform efficiently across varying conditions in real application. This is the reason why the success of deep learning has predominantly occurred in fields such as computer vision or natural language processes where vast amount of data are readily available [209]. However, in most of the real-world applications, including PEMFC fault diagnosis, the size of the available datasets is often modest, and the data collection procedure are expensive as well as extremely time-consuming. In such contexts, a promising solution is to leverage data from related domains and adapt them to the required field of application through transfer learning. Transfer learning is a family of machine learning approaches that relax the assumption that training and test data must originate from the same distribution, by enabling algorithms to transfer knowledge from one domain (source) to another (target) [210]. The source domain may consist of data from a different system, unit, operating conditions, or a high-fidelity simulation, while the target domain typically represents the specific real-world system where the algorithm is to be deployed, for which data are rare, expensive to collect and label, or inaccessible [211]. In essence, transfer learning addresses the challenge of learning in data-limited scenarios by allowing information to be reused across different environments, including between physics-based simulation models aand real systems [212]. This allows to alleviate the need for large datasets by exploiting the information embedded in different but similar fields [213]. Notwithstanding, transferring knowledge across domains is not trivial because a distributional shift is often present between the two domains, and dedicated techniques are required [214]. The most widely used and straightforward approach is fine-tuning, where a neural network pretrained on a source domain is adapted to the target domain using a smaller set of labeled target data to update its parameters and specialize it to the target distribution. This method has been successfully applied in the context of PEMFC, where a pretrained auxiliary network has been fine-tuned on task-specific data to improve the fault diagnosis performance [215]. Beyond fine-tuning, domain adaptation and domain generalization have emerged as two notable and more flexible approaches [216]. Domain adaptation aims at learning algorithms that can bridge the distributional shift between source and target domain, while domain generalization aims at learning algorithms that can generalize to an unseen target domain by leveraging information from several source domains [209]. Notably, domain adaptation has gained attention in the PHM community, as it offers a systematic approach to exploit simulation data or data from similar machines to build robust algorithms for real-world deployment of fault diagnosis [217].

5.2.1 Domain adaptation

Domain adaptation is an ensemble of techniques that can support the knowledge transfer from a source domain to a target domain. It attempts to solve the issue of enabling an algorithm trained on a certain dataset, i.e., source domain, to perform well on a differently distributed dataset, i.e., target domain [214]. To address this, domain adaptation techniques aim to learn a representation that is both discriminative for the source task and invariant to the domain-specific features, such that the algorithm can generalize well on the target data despite the domain distribution shift [218]. For example, it can alleviate this shift, by constructing a domain invariant feature space

in which the distributions of source and target domain are aligned. This domain shift problem is illustrated in Figure 5.2.

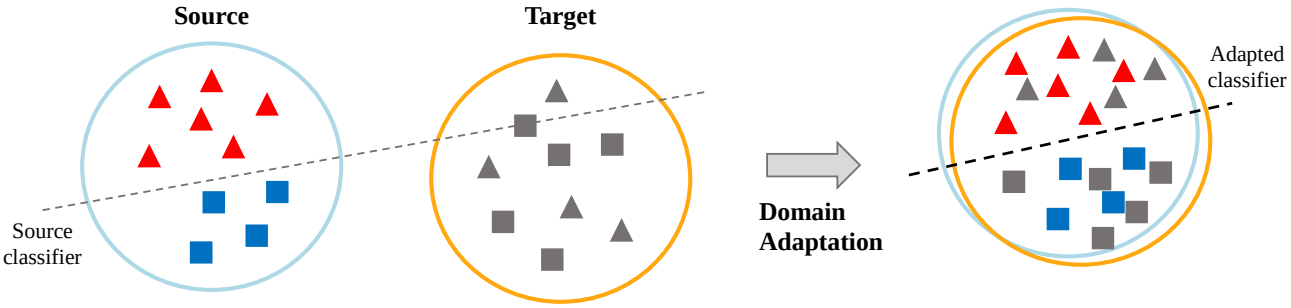


Figure 5.2: Domain adaptation: illustration of the distribution shift between source and target domain.

The source domain has rich knowledge, which can be used to enhance the identification accuracy of the target domain, in which the knowledge is limited. Specifically, the general assumption in domain adaptation problems is that the data in the target domain are not labeled. Thus, domain adaptation is usually formulated as an unsupervised problem. We introduce a mathematic description of the problem. A domain $\mathcal{D}$ consists of an input space $\mathcal{X}$ and a marginal probability distribution $P(X)$, from which a dataset $X \in \mathcal{X}$ can be obtained, so that $D = \{X, P(X)\}$. For a given domain $\mathcal{D}$, a task (e.g., classification) can be defined to map the input space $\mathcal{X}$ into an output space $\mathcal{Y}$. A dataset sampled from the domain $\mathcal{D}$ is defined by $D = \{X, Y\} = \{(\underline{x}_i, y_i)\}_{i=1}^{n}$, where $X = \{\underline{x}_i\}_{i=1}^{n} \in \mathcal{X}$ is a set of n samples (e.g., multivariate time series), and $Y = \{y_i\}_i^n \in \mathcal{Y}$ are the corresponding labels (e.g., fault classes). In machine learning, the dataset D is used to learn the mapping $f(\cdot)$ between $\mathcal{X}$ and $\mathcal{Y}$, in order to achieve the task. For the purpose of fault diagnosis, in the following discussion we will focus on the specific case of classification tasks. In domain adaptation settings we define:

- a labeled source domain dataset

$$D_s = \{X_s, Y_s\} = \{(\underline{x}_{s,i_s}, y_{s,i_s})\}_{i_s=1}^{n_s} \tag{5.1}$$

  with marginal probability distribution $P(X_s)$ and n_s samples;

- an unlabeled target domain dataset

$$D_t = \{X_t\} = \{\underline{x}_{t,i_t}\}_{i_t=1}^{n_t} \tag{5.2}$$

  with marginal probability distribution $P(X_t)$ and n_t samples.

The marginal probability distributions for the two domains are different, i.e., $P(X_s) \neq P(X_t)$. This can be related to noise in the measurements, different environmental conditions or variations in the data collection processes. Then, the objective of domain adaptation is to improve the learning of a mapping function $f_t(\cdot)$ in the target domain $\mathcal{D}_t$ using the knowledge available in the source domain $\mathcal{D}_s$. Numerous domain adaptation methods have been proposed over the past decade, especially in deep learning. The appeal of domain adaptation approaches is the ability to learn mappings between domains in situations in which target domain data are either fully unlabeled (unsupervised) or have few labeled samples (semi-supervised). Hence, most of the applications of domain adaptation to machinery fault diagnosis has often been proposed to address unsupervised problems, given the assumption that labeled datasets are rarely available [219]. Some techniques aim to map the feature distributions of the two domains using statistical distances such as the maximum mean discrepancy (MMD) or correlation alignment (CORAL). Others, like adaptive batch normalization (AdaBN), adjust the internal network statistics to reduce covariate shift between domains [220]. A widely adopted family of methods is based on adversarial learning, such as the domain adversarial neural network (DANN), which introduces a domain discriminator to encourage the network to learn domain-invariant features. The success of (DANN)s is largely attributed

to the ability to learn domain-invariant representations of the input data, which substantially improves model performance in real-world scenarios. Motivated by the promising results reported in the literature, this work focuses on exploring the applications of DANN to fault diagnosis tasks.

5.2.2 Domain adversarial neural network

The idea behind DANN is to combine domain adaptation and deep feature learning within one training process [221]. It was inspired by the generative adversarial network (GAN) proposition, in which synthetic data can be generated by training the algorithm in an adversarial manner. DANN instead, while having a similar structure, has been proposed to address transfer learning problems [222]. The objective is to learn a feature space from the input data that is both discriminative for the final task, e.g., classification, and invariant with respect to the domain. The latter is achieved by adversarially learning a feature representation of the input space where the shift in the marginal distributions of source and target domain, i.e., $P(X_s)$ and $P(X_t)$, is reduced.
The original formulation of DANN comprises a feature extractor $G_f(\cdot;\underline{\theta}_f)$ with parameters $\underline{\theta}_f$, a domain discriminator $G_d(\cdot;\underline{\theta}_d)$ with parameters $\underline{\theta}_d$ and a classifier $G_y(\cdot;\underline{\theta}_y)$ with parameters $\underline{\theta}_y$. The feature extractor G_f transforms the input data into a feature representation that is intended to be domain-invariant as well as relevant to the classification task. The domain discriminator G_d serves as domain label predictor, $z = 0$ for source domain and $z = 1$ for target domain, while the classifier G_y serves as predictor for the fault label y. The adversarial learning is implemented between the feature extractor and the domain discriminator by means of a gradient reversal layer (GRL): in the feed forward pass of the network, GRL serves as identity mapping, in the back-propagation GRL applies a flipping sign [221].

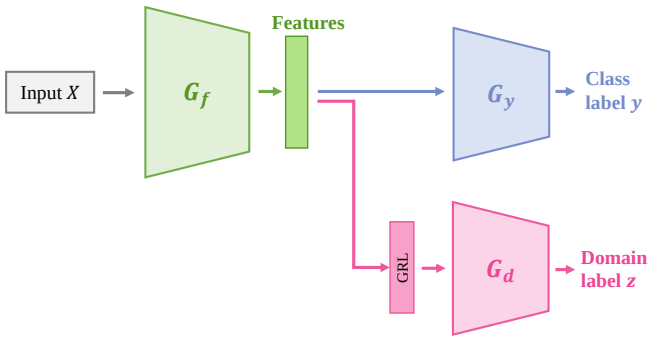


Figure 5.3: Scheme of unsupervised DANN (adapted from [221]).

Then, a labeled dataset from source domain $D_s = \{X_s, Y_s\}$ and an unlabeled data from target domain $D_t = \{X_t\}$ are used for training. The adversarial loss is:

$$\mathcal{L}(\underline{\theta}_f, \underline{\theta}_y, \underline{\theta}_d) = \frac{1}{n_s} \sum_{\underline{x}_i \in D_s} \mathcal{L}_y(G_y(G_f(\underline{x}_i)), y_i) - \lambda_d \frac{1}{n_s + n_t} \sum_{\underline{x}_i \in \{D_s, D_t\}} \mathcal{L}_d(G_d(G_f(\underline{x}_i)), z_i) \tag{5.3}$$

where $\mathcal{L}_y$ is the loss function of the classifier, $\mathcal{L}_d$ is the loss function on the discriminator and λ_d is a weighting parameter for the discriminator loss. Hence, this loss minimizes the classification error and maximizes the domain discriminator error, so that the extracted features capture the underlying structure of the data relevant to the classification while being agnostic to the domain from which the data originate.

5.2.3 Open issues

DANN has found widespread applications in fields like images or speech recognition in which abundant data are available. Some studies extended DANN to field of application related to time series data, proposing dedicated architecture with convolutional networks [79] or recurrent networks [223, 224]. More recently this methodology has received increasing interest in the PHM community, both for solving diagnosis and prognosis tasks [225]. However, especially in the field of PHM, and fault diagnosis in particular, several works have proposed variations of its implementation, demonstrating that, while DANN provides a general framework, there is the necessity to

tailor the methodology to the specific use case [226]. Indeed, DANN, like most of the works published on domain adaptation, is developed in an unsupervised framework, since not requiring labeled data in the target domain is an attractive feature [154]. While the underlying assumption in DANN is that by aligning the marginal distributions of $\mathcal{D}_s$ and $\mathcal{D}_t$ in the feature space, the classifier can operate effectively on both domains, this is not always true in practice. Its performance depends on the level of discrepancy between the two domains along with the amount of available data in the target domain. In practice, when the datasets available are imbalanced or the data are not sufficient to capture the overall distribution, aligning solely the marginal distributions of the two domains can lead to misclassification issues. This is known as negative transfer, which causes unsatisfactory performances for the classification task in the target domain [227]. Negative transfer can be particularly pronounced in unsupervised framework. Indeed, if the target dataset is not sufficient to represent the underlying domain distribution, it is difficult to preserve class relevant information while aligning the target marginal distribution to the source marginal distribution. As a consequence, the class-conditional distributions might be not correctly aligned in the feature space, i.e., $P(X_s \mid Y_s) \nsim P(X_t \mid Y_t)$. While being domain invariant, the extracted features are aligned in the wrong direction and, thus, they are not discriminative for the classification task in the target domain [228]. This results in an algorithm that is not able to perform the desired task in the target domain. In classification problems, it implies that data in the target domain could be incorrectly classified by the decision boundaries learned in the source domain. In practice, unsupervised domain adaptation methods require a considerable amount of data in the target domain to be effective, and this is emphasized even more when using deep learning algorithms including DANN.
Propositions to address this issue include: using a measure of entropy to exclude out-of-distribution or hard-to-transfer samples that make the distribution alignment harder to achieve [229]; introducing class reweighting as a measure to counteract the negative transfer [230]; use multiple discriminators to correctly align class-conditional distributions [226]; combine domain adaptation with GAN to generate unseen faulty conditions [231]; use domain generalization to better capture highly generalizable features exploiting more datasets [232]. Contrastive learning has also been presented as a possible alternative to adversarial learning [233]. To find domain invariant feature spaces by means of DANN in fields of applications where large datasets are unavailable, methods revolving towards weak supervision in DANN algorithm have been presented [234]. Similar works suggested that including supervised learning in the framework of DANN could indeed be beneficial, since it allows achieving better algorithm generalization through the source domain as well as better discriminative capabilities in the target domain, thus leading to a better performing algorithm.

5.3 The methodology

In this study, we propose a supervised DANN-based method for scenarios with scarce data availability, focusing on PEMFC fault diagnosis. It consists of two steps. First, a physics-based model is used to generate a simulated dataset of the PEMFC stack operating with dynamic load under several faulty conditions. Second, an adversarial neural network is trained together with simulated and real data to obtain a fault diagnosis algorithm. Indeed, a dataset generated with a physics-based model integrates physical knowledge, and it can be representative of several operating conditions not available in the real dataset, boosting the generalization of the fault diagnosis algorithm. Moreover, to better capture the temporal dependency embedded in the PEMFC monitoring signals, we introduce a RNN in the DANN architecture. To overcome the challenges related to the negative transfer and insufficient target data, we present here a supervised variation of DANN. Therefore, we operate under the hypothesis that labels are available in the target domain, allowing us to formulate the problem in a supervised setting. The supervised DANN framework is thoroughly described in the following.
We validate the effectiveness of the methodology in exploiting the knowledge embedded in a physics-based model for PEMFC fault diagnosis. In particular, scenarios with limited data in the target domain are investigated, demonstrating that the supervised DANN-based method allows to considerably reduce the need for large PEMFC faulty datasets compared to pure data-driven algorithms.

5.3.1 Problem description

The transfer from a physics-based simulated dataset to a real monitoring dataset of PEMFC stack is here investigated for fault diagnosis. In this work, we refer to the simulated dataset as source domain and to the real dataset as target domain. The study is carried out under the following assumptions:

- Sufficient labeled data are available from the source domain;
- Few labeled data are available from the target domain;
- Fault classes are the same in the two domains.

The source domain dataset D_s is given in Equation (5.1), while the target domain dataset D_t becomes:

$$D_t = \{(\underline{x}_{t,i_t}, y_{t,i_t})\}_{i_t=1}^{n_s} \sim P(X_t) \tag{5.4}$$

where the fault label y_{t,i_t} has been included. In both datasets, $\underline{x} \in \mathbb{R}^{m \times L}$ represents the input multivariate time series constituted by m signals of L consecutive time steps and $y \in \{0, \dots, K-1\}$ represents the corresponding fault label, K being the number of fault classes. Additionally, we define $P(X_s \mid Y_s)$ and $P(X_t \mid Y_t)$ as the class-conditional distribution for the source and target domain, respectively.
The objective of this work is to build a fault diagnosis classifier which correctly classifies the input time series in the target domain, i.e., $\underline{x}_{t,i_t}$, to the corresponding fault label y_{t,i_t}. We focus on situations in which few data are available in the target domain, which implies $n_t \ll n_s$. We aim at aligning the marginal distributions as well as the class-conditional distributions of the source and target domain in the feature space. Indeed, it is expected that the marginal distributions for the source and target domain are not aligned: there is always a modeling error associated with the physics-based model used to generate the simulated dataset and, on the other hand, the real dataset is always affected by noise. Moreover, by using all the supervised information available, we expect to avoid the negative transfer issue by correctly aligning the class-conditional distributions over the feature space, i.e., $P(X_s \mid Y_s) \simeq P(X_t \mid Y_t)$. This will guarantee the correct discrimination of the faults in the target domain.

5.3.2 Supervised DANN

A supervised fault diagnosis algorithm inspired by DANN [221] architecture is here proposed. In contrast to the original algorithm, which had no supervision on the target domain, we introduce supervision in the adversarial alignment. Since supervised domain adaptation approaches can better ensure the correct alignment of the source and target distributions in the feature space, considering the class labels in the target domain can be beneficial. Indeed, the straightforward implementation of DANN aligns marginal distributions, but does not guarantee the preservation of the class separability in the target domain. Under this circumstance, we resort to the fault labels information in the target domain to facilitate the extraction of fault relevant information by adding a term in the loss function for the target domain classification. This allows learning a feature space with better intra class cohesion and interclass separability, in which features are domain invariant as well as discriminative for the fault diagnosis task.
The architecture of the supervised DANN is shown in Figure 5.4. It consists of three modules: a feature extractor G_f, a fault classifier G_y and a domain discriminator G_d, whose parameters are denoted as $\underline{\theta}_f$, $\underline{\theta}_y$ and $\underline{\theta}_d$, respectively.

- The feature extractor G_f consists of 3 RNN layers, which have been selected for being suitable in dealing with temporal data. The RNN-based feature extractor aims at finding a latent feature space in which the feature representation of the source and target domain have effective overlapping distributions, i.e., $P(X_s \mid Y_s) \simeq P(X_t \mid Y_t)$, and are discriminative with respect to the fault class.
- The fault classifier G_y receives in input the feature representation $G_f(\underline{x})$ and produces in output the K-dimensional vector $\underline{\hat{y}} = [\hat{y}_0, \dots, \hat{y}_{K-1}]$, where $\hat{y}_k$ indicates the probability that $\underline{x}$ belongs to the k-th fault class. The fault classifier consists of a fully connected layer, followed by a softmax activation

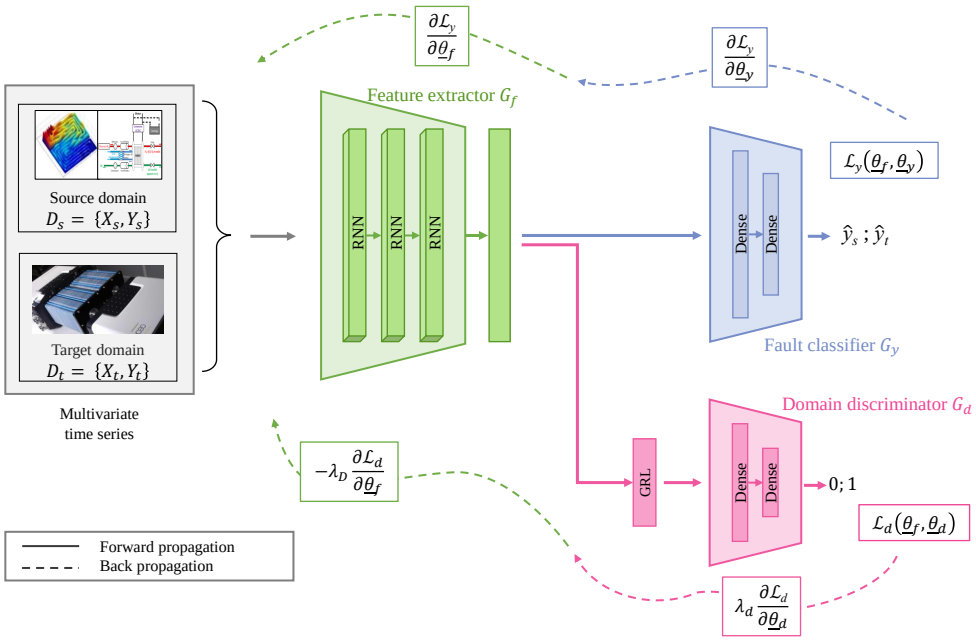


Figure 5.4: Scheme of supervised DANN: feature extractor G_f, fault classifier G_y and domain discriminator G_d. A RNN network is used in G_f to effectively process the multivariate time series.

function. The loss function for training G_y, $\mathcal{L}_y(\underline{\theta}_f, \underline{\theta}_y)$, is the sparse categorical cross-entropy. It guides the classifier to improve its classification performance and make accurate predictions.

- The domain discriminator G_d receives in input the feature representation $G_f(\underline{x})$ and produces in output a 2-dimensional vector $\underline{\hat{z}} = [\hat{z}_0, \hat{z}_0]$, where $\hat{z}_0$ indicates the probability that $\underline{x}$ belongs to the source domain, while $\hat{z}_1$ indicates the probability that $\underline{x}$ belongs to the target domain. It consists of fully connected layers, followed by a softmax activation function. The loss function for training G_d, $\mathcal{L}_d(\underline{\theta}_f, \underline{\theta}_d)$, is the binary cross-entropy, which quantifies the discrepancy between the source and target domain labels. Its objective is to ensure that the features extracted by the feature extractor are distinguishable by the domain discriminator.

The GRL layer has been specifically included in the DANN architecture between G_f and G_d to enable a synchronized adversarial training of the network [221]. Specifically, the parameters $\underline{\theta}_f$ and $\underline{\theta}_y$ are optimized to extract discriminative feature representations for the fault diagnosis task, by minimizing the classification loss $\mathcal{L}_y(\underline{\theta}_f, \underline{\theta}_y)$ on the data of the source domain and on the few data of the target domain. The parameters $\underline{\theta}_d$ are first optimized by minimizing the discriminator loss $\mathcal{L}_d(\underline{\theta}_f, \underline{\theta}_d)$. Then, they are adversarially optimized over $\underline{\theta}_y$ to obtain domain-invariant features using the overall loss function given in equation (5.5):

$$\begin{aligned}\mathcal{L}(\underline{\theta}_f, \underline{\theta}_y, \underline{\theta}_d) = &\frac{1}{n_s} \sum_{\underline{x}_i \in D_s} \mathcal{L}_y(G_y(G_f(\underline{x}_i)), y_i) \\ &+\lambda_y \frac{1}{n_t} \sum_{\underline{x}_i \in D_t} \mathcal{L}_y(G_y(G_f(\underline{x}_i)), y_i) \\ &-\lambda_d \frac{1}{n_s + n_t} \sum_{\underline{x}_i \in \{D_s, D_t\}} \mathcal{L}_d(G_d(G_f(\underline{x}_i)), z_i)\end{aligned} \tag{5.5}$$

where z_i is the domain label, $z_i = 0$ for the source domain and $z_i = 1$ for the target domain. Differently from Equation (5.3), supervision on the target domain through the available fault labels y_t is included: the second term in Equation (5.5) represents the supervised classification loss for the target domain, which supports the

alignment of the class-conditional distributions. The hyperparameters λ_y and λ_d are introduced to weight the loss functions. In the adversarial framework, the optimization of the overall loss in equation (5.5) is equivalent to the following two-players min-max game:

$$\hat{\underline{\theta}}_f, \hat{\underline{\theta}}_y = \arg\min_{\underline{\theta}_f, \underline{\theta}_y} \mathcal{L}(\underline{\theta}_f, \underline{\theta}_y, \hat{\underline{\theta}}_d) \tag{5.6}$$

$$\hat{\underline{\theta}}_d = \arg\max_{\underline{\theta}_d} \mathcal{L}(\hat{\underline{\theta}}_f, \hat{\underline{\theta}}_y, \underline{\theta}_d) \tag{5.7}$$

The feature extractor G_f is forced to extract features that confuse the discriminator, thereby maximizing the discriminator loss, while the discriminator tries to correctly distinguish between source and target domain, minimizing its own loss. By so doing, a domain invariant feature space is learned. At the same time, G_f is simultaneously forced to extract features that are well discriminative for both domains by minimizing the classification loss on the source as well as on the target domain. Differently from the unsupervised DANN, in this supervised formulation the parameters $\underline{\theta}_y$ are modified using also the observations in the target dataset. The training is detailed in Algorithm 3.

Algorithm 3 Training of the supervised DANN

Require: Source domain dataset $D_s = \{(\underline{x}_{s,i_s}, y_{s,i_s})\}_{i_s=1}^{n_s}$; Target domain dataset $D_t = \{(\underline{x}_{t,i_t}, y_{t,i_t})\}_{i_t=1}^{n_t}$; Loss weight parameters λ_d, λ_y; batch size h; learning rate μ; max epochs $n_{\text{epochs,max}}$

Ensure: Trained networks $G_f(\underline{\theta}_f)$, $G_y(\underline{\theta}_y)$, $G_d(\underline{\theta}_d)$

Initialize network parameters $\underline{\theta}_f, \underline{\theta}_y, \underline{\theta}_d$
for epoch = 1 to $n_{\text{epochs,max}}$ **do**
  $n_{\text{batch,s}} = n_s \,/\, h$; $n_{\text{batch,t}} = n_t \,/\, h$
  while $n_{\text{batch}} < \max(n_{\text{batch,s}}, n_{\text{batch,t}})$ **do**
    Sample source batch $\{(\underline{x}_{s,i_s}, y_{s,i_s})\}_{i_s=1}^{h}$ from D_s
    Sample target batch $\{(\underline{x}_{t,i_t}, y_{t,i_t})\}_{i_t=1}^{h}$ from D_t
    Update parameters $\hat{\underline{\theta}}_f, \hat{\underline{\theta}}_y, \hat{\underline{\theta}}_d$ by:

$$\hat{\underline{\theta}}_f \leftarrow \underline{\theta}_f - \mu\left(\frac{\partial L_{y,s}}{\partial \underline{\theta}_f} + \lambda_c \frac{\partial L_{y,t}}{\partial \underline{\theta}_f} - \lambda_d \frac{\partial L_d}{\partial \underline{\theta}_f}\right)$$

$$\hat{\underline{\theta}}_y \leftarrow \underline{\theta}_y - \mu\left(\frac{\partial L_{y,s}}{\partial \underline{\theta}_y} + \frac{\partial L_{y,t}}{\partial \underline{\theta}_y}\right)$$

$$\hat{\underline{\theta}}_d \leftarrow \underline{\theta}_d - \mu \frac{\partial L_d}{\partial \underline{\theta}_d}$$

  end while
end for

5.3.3 Comprehensive framework

In Figure 5.5 we provide an overview of the comprehensive framework for the fault diagnosis methodology. First, the simulated dataset is generated with the physics-based model, which provides a large and representative set of labeled samples of monitoring variables for nominal and faulty operations of PEMFC. This constitutes the source domain dataset. In parallel, the experimental data are collected from the real PEMFC stack, and used as target domain dataset, since they represent the actual conditions in which the fault diagnosis algorithm will be deployed. Then, target domain dataset undergoes the full pre-processing procedure described in section 3.4: normalization, train-validation-test split, windowing and class balancing. For the source domain dataset, the train-validation-test split is not performed, since all the data can be used for training. Furthermore, class balancing is not necessary, since the simulated dataset is already generated balanced. Then, the architectures of the networks used in the supervised DANN, i.e., G_f, G_y and G_d, are selected. Specifically, the choice has been made to optimize the hyperparameters for the feature extractor G_f, i.e., number of layers and number of neurons of the network, using Optuna [235]. This is a Bayesian optimization framework that efficiently tunes the hyperparameters of a model using Tree-structured Parzen Estimator (TPE). The optimization of the hyperparameters is done on the validation set performance. Eventually, the training of the supervised DANN is performed adversarially using

the source domain dataset and the train set of the target domain dataset. Once the overall DANN is trained, and its performance validated, the features extractor and the classifier are used for the fault diagnosis task, while the domain discriminator is discarded.

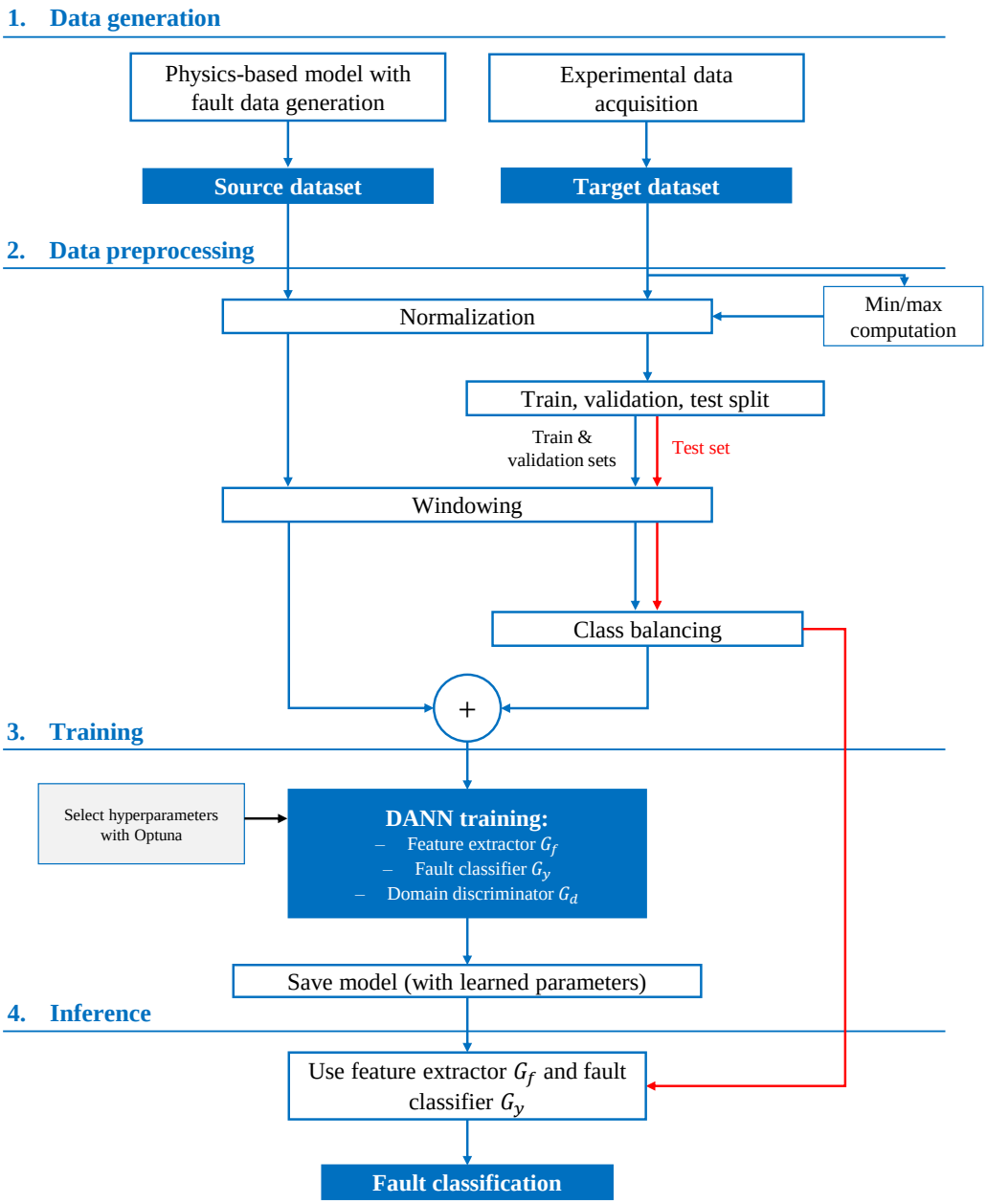


Figure 5.5: Flowchart of the proposed methodology.

5.4 Implementation details and results

5.4.1 Physics-based simulated dataset

The PEMFC physics-based model is used to generate a simulated dataset in Matlab/Simulink. After calibration and validation, extensively described in chapter 3, the generation of the simulated dataset is straightforward. First we select a load to be imposed to the stack, between a static current, a step current profile or a current profile corresponding to the FC-DLC cycle. Then, the fault injection procedure is applied for each fault class, i.e., nominal (0), flooding (1), drying (2), air starvation (3), hydrogen starvation (4), by modifying the nominal operating conditions. For each load profile, flooding is simulated setting high relative humidity for the input gases, while drying setting low relative humidity. Alternatively, low coolant temperature is used for flooding,

while high coolant temperature for drying. Air and hydrogen starvation are injected by lowering inlet air and hydrogen stoichiometry, respectively. More concretely, the fuel cell model receives in input the desired current load, and the operating conditions representative of the specific fault. This procedure is repeated multiple times for each fault class, as well as for the nominal class, to create a balanced dataset. To introduce variability, noisy inputs and different aging levels are used, resulting in a dataset that reflects a broad range of operating scenarios that are encountered in real-world applications. Figure 5.6 shows the kernel density estimation, through KDE plots of voltage, anode and cathode pressure drops, for a simulation with an average current load of 0.5 A/cm^2, where noise has been added.

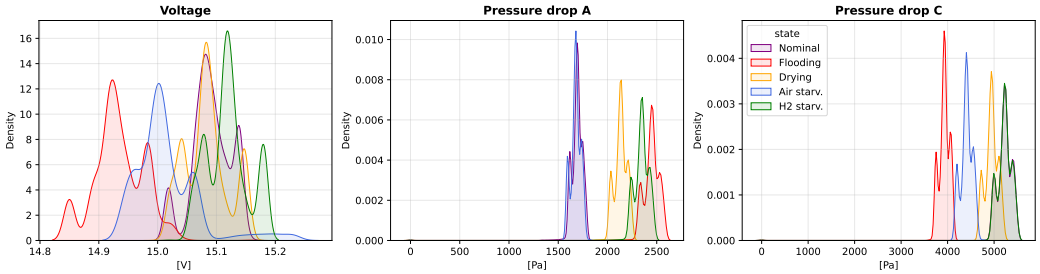


Figure 5.6: Kernel density estimation (KDE) plots of voltage, anode and cathode pressure drop, for a simulated average current load of 0.5 A/cm^2, with added noise. The nominal class and the four fault classes are represented.

5.4.2 Insights on the sim-to-real transfer

In the following, we aim to demonstrate the importance of domain adaptation and, in particular, of supervised domain adaptation methods, to transfer the knowledge from a simulated dataset (source) to a real, experimental dataset (target), with emphasis on the DANN algorithm. For this study, we use only the first 100 hours of stack operation as target dataset. We consider three main scenarios for training the fault diagnosis algorithm:

i) No adaptation. We train a fault classification algorithm using solely the source domain, and we test the performance directly on the target domain, without the support of any domain adaptation method;

ii) Unsupervised DANN. We implement the original formulation of DANN, as proposed in [221], to align the marginal distributions of source and target domain in the feature space. The datasets are $D_s = \{X_s, Y_s\}$ and $D_t = \{X_t\}$, i.e., the fault labels in the target domain are not used;

iii) Supervised DANN network. We implement the proposed supervised DANN, in which supervised training on the target domain is included. Hence, the datasets are $D_s = \{X_s, Y_s\}$ and $D_t = \{X_t, Y_t\}$.

In case ii) and iii), the network comprises a feature extractor G_f, a fault classifier G_y and a domain discriminator G_d, while in case i) the network comprises only a feature extractor G_f and a fault classifier G_y. The same architecture is considered for the three cases: G_f consist of 3 LSTM layers, with 16 recurrent neurons each; G_y consists of 2 Dense layers, with, respectively, 64 and 5 neurons (one per fault class); the domain discriminator G_d consists of 2 Dense layers, with, respectively, 32 and 2 neurons (one per domain).
In this work, we do not aim to perform a quantitative evaluation of the algorithms' performance. Instead, our objective is to qualitatively examine the challenges associated with each methodology and to motivate the necessity of using the supervised DANN. To this end, we employ t-SNE as a visualization tool to provide an intuitive 2D representation of the high-dimensional feature distributions. It is important to note that t-SNE should not be interpreted as a precise measure of the distributional distance between the source and target domains, and that the axes used in the visualization (x_1 and x_2) do not correspond to actual features. Indeed, while t-SNE effectively preserves local neighborhood structures and can group similar data points into clusters, it does not maintain the true global distances of the original feature space. Nonetheless, it serves as a useful tool to illustrate the presence of domain shift. In our context, it can reveal distinct clustering patterns and limited overlap between the source and target domains in the projected space, thereby qualitatively highlighting the presence of

distribution shift or of a correct distribution alignment. Hence, in Figure 5.7 we present the t-SNE visualization of the input spaces for the source (X_s) and target (X_t) domain. This provides an intuitive representation of the domain shift between their respective marginal distributions.

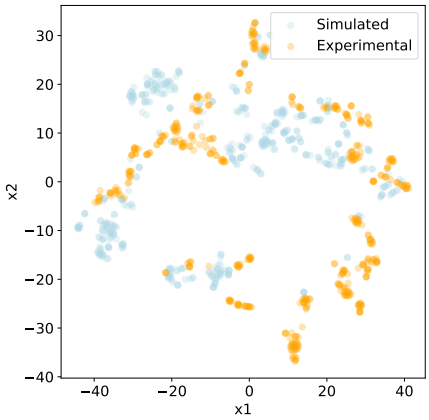
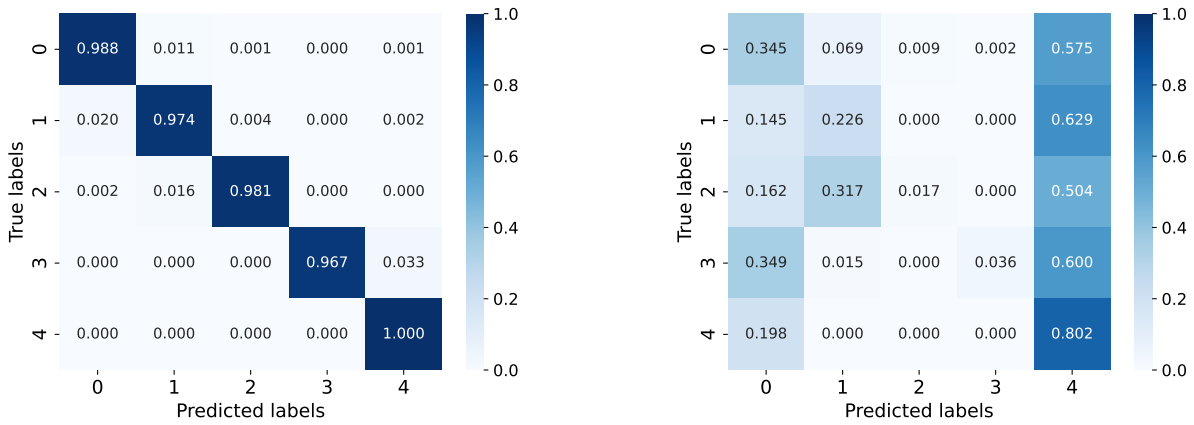


Figure 5.7: t-SNE visualization of the input space for source (simulated, light-blue) and target (experimental, orange) domain. For the target domain, we use only the first 100 hours of stack operation from the experimental dataset.

Indeed, let's consider case i). We train the network, comprising G_f and G_y, using the source dataset, and we test its performance in classifying faults on the target domain: the accuracy is 0.274. The confusion matrices in Figure 5.8, source domain (a) and target domain (b), reveal a clear reduction in the algorithm ability to correctly classify faults when applied in the target domain.



(a) (b)

Figure 5.8: Confusion matrices in the case of no adaptation i) for the source (a) and target (b) domain.

This performance degradation indicates that the algorithm, when trained only on source domain, fails to effectively generalize to the target domain. Consequently, direct transfer is not feasible, and a dedicated domain adaptation strategy is necessary. Accordingly, we consider case ii), in which domain adaptation is realized with DANN in the original unsupervised formulation, using source and target dataset for training the algorithm. In Figure 5.9 we present the t-SNE visualization of the feature space, once the network has been trained. On the left, the marginal distributions of the features extracted from the source $G_f(X_s)$ and target $G_f(X_t)$ domains appear well aligned. However, although five clusters can be identified, the class-conditional distributions remain misaligned, i.e., $P(G_f(X_s \mid Y_s)) \neq P(G_f(X_t \mid Y_t))$. As shown on the right, the fault classes are not clearly separated into distinct clusters. This hinders the ability of the algorithm to detect and diagnose faults in the target domain, resulting in an accuracy of 0.377, compared to 0.991 in the source domain.
The confusion matrices for the source and target datasets are presented in Figure 5.10. They reveal excellent

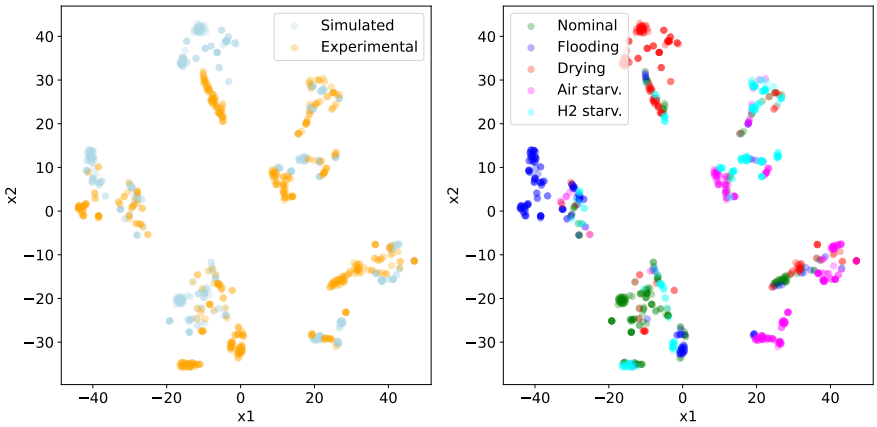
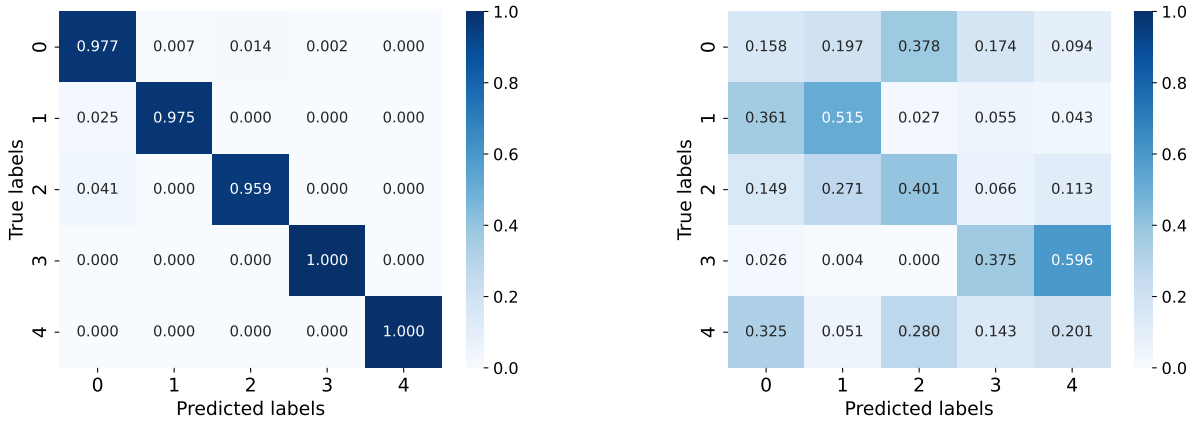


Figure 5.9: t-SNE visualization of the features extracted from source $G_f(X_s)$ and target $G_f(X_t)$ dataset with the unsupervised DANN, case ii). Left: source and target data are shown in light blue and orange, respectively. Right: features are colored according to the fault classes: nominal (green), flooding (blue), drying samples (red), air starvation (cyan), hydrogen starvation (magenta).

classification performance in the source domain, but persistent misclassifications in the target domain, indicating that the decision boundaries learned from the source data do not transfer effectively to the target domain.



(a) (b)

Figure 5.10: Confusion matrices in the case of unsupervised DANN ii) for the source (a) and target (b) domain.

Eventually, we consider the supervised DANN, case iii). In Figure 5.11 we visualize the feature space using t-SNE: the marginal distributions of the source and target domains as well as the class-conditional distributions are correctly aligned in the feature space, i.e., $P(G_f(X_s \mid Y_s)) \simeq P(G_f(X_t \mid Y_t))$. Indeed, five distinct clusters are clearly visible, each corresponding to a specific fault class. This indicates that the algorithm can successfully discriminate faults in the target domain, as supported by the classification accuracies of 0.991 (source) and 0.985 (target), and further confirmed by the confusion matrices in Figure 5.12, which show that sim-to-real transfer has been achieved and that the model can effectively classify faults in the target domain.

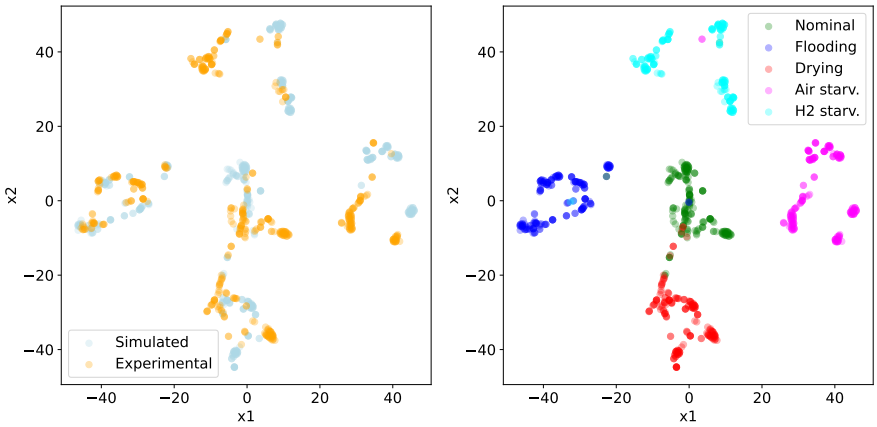
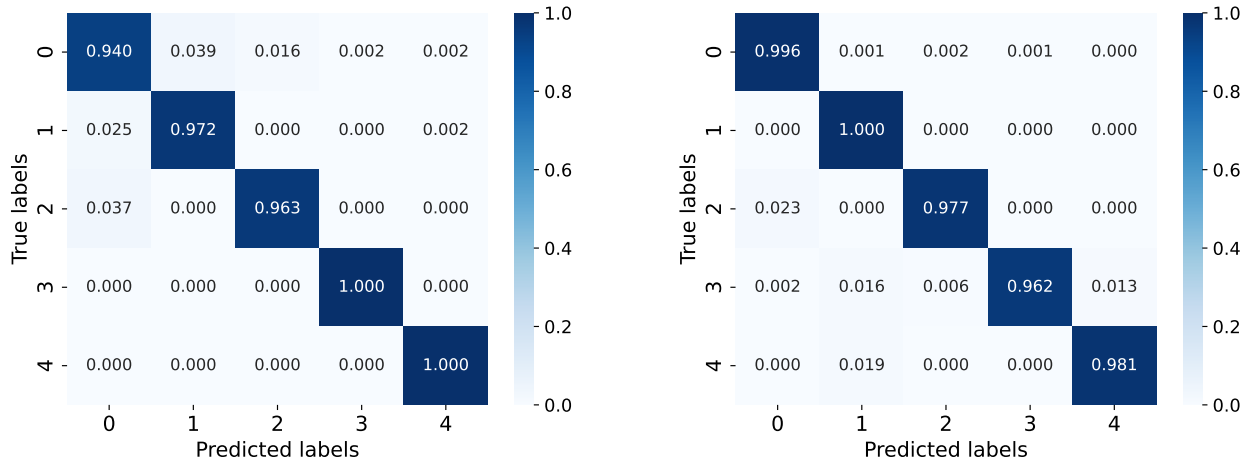


Figure 5.11: t-SNE visualization of the features extracted from source $G_f(X_s)$ and target $G_f(X_t)$ dataset with the supervised DANN, case iii). Left: source and target data are shown in light blue and orange, respectively. Right: features are colored according to the fault classes: nominal (green), flooding (blue), drying samples (red), air starvation (cyan), hydrogen starvation (magenta).



(a) (b)

Figure 5.12: Confusion matrices in the case of supervised DANN iii) for the source (a) and target (b) domain.

5.4.3 Details on the source and target domain datasets

This section reports on the characteristics of the datasets used to train the supervised DANN for PEMFC fault diagnosis: the source domain dataset is given by the simulated dataset described in section 5.4.1, and the target domain dataset is the experimental dataset presented in section 3.2.

For what concerns the experimental dataset, since it comprises more than a thousand hours of operation, we decided to divide it in different sub-datasets. Specifically, as illustrated in Figure 5.13, we generate three sub-datasets from the original experimental dataset: a sub-dataset comprising the first 100 hours of operation of the stack, representing the stack at beginning of life; a sub-dataset comprising the first 500 hours of operation of the stack, representing the stack with moderate aging level; a sub-dataset comprising 1000 hours of operation of the stack, representing the stack with advanced aging level. Note that the latter corresponds to the entire experimental dataset; however, for coherence, we will refer to it as sub-dataset 3. The three sub-datasets allow evaluating the performance of the fault diagnosis algorithm under increasing levels of aging and degradation of the monitoring signals. The same diagnosis variables are used in the source and target datasets, specifically: voltage, pressure drop at anode and cathode side, current load and outlet coolant temperature. Eventually, the number of samples for the source dataset and the three sub-datasets of the target domain is given in Table 5.1. The target domain sub-datasets are clearly imbalanced, as it has been highlighted in section 3.2.1. Evidently, the number of samples in sub-dataset 1 is the lowest, since it comprises a short range of operation time, while it is

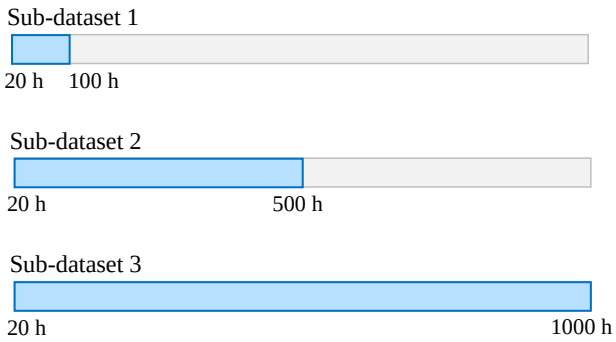


Figure 5.13: Schematic representation of the three sub-datasets considered as the target domain.

much larger in sub-dataset 3, as it comprises the entire range of operation time.

Dataset		Number of samples per fault class				
		Nominal	Flooding	Drying	Air starv.	H2 starv.
Source		46325	46325	46325	46325	46325
Target	sub-dataset 1	7289	6852	3191	2959	3101
	sub-dataset 2	40834	16484	10783	3909	3998
	sub-dataset 3	59152	24156	16684	6090	7084

Table 5.1: Number of samples per fault class, for the source dataset and the sub-datasets of the target domain.

5.4.4 Results for the supervised DANN

As demonstrated in the previous section, the supervised DANN is the only method capable of achieving effective transfer from the source to target domain, enabling the transfer of knowledge from a simulated dataset to the real PEMFC for fault diagnosis. Therefore, in this section we provide a quantitative evaluation of the performance of the proposed DANN-based method for PEMFC fault diagnosis. We compare the supervised DANN, trained with source and target domain datasets, to a pure data-driven algorithm, comprising a feature extractor G_f and a fault classifier G_y, trained solely with the target domain dataset. Three key aspects are addressed in the following:

- Flexibility of the methodology, demonstrated through a comparative analysis of different RNN architectures, each optimized using Optuna;
- Resilience under scarce data availability, validated by the method ability to achieve high performance even with limited training data, and compared against the pure data-driven counterpart;
- Robustness to aging, assessed through validation on the three sub-datasets of the target domain, confirming the stability and consistency of the results.

The following types of deep learning neural networks, commonly used for time series, are tested for the feature extractor: GRU, LSTM and BiLSTM. For a fair comparison of the performance, the number of layers and the number of neurons of the feature extractor network are optimized using Optuna, independently for the supervised DANN and the data-driven algorithms. The search space was given between 2 and 4 layers, 4 and 16 neurons, as a compromise between the number of data and model complexity. For consistency across experiments, the hyperparameters of the fault classifier and the domain discriminator are fixed to empirically validated values. These are selected based on preliminary exploratory trials, and maintained identical for all algorithms under consideration. The optimized architectures for the data-driven and DANN algorithms are given in Table 5.2.
To further ensure a fair comparison, we set 200 epochs as maximum number of iterations for model training, with a batch size of 128 and a learning rate of 0.001, for both the baseline data-driven algorithm and the proposed supervised DANN counterpart.

	GRU	DANN GRU	LSTM	DANN LSTM	BiLSTM	DANN BiLSTM
G_f	[16,16,16]	[16,16,16,16]	[16,16,16]	[16,16,16,16]	[8,8]	[16,16,16,16]
G_y - Dense	[64,5]	[64,5]	[64,5]	[64,5]	[64,5]	[64,5]
G_d - Dense	-	[32,1]	-	[32,1]	-	[32,1]
# params	5781	7990	7622	9734	4486	25158

Table 5.2: Optimize algorithms architectures. Optuna is used to optimize the feature extractor, while the fault classifier and domain discriminator are empirically optimized.

First result: resilience to scarce data availability

To show the effectiveness of the proposed supervised DANN method in addressing scenarios in which the available experimental data to build a fault diagnosis algorithm is limited, we first consider sub-dataset 1 as target domain, and we analyze the performance of the algorithms. An ablation study is conducted using progressively smaller fractions of the volume of the target dataset used for training and validating the algorithms: 0.8, 0.5, 0.1, 0.05, 0.01 and 0.005. To account for variability due to the stochastic training effects, each network was trained 10 times with different random seed. The mean accuracy over these runs is presented in Table 5.3. The overall fault classification performance has been considerably improved by the integration of physics-based simulation dataset through the supervised DANN method.

Target dataset used		GRU	DANN GRU	LSTM	DANN LSTM	BiLSTM	DANN BiLSTM
Fraction	# samples						
Sub-dataset 1: Beginning of life							
0.8	26090	**0.980**	0.963	**0.974**	0.965	0.970	**0.976**
0.5	16310	0.970	**0.971**	**0.978**	0.972	0.956	**0.967**
0.1	3300	0.778	**0.960**	0.860	**0.964**	0.877	**0.970**
0.05	1725	0.692	**0.949**	0.804	**0.955**	0.844	**0.968**
0.01	350	0.441	**0.930**	0.569	**0.895**	0.754	**0.918**
0.005	180	0.291	**0.901**	0.278	**0.852**	0.395	**0.905**

Table 5.3: Accuracy results for sub-dataset 1, averaged over 10 independent training runs with different random seeds. We compare the supervised DANN method with the data-driven algorithms, for GRU, LSTM and BiLSTM as feature extractor networks. For each experiment, the best between DANN and data-driven is in bold.

While the baseline data-driven algorithms have good accuracy when a large fraction of the target dataset is used, their performance drops when the target dataset size decreases. Indeed, when the fraction of the target train set used is 0.8, the data-driven algorithms demonstrate good accuracy, even higher than the DANN counterpart for GRU and LSTM networks. However, as this fraction decreases, the performance of the data-driven algorithms rapidly drops, while the DANN one remains high. In fact, in the supervised DANN algorithms, the simulated dataset supports the training, and even when 0.5 % of the original target train set is used the accuracy in fault diagnosis remains above 0.8. This advantage holds for all the recurrent networks tested, i.e., GRU, LSTM and BiLSTM, as it can be clearly observed in Figure 5.14. This result confirms that the DANN method can be a viable option independently of the algorithm used, which demonstrates a high flexibility.

Second result: validation under PEMFC aging

To evaluate the performance of the algorithms under different aging levels of the PEMFC stack, we extend the study to sub-dataset 2, which comprises the first 500 hours of operation of the stack, representing the stack with moderate aging level, and sub-dataset 3, which comprises 1000 hours of operation of the stack, representing the stack with advanced aging level. Table 5.4 and Table 5.5 show the results for the ablation study over the moderate and advanced aging target datasets.

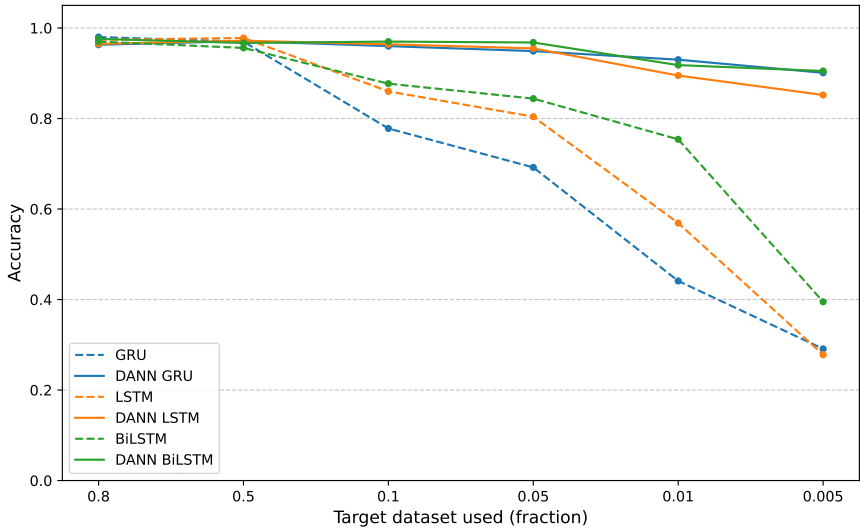


Figure 5.14: Accuracy vs target dataset (sub-dataset 1) fraction for data-driven and supervised DANN algorithms. Different feature extractor networks are tested: GRU, LSTM, BiLSTM.

	Target dataset used		**GRU**	**DANN GRU**	**LSTM**	**DANN LSTM**	**BiLSTM**	**DANN BiLSTM**
	Fraction	# samples						
Sub-dataset 2: Moderate aging								
	0.8	62655	**0.954**	0.908	**0.966**	0.946	0.946	**0.963**
	0.5	39715	**0.951**	0.928	**0.961**	0.925	0.935	**0.958**
	0.1	7935	**0.890**	0.881	0.912	**0.914**	0.897	**0.959**
	0.05	3935	0.765	**0.915**	0.850	**0.920**	0.862	**0.950**
	0.01	795	0.557	**0.875**	0.664	**0.856**	0.732	**0.912**
	0.005	415	0.441	**0.843**	0.424	**0.823**	0.650	**0.876**

Table 5.4: Accuracy results for sub-dataset 2, averaged over 10 independent training runs with different random seeds. We compare the supervised DANN method with the data-driven algorithms, for GRU, LSTM and BiLSTM as feature extractor networks. For each experiment, the best between DANN and data-driven is in bold.

	Target dataset used		**GRU**	**DANN GRU**	**LSTM**	**DANN LSTM**	**BiLSTM**	**DANN BiLSTM**
	Fraction	# samples						
Sub-dataset 3: Advanced aging								
	0.8	92135	**0.965**	0.920	**0.968**	0.948	0.952	**0.967**
	0.5	57750	**0.954**	0.916	**0.964**	0.934	0.946	**0.961**
	0.1	11600	**0.905**	0.902	**0.924**	0.920	0.897	**0.957**
	0.05	5810	0.877	**0.885**	0.881	**0.918**	0.854	**0.950**
	0.01	1200	0.629	**0.838**	0.658	**0.872**	0.800	**0.909**
	0.005	595	0.524	**0.851**	0.662	**0.838**	0.705	**0.881**

Table 5.5: Accuracy results for sub-dataset 3, averaged over 10 independent training runs with different random seeds. We compare the supervised DANN method with the data-driven algorithms, for GRU, LSTM and BiLSTM as feature extractor networks. For each experiment, the best between DANN and data-driven is in bold.

The results obtained for sub-dataset 1 are confirmed: the overall fault classification performance is considerably improved by the integration of physics-based, simulated dataset through the supervised DANN method, specifically if we consider conditions in which few samples are available in the target dataset. These results promisingly

suggest that the method is robust to the stack aging. Indeed, despite the variability introduced in the diagnosis signals by the aging phenomena, the DANN networks are still capable of extracting domain-invariant and discriminative features for fault diagnosis. However, we notice that the data-driven algorithms outperform the DANN counterpart when the fraction of target dataset available is larger. Moreover, data-driven algorithms have higher performances with aged data than with data at beginning of life. This is because the number of samples used for training the algorithms is higher (larger operation time): 62655 samples for sub-dataset 2, only 26090 samples for sub-dataset 1. Accordingly, to have an unbiased discussion, in Figure 5.15 we compare the performance on the 3 sub-datasets using the number of samples in the target train set, for the LSTM network.

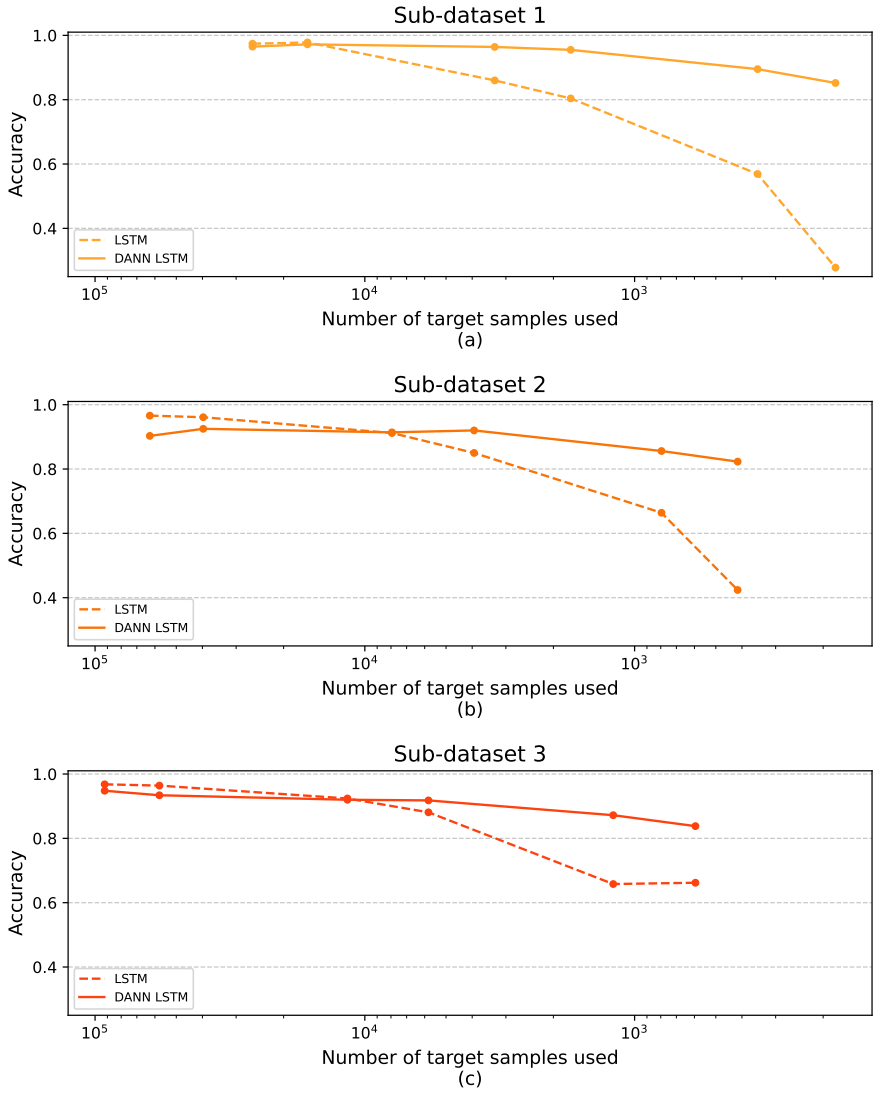


Figure 5.15: Accuracy of supervised DANN and data-driven algorithm vs number of samples for sub-dataset 1, sub-dataset 2, sub-dataset 3. The LSTM architecture is used.

First, we observe that the accuracy tends to decrease with increasing aging, which is due to the degrading effect of aging on the monitoring signals. Second, we observe that for the three sub-datasets, the supervised DANN LSTM has always an accuracy above 0.8, while the accuracy of the data-driven LSTM drops when the quantity of data is limited. Eventually, Figure 5.15 clearly suggests that the advantage of using the DANN method becomes prominent when the number of target samples falls below 10^4.

Analysis of the influence of algorithm hyperparameters

As we have seen in Table 5.2, the optimal number of hyperparameters differs according to the type of algorithm selected, either data-driven or supervised DANN. This difference arises from the higher complexity of the

learning objective, i.e., loss function, in the DANN network. Specifically, a DANN must simultaneously learn features that are discriminative for the fault diagnosis task, and features that are invariant across different domains. Furthermore, DANN algorithms are trained on a larger volume of data, as the source domain provides a massive amount of faulty samples to support training. This increased data availability not only influences the choice of hyperparameters, but also affects the stability of adversarial training, necessitating careful parameters selection and regularization strategies. Consequently, the number of hyperparameters has been adapted by the Optuna optimization procedure. Nonetheless, one could erroneously conclude that the increased performance of the DANN algorithms is primarily attributable to its larger network architecture, rather than to the prior knowledge embedded in the simulated dataset generated by the physics-based model. To rule out this possibility, we focus on the BiLSTM architecture, which exhibits the largest performance discrepancy between the two approaches. For this case, in Table 5.6 we report the results of the data-driven BiLSTM, of the DANN BiLSTM network configured with the same network architecture as the data-driven counterpart and of the DANN BiLSTM with the architecture optimized by Optuna, thereby isolating the effect of the prior knowledge from that of architectural differences. Sub-dataset 1 is used as target domain. These results demonstrate that, even when the number of hyperparameters in the BiLSTM based feature extractor is held approximately constant between the data-driven and the DANN network, the DANN methods achieve superior performance in low-data regimes. This finding holds despite not using a DANN architecture optimized on the specific loss function.

Target dataset used	**BiLSTM**	**DANN BiLSTM**	**DANN BiLSTM**
fraction	$\#G_f$ params: 2496	$\#G_f$ params: 2496	$\#G_f$ params: 23135
0.8	0.970	0.862	0.976
0.5	0.956	0.899	0.967
0.1	0.877	0.883	0.970
0.05	0.844	0.921	0.968
0.01	0.754	0.856	0.918
0.005	0.395	0.755	0.905

Table 5.6: Accuracy results, averaged over 10 independent training runs with different random seeds. We compare the data-driven and supervised DANN BiLSTM with the optimized architecture, and the supervised DANN BiLSTM with the same G_f as the data driven.

In other words, even in an intentionally non-favorable setting, where the DANN is constrained to use a comparable number of parameters as the data-driven baseline, it still exhibits a clear advantage. This implies that the performance gains of the DANN arise not merely from the algorithm complexity, but rather from its ability to leverage domain-invariant information derived by the prior knowledge embedded in the simulated, physics-based dataset.

5.5 Summary

In this work, we addressed the challenge of transferring knowledge from a simulated dataset to a real dataset in order to compensate for the limited availability of experimental data from PEMFC stacks. An adversarial neural network is successfully applied to integrate physics information in a fault diagnosis algorithm for PEMFC fault diagnosis. Specifically, a supervised DANN-based method has been proposed. Our method consistently outperforms a baseline purely data-driven approach by effectively exploiting the information extracted from a physics-based simulated dataset, used as source domain, to mitigate data scarcity in the experimental dataset, i.e., target domain. Importantly, the methodology proves robust to errors in the physics-based PEMFC model, which hinder the direct sim-to-real transfer, in the case no adaptation is implemented. Experimental results demonstrate that the supervised DANN achieves high diagnosis accuracy even when the number of available labeled samples from the real PEMFC stack is extremely limited, a scenario in which pure data-driven algorithms fail to generalize effectively. Despite the presence of substantial modeling errors in certain outputs of the physics-based model, the simulated dataset remains valuable because it encodes diverse fault trajectories and degradation patterns. The consequent shift between source and target domain distributions is effectively reduced by the supervised DANN.

Additionally, the 1000 hours faulty dataset collected on a PEMFC for automotive applications enabled the validation of the methodology on a wide range of operating conditions and different degradation levels, which are those experienced in real-world scenarios. In fact, high accuracy is achieved with aged data, proving the robustness of the proposed method to the aging of the stack.
The key contributions of this study can be summarized as follows:

1. Effective simulation-to-real transfer. The proposed domain-adversarial framework successfully mitigates distributional shifts between simulated and real data, leading to higher classification accuracy compared to non-adapted methods. Moreover, the mean percentage error in stack voltage prediction (5.97 %) is consistent with values reported in prior studies for similar fuel cell models [31, 61, 196], suggesting that the methodology could be generalized to other PEMFC designs.

2. High accuracy with data scarcity. The method achieves strong fault classification performance with very small amounts of labeled real data. This significantly reduces the practical burden of data collection. Notice that it is far more time-consuming and cost intensive to acquire an unlabeled but very large dataset on instrumented PEMFC. Instead, obtaining a small dataset of labeled samples from a real system is not extremely costly and could be reasonably achieved in several industrial applications.

3. Robustness to system aging. The proposed method demonstrates strong generalization across different aging levels of PEMFC degradation. Tests conducted on the different sub-datasets of the target domain show that the extracted features are not only domain-invariant and discriminative with respect to faults, but also invariant to aging effects. This suggests that the approach can maintain high diagnosis accuracy over the lifetime of the system, making it well-suited for real world application. This aging robustness could be further enhanced by incorporating test-time adaptation methods [236].

In summary, this study demonstrates that domain-adversarial learning, combined with physics-based simulation, offers a powerful and practical solution for fault diagnosis in PEMFC under scarce data availability, with strong potential for adaptation to other energy systems and industrial prognosis applications.

6

An Integrated Framework for Fault Diagnosis and SoH estimation

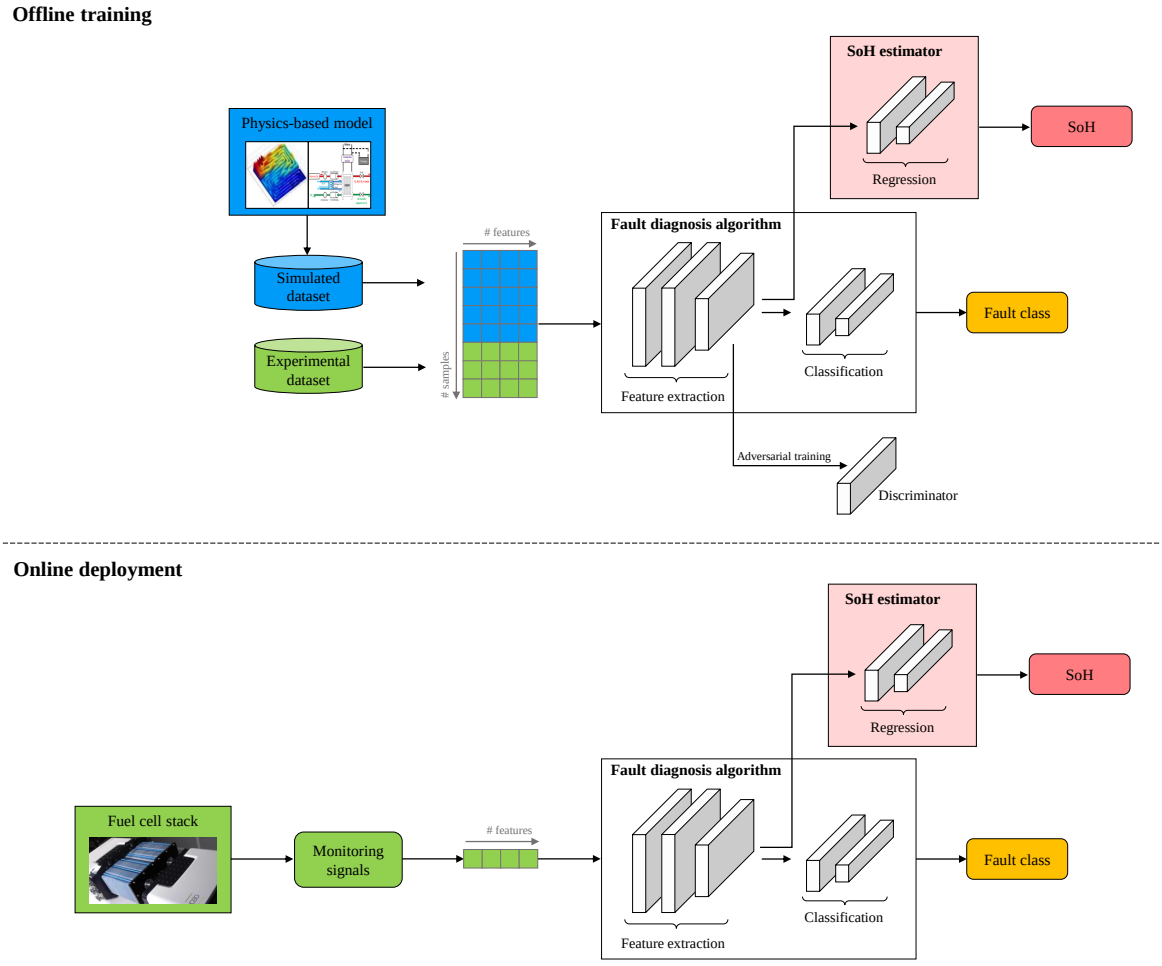


Figure 6.1: **Multitask domain adversarial neural network.** Starting from the methodology proposed in chapter 5, we develop a multitask network to: *i*) perform fault diagnosis and *ii*) estimate the state of health of the fuel cell stack. The physics-based model is used to generate a simulated dataset that supports the training of the deep learning algorithm; at testing time, only the monitoring signals measured on the fuel cell stack are used to serve the twofold objective.

Contents

6.1 Introduction 91
6.2 Background on multitask learning 91
6.3 SoH identification in PEMFC 92
6.3.1 SoH estimation: reference definition and faulty dataset annotation 93
6.4 The multitask methodology 95
6.4.1 Problem statement 95
6.4.2 The multitask DANN network 96
6.5 Implementation details and results 97
6.5.1 Datasets 97
6.5.2 Hyperparameter settings and multitask DANN training . . . 97
6.5.3 Evaluation metrics 98
6.5.4 Performance evaluation 99
6.6 Summary 101

6.1 Introduction

In this chapter we aim at extending the adversarial learning algorithm presented in chapter 5 using multitask learning, to simultaneously address fault diagnosis and state of health (SoH) estimation, as a step toward an integrated monitoring framework for PEMFC systems.
In many real-world monitoring applications, fault mechanisms and system degradation are not independent phenomena, but closely related. This affects also PEMFC systems, in which aging impacts not only the stack performance but also the way faults manifest and evolve, as observed in chapter 3. Fault diagnosis focuses on detecting and classifying abnormal operating conditions, SoH estimation provides a measure of the system degradation level. These two tasks have often been addressed independently, disregarding the inherent coupling between them. Indeed, the physical degradation of the PEMFC stack alters the system performance and, consequently, the signals used to monitor the system, e.g., the voltage.
Building upon this, in this chapter we aim at exploring a unified framework that integrates fault diagnosis and SoH estimation within a single deep learning architecture. Unlike prognosis, our objective is not to forecast the remaining useful life or predict future voltage loss, but rather to provide an accurate estimate of the current degradation state of the PEMFC. This information is valuable in itself, as it supports more informed decision-making for fault interpretation, maintenance planning, and potential corrective control strategies. To this end, we extend the supervised DANN approach introduced in the previous chapter by incorporating multitask learning. Multitask learning is a machine learning paradigm designed to solve multiple related tasks simultaneously. The multitask DANN that we present here is an innovative framework capable of jointly addressing fault diagnosis and SoH estimation, by enforcing the extraction of fault relevant as well as degradation relevant features from the monitoring signals. To the best of our knowledge, a similar integrated framework has never been proposed in the field.
The objectives of this chapter are threefold:

- Develop a SoH estimator, so as to provide a quantitative measure of the stack degradation level;
- Enforce more informative feature extraction for PEMFC fault diagnosis, by guiding the algorithm to learn features that are more sensitive to the degradation of the system;
- Move towards an integrated PHM framework, which facilitates algorithms deployment, enables coherent decision-making and more effective control strategy.

This integrated framework can offer several advantages. First, shared feature representations encourage capturing information relevant both for fault diagnosis and degradation, thus improving the robustness of learned features and reducing the fault misclassification. Second, multitask learning acts as a regularizer, since learning across two related objectives prevents the algorithm from overfitting to one specific task, particularly under limited training data. Finally, combining fault diagnosis and SoH estimation ensures consistent outputs across tasks, providing a more reliable PEMFC fault diagnosis across a wider range of operating conditions.

6.2 Background on multitask learning

Multitask learning [237] is an approach in machine learning in which a single algorithm is trained to perform several tasks simultaneously by exploiting shared information among them. The underlying intuition is that related tasks can reinforce each other, so that what is learned in one task can provide additional knowledge that improves learning in another. In contrast to domain generalization [238], which aims to adapt an algorithm to entirely new tasks or domains, multitask learning focuses on leveraging related information between known, related tasks. This is typically achieved through parameters sharing, which can serve as a form of regularization that promotes the extraction of features that are robust and more generalizable [209]. Indeed, a common strategy is to design architectures with a shared backbone network that captures a rich representation of the input data, followed by multiple task-specific heads, as illustrated in Figure 6.2. The shared layers facilitate knowledge transfer across tasks, while the task-specific layers preserve task specialization.

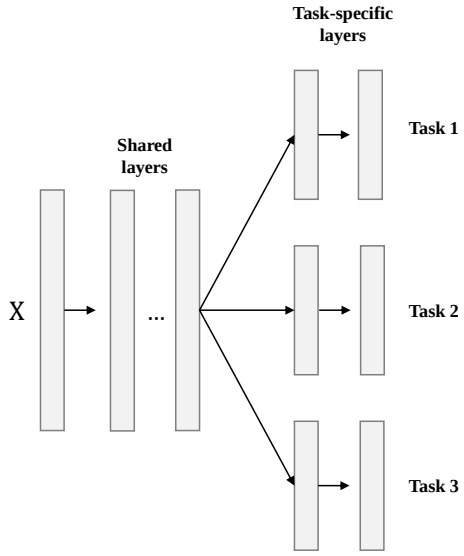


Figure 6.2: General architecture of multitask learning.

This architecture comes with practical challenges in the joint optimization of the tasks. The shared network must learn to satisfy multiple objectives simultaneously, and, consequently, the task specific loss functions are often combined into a single aggregated loss function [237]. However, while this joint training procedure can reduce the risk of overfitting and promote the learning of more generalizable features, it also introduces significantly greater complexity in the training process. Eventually, multitask learning can enhance the generalization of an algorithm on a specific task by embedding knowledge drawn from related ones, and it can improve the overall performance [239].

Tasks may be related in different ways. For example, one task may share features in the input space, or a task may serve as auxiliary to inform or constrain the learning on another task [212]. Multitask learning has been widely adopted in domains such as natural language processing and computer vision, where it has also been complemented with domain adaptation techniques to achieve robust feature extraction. For example, [240] suggested that combining multitask learning with adversarial learning can be beneficial for text classification. More recently, research in PHM has begun to adopt multitask learning, due to the natural interdependencies between tasks in this field. In this context, multitask learning has been successfully applied to jointly estimate the remaining useful life and the health index of turbofan engines [239]. Similarly, when integrated with domain adaptation, multitask learning has proven effective for SoH prediction of batteries [241]. These developments highlight the potentiality of multitask learning in PHM. Building on this, in this chapter we explore a framework that integrates multitask learning and adversarial learning to link fault diagnosis and SoH estimation for PEMFC, marking a novel contribution in the field.

6.3 SoH identification in PEMFC

The identification of the current degradation state of an industrial component, and the prediction of its future evolution, is essential for developing condition-based maintenance strategies. For this scope, in PHM a SoH indicator is defined such that a value of 1 corresponds to perfect health (no degradation), a value of 0 indicates a failed state (full degradation), and intermediate values represent degraded but still functional conditions [242]. In general, a valid SoH indicator is expected to exhibit monotonicity, trendability, and robustness to noise and operating variations [243].

As we have seen in section 2.2.2, defining a criterion for PEMFC performance level, and, subsequently, a suitable SoH indicator, is not straightforward. This challenge derives from the high sophistication of PEMFC systems [244]. Characterization techniques, including polarization curves, cyclic voltammetry and EIS can provide valuable insights on the health state of PEMFC, but their acquisition requires specific experimental

devices difficult to integrate online. Accordingly, the use of these techniques implies pausing operation, and it is consequently limited to maintenance phases [245–247]. For continuous monitoring, direct measurements are necessary. For this reason, the most used SoH indicators are stack voltage and stack power, as they can be easily measured online through sensors. However, while being effective under static current load, voltage and power cannot provide a clear degradation trend when dynamic mission profiles, e.g., FC-DLC, are considered. Indeed, the effects of load variation and degradation, as well as the effects of the faults, cumulate on the voltage signal making it hard to decouple the different contributions. To overcome this issue, in [248] the relative power loss rate (RPLR) has been proposed as a PEMFC health indicator that can be more robust to load variations. A different strategy would be to resort to indirect methods, which rely on physical or semiempirical models [249]. Despite being robust to dynamic load and varying operating conditions, modeling PEMFC degradation mechanisms, which combine loss of active surface, membrane degradation, and several other phenomena, is complex. Thus, exhaustive degradation models are difficult to obtain. In this scenario, machine learning approaches, often based on RNN neural networks, can provide a promising alternative [250–252]. By learning functional mappings between the measured signals and the degradation state, i.e., the SoH, these algorithms can efficiently capture the underlying degradation phenomena. Building on this perspective, and in line with our work on fault diagnosis, we adopt a deep learning approach to estimate PEMFC SoH from measured signals, which will be integrated in the multitask framework.

6.3.1 SoH estimation: reference definition and faulty dataset annotation

For developing a deep learning based SoH estimator using a supervised approach, it is necessary to provide the algorithm with training data, including a reliable reference for the SoH. This reference will serve for the algorithm to learn the mapping between input features and SoH during training.
We begin by analyzing in detail the degradation trend in the faulty dataset. Starting from the RPLR definition proposed in [248], we adopt a similar formulation of the concept based on the voltage, the relative voltage loss rate (RVLR). For the current density i_d, it is defined as:

$$RVLR_{i_d} = \frac{V(t)_{i_d}}{V(0)_{i_d}} \tag{6.1}$$

where $V(t)_{i_d}$ is the voltage at time t, while $V(0)_{i_d}$ is the reference voltage measured at beginning of life. In Figure 6.3 we provide the RVLR at $i_d = 1A/cm^2$ for the dataset presented in chapter 3.

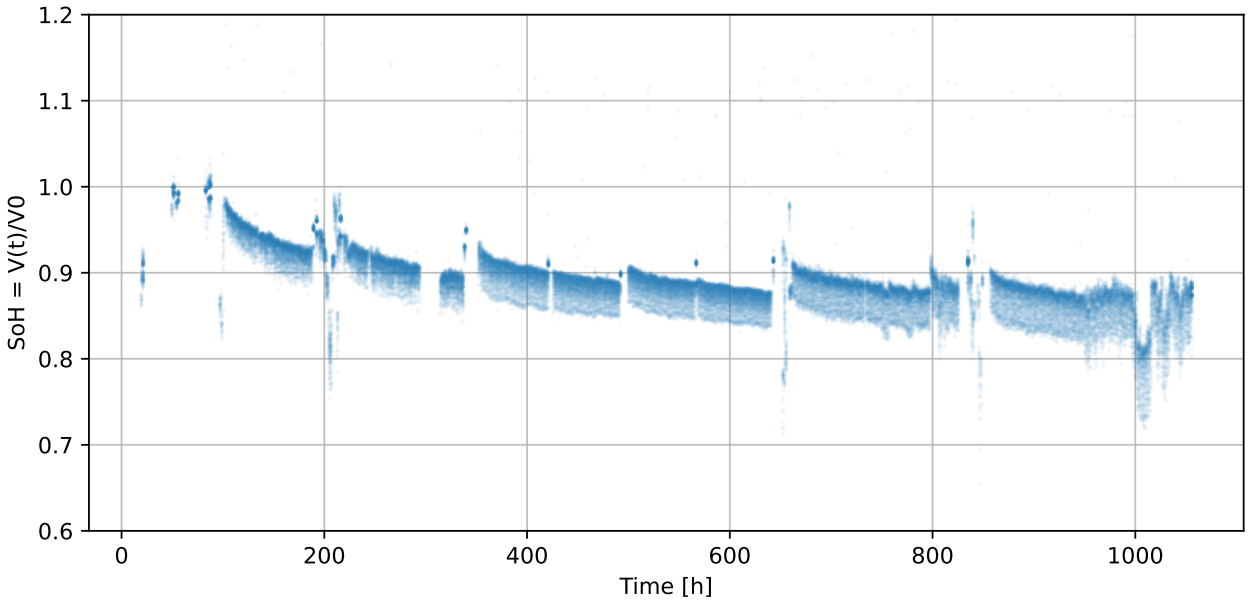


Figure 6.3: Relative voltage loss rate (RVLR) at $1A/cm^2$.

The global degradation trend can be clearly distinguished. However, a lot of oscillations can be observed locally, as well as the occurrence of regeneration phenomena related to the specific experimental campaign, hindering a clear monotonic trend. Such non-monotonic behavior can have a negative impact on the training and convergence of deep learning algorithms. A possible solution could be to adopt filtering techniques, such as moving average, to mitigate this effect. Instead, we adopt a different strategy for obtaining the reference SoH for the dataset: we compute the RVLR at $1A/cm^2$ using the polarization curves measured during the experimental campaign, in nominal condition, and we use it to obtain an empirical model of the SoH. The resulting RVLR values are reported in Table 6.1.

	Avg cell voltage [V]	RVLR [-]
21 h	0.550	0.921
83 h	0.594	0.995
86 h	**0.597**	**1**
188 h	0.563	0.943
192 h	**0.568**	**0.951**
216 h	0.570	0.955
339 h	**0.577**	**0.932**
421 h	0.539	0.903
492 h	0.536	0.897
566 h	0.536	0.898
643 h	**0.544**	**0.910**
835 h	0.543	0.909
1055 h	**0.526**	**0.881**

Table 6.1: RVLR at 1 A/cm^2 for the polarization curves in nominal condition. The reference value is highlighted in blue, and in bold the polarization curves with monotonic trend.

Note that the polarization curve measured at 86 hours is considered as the reference, i.e., $V(0)_{i_d}$. This choice is motivated by the observation that fuel cell performance improved during the initial hours of operation. Although the exact causes have not been fully established, this effect is likely due to an incomplete conditioning phase. Moreover, to ensure that the SoH reference exhibits a monotonic trend, only the polarization curves displaying consistent decreasing behavior are considered. Eventually, an empirical SoH model is derived by applying logarithmic regression, which provides a representation of the underlying degradation trend. Before 86 hours, the system is assumed to be in a healthy state. Hence, the resulting empirical model is given by:

$$SoH(t) = \begin{cases} 1, & \text{if } t \leq 86\,\text{h} \\ -0.045 \cdot \log(t \cdot 3600) + 1.565, & \text{if } t > 86\,\text{h} \end{cases} \tag{6.2}$$

where t is the operating time in seconds. This empirical SoH model is illustrated in Figure 6.4, and the data points used for the regression are highlighted. This provides the SoH reference that will eventually be used for training the multitask network.

We emphasize that the choice of SoH reference adopted here is not restrictive, and the multitask framework that we propose remains applicable and valid even if alternative SoH formulations are used for training the algorithm. During our study, an alternative model for the SoH, based on a simplified physical descriptor of the loss of active surface area, has been investigated. Nonetheless, this choice only affects the annotation of the dataset, while the overall methodology and outcome remain unchanged. Therefore, for clarity and consistency, in this thesis we decided to take as SoH reference the empirical model derived using the RVLR definition.

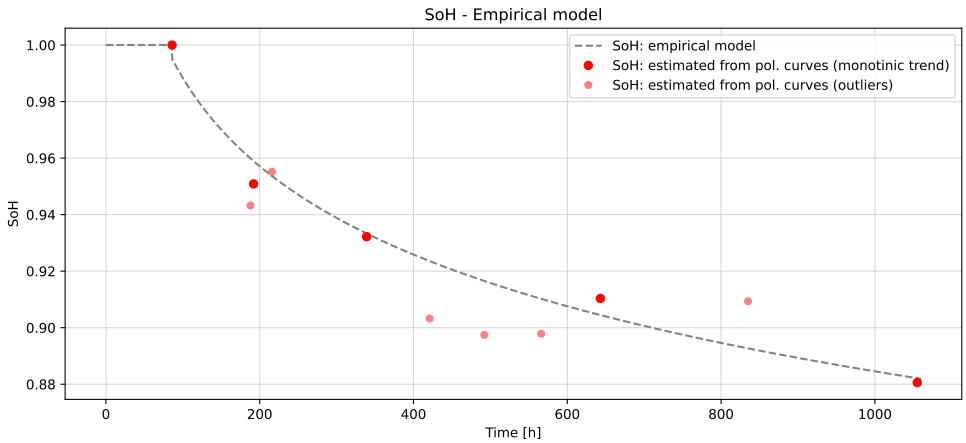


Figure 6.4: Derived empirical SoH model for the faulty dataset.

6.4 The multitask methodology

The objective of this chapter is to extend the supervised DANN method for PEMFC fault diagnosis presented in chapter 5 by integrating the estimation of the SoH as a complementary task. So far, we focused on the classification of faults by leveraging a simulated dataset and an adversarial learning algorithm. Here, we formulate the problem in a multitask framework to obtain a comprehensive algorithm which includes a SoH estimator.

6.4.1 Problem statement

Let's consider the input data $X = \{\underline{x}_i\}_{i=1}^{n}$, where $\underline{x}_i \in \mathbb{R}^{m \times L}$ is the input multivariate time series constituted by m signals of L consecutive time steps, and n is the number of samples. The input data X are associated with:

- a fault label $Y = \{y_i\}_{i=1}^{n}$, where $y_i \in \{0, \ldots, K-1\}$, K being the number of fault classes;
- a SoH indicator $H = \{h\}_{i=1}^{n}$, where $h_i \in [0, 1]$.

Then, the aim of multitask learning is to simultaneously achieve the two following tasks:

- a classification task for fault diagnosis (task 1), by learning the mapping function $f_y : X \rightarrow Y$;
- a regression task for SoH estimation (task 2), by learning the mapping function $f_h : X \rightarrow H$.

To leverage related information, in multitask learning a shared feature representation is enforced so that the input-output functions are partially shared between the tasks, i.e., $f_y(\cdot) = g_y(f(\cdot))$ and $f_h(\cdot) = g_h(f(\cdot))$, f being the shared feature mapping, g_y the task-specific mapping for fault classification and g_h the task-specific mapping for SoH estimation.
Eventually, the source dataset becomes:

$$D_s = \{X_s, Y_s, H_s\} = \{(\underline{x}_{s,i_s}, y_{s,i_s}, h_{s,i_s})\}_{i_s=1}^{n_s} \tag{6.3}$$

where n_s is the total number of samples in the source domain, and the target dataset becomes:

$$D_t = \{X_t, Y_t, H_t\} = \{(\underline{x}_{t,i_t}, y_{t,i_t}, h_{t,i_t})\}_{i_t=1}^{n_t} \tag{6.4}$$

where n_t is the total number of samples in the target domain.
Therefore, the objective is to learn a feature space that is representative for both tasks, i.e., the features have to be discriminative for the faults and relevant for capturing the degradation of the system, while preserving the correct alignment between the fault conditional distributions of source and target domain.

6.4.2 The multitask DANN network

The proposed multitask DANN network is presented in Figure 6.5. It consists of a shared feature extractor G_f, a discriminator G_d, and two task-specific heads: a fault classifier G_y and a SoH estimator G_h.

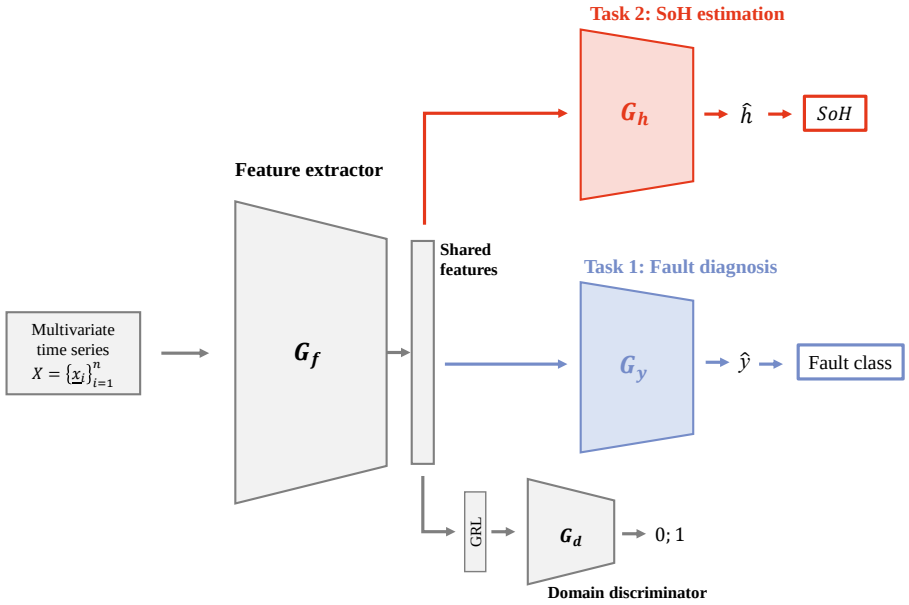


Figure 6.5: Multitask supervised DANN.

- The feature extractor G_f, with parameters $\underline{\theta}_f$, is the shared network with the purpose of finding a latent feature space in which the representation $G_f(\underline{x})$ of the input data is simultaneously discriminative for fault classification and SoH estimation, while being invariant between the source and target domain. This encourages knowledge sharing and it prevents overfitting to task-specific patterns that could hinder the robustness of the algorithm.

- The fault classifier G_y, with parameters $\underline{\theta}_y$, consists of fully connected layers which process the extracted features $G_f(\underline{x})$. The objective is to identify the optimal decision boundaries, which integrate the knowledge about system degradation and faults, and to provide in output the predicted fault class $\hat{y} \in \{0, \dots, K-1\}$. The loss function $\mathcal{L}_y(\underline{\theta}_f, \underline{\theta}_y)$ used for training G_y is the sparse categorical cross-entropy.

- The SoH estimator G_h, with parameters $\underline{\theta}_h$, receives in input the features $G_f(\underline{x})$, and, through a series of fully connected layers, processes them to emphasize the information relevant for the degradation state of the system. The output layer of G_h provides the predicted SoH $\hat{h} \in [0, 1]$. The loss function $\mathcal{L}_h(\underline{\theta}_f, \underline{\theta}_h)$ used for training G_h is the mean absolute error.

- Eventually, the domain discriminator G_d, consisting of fully connected layers, is trained to distinguish the domain of input data i.e., $\hat{z} = 0$ for source domain and $\hat{z} = 1$ for target domain, while being fooled by the feature extractor G_f. The loss function $\mathcal{L}_d(\underline{\theta}_f, \underline{\theta}_d)$ used for training G_d is the binary cross-entropy. This component of the network remains of fundamental importance to integrate the physics-based simulated dataset in the training of the algorithm.

Hence, the multitask DANN algorithm serves three objectives: align source and target domain, correctly discriminate fault patterns for fault diagnosis (task 1), and capture the degradation trend in the monitoring signals for accurate estimation of the stack SoH (task 2). The training of the algorithm is achieved by minimizing the loss on fault classification $\mathcal{L}_y$ and maximizing the discriminator loss $\mathcal{L}_d$, in order to enforce the extraction of domain invariant features, as it is done in equation (5.5) for the supervised DANN, and by minimizing the loss on

the estimation of the SoH $\mathcal{L}_h$. To simultaneously train the network with these requirements, the task specific loss functions, i.e., $\mathcal{L}_y$, $\mathcal{L}_d$ and $\mathcal{L}_h$, are grouped into a single multitask objective $\mathcal{L}(\underline{\theta}_f, \underline{\theta}_y, \underline{\theta}_h, \underline{\theta}_d)$:

$$\mathcal{L}(\underline{\theta}_f, \underline{\theta}_y, \underline{\theta}_h, \underline{\theta}_d) = \mathcal{L}_y(\underline{\theta}_f, \underline{\theta}_y) - \lambda_d \mathcal{L}_d(\underline{\theta}_f, \underline{\theta}_d) + \lambda_h \mathcal{L}_h(\underline{\theta}_f, \underline{\theta}_h) = \mathcal{L}_{DANN,s}(\underline{\theta}_f, \underline{\theta}_y, \underline{\theta}_d) + \lambda_h \mathcal{L}_h(\underline{\theta}_f, \underline{\theta}_h) \tag{6.5}$$

where the first term $\mathcal{L}_{DANN,s}(\underline{\theta}_f, \underline{\theta}_y, \underline{\theta}_d)$ is the loss function described in equation (5.5) for the supervised DANN, and the second term is introduced to force the network to learn degradation patterns in the input data, with λ_h being a hyperparameter used to weight $\mathcal{L}_h$. The different terms are competing during the training of the feature extractor. Hence, to mitigate potential convergence issues due to gradient conflicts or uneven tasks complexity, it is essential to adapt the learning rate for each task.
In this way, the feature extractor G_f learns a shared feature representation that is both task-relevant and domain-invariant, enabling accurate fault classification and SoH estimation across domains. By jointly learning both tasks in an adversarial multitask framework, the shared feature extractor is encouraged to capture more robust and generalizable patterns than in the single-task setting, adequately modifying the decision boundaries of the fault classifier with respect to the supervised DANN presented in chapter 5.

6.5 Implementation details and results

6.5.1 Datasets

The algorithm is developed using the experimental dataset described in section 3.2, and the simulated dataset presented in section 5.4.1. The experimental dataset, which consists of measurements of PEMFC signals collected over 1055 hours of stack operation, is representative of the target domain, while the simulated dataset constitutes the source domain that supports the training. The procedure illustrated in section 3.4 is used to preprocess the two datasets, using a sliding window of 20 s. The total number of samples available is reported in Table 5.1. The input X of the multitask network are the following monitoring signals: stack voltage, current, anode pressure drop, cathode pressure drop and outlet coolant temperature. For task 1, the network is trained using the true fault labels y; for task 2, the SoH annotations, estimated according to the empirical model in equation (6.2), are used.

6.5.2 Hyperparameter settings and multitask DANN training

In this study, the proposed multitask DANN network is evaluated against a pure data-driven and a supervised DANN algorithm, which represent benchmarks for comparing fault classification performances. For these two benchmarks, we selected the best performing architectures, i.e., BiLSTM networks, coherently with the results obtained in chapter 5 using the Optuna hyperparameter optimization framework. In contrast, the multitask DANN network was not optimized through this systematic procedure, owing to the computational burden of its training process. Instead, the optimized architecture of the supervised DANN network has been preserved also for the multitask algorithm, and the hyperparameters of the additional component, i.e., SoH estimator, were selected through a trial-and-error procedure to ensure sufficiently low training error and satisfactory accuracy. The algorithms architectures are detailed in Table 6.2.

	Pure data-driven	**Supervised DANN**	**Multitask DANN**
G_f - BiLSTM	[8,8]	[16,16,16,16]	[16,16,16,16]
G_y - Dense	[64,5]	[64,5]	[64,5]
G_d - Dense	-	[32,1]	[32,1]
G_h - Dense	-	-	[32,1]
# params	4486	25158	26247

Table 6.2: Architectures of the two comparative algorithms and of the multitask DANN network.

The multitask DANN consists of a feature extractor G_f composed of four BiLSTM layers with 16 cells each. Network weights are initialized using an orthogonal initializer to preserve variance across layers and improve

training stability. The fault classifier G_y includes one dense layer with 64 neurons, followed by a dense layer with 5 neurons and softmax activation. The SoH estimator G_h comprises one dense layer with 32 neurons, followed by a dense layer with sigmoid activation. The domain discriminator G_d consists of one dense layer with 32 neurons, followed by a dense layer with sigmoid activation. The dense layers have ReLU activation, and the weights are initialized with He initialization [253]. A dropout rate of 0.2 is applied throughout the network to enhance regularization and limit overfitting. Error backpropagation is performed using the Adam optimizer. To account for differing learning dynamics across tasks in the multitask framework and to promote a balanced optimization, the learning rates are separately assigned to each network: 0.001 for task 1 (G_y) and 0.0001 for task 2 (G_h). The hyperparameter λ_h is set to 1 after exploring the values [0.1, 1, 10], resulting in a balanced contribution of task 2 during training. For the multitask DANN, and the two comparative algorithms, the batch size is set to 256, and the training is conducted for a maximum of 2000 epochs, enabling the networks to accurately capture the complexity of both tasks. All experiments are conducted in Python using the TensorFlow library. The training history for the multitask DANN is provided in Figure 6.6, where the three loss functions, $\mathcal{L}_y$, $\mathcal{L}_h$ and $\mathcal{L}_d$, are illustrated.

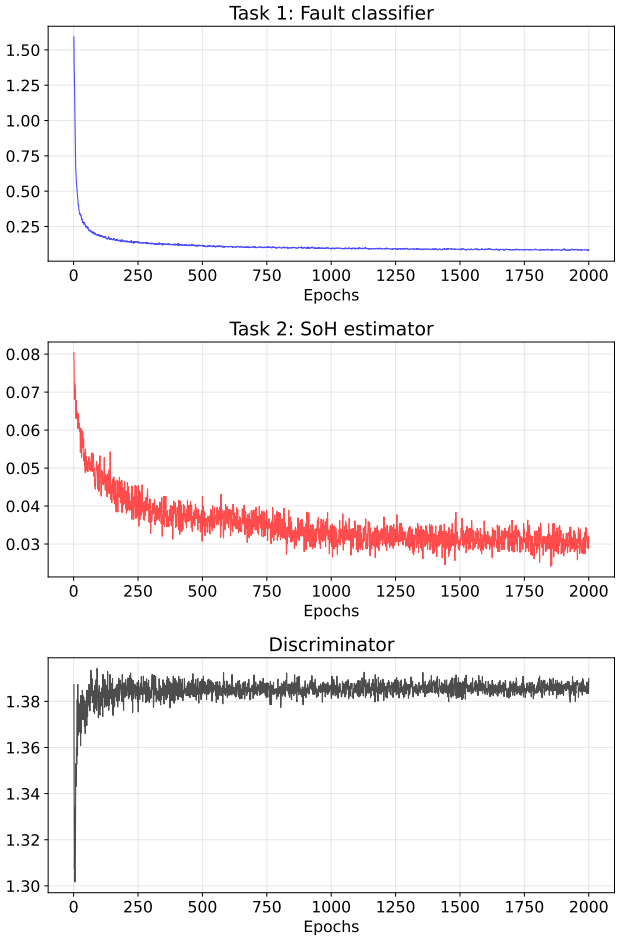


Figure 6.6: Training history of the multitask DANN. From top to bottom: $\mathcal{L}_y$ (task 1), $\mathcal{L}_h$ (task 2), $\mathcal{L}_d$ (discriminator).

6.5.3 Evaluation metrics

For each task, an adequate metric is defined. For the SoH estimation, the MAE is used to compare the predicted with the reference SoH. For fault classification, accuracy is used. Additionally, the entropy measure on the softmax output of the fault classifier is computed, in order to provide a more robust evaluation of the classification performance. Since the softmax function provides probability distributions over the fault classes, p_i, $i =, \ldots, K$, the entropy can measure the confusion over these distributions. The normalized entropy, used as entropy measure, is defined as:

$$H_n = \frac{-\sum_{i=1}^{K} p_i \log(p_i)}{\log K} \tag{6.6}$$

Then, the normalized entropy over the fault classes measures the confidence on the algorithm estimates: an entropy measure close to 0 indicates high confidence, an entropy measure close to 1 indicates large confusion [254]. Hence, it allows assessing of the confidence and reliability in the decisions of the algorithm.

6.5.4 Performance evaluation

Ablation study

To evaluate the validity of the proposed multitask DANN method, we consider the entire experimental dataset, comprising 1055 hours of operation, as target dataset. This allows the multitask network to be tested across a wide range of degradation levels of the PEMFC stack. After dividing the experimental dataset into training, validation, and test sets, an ablation study is performed by progressively reducing the fraction of the target dataset used for training: 0.8, 0.5, 0.1, and 0.05. In this way we can assess the robustness of the method under limited data availability. Table 6.3 provides the results of the proposed multitask DANN for the fault classification and SoH estimation, and of the two algorithms for comparison. Moreover, Figure 6.7 presents a bar plot comparing the fault diagnosis accuracy of the three methods, allowing for a clear visual evaluation of their relative performance.

Target dataset used		**Pure data-driven**		**Supervised DANN**		**Multitask DANN**		
Fraction	# samples	Accuracy	H_n	Accuracy	H_n	Accuracy	H_n	MAE SoH
0.8	92135	0.968	0.380	0.979	0.147	0.979	0.157	0.031
0.5	57750	0.963	0.373	0.979	0.131	0.971	0.158	0.025
0.1	11600	0.954	0.386	0.970	0.153	0.974	0.148	0.028
0.05	5810	0.903	0.476	0.965	0.147	0.963	0.145	0.023

Table 6.3: Comparison of fault diagnosis performance in terms of accuracy and prediction confidence for the data-driven, the supervised DANN, and the proposed multitask DANN, for which the MAE for SoH estimation is also reported. A BiLSTM-based feature extractor is used. H_n is the normalized entropy.

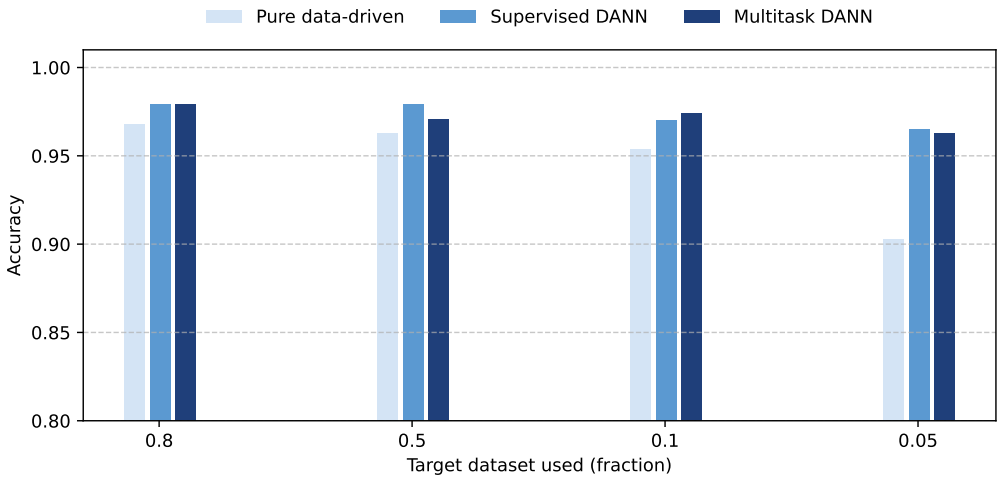


Figure 6.7: Comparison of the accuracy for the data-driven, the supervised DANN, and the proposed multitask DANN algorithms. The fraction of the target dataset used for training is progressively reduced: 0.8, 0.5, 0.1, 0.05.

The fault diagnosis performances confirm that the multitask DANN can achieve accurate classification of the faults, with accuracy scores coherent with those of the supervised DANN. Consistent with the findings for the supervised DANN, the ablation study, which progressively reduces the fraction of experimental data used for training, confirms the importance of the simulated dataset and the domain adaptation setting in maintaining high diagnosis performance when the experimental data are limited. As illustrated in Figure 6.7, when only 5% of

the experimental data is available for training, both the supervised DANN and the multitask DANN retain high fault classification performance, whereas the pure data-driven model exhibits a loss of accuracy. Furthermore, the entropy measure for the supervised DANN (0.147) and multitask DANN (0.145) is substantially lower than that of the data-driven algorithm (0.476) under the 5% dataset condition. This indicates that the predictions of the DANN-based methods are more reliable and that their decision boundaries for fault diagnosis better discriminate between fault classes, enhancing intra-class coherence and inter-class separability. In addition to fault diagnosis, the multitask framework provides a reliable estimate of the PEMFC degradation directly from the monitoring signals, without requiring a dedicated preprocessing procedure. Indeed, the highest observed error in SoH estimation is 0.031, demonstrating that the model can effectively capture the stack degradation level. Since the dataset spans across the entire experimental campaign, we can conclude that the multitask DANN can successfully disentangle the effect of aging from that induced by the faults in the measured signals. The task specific heads, one for fault classification and one for SoH estimation, enable the network to jointly learn features representations that serve the two objectives.

Validation across different aging levels

To strengthen the validation of the methodology and support more general conclusions, we propose a series of experiments on different dataset partitions. From the experimental dataset, a frame [20 h, t] is selected, t being the operation time in hours. This is divided in train-val-test sets, as detailed in section section 3.4, ensuring a rigorous procedure and avoiding data leakage. The ratio 0.7-0.1-0.2 is fixed. No ablation study is conducted in this setting. Instead, five experiments are performed, progressively extending the portion of the dataset used for training: t = 100, 250, 400, 500, 650 hours. For each dataset partition, three distinct test configurations are considered:

- Test 1: evaluation on the time interval [20 h, t], i.e., within the same degradation range as the training data. Here, the 20% test set defined by the splitting procedure is used to evaluate algorithms' performance;
- Test 2: evaluation on the subsequent 300 h, i.e., on the time interval [t, t+300 h], to assess the algorithms' ability to generalize to degradation states not included in training;
- Test 3: evaluation on the whole remaining dataset, i.e., [t, 1055 h], providing a more challenging setting that covers the full operation of the stack.

We emphasize that tests 2 and 3 enable the assessment of the methodology under aging conditions not encountered during training, thus providing a more rigorous testing of the generalization capability of the algorithms. The testing protocol is illustrated in Figure 6.8.

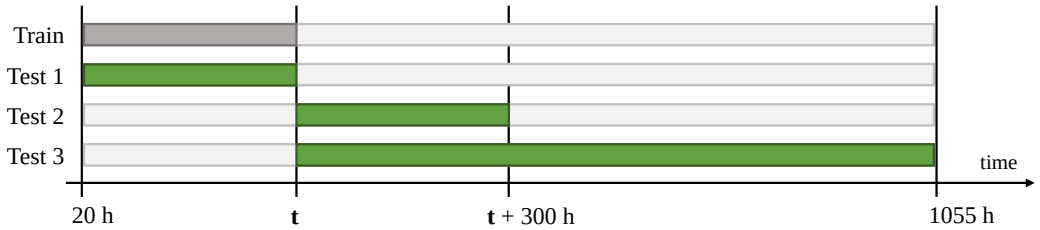


Figure 6.8: Illustration of the dataset partitioning. In dark gray, the interval of operating time used for training and validation; in green, the intervals used for evaluating algorithms' performance in test 1, test 2, and test 3.

The results, reported in Table 6.4, are consistent with those given in Table 6.3. The proposed multitask DANN network provides accurate SoH estimates. The MAE for the SoH estimation remains below 0.05, indicating that the health of the fuel cell can be predicted with an error of less than 5%. Although a clear trend is difficult to identify, the error appears to increase in tests 2 and 3. Overall, the multitask network demonstrates a good capability to estimate the current health state of the fuel cell stack. Simultaneously, as it is for the supervised DANN, it yields significantly more reliable fault classification performance compared to the pure data-driven model, as indicated by the lower entropy measure. However, although the accuracy is satisfactory across all tests,

no clear trend emerges across the performance of the five experiments. Specifically, in Figure 6.9 we present the bar plots for the fault classification accuracy in the experiments with t = 100 h (left) and t = 250 h (right). In the first case (t = 100 h), for tests 2 and 3, we see a clear benefit of the DANN based methods compared to the data-driven one, suggesting a greater robustness to aging and unseen conditions. On the contrary, in the second case (t = 250 h), the trend is less pronounced. Nonetheless, despite the complex task, good accuracy scores are achieved, above 0.8. This effect may be explained by the fact that, when the training dataset starts covering more advanced aging of the PEMFC stack, even pure data-driven deep learning algorithms become capable of capturing the underlying degradation trend, thereby reducing the advantage of the multitask DANN method.

Experiment		**Pure data-driven**		**Supervised DANN**		**Multitask DANN**		
t [h]	Test:	Accuracy	H_n	Accuracy	H_n	Accuracy	H_n	MAE SOH
100	# 1	0.978	0.334	0.982	0.140	0.979	0.142	0.023
	# 2	0.530	0.472	0.665	0.175	0.674	0.176	0.056
	# 3	0.493	0.475	0.639	0.172	0.644	0.178	0.051
250	# 1	0.969	0.376	0.974	0.140	0.976	0.140	0.029
	# 2	0.806	0.437	0.821	0.170	0.846	0.155	0.046
	# 3	0.844	0.420	0.817	0.168	0.833	0.159	0.041
400	# 1	0.961	0.406	0.976	0.136	0.971	0.140	0.040
	# 2	0.927	0.403	0.956	0.138	0.942	0.157	0.021
	# 3	0.916	0.409	0.904	0.142	0.912	0.153	0.023
500	# 1	0.959	0.375	0.970	0.131	0.980	0.144	0.020
	# 2	0.939	0.419	0.981	0.135	0.969	0.146	0.019
	# 3	0.905	0.409	0.914	0.136	0.916	0.151	0.032
650	# 1	0.972	0.400	0.976	0.151	0.983	0.144	0.051
	# 2	0.903	0.418	0.891	0.161	0.914	0.151	0.025
	# 3	0.903	0.418	0.891	0.161	0.914	0.151	0.025

Table 6.4: Fault classification accuracy and SoH estimation performance for different dataset partitions, for the experimental testing protocol illustrated. Comparison between data-driven, supervised DANN and multitask DANN. H_n is the normalized entropy.

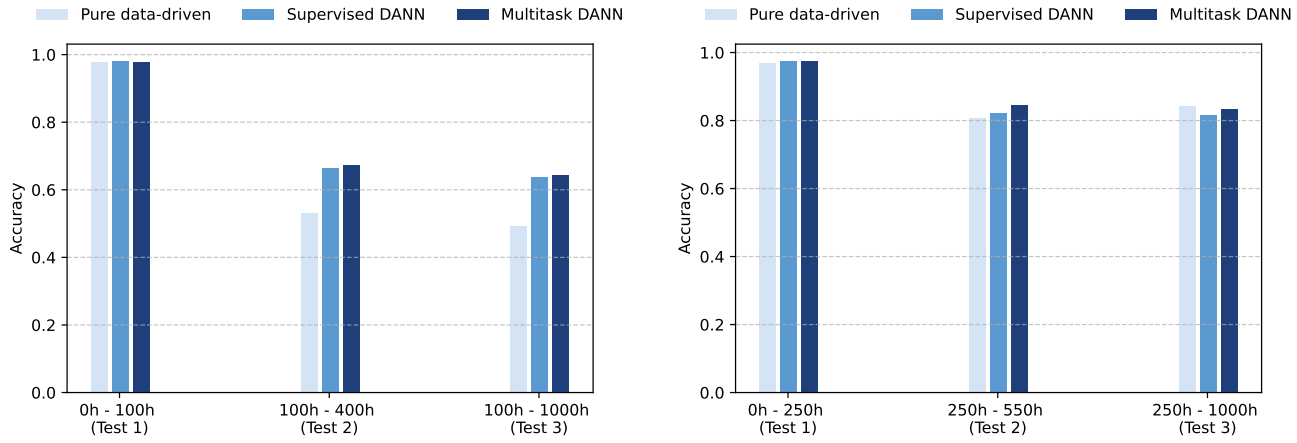


Figure 6.9: Fault classification accuracy for t = 100 h (left) and t = 250 h (right), for tests illustrated in Figure 6.8.

6.6 Summary

In this chapter we investigated the formulation of the supervised DANN method within a multitask framework, in which fault diagnosis and SoH estimation are jointly addressed to exploit shared information between the task. The novelty of the proposed method is that by jointly training the feature extractor with two objectives, it is possible to find a new feature space relevant for capturing fault patters as well as degradation trends. Although the

availability of a single dataset limits the possibility of predicting degradation trajectories or performing additional types of validation, the series of experiments conducted allow us to draw the following general conclusions:

- The multitask DANN can provide consistent and accurate fault classification across different training partitions, demonstrating robustness and high confidence even under limited or partially representative data;
- SoH estimation can be simultaneously achieved directly from monitoring signals, exhibiting low estimation error throughout the operational range, without experiencing negative transfer;
- Integrating a physics-based simulated dataset by means of adversarial learning, enables the model to learn physically meaningful information, which, in turn, can boost the generalization and mitigates the challenges of scarce data availability.

Nonetheless, the multitask framework enables a better understanding of the system, opening to new opportunities for integrated PHM strategies and more informed decision for the control and maintenance strategies. Moreover, as a next step, we plan to explore different approaches to integrate SoH information into the fault diagnosis algorithm, such as including SoH as an input feature or conditioning the classification boundaries on the estimated SoH. These approaches are still under investigation, but they represent promising directions to achieve a more robust and informed fault diagnosis for PEMFC systems.

7

Conclusions

Contents

7.1 Scientific contributions 104

7.2 Key findings 105

7.3 Limitations and future perspectives 105

7.1 Scientific contributions

This PhD thesis deals with the use of physics-informed deep learning for advancing fuel cell fault diagnosis and addressing major constraints in the field. The theoretical background of fault diagnosis, and the current state of the art for applications to fuel cell systems, have been explored in chapter 2. The presented work is grounded on two main building blocks, which have been introduced in chapter 3: an experimental faulty dataset and a physics-based fuel cell model. First, the experimental dataset, which embeds 1000 hours of fuel cell operation in optimal and faulty conditions, has been analyzed. The remarked imbalance of the faults all along the dataset is distinctive of most faulty datasets across industrial systems, and the shift of the statistical distributions of the signals underlies the inherent degradation process characterizing fuel cell systems. Therefore, it is a valid descriptor of typical scenarios. Handling temporal faulty datasets, in which faults are not evenly distributed, and for which it is fundamental to retain time characteristics, requires careful considerations. Hence, a systematic processing procedure for temporal faulty datasets is also provided. Then, the physics-based fuel cell model, which relies on fundamental laws describing the complex physical and chemical processes inside a fuel cell, has been described, elucidating the details of model calibration. The model accuracy on the reported observable estimates, pertain to the specific fuel cell considered. Nonetheless, the ability in describing the physics of failures, modeling major faults, and representing variations in the internal states of the fuel cells remain valid regardless of the fuel cell technology considered.

An integration of the physics-based model within the fault diagnosis framework is first proposed in chapter 4. Embedding the model online, enables the access to the estimates of internal states of the fuel cell, reducing the need for embedding dedicated sensors. Membrane resistance, water content, current density distribution are among the relevant variables derived from physical laws that, combined with the signals measured using traditional sensors, guide the deep learning algorithm towards a correct discrimination of the faults. This physics-guided method has been tested and compared to a pure data-driven framework, using traditional machine learning algorithms and a LSTM network. The study of the accuracy score is particularly relevant: it demonstrated that the use of deep learning yields improved performance, and confirmed that additional physical knowledge improves fault discrimination in the fuel cell. The increased robustness and generalization ability of the physics-guided method result in an accuracy increase of 55.9 % when tested on a fuel cell that experienced degradation.

In chapter 5, with the objective of relaxing the requirements dictated by the physics-guided method for online deployment of the fuel cell model, we explored a different approach for integrating the physics-based model into deep learning algorithms. The physics-based model has been used offline, to obtain a simulated dataset that can assist the training of deep learning algorithms. However, directly transferring knowledge from the physics-based model is infeasible, and, as demonstrated in the study, it resulted in inadequate accuracy of the fault diagnosis algorithms. Accordingly, we resorted to domain adaptation. The methodology has been tested with several deep learning algorithms that are conventionally used to handle time series signals. By enabling increased data availability, provided by the physics-based model, we could remarkably reduce the requirements on experimental data while maintaining excellent diagnosis accuracy, also when degradation is observed in the fuel cell. This is a major achievement since it reduces the practical burden of data collection and the cost associated with long experimental campaigns.

Finally, a study investigating the possibility and relevance of extending the method presented in chapter 5, by including SoH estimation, has been presented in chapter 6. In particular, based on the premise that fault diagnosis and health estimation are complementary tasks, we proposed a novel architecture implementing a multitask network. We demonstrated that the proposed framework can achieve high performance on both tasks without running into negative transfer. This represents a step toward an aging-aware fault diagnosis and health monitoring framework for fuel cell systems.

7.2 Key findings

The aim of this research was to explore methods that can effectively integrate physics knowledge with deep learning to address the state-of-the-art limitations in fuel cell fault diagnosis. The major findings allowed to answer to the preliminary research questions, as summarized here:

1. We demonstrated that, by using deep learning algorithms dedicated to temporal data analysis, it is possible to achieve automatic feature extraction from traditional on-board monitoring signals readily available in automotive applications, without resorting to dedicated measurements techniques or ad-hoc feature extraction procedure.

2. The scarce availability of faulty datasets for fuel cell technologies, particularly in degraded conditions of the stack, has a negative impact on the robustness and generalization of deep learning algorithms. However, this limitation can be mitigated through the use of a physics-based model.

3. There is no single optimal method for combining a physics-based model with deep learning; multiple strategies can be effective. This leads us to conclude that the selection of a specific approach depends on the context, including the availability of the data, the computational resources, and the objectives of the task.

4. Ensuring the reliability of a fault diagnosis algorithm throughout the life of the fuel cell system is possible. Deep learning algorithms can effectively distinguish the effects of faults from continuous degradation within the monitoring signals. Nonetheless, to consistently guarantee excellent performance, the training data must include representative samples of the system behavior under aging conditions.

7.3 Limitations and future perspectives

Although this study makes relevant contributions to fuel cell fault diagnosis, advancing the practical applicability of deep learning in the field, there is still room for improvement. Certain challenges remain.

Limited labeled data and unseen faults. Despite the effectiveness of the proposed methods, as they rely on supervised learning, labeled datasets are required for the implementation. These datasets are extremely expensive and time-consuming to obtain. Therefore, unsupervised or semi-supervised learning solutions should be explored. Moreover, supervised learning inherently restricts the faults that can be diagnosed by the algorithm. Strategies that allow the inclusion of previously unseen faults, or that leverage physics-based models to generate synthetic fault data, would enable the algorithms to capture a broader range of fault types, with varying levels of severity.

Online implementation challenges. The deployment of a physics-based model online in embedded automotive applications poses practical challenges, particularly regarding computational efficiency and real-time performance. Solutions such as reduced-order modeling could help provide real-time, virtual state estimations without compromising efficiency. Exploring these approaches is essential to ensure that the physics-guided fault diagnosis can operate effectively during fuel cell system monitoring. Moreover, future works could investigate techniques such as weight quantization to enhance the compatibility of the developed deep learning algorithms with embedded hardware capabilities. Similar strategies can significantly decrease memory usage and computational requirements, thereby facilitating efficient implementation of more complex and sophisticated algorithms.

Experimental validation in fault-tolerant control. To assess the concrete benefits of the developed fault diagnosis algorithms, it is critical to test their integration within control strategies and evaluate their impact on fuel cell degradation. Fault-tolerant control allows modifying the system operation in real time, in a self-adaptive way, enabling immediate response to detected faults and preventing further damage. This would provide insights into the practical advantages of the methods here presented, including improved durability, which cannot be fully

captured through offline analysis alone.

State of health and fault diagnosis. The multitask network introduced in chapter 6 suggested that including information about the health state of the fuel cell could enhance the robustness of the fault diagnosis. Different strategies could be explored to further benefit from integration. For instance, the health indicator could be incorporated through attention mechanisms, allowing the network to focus on relevant degradation information. Alternatively, conditioning the learned features on the state of health could help the algorithm to better adapt to varying aging levels. Eventually, to ensure that the fault diagnosis remains accurate and reliable over time, adaptive model updates, that account for evolving conditions in the fuel cell stack, could also be tested.

Cross-dataset generalization. Fuel cells technology is still progressing, which implies that only small datasets are typically available across multiple technologies. Exploring how these limited datasets can be effectively exploited to develop highly generalizable fault diagnosis algorithms represents a promising research direction. Techniques such as domain generalization could enable knowledge transfer across datasets, facilitating the extraction of useful information from scarce or heterogeneous data sources. Looking ahead, it is reasonable to anticipate that, for a given fuel cell technology, there will be fleet of systems operating under a wide range of conditions and environments, and, consequently, multiple datasets will become available. This scenario will enable training across several fuel cell stacks and subsequent testing on entirely new ones, ultimately accelerating the development of robust and generalizable fault diagnosis algorithms that could be safely deployed.

List of Publications

Scientific papers

- C. Pettorossi, V. Heiries, R. Morvillier, S. Rosini and M. Gerard, "PEM Fuel Cell Fault Diagnosis in Automotive Applications Under Aging Conditions: A Physics-Informed Deep Learning Method," 2025 IEEE/AIAA Transportation Electrification Conference and Electric Aircraft Technologies Symposium (ITEC+EATS), Anaheim, CA, USA, 2025, pp. 1-6, doi: 10.1109/ITEC63604.2025.11098144.
- C. Pettorossi, R. Morvillier, V. Heiries, S. Rosini, and M. Gerard, "Physics-guided fault diagnosis method for proton exchange membrane fuel cells based on LSTM neural network," J. Power Sources, vol. 626, p. 235696, Jan. 2025, doi: 10.1016/J.JPOWSOUR.2024.235696.
- C. Pettorossi, V. Heiries, M. Gerard and S. Rosini, "Addressing Data Scarcity in PEMFC Fault Diagnosis Using Adversarial Learning," 2024 IEEE International Conference on Prognostics and Health Management (ICPHM), Spokane, WA, USA, 2024, pp. 393-398, doi: 10.1109/ICPHM61352.2024.10627464.
- C. Pettorossi, V. Heiries, R. Morvillier, S. Rosini and M. Gerard, "An Integrated Framework for Fault Diagnosis and SoH estimation in Fuel Cell Systems", under internal review.

Non acted conferences and other communications

Talks

- C. Pettorossi, M. Gerard, V. Heiries, S. Rosini. "A novel hybrid method for PEMFC systems fault diagnosis," European Fuel Cells and Hydrogen Pietro Lunghi Conference (EFC23), Capri, Naples, Italy, 2023.
- C. Pettorossi, "Enhancing PEMFC Durability through Hybrid Fault Diagnosis: Combining Physics-based Models and Deep Learning," Fédération HYDROGENE (FRH2) - 4ème réunion des plénières, La Grande Motte, France, 2024.

Posters

- C. Pettorossi, "How physics knowledge can support AI: the case of PEM fuel cell fault diagnosis," PEPR-H2 Day, Nantes, France, 2025.
- C. Pettorossi, "Addressing Data Scarcity in PEMFC Fault Diagnosis using Domain Adaptation," Fédération HYDROGENE (FRH2), La Grande Motte, France, 2024.
- C. Pettorossi, "Increasing PEMFC Durability through Prognostics and Health Management," PEPR-H2 Day, Grenoble, France 2024.

Acronyms

acc	Accuracy
AdaBN	Adaptive Batch Normalization
ANN	Artificial Neural Network
BiLSTM	Bidirectional Long Short-Term Memory
BoL	Beginning of Life
BP	Bipolar plate
CL	Catalyst layer
CNN	Convolutional Neural Network
CORAL	Correlation Alignment
CV	Cyclic Voltammetry
DANN	Domain Adversarial Neural Network
DoE	U.S. Department of Energy
DRT	Distribution of relaxation times
ECSA	Electrochemically Active Surface Area
EIS	Electrochemical Impedance Spectroscopy
EoL	End of Life
FC-DLC	Fuel Cell Dynamic Load Cycle
FCEV	Fuel Cell Electric Vehicles
FCMS	Fuel Cell Management System
FDA	Fisher Discriminant Analysis
FFT	Fast Fourier Transform
FMEA	Failure Mode and Effect Analysis
FNN	Feedforward Neural Network
GAN	Generative Adversarial Network
GB	Gradient Boosting
GDL	Gas diffusion layer
GNB	Gaussian Naïve Bayes
GRL	Gradient Reversal Layer
GRU	Gated Recurrent Unit
HT	Hilbert Transform
k-NN	k-Nearest Neighbors
KDE	Kernel Distribution Estimation
KF	Kalman Filter
KL	Kullback-Leibler

LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MEA	Membrane electrode assembly
MMD	Maximum Mean Discrepancy
PCA	Principal Component Analysis
PEM	Proton exchange membrane
PEMFC	Proton Exchange Membrane Fuel Cell
PHM	Prognostics and Health Management
PINN	Physics-Informed Neural Network
prec	Precision
RC	Reservoir Computing
rec	Recall
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RPLR	Relative Power Loss Rate
RVLR	Relative Voltage Loss Rate
SoH	State of Health
SVM	Support Vector Machine
t-SNE	t-distributed Stochastic Neighbor Embedding
THD	Total harmonic distortion
UKF	Unscented Kalman Filter
WT	Wavelet Transform

Bibliography

References

[1] Transport routier de marchandises : analyse des technologies alternatives aux poids lourds diesel, 2025. Direction générale des Entreprises, URL: https://www.entreprises.gouv.fr/la-dge/publications/transport-routier-de-marchandises-analyse-des-technologies-alternatives-aux.

[2] Hua, Z., Zheng, Z., Pahon, E., Péra, M.C., Gao, F.. A review on lifetime prediction of proton exchange membrane fuel cells system. Journal of Power Sources , vol. 529. doi:10.1016/j.jpowsour.2022.231256, 2022.

[3] Zio, E.. Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice. Reliability Engineering and System Safety , vol. 218. doi:10.1016/j.ress.2021.108119, 2022.

[4] Schopen, O., Shah, N., Esch, T., Shabani, B.. Critical quantitative evaluation of integrated health management methods for fuel cell applications. International Journal of Hydrogen Energy , vol. 70. , pp. 370–388. doi:10.1016/j.ijhydene.2024.05.156, 2024.

[5] Chanal, D.. Diagnosis and prognosis of proton exchange membrane fuel cells by machine learning. PhD thesis. Universitè de Franche-Comte. URL: http://www.theses.fr/2024UBFCD035. thèse de doctorat dirigée par Péra, Marie-CécileYousfi Steiner, Nadia et Chamagne, Didier Énergétique Bourgogne Franche-Comté 2024, 2024.

[6] Fink, O., Wang, Q., Svensén, M., Dersin, P., Lee, W.J., Ducoffe, M.. Potential, challenges and future directions for deep learning in prognostics and health management applications. Engineering Applications of Artificial Intelligence , vol. 92. doi:10.1016/j.engappai.2020.103678, 2020.

[7] Zuo, J., Cadet, C., Li, Z., Bérenguer, C., Outbib, R.. A deterioration-aware energy management strategy for the lifetime improvement of a multi-stack fuel cell system subject to a random dynamic load. Reliability Engineering and System Safety , vol. 241. doi:10.1016/j.ress.2023.109660, 2024.

[8] Dirkes, S., Leidig, J., Fisch, P., Pischinger, S.. Prescriptive Lifetime Management for PEM fuel cell systems in transportation applications, Part I: State of the art and conceptual design. Energy Conversion and Management , vol. 277. doi:10.1016/j.enconman.2022.116598, 2023.

[9] Gibey, G., Pahon, E., Zerhouni, N., Hissel, D.. Diagnostic and prognostic for prescriptive maintenance and control of PEMFC systems in an industrial framework. Journal of Power Sources , vol. 613. doi:10.1016/j.jpowsour.2024.234864, 2024.

[10] Cano, Z., Banham, D., Ye, S., Hintennach, A., Lu, J., Fowler, M., Chen, Z.. Batteries and fuel cells for emerging electric vehicle markets. Nature Energy , vol. 3. , pp. 279–289. doi:10.1038/s41560-018-0108-1, 2018.

[11] Lorenzo, C., Bouquain, D., Hibon, S., Hissel, D.. Synthesis of degradation mechanisms and of their impacts on degradation rates on proton-exchange membrane fuel cells and lithium-ion nickel–manganese–cobalt batteries in hybrid transport applications. Reliability Engineering and System Safety , vol. 212. doi:10.1016/j.ress.2020.107369, 2021.

[12] Abedin, T., Pasupuleti, J., Paw, J., Tak, Y., Mahmud, M., Abdullah, M., Nur-E-Alam, M.. Proton exchange membrane fuel cells in electric vehicles: Innovations, challenges, and pathways to sustainability. Journal of Power Sources , vol. 640. doi:10.1016/j.jpowsour.2025.236769, 2025.

[13] IEA (2025), Global Hydrogen Review 2025, 2025. IEA, URL: https://www.iea.org/reports/global-hydrogen-review-2025.

[14] Meraghni, S., Terrissa, L., Yue, M., Ma, J., Jemei, S., Zerhouni, N.. A data-driven digital-twin prognostics method for proton exchange membrane fuel cell remaining useful life prediction. International Journal of Hydrogen Energy , vol. 46. , pp. 2555–2564. doi:10.1016/j.ijhydene.2020.10.108, 2021.

[15] Kleen, G., Elliot, P.. Durability adjusted fuel cell cost., 2021. U.S. DOE Hydrogen and Fuel Cell Technologies Office, URL: https://www.hydrogen.energy.gov/docs/hydrogenprogramlibraries/pdfs/21001-durability-adjusted-fcs-cost.pdf.

[16] U.S. Department of Energy. Fuel Cell Systems. URL: https://www.energy.gov/eere/fuelcells/fuel-cell-systems, 2021.

[17] Hydrogen and Fuel Cell Technologies Office Multi-Year Program Plan, 2024. U.S. Departement of Energy, URL: https://www.energy.gov/eere/fuelcells/hydrogen-and-fuel-cell-technologies-office-multi-year-program-plan.

[18] Jouin, M., Gouriveau, R., Hissel, D., Péra, M.C., Zerhouni, N.. Prognostics and Health Management of PEMFC – State of the art and remaining challenges. International Journal of Hydrogen Energy , vol. 38. , pp. 15307–15317. doi:10.1016/J.IJHYDENE.2013.09.051, 2013.

[19] Liu, H., Chen, J., Hissel, D., Lu, J., Hou, M., Shao, Z.. Prognostics methods and degradation indexes of proton exchange membrane fuel cells: A review. Renewable and Sustainable Energy Reviews , vol. 123. doi:10.1016/j.rser.2020.109721, 2020.

[20] Pettorossi, C., Morvillier, R., Heiries, V., Rosini, S., Gerard, M.. Physics-guided fault diagnosis method for proton exchange membrane fuel cells based on LSTM neural network. Journal of Power Sources , vol. 626. , pp. 235696. doi:10.1016/J.JPOWSOUR.2024.235696, 2025.

[21] Li, Z., Zheng, Z., Gao, F.. Diagnosis and Prognosis of Proton Exchange Membrane Fuel Cells, in: Electrical Systems 2. (eds A. Soualhi and H. Razik). doi:https://doi.org/10.1002/9781119720584.ch5, 2020.

[22] Zhao, J., Li, X.. A review of polymer electrolyte membrane fuel cell durability for vehicular applications: Degradation modes and experimental techniques. Energy Conversion and Management , vol. 199. doi:10.1016/j.enconman.2019.112022, 2019.

[23] Wang, L., Li, X., Guo, P., Guo, S., Yang, Z., Pei, P.. Bibliometric analysis of prognostics and health management (PHM) in hydrogen fuel cell engines. International Journal of Hydrogen Energy , vol. 47. , pp. 34216–34243. doi:10.1016/J.IJHYDENE.2022.08.024, 2022.

[24] Marrony, M., Barrera, R., Quenet, S., Ginocchio, S., Montelatici, L., Aslanides, A.. Durability study and lifetime prediction of baseline proton exchange membrane fuel cell under severe operating conditions. Journal of Power Sources , vol. 182. , pp. 469–475. doi:10.1016/j.jpowsour.2008.02.096, 2008.

[25] Cheikh, A., Steiner, N.Y., Pahon, E., Damour, C., Benne, M., Hissel, D.. Bayesian Neural Networks with semi-empirical signal processing for PEM Fuel Cell Fault Diagnosis, in: 2024 IEEE Vehicle Power and Propulsion Conference (VPPC), pp. 1–6. doi:10.1109/VPPC63154.2024.10755308, 2024.

[26] Gerard, M.. Optimisation de la durée de vie des systèmes électrogènes : Une approche basée système par modélisation multi-physiques appliquée aux piles à combustible PEMFC. HDR. Energie électrique. Université Grenoble - Alpes, 2021. URL: https://hal.science/tel-04813941, .

[27] Wallnöfer-Ogris, E., Poimer, F., Köll, R., Macherhammer, M.G., Trattner, A.. Main degradation mechanisms of polymer electrolyte membrane fuel cell stacks – Mechanisms, influencing factors, consequences, and mitigation strategies. International Journal of Hydrogen Energy , vol. 50. , pp. 1159–1182. doi:10.1016/J.IJHYDENE.2023.06.215, 2024.

[28] Robin, C.. Développement d'un modèle prédictif de durée de vie d'une pile PEMFC pour une application aéronautique : étude des interactions entre le coeur de pile et les conditions d'opération du système. PhD thesis. Université de Grenoble, 2015.

[29] Nandjou, F., Poirot-Crouvezier, J.P., Chandesris, M., Blachot, J.F., Bonnaud, C., Bultel, Y.. Impact of heat and water management on proton exchange membrane fuel cells degradation in automotive application. Journal of Power Sources , vol. 326. , pp. 182–192. doi:10.1016/j.jpowsour.2016.07.004, 2016.

[30] Shateri, M., Torabi, F.. Influence of liquid water accumulation on the impedance of a PEM fuel cell operating in dead end mode: Physical modeling and experimental validation. Electrochimica Acta , vol. 443. , pp. 141940. doi:10.1016/j.electacta.2023.141940, 2023.

[31] Nandjou, F., Poirot-Crouvezier, J.P., Chandesris, M., Bultel, Y.. A pseudo-3D model to investigate heat and water transport in large area PEM fuel cells – Part 1: Model development and validation. International Journal of Hydrogen Energy , vol. 41. , pp. 15545–15561. doi:10.1016/J.IJHYDENE.2016.05.117, 2016.

[32] Moçotéguy, P., Ludwig, B., Yousfi Steiner, N.. Application of current steps and design of experiments methodology to the detection of water management faults in a proton exchange membrane fuel cell stack. Journal of Power Sources , vol. 303. , pp. 126–136. doi:10.1016/j.jpowsour.2015.10.078, 2016.

[33] Sorrentino, A., Sundmacher, K., Vidakovic-Koch, T.. Polymer electrolyte fuel cell degradation mechanisms and their diagnosis by frequency response analysis methods: A review. Energies , vol. 13. doi:10.3390/en13215825, 2020.

[34] Pei, P., Chen, H.. Main factors affecting the lifetime of Proton Exchange Membrane fuel cells in vehicle applications: A review. Applied Energy , vol. 125. , pp. 60–75. doi:10.1016/j.apenergy.2014.03.048, 2014.

[35] Zhou, S., Shearing, P., Brett, D., Jervis, R.. Machine learning as an online diagnostic tool for proton exchange membrane fuel cells. Current Opinion in Electrochemistry , vol. 31. doi:10.1016/j.coelec.2021.100867, 2022.

[36] Bahrami, M., Martin, J.P., Maranzana, G., Pierfederici, S., Weber, M., Didierjean, S.. Fuel cell management system: An approach to increase its durability. Applied Energy , vol. 306. doi:10.1016/j.apenergy.2021.118070, 2022.

[37] Andrade, P., Laadjal, K., Alcaso, A., Cardoso, A.. A Comprehensive Review on Condition Monitoring and Fault Diagnosis in Fuel Cell Systems: Challenges and Issues. Energies , vol. 17. doi:10.3390/en17030657, 2024.

[38] Ijaodola, O.S., El- Hassan, Z., Ogungbemi, E., Khatib, F.N., Wilberforce, T., Thompson, J., Olabi, A.G.. Energy efficiency improvements by investigating the water flooding management on proton exchange membrane fuel cell (PEMFC). Energy , vol. 179. , pp. 246–267. doi:10.1016/J.ENERGY.2019.04.074, 2019.

[39] Aubry, J., Steiner, N., Morando, S., Zerhouni, N., Hissel, D.. Fuel cell diagnosis methods for embedded automotive applications. Energy Reports , vol. 8. , pp. 6687–6706. doi:10.1016/j.egyr.2022.05.036, 2022.

[40] Ma, T., Zhang, Z., Lin, W., Yang, Y., Yao, N.. A Review on Water Fault Diagnosis of a Proton Exchange Membrane Fuel Cell System. Journal of Electrochemical Energy Conversion and Storage , vol. 18. doi:10.1115/1.4050587, 2021.

[41] Xiao, F., Chen, T., Zhang, J., Zhang, S.. Water management fault diagnosis for proton-exchange membrane fuel cells based on deep learning methods. International Journal of Hydrogen Energy , vol. 48. , pp. 28163–28173. doi:10.1016/J.IJHYDENE.2023.03.097, 2023.

[42] Xue, Q., Shan, Z., Wang, J.. Humidity impact on polarization dynamics in polymer electrolyte membrane fuel cells through distribution of relaxation times analysis. Journal of Power Sources , vol. 609. doi:10.1016/j.jpowsour.2024.234655, 2024.

[43] Aubry, J., Yousfi Steiner, N., Morando, S., Zerhouni, N., Van der Linden, F., Hissel, D.. Fuel Cell prognosis using particle filter: application to the automotive sector, in: IEEE International Symposium on Industrial Electronics, pp. 360–365. doi:10.1109/ISIE51582.2022.9831770, 2022.

[44] Hissel, D., Pera, M.. Diagnostic & health management of fuel cell systems: Issues and solutions. Annual Reviews in Control , vol. 42. , pp. 201–211. doi:10.1016/j.arcontrol.2016.09.005, 2016.

[45] Pahon, E., Yousfi Steiner, N., Jemei, S., Hissel, D., Moçoteguy, P.. A signal-based method for fast PEMFC diagnosis. Applied Energy , vol. 165. , pp. 748–758. doi:10.1016/J.APENERGY.2015.12.084, 2016.

[46] Zuo, J., Steiner, N., Li, Z., Cadet, C., Bérenguer, C., Hissel, D.. Reinforcement learning-based maintenance scheduling for a stochastic deteriorating fuel cell considering stack-to-stack heterogeneity. Reliability Engineering and System Safety , vol. 256. doi:10.1016/j.ress.2024.110700, 2025.

[47] Piffard, M., Da Fonseca, R., Massioni, P., Bideaax, E., Gerard, M.. Fuel cell management system: PEMFC lifetime optimization by model based approach, in: ECS Transactions, pp. 25–35. doi:10.1149/08613.0025ecst, 2018.

[48] Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., Nandi, A.K.. Applications of machine learning to machine fault diagnosis: A review and roadmap. Mechanical Systems and Signal Processing , vol. 138. , pp. 106587. doi:10.1016/J.YMSSP.2019.106587, 2020.

[49] ZHANG, S., SU, L., GU, J., LI, K., ZHOU, L., PECHT, M.. Rotating machinery fault detection and diagnosis based on deep domain adaptation: A survey. Chinese Journal of Aeronautics , vol. 36. , pp. 45–74. doi:10.1016/j.cja.2021.10.006, 2023.

[50] Chanal, D.. Diagnosis and prognosis of proton exchange membrane fuel cells by machine learning. PhD thesis. Université Bourgogne Franche-Comté. URL: tel-04903816, 2024.

[51] Compare, M., Zio, E.. Opportunities and risks of artificial intelligence for industry 5.0 in the context of reliability and maintenance engineering. Journal of Reliability Science and Engineering , vol. 1. doi:10.1088/3050-2454/ad9f63, 2024.

[52] Escobet, T., Feroldi, D., de Lira, S., Puig, V., Quevedo, J., Riera, J., Serra, M.. Model-based fault diagnosis in PEM fuel cell systems. Journal of Power Sources , vol. 192. , pp. 216–223. doi:10.1016/J.JPOWSOUR.2008.12.014, 2009.

[53] Gao, Z., Cecati, C., Ding, S.. A survey of fault diagnosis and fault-tolerant techniques-part I: Fault diagnosis with model-based and signal-based approaches. IEEE Transactions on Industrial Electronics , vol. 62. , pp. 3757–3767. doi:10.1109/TIE.2015.2417501, 2015.

[54] Won, J., Oh, H., Hong, J., Kim, M., Lee, W.Y., Choi, Y.Y., Han, S.B.. Hybrid diagnosis method for initial faults of air supply systems in proton exchange membrane fuel cells. Renewable Energy , vol. 180. , pp. 343–352. doi:10.1016/j.renene.2021.07.079, 2021.

[55] Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A., Wu, B., 2007. Intelligent Fault Diagnosis and Prognosis for Engineering Systems. doi:10.1002/9780470117842.

[56] Petrone, R., Zheng, Z., Hissel, D., Péra, M.C., Pianese, C., Sorrentino, M., Becherif, M., Yousfi-Steiner, N.. A review on model-based diagnosis methodologies for PEMFCs. International Journal of Hydrogen Energy , vol. 38. , pp. 7077–7091, 2013.

[57] Wilhelm, Y., Reimann, P., Gauchel, W., Mitschang, B.. Overview on hybrid approaches to fault detection and diagnosis: Combining data-driven, physics-based and knowledge-based models, in: Procedia CIRP, pp. 278–283. doi:10.1016/j.procir.2021.03.041, 2021.

[58] Frank, P.. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy. A survey and some new results. Automatica , vol. 26. , pp. 459–474. doi:10.1016/0005-1098(90)90018-D, 1990.

[59] Pohjoranta, A., Sorrentino, M., Pianese, C., Amatruda, F., Hottinen, T.. Validation of neural network-based fault diagnosis for multi-stack fuel cell systems: Stack voltage deviation detection, in: Energy Procedia, pp. 173–181. doi:10.1016/j.egypro.2015.12.072, 2015.

[60] Ariza, E., Correcher, A., Vargas-Salgado, C.. PEMFCs Model-Based Fault Diagnosis: A Proposal Based on Virtual and Real Sensors Data Fusion. Sensors , vol. 23. doi:10.3390/s23177383, 2023.

[61] Jullian, G., Cadet, C., Rosin, S., Gerard, M., Heiries, V., Bérenguer, C.. Fault Detection and Isolation for Proton Exchange Membrane Fuel Cell Using Impedance Measurements and Multiphysics Modeling. Fuel Cells , vol. 20. , pp. 558–569, 2020.

[62] Zheng, Z., Petrone, R., Péra, M., Hissel, D., Becherif, M., Pianese, C., Yousfi Steiner, N., Sorrentino, M.. A review on non-model based diagnosis methodologies for PEM fuel cell stacks and systems. International Journal of Hydrogen Energy , vol. 38. , pp. 8914–8926. doi:10.1016/j.ijhydene.2013.04.007, 2013.

[63] Koo, T., Ko, R., Ha, D., Han, J.. Development of Model-Based PEM Water Electrolysis HILS (Hardware-in-the-Loop Simulation) System for State Evaluation and Fault Detection. Energies , vol. 16. doi:10.3390/en16083379, 2023.

[64] Hernandez, A., Hissel, D., Outbib, R.. Modeling and fault diagnosis of a polymer electrolyte fuel cell using electrical equivalent analysis. IEEE Transactions on Energy Conversion , vol. 25. , pp. 148–160. doi:10.1109/TEC.2009.2016121, 2010.

[65] Buchholz, M., Eswein, M., Krebs, V.. Modelling PEM fuel cell stacks for FDI using linear subspace identification, in: Proceedings of the IEEE International Conference on Control Applications, pp. 341–346. doi:10.1109/CCA.2008.4629629, 2008.

[66] Zhong, Z.D., Zhu, X.J., Cao, G.Y., Shi, J.H.. A hybrid multi-variable experimental model for a PEMFC. Journal of Power Sources , vol. 164. , pp. 746–751. doi:10.1016/J.JPOWSOUR.2006.11.030, 2007.

[67] Yousfi Steiner, N., Candusso, D., Hissel, D., Moçtéguy, P.. Model-based diagnosis for proton exchange membrane fuel cells. Mathematics and Computers in Simulation , vol. 81. , pp. 158–170. doi:10.1016/j.matcom.2010.02.006, 2010.

[68] Yousfi Steiner, N., Hissel, D., Moçtéguy, P., Candusso, D.. Diagnosis of polymer electrolyte fuel cells failure modes (flooding & drying out) by neural networks modeling. International Journal of Hydrogen Energy , vol. 36. , pp. 3067–3075. doi:10.1016/J.IJHYDENE.2010.10.077, 2011.

[69] Young Park, J., Seop Lim, I., Ho Lee, Y., Lee, W.Y., Oh, H., Soo Kim, M.. Severity-based fault diagnostic method for polymer electrolyte membrane fuel cell systems. Applied Energy , vol. 332. , pp. 120486. doi:10.1016/j.apenergy.2022.120486, 2023.

[70] De Lira, S., Puig, V., Quevedo, J.. Robust LPV model-based sensor fault diagnosis and estimation for a PEM fuel cell system, in: Conference on Control and Fault-Tolerant Systems, SysTol'10 - Final Program and Book of Abstracts, pp. 819–824. doi:10.1109/SYSTOL.2010.5676000, 2010.

[71] Dijoux, E., Steiner, N., Benne, M., Péra, M.C., Pérez, B.. A review of fault tolerant control strategies applied to proton exchange membrane fuel cell systems. Journal of Power Sources , vol. 359. , pp. 119–133. doi:10.1016/j.jpowsour.2017.05.058, 2017.

[72] Liu, J., Luo, W., Yang, X., Wu, L.. Robust Model-Based Fault Diagnosis for PEM Fuel Cell Air-Feed System. IEEE Transactions on Industrial Electronics , vol. 63. , pp. 3261–3270. doi:10.1109/TIE.2016.2535118, 2016.

[73] Diez-Olivan, A., Del Ser, J., Galar, D., Sierra, B.. Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. Information Fusion , vol. 50. , pp. 92–111. doi:10.1016/j.inffus.2018.10.005, 2019.

[74] Gao, Z., Cecati, C., Ding, S.. A survey of fault diagnosis and fault-tolerant techniques-part II: Fault diagnosis with knowledge-based and hybrid/active approaches. IEEE Transactions on Industrial Electronics , vol. 62. , pp. 3768–3774. doi:10.1109/TIE.2015.2419013, 2015.

[75] Gjorgiev, B., Das, L., Merkel, S., Rohrer, M., Auger, E., Sansavini, G.. Simulation-driven deep learning for locating faulty insulators in a power line. Reliability Engineering and System Safety , vol. 231. doi:10.1016/j.ress.2022.108989, 2023.

[76] Isermann, R., 2006. Fault-diagnosis systems: An introduction from fault detection to fault tolerance. doi:10.1007/3-540-30368-5.

[77] Khemani, V., Azarian, M., Pecht, M.. WavePHMNet: A comprehensive diagnosis and prognosis approach for analog circuits. Advanced Engineering Informatics , vol. 59. doi:10.1016/j.aei.2023.102323, 2024.

[78] Khan, S., Yairi, T.. A review on the application of deep learning in system health management. Mechanical Systems and Signal Processing , vol. 107. , pp. 241–265. doi:10.1016/J.YMSSP.2017.11.024, 2018.

[79] Wilson, G., Doppa, J., Cook, D.. Multi-Source Deep Domain Adaptation with Weak Supervision for Time-Series Sensor Data, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1768–1778. doi:10.1145/3394486.3403228, 2020.

[80] Chao, M., Kulkarni, C., Goebel, K., Fink, O.. Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics. Data , vol. 6. , pp. 1–14. doi:10.3390/data6010005, 2021.

[81] Nielsen, M., 2019. Neural networks and deep learning. URL: http://neuralnetworksanddeeplearning.com/.

[82] Arias Chao, M., Kulkarni, C., Goebel, K., Fink, O.. Fusing physics-based and deep learning models for prognostics. Reliability Engineering and System Safety , vol. 217. doi:10.1016/j.ress.2021.107961, 2022.

[83] Morizet, N., Desforges, P., Geissler, C., Pahon, E., Jemei, S., Hissel, D.. Time to market reduction for hydrogen fuel cell stacks using Generative Adversarial Networks. Journal of Power Sources , vol. 579. doi:10.1016/j.jpowsour.2023.233286, 2023.

[84] Pahon, E., Hissel, D., Jemei, S., Steiner, N.. Signal-based diagnostic approach to enhance fuel cell durability. Journal of Power Sources , vol. 506. doi:10.1016/j.jpowsour.2021.230223, 2021.

[85] Yousfi Steiner, N., Hissel, D., Moçtéguy, P., Candusso, D.. Non intrusive diagnosis of polymer electrolyte fuel cells by wavelet packet transform. International Journal of Hydrogen Energy , vol. 36. , pp. 740–746. doi:10.1016/j.ijhydene.2010.10.033, 2011.

[86] Benouioua, D., Candusso, D., Harel, F., Oukhellou, L.. PEMFC stack voltage singularity measurement and fault classification. International Journal of Hydrogen Energy , vol. 39. , pp. 21631–21637. doi:10.1016/j.ijhydene.2014.09.117, 2014.

[87] Pahon, E., Yousfi Steiner, N., Jemei, S., Hissel, D., Péra, M., Wang, K., Moçoteguy, P.. Solid oxide fuel cell fault diagnosis and ageing estimation based on wavelet transform approach. International Journal of Hydrogen Energy , vol. 41. , pp. 13678–13687. doi:10.1016/j.ijhydene.2016.06.143, 2016.

[88] Cheikh, A., Steiner, N., Pahon, E., Damour, C., Benne, M., Hissel, D.. Proton Exchange Membrane Fuel Cell Signal-Based Diagnostics Using Empirical Fourier Transform, in: 2022 IEEE Vehicle Power and Propulsion Conference, VPPC 2022 - Proceedings, pp. 1–6. doi:10.1109/VPPC55846.2022.10003412, 2022.

[89] Li, Z., Outbib, R., Giurgea, S., Hissel, D.. Fault Diagnosis for PEMFC Systems in Consideration of Dynamic Behaviors and Spatial Inhomogeneity. IEEE Transactions on Energy Conversion , vol. 34. , pp. 3–11. doi:10.1109/TEC.2018.2824902, 2019.

[90] Morando, S., Pera, M., Steiner, N., Jemei, S., Hissel, D., Larger, L.. Fuel Cells Fault Diagnosis under Dynamic Load Profile Using Reservoir Computing, in: 2016 IEEE Vehicle Power and Propulsion Conference, VPPC 2016 - Proceedings, pp. 1–6. doi:10.1109/VPPC.2016.7791693, 2016.

[91] Zheng, Z., Morando, S., Pera, M.C., Hissel, D., Larger, L., Martinenghi, R., Baylon Fuentes, A.. Brain-inspired computational paradigm dedicated to fault diagnosis of PEM fuel cell stack. International Journal of Hydrogen Energy , vol. 42. , pp. 5410–5425. doi:10.1016/j.ijhydene.2016.11.043, 2017.

[92] Lu, H., Chen, J., Yan, C., Liu, H.. On-line fault diagnosis for proton exchange membrane fuel cells based on a fast electrochemical impedance spectroscopy measurement. Journal of Power Sources , vol. 430. , pp. 233–243. doi:10.1016/j.jpowsour.2019.05.028, 2019.

[93] Ritzberger, D., Jakubek, S.. Nonlinear data-driven identification of polymer electrolyte membrane fuel cells for diagnostic purposes: A Volterra series approach. Journal of Power Sources , vol. 361. , pp. 144–152. doi:10.1016/j.jpowsour.2017.06.068, 2017.

[94] Steffy, N., Selvaganesh, S., Kumar L, M., Sahu, A.. Online monitoring of fuel starvation and water management in an operating polymer electrolyte membrane fuel cell by a novel diagnostic tool based on total harmonic distortion analysis. Journal of Power Sources , vol. 404. , pp. 81–88. doi:10.1016/j.jpowsour.2018.10.012, 2018.

[95] Detti, A., Jemei, S., Morando, S., Steiner, N.. Classification Based Method Using Fast Fourier Transform (FFT) and Total Harmonic Distortion (THD) Dedicated to Proton Exchange Membrane Fuel Cell (PEMFC) Diagnosis, in: 2017 IEEE Vehicle Power and Propulsion Conference, VPPC 2017 - Proceedings, pp. 1–6. doi:10.1109/VPPC.2017.8331040, 2018.

[96] Rubio, M., Sanchez, D., Gazdzicki, P., Friedrich, K., Urquia, A.. Failure mode diagnosis in proton exchange membrane fuel cells using local electrochemical noise. Journal of Power Sources , vol. 541. doi:10.1016/j.jpowsour.2022.231582, 2022.

[97] Ao, Y., Li, Z., Laghrouche, S., Depernet, D., Candusso, D., Zhao, K.. Stack-level diagnosis of proton exchange membrane fuel cell by the distribution of relaxation times analysis of electrochemical impedance spectroscopy. Journal of Power Sources , vol. 603. doi:10.1016/j.jpowsour.2024.234420, 2024.

[98] Nasarre Artigas, S., Xu, H., Mack, F.. Use of distribution of relaxation times analysis as an in-situ diagnostic tool for water management in PEM fuel cells applications. Journal of Power Sources , vol. 600. doi:10.1016/j.jpowsour.2024.234179, 2024.

[99] Zhao, X., Xu, L., Li, J., Fang, C., Ouyang, M.. Faults diagnosis for PEM fuel cell system based on multi-sensor signals and principle component analysis method. International Journal of Hydrogen Energy , vol. 42. , pp. 18524–18531. doi:10.1016/j.ijhydene.2017.04.146, 2017.

[100] Zhang, C.L., Zhang, B.X., Xu, J.H., Chen, Z.L., Zheng, X.Y., Zhu, K.Q., Bo, Z., Yang, Y.R., Wang, X.D.. Fault diagnosis of the hybrid system composed of high-power PEMFCs and ammonia-hydrogen fueled internal combustion engines using ensemble deep learning methods. International Journal of Hydrogen Energy , vol. 92. , pp. 1215–1235. doi:10.1016/j.ijhydene.2024.10.332, 2024.

[101] Li, Z., Outbib, R., Giurgea, S., Hissel, D., Li, Y.. Fault detection and isolation for Polymer Electrolyte Membrane Fuel Cell systems by analyzing cell voltage generated space. Applied Energy , vol. 148. , pp. 260–272. doi:10.1016/j.apenergy.2015.03.076, 2015a.

[102] Li, Z., Outbib, R., Giurgea, S., Hissel, D.. Diagnosis for PEMFC Systems: A Data-Driven Approach With the Capabilities of Online Adaptation and Novel Fault Detection. IEEE Transactions on Industrial Electronics , vol. 62. , pp. 5164–5174. doi:10.1109/TIE.2015.2418324, 2015b.

[103] Li, Z., Outbib, R., Giurgea, S., Hissel, D., Giraud, A., Couderc, P.. Fault diagnosis for fuel cell systems: A data-driven approach using high-precise voltage sensors. Renewable Energy , pp. 1435–1444. doi:10.1016/j.renene.2018.09.077, 2019.

[104] Lin, R.H., Pei, Z.X., Ye, Z.Z., Guo, C.C., Wu, B.D.. Hydrogen fuel cell diagnostics using random forest and enhanced feature selection. International Journal of Hydrogen Energy , vol. 45. , pp. 10523–10535. doi:10.1016/j.ijhydene.2019.10.127, 2020.

[105] Chanal, D., Steiner, N.Y., Petrone, R., Chamagne, D., Péra, M.C.. Online Diagnosis of PEM Fuel Cell by Fuzzy C-Means Clustering. Encyclopedia of Energy Storage , vol. Oxford: El. , pp. 359–393. doi:10.1016/B978-0-12-819723-3.00099-8, 2022.

[106] Yu, D., Li, X., Araya, S., Sahlin, S., Zhou, F., Liso, V.. A novel method of EIS application in online fault diagnosis of high-temperature PEMFC with CNN. Journal of Power Sources , vol. 639. doi:10.1016/j.jpowsour.2025.236663, 2025.

[107] Li, Z., Outbib, R., Giurgea, S., Hissel, D., Jemei, S., Giraud, A., Rosini, S.. Online implementation of SVM based fault diagnosis strategy for PEMFC systems. Applied Energy , vol. 164. , pp. 284–293. doi:10.1016/J.APENERGY.2015.11.060, 2016.

[108] Pahon, E., Jemei, S., Hissel, D.. Supervised classification approach dedicated to proton exchange membrane fuel cell diagnostic, in: Proceedings of the 2019 IEEE 12th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives, SDEMPED 2019, pp. 446–451. doi:10.1109/DEMPED.2019.8864841, 2019.

[109] Petrone, R., Chamagne, D., Pera, M.C., Hissel, D.. Data-driven multi-fault diagnosis for H2/O2 and H2/air PEMFCs, in: Lecture Notes in Electrical Engineering. volume 697, pp. 3–12. doi:10.1007/978-3-030-56970-9_1, 2020.

[110] Franić, N., Pivac, I., Barbir, F.. A review of machine learning applications in hydrogen electrochemical devices. International Journal of Hydrogen Energy , vol. 102. , pp. 523–544. doi:10.1016/J.IJHYDENE.2025.01.070, 2025.

[111] Zhang, X., Guo, X.. Fault diagnosis of proton exchange membrane fuel cell system of tram based on information fusion and deep learning. International Journal of Hydrogen Energy , vol. 46. , pp. 30828–30840. doi:10.1016/J.IJHYDENE.2021.05.137, 2021.

[112] Gu, X., Hou, Z., Cai, J.. Data-based flooding fault diagnosis of proton exchange membrane fuel cell systems using LSTM networks. Energy and AI , vol. 4. doi:10.1016/j.egyai.2021.100056, 2021.

[113] Liu, J., Li, Q., Yang, H., Han, Y., Jiang, S., Chen, W.. Sequence Fault Diagnosis for PEMFC Water Management Subsystem Using Deep Learning with t-SNE. IEEE Access , vol. 7. , pp. 92009–92019. doi:10.1109/ACCESS.2019.2927092, 2019.

[114] Zuo, B., Zhang, Z., Cheng, J., Huo, W., Zhong, Z., Wang, M.. Data-driven flooding fault diagnosis method for proton-exchange membrane fuel cells using deep learning technologies. Energy Conversion and Management , vol. 251. doi:10.1016/j.enconman.2021.115004, 2022.

[115] Zhou, S., Lu, Y., Bao, D., Wang, K., Shan, J., Hou, Z.. Real-time data-driven fault diagnosis of proton exchange membrane fuel cell system based on binary encoding convolutional neural network. International Journal of Hydrogen Energy , vol. 47. , pp. 10976–10989. doi:10.1016/j.ijhydene.2022.01.145, 2022.

[116] Liu, Z., Mao, L., Hu, Z., Huang, W., Wu, Q., Jackson, L.. A novel densely connected neural network for proton exchange membrane fuel cell fault diagnosis. International Journal of Hydrogen Energy , vol. 47. , pp. 40041–40053. doi:10.1016/j.ijhydene.2022.09.158, 2022.

[117] Kim, K., Kim, J., Choi, H., Kwon, O., Jang, Y., Ryu, S., Lee, H., Shim, K., Park, T., Cha, S.. Pre-diagnosis of flooding and drying in proton exchange membrane fuel cells by bagging ensemble deep learning models using long short-term memory and convolutional neural networks. Energy , vol. 266. doi:10.1016/j.energy.2022.126441, 2023.

[118] Liu, Z., Sun, Y., Tang, X., Mao, L.. Enabling unsupervised fault diagnosis of proton exchange membrane fuel cell stack: Knowledge transfer from single-cell to stack. Applied Energy , vol. 360. doi:10.1016/j.apenergy.2024.122814, 2024.

[119] Gao, S., Sun, Z., Wang, Y., Chen, Z.. A knowledge transfer method for water faults diagnosis of proton exchange membrane fuel cell based on sample re-weighting. Applied Energy , vol. 386. , pp. 125575. doi:10.1016/J.APENERGY.2025.125575, 2025.

[120] Tidriri, K., Chatti, N., Verron, S., Tiplica, T.. Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges. Annual Reviews in Control , vol. 42. , pp. 63–81. doi:10.1016/J.ARCONTROL.2016.09.008, 2016.

[121] Chinesta, F., Cueto, E.. Empowering engineering with data, machine learning and artificial intelligence: a short introductive review. Advanced Modeling and Simulation in Engineering Sciences , vol. 9. doi:10.1186/s40323-022-00234-8, 2022.

[122] Gao, S., Feng, C., Wang, B., Yan, T., He, B., Zio, E.. Data and Model Combined Unsupervised Fault Detection and Assessment Framework for Underwater Thruster. IEEE Transactions on Industrial Informatics , vol. 20. , pp. 8229–8238. doi:10.1109/TII.2024.3352269, 2024.

[123] Xu, Y., Kohtz, S., Boakye, J., Gardoni, P., Wang, P.. Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges. Reliability Engineering and System Safety , vol. 230. doi:10.1016/j.ress.2022.108900, 2023.

[124] Faghih Niresi, K., Bissig, H., Baumann, H., Fink, O.. Physics-Enhanced Graph Neural Networks for Soft Sensing in Industrial Internet of Things. IEEE Internet of Things Journal , vol. 11. , pp. 34978–34990. doi:10.1109/JIOT.2024.3434732, 2024.

[125] Sepe, M., Graziano, A., Badora, M., Di Stazio, A., Bellani, L., Compare, M., Zio, E.. A physics-informed machine learning framework for predictive maintenance applied to turbomachinery assets. Journal of the Global Power and Propulsion Society , vol. 2021. doi:10.33737/jgpps/134845, 2021.

[126] Li, C., Li, S., Feng, Y., Gryllias, K., Gu, F., Pecht, M.. Small data challenges for intelligent prognostics and health management: a review. Artificial Intelligence Review , vol. 57. , pp. 214. doi:10.1007/s10462-024-10820-4, 2024.

[127] Achouch, M., Dimitrova, M., Ziane, K., Sattarpanah Karganroudi, S., Dhouib, R., Ibrahim, H., Adda, M.. On Predictive Maintenance in Industry 4.0: Overview, Models, and Challenges. Applied Sciences (Switzerland) , vol. 12. doi:10.3390/app12168081, 2022.

[128] Cuomo, S., Di Cola, V., Giampaolo, F., Rozza, G., Raissi, M., Piccialli, F.. Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What's Next. Journal of Scientific Computing , vol. 92. doi:10.1007/s10915-022-01939-z, 2022.

[129] Meng, H., Geng, M., Xing, J., Zio, E.. A hybrid method for prognostics of lithium-ion batteries capacity considering regeneration phenomena. Energy , vol. 261. doi:10.1016/j.energy.2022.125278, 2022.

[130] Zhu, Y., Zabaras, N., Koutsourelakis, P.S., Perdikaris, P.. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. Journal of Computational Physics , vol. 394. , pp. 56–81. doi:10.1016/J.JCP.2019.05.024, arXiv:1901.06314, 2019.

[131] Jia, X., Karpatne, A., Willard, J., Steinbach, M., Read, J., Hanson, P.C., Dugan, H.A., Kumar, V.. Physics Guided Recurrent Neural Networks For Modeling Dynamical Systems: Application to Monitoring Water Temperature And Quality In Lakes doi:10.48550/arxiv.1810.02880, arXiv:1810.02880, 2018.

[132] Gao, S., Liu, J., Zhang, Z., Feng, C., He, B., Zio, E.. Physics-Guided Generative Adversarial Networks for fault detection of underwater thruster. Ocean Engineering , vol. 286. doi:10.1016/j.oceaneng.2023.115585, 2023.

[133] Wang, C., Liu, J., Zio, E.. A Modified Generative Adversarial Network for Fault Diagnosis in High-Speed Train Components with Imbalanced and Heterogeneous Monitoring Data. Journal of Dynamics, Monitoring and Diagnostics , vol. 1. , pp. 84–92. doi:10.37965/jdmd.2022.68, 2022.

[134] He, Y., Su, H., Zio, E., Fan, L., Zhang, J.. A Physics-Informed Training Approach for Data-Driven Method in Remaining Useful Life Estimation, in: 2022 6th International Conference on System Reliability and Safety, ICSRS 2022, pp. 500–504. doi:10.1109/ICSRS56243.2022.10067722, 2022.

[135] Karniadakis, G., Kevrekidis, I., Lu, L., Perdikaris, P., Wang, S., Yang, L.. Physics-informed machine learning. Nature Reviews Physics , vol. 3. , pp. 422–440. doi:10.1038/s42254-021-00314-5, 2021.

[136] Chakraborty, S.. Transfer learning based multi-fidelity physics informed deep neural network. Journal of Computational Physics , vol. 426. , pp. 109942. doi:10.1016/J.JCP.2020.109942, arXiv:2005.10614, 2021.

[137] Aykol, M., Gopal, C.B., Anapolsky, A., Herring, P.K., van Vlijmen, B., Berliner, M.D., Bazant, M.Z., Braatz, R.D., Chueh, W.C., Storey, B.D.. Perspective—Combining Physics and Machine Learning to Predict Battery Lifetime. Journal of The Electrochemical Society , vol. 168. , pp. 030525. doi:10.1149/1945-7111/abec55, 2021.

[138] Jullian, G., Rosini, S., Gérard, M., Cadet, C., Bérenguer, C., Heiries, V.. Model-based fault detection using analytical redundancy for automotive proton exchange membrane fuel cell, in: Safety and Reliability - Theory and Applications - Proceedings of the 27th European Safety and Reliability Conference, ESREL 2017, pp. 905–912. doi:10.1201/9781315210469-117, 2017.

[139] Lee, W.Y., Oh, H., Kim, M., Choi, Y.Y., Sohn, Y.J., Kim, S.G.. Hierarchical fault diagnostic method for a polymer electrolyte fuel cell system. International Journal of Hydrogen Energy , vol. 45. , pp. 25733–25746. doi:10.1016/j.ijhydene.2019.10.145, 2020.

[140] Mohammadi, A., Djerdir, A., Yousfi Steiner, N., Khaburi, D.. Advanced diagnosis based on temperature and current density distributions in a single PEMFC. International Journal of Hydrogen Energy , vol. 40. , pp. 15845–15855. doi:10.1016/j.ijhydene.2015.04.157, 2015.

[141] Gong, Z., Wang, B., Xing, Y., Xu, Y., Qin, Z., Chen, Y., Zhang, F., Gao, F., Li, B., Yin, Y., Du, Q., Jiao, K.. High-precision and efficiency diagnosis for polymer electrolyte membrane fuel cell based on physical mechanism and deep learning. eTransportation , vol. 18. , pp. 100275. doi:10.1016/J.ETRAN.2023.100275, 2023.

[142] Laribi, S., Mammar, K., Sahli, Y., Koussa, K.. Analysis and diagnosis of PEM fuel cell failure modes (flooding & drying) across the physical parameters of electrochemical impedance model: Using neural networks method. Sustainable Energy Technologies and Assessments , vol. 34. , pp. 35–42. doi:10.1016/j.seta.2019.04.004, 2019.

[143] Shi, Y., He, W., Xie, B., Du, Y., Yang, X., Su, J.. PEMFC Fault Diagnosis Based on an Equivalent Circuit and OS-ELM Framework. IEEE Transactions on Industry Applications doi:10.1109/TIA.2023.3313134, 2023.

[144] Khan, S., Shareef, H., Ibrahim, A.. Improved Semi-empirical Model of Proton Exchange Membrane Fuel Cell Incorporating Fault Diagnostic Feature. Journal of Modern Power Systems and Clean Energy , vol. 9. , pp. 1566–1573. doi:10.35833/MPCE.2019.000179, 2021.

[145] Lv, J., Yu, Z., Sun, G., Liu, J.. Deep learning-based fault diagnosis and Electrochemical Impedance Spectroscopy frequency selection method for Proton Exchange Membrane Fuel Cell. Journal of Power Sources , vol. 591. doi:10.1016/j.jpowsour.2023.233815, 2024.

[146] Nascimento, R.G., Viana, F.A.C.. Fleet Prognosis with Physics-informed Recurrent Neural Networks doi:10.48550/arxiv.1901.05512, arXiv:1901.05512, 2019.

[147] Liu, Z., Sun, H., Xu, L., Mao, L., Hu, Z., Li, J.. Toward low-data and real-time PEMFC diagnostic: Multi-sine stimulation and hybrid ECM-informed neural network. Applied Energy , vol. 391. doi:10.1016/j.apenergy.2025.125959, 2025.

[148] Zhang, D., Baraldi, P., Cadet, C., Yousfi-Steiner, N., Bérenguer, C., Zio, E.. An ensemble of models for integrating dependent sources of information for the prognosis of the remaining useful life of Proton Exchange Membrane Fuel Cells. Mechanical Systems and Signal Processing , vol. 124. , pp. 479–501. doi:10.1016/j.ymssp.2019.01.060, 2019.

[149] Thelen, A., Zhang, X., Fink, O., Lu, Y., Ghosh, S., Youn, B., Todd, M., Mahadevan, S., Hu, C., Hu, Z.. A comprehensive review of digital twin—part 2: roles of uncertainty quantification and optimization, a battery digital twin, and perspectives. Structural and Multidisciplinary Optimization , vol. 66. doi:10.1007/s00158-022-03410-x, 2023.

[150] Thelen, A., Lui, Y., Shen, S., Laflamme, S., Hu, S., Ye, H., Hu, C.. Integrating physics-based modeling and machine learning for degradation diagnostics of lithium-ion batteries. Energy Storage Materials , vol. 50. , pp. 668–695. doi:10.1016/j.ensm.2022.05.047, 2022.

[151] Costamagna, P., De Giorgi, A., Moser, G., Serpico, S., Trucco, A.. Data-driven techniques for fault diagnosis in power generation plants based on solid oxide fuel cells. Energy Conversion and Management , vol. 180. , pp. 281–291. doi:10.1016/j.enconman.2018.10.107, 2019.

[152] Lu, Y., Zhou, S., Yin, D., Fan, L., Zhang, G., Gao, J.. Simultaneous fault diagnosis of proton exchange membrane fuel cell systems based on an Incremental Multi-label Classification Network. International Journal of Hydrogen Energy , vol. 47. , pp. 24963–24977. doi:10.1016/J.IJHYDENE.2022.05.231, 2022.

[153] Sun, Z., Wang, Y., Chen, Z.. Fault diagnosis method for proton exchange membrane fuel cell system based on digital twin and unsupervised domain adaptive learning. International Journal of Hydrogen Energy , vol. 50. , pp. 1207–1219. doi:10.1016/j.ijhydene.2023.10.148, 2024.

[154] Gong, Z., Wang, B., Benbouzid, M., Li, B., Xu, Y., Yang, K., Bao, Z., Amirat, Y., Gao, F., Jiao, K.. Cross-domain diagnosis for polymer electrolyte membrane fuel cell based on digital twins and transfer learning network. Energy and AI , vol. 17. , pp. 100412. doi:10.1016/J.EGYAI.2024.100412, 2024.

[155] Hong, J., Yang, J., Weng, Z., Ma, F., Liang, F., Zhang, C.. Review on proton exchange membrane fuel cells: Safety analysis and fault diagnosis. Journal of Power Sources , vol. 617. doi:10.1016/j.jpowsour.2024.235118, 2024.

[156] Pourrahmani, H., Yavarinasab, A., Siavashi, M., Matian, M., Van herle, J.. Progress in the proton exchange membrane fuel cells (PEMFCs) water/thermal management: From theory to the current challenges and real-time fault diagnosis methods. Energy Reviews , vol. 1. doi:10.1016/j.enrev.2022.100002, 2022.

[157] Lukaszewski, M., Soszko, M., Czerwiński, A.. Electrochemical methods of real surface area determination of noble metal electrodes - an overview. International Journal of Electrochemical Science , vol. 11. , pp. 4442–4469. doi:10.20964/2016.06.71, 2016.

[158] Sun, Y., Wang, H., Li, R., Hu, B., Fang, J., Lei, Y., Li, Z., Hissel, D.. A Low-Cost High-Accuracy Online Electrochemical Impedance Spectroscopy Measurement Strategy for Fuel Cell Electric Vehicle Application. IEEE Transactions on Instrumentation and Measurement , vol. 74. doi:10.1109/TIM.2025.3557105, 2025.

[159] Plait, A., Giurgea, S., Hissel, D., Espanet, C.. New magnetic field analyzer device dedicated for polymer electrolyte fuel cells noninvasive diagnostic. International Journal of Hydrogen Energy , vol. 45. , pp. 14071–14082. doi:10.1016/J.IJHYDENE.2020.03.082, 2020.

[160] Ifrek, L., Chadebec, O., Rosini, S., Cauffet, G., Bultel, Y., Bannwarth, B.. Fault Identification on a Fuel Cell by 3-D Current Density Reconstruction from External Magnetic Field Measurements. IEEE Transactions on Magnetics , vol. 55. doi:10.1109/TMAG.2019.2895755, 2019.

[161] Aubry, J.. Diagnostic, pronostic, contrôle tolérant aux défauts et au vieillissement d'une pile à combustible à membrane échangeuse de protons, appliqués à l'automobile. PhD thesis. Universitè de Franche-Comte, 2022.

[162] Lee, Y.H., Kim, J., Yoo, S.. On-line and real-time diagnosis method for proton membrane fuel cell (PEMFC) stack by the superposition principle. Journal of Power Sources , vol. 326. , pp. 264–269. doi:10.1016/J.JPOWSOUR.2016.06.113, 2016.

[163] Pei, P., Li, Y., Xu, H., Wu, Z.. A review on water fault diagnosis of PEMFC associated with the pressure drop. Applied Energy , vol. 173. , pp. 366–385. doi:10.1016/j.apenergy.2016.04.064, 2016.

[164] Ziane, M.A., Join, C., Benne, M., Damour, C., Steiner, N.Y., Pera, M.C.. A new concept of water management diagnosis for a PEM fuel cell system. Energy Conversion and Management , vol. 285. , pp. 116986. doi:10.1016/J.ENCONMAN.2023.116986, 2023.

[165] Jullian, G.. Diagnostic robuste de la pile à combustible PEM par modélisation physique et mesures d'impédance: prise en compte de conditions dynamiques et du vieillissement. PhD thesis. Université Grenoble Alpes, 2016.

[166] seaborn.kdeplot. URL: https://seaborn.pydata.org/generated/seaborn.kdeplot.html, .

[167] Kullback, S., Leibler, R.A.. On information and sufficiency. Annals of Mathematical Statistics , vol. 22. , pp. 79–86. doi:https://doi.org/10.1214/aoms/1177729694, 1951.

[168] Poirot-Crouvezier, J.P.. Modélisation dynamique des phénomènes hydrauliques, thermiques et électriques dans un groupe électrogène à pile à combustible destiné à l'application automobile. PhD thesis. Institut National Polytechnique de Grenoble. doi:https://hal.science/tel-02907129v1, 2000.

[169] Chandesris, M., Médeau, V., Guillet, N., Chelghoum, S., Thoby, D., Fouda-Onana, F.. Membrane degradation in PEM water electrolyzer: Numerical modeling and experimental evidence of the influence of temperature and current density. International Journal of Hydrogen Energy , vol. 40. , pp. 1353–1366. doi:10.1016/j.ijhydene.2014.11.111, 2015.

[170] Parthasarathy, P., Virkar, A.V.. Electrochemical Ostwald ripening of Pt and Ag catalysts supported on carbon. Journal of Power Sources , vol. 234. , pp. 82–90. doi:10.1016/J.JPOWSOUR.2013.01.115, 2013.

[171] Robin, C., Gerard, M., Franco, A., Schott, P.. Multi-scale coupling between two dynamical models for PEMFC aging prediction. International Journal of Hydrogen Energy , vol. 38. , pp. 4675–4688. doi:10.1016/j.ijhydene.2013.01.040, 2013.

[172] Tu, H., Moura, S., Wang, Y., Fang, H.. Integrating physics-based modeling with machine learning for lithium-ion batteries. Applied Energy , vol. 329. doi:10.1016/j.apenergy.2022.120289, 2023.

[173] Rai, R., Sahu, C.. Driven by Data or Derived through Physics? A Review of Hybrid Physics Guided Machine Learning Techniques with Cyber-Physical System (CPS) Focus. IEEE Access , vol. 8. , pp. 71050–71073. doi:10.1109/ACCESS.2020.2987324, 2020.

[174] Thelen, A., Zhang, X., Fink, O., Lu, Y., Ghosh, S., Youn, B.D., Todd, M.D., Mahadevan, S., Hu, C., Hu, Z.. A comprehensive review of digital twin — part 1: modeling and twinning enabling technologies. Structural and Multidisciplinary Optimization , vol. 65. doi:10.1007/s00158-022-03425-4, 2022.

[175] Robin, C., Gerard, M., D'Arbigny, J., Schott, P., Jabbour, L., Bultel, Y.. Development and experimental validation of a PEM fuel cell 2D-model to study heterogeneities effects along large-area cell surface. International Journal of Hydrogen Energy , vol. 40. , pp. 10211–10230. doi:10.1016/j.ijhydene.2015.05.178, 2015.

[176] Yuan, H., Dai, H., Wei, X., Ming, P.. Model-based observers for internal states estimation and control of proton exchange membrane fuel cell system: A review. Journal of Power Sources , vol. 468. doi:10.1016/j.jpowsour.2020.228376, 2020.

[177] Zhao, H., Sun, S., Jin, B.. Sequential Fault Diagnosis Based on LSTM Neural Network. IEEE Access , vol. 6. , pp. 12929–12939. doi:10.1109/ACCESS.2018.2794765, 2018.

[178] Yan, P., Abdulkadir, A., Luley, P.P., Rosenthal, M., Schatte, G.A., Grewe, B.F., Stadelmann, T.. A Comprehensive Survey of Deep Transfer Learning for Anomaly Detection in Industrial Time Series: Methods, Applications, and Directions. IEEE Access , vol. 12. , pp. 3768 – 3789. doi:10.1109/ACCESS.2023.3349132, 2024.

[179] Hu, B., Chen, Y., Keogh, E.. Time series classification under more realistic assumptions, in: Proceedings of the 2013 SIAM International Conference on Data Mining, SDM 2013, pp. 578–586. doi:10.1137/1.9781611972832.64, 2013.

[180] Weerakody, P.B., Wong, K.W., Wang, G., Ela, W.. A review of irregular time series data handling with gated recurrent neural networks. Neurocomputing , vol. 441. , pp. 161–178. doi:10.1016/J.NEUCOM.2021.02.046, 2021.

[181] Li, G., Jung, J.J.. Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. Information Fusion , vol. 91. , pp. 93–102. doi:10.1016/J.INFFUS.2022.10.008, 2023.

[182] Karim, F., Majumdar, S., Darabi, H., Harford, S.. Multivariate LSTM-FCNs for time series classification. Neural Networks , vol. 116. , pp. 237–245. doi:10.1016/j.neunet.2019.04.014, 2019.

[183] Orsenigo, C., Vercellis, C.. Combining discrete SVM and fixed cardinality warping distances for multivariate time series classification. Pattern Recognition , vol. 43. , pp. 3787–3794. doi:10.1016/J.PATCOG.2010.06.005, 2010.

[184] Zhang, C., Song, D., Chen, Y., Feng, X., Lumezanu, C., Cheng, W., Ni, J., Zong, B., Chen, H., Chawla, N.V.. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data, in: 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, pp. 1409–1416. doi:10.1609/aaai.v33i01.33011409, 2019.

[185] Richard, G., Grossin, B., Germaine, G., Hébrail, G., de Moliner, A.. Autoencoder-based time series clustering with energy applications, in: Conférence sur l'Apprentissage Automatique. doi:https://doi.org/10.48550/arXiv.2002.03624, 2018.

[186] Wang, Z., Yan, W., Oates, T.. Time series classification from scratch with deep neural networks: A strong baseline, in: Proceedings of the International Joint Conference on Neural Networks. doi:10.1109/IJCNN.2017.7966039, 2017.

[187] Nguyen, H.P., Liu, J., Zio, E.. A long-term prediction approach based on long short-term memory neural networks with automatic parameter optimization by Tree-structured Parzen Estimator and applied to time-series data of NPP steam generators. Applied Soft Computing Journal , vol. 89. doi:10.1016/j.asoc.2020.106116, 2020.

[188] Lee, J., Lee, Y., Kim, J.. Fault detection based on one-class deep learning for manufacturing applications limited to an imbalanced database. Journal of Manufacturing Systems , vol. 57. , pp. 357–366. doi:10.1016/j.jmsy.2020.10.013, 2020.

[189] Bianchi, F.M., Scardapane, S., Lokse, S., Jenssen, R.. Reservoir Computing Approaches for Representation and Classification of Multivariate Time Series. IEEE Transactions on Neural Networks and Learning Systems , vol. 32. , pp. 2169 – 2179. doi:10.1109/TNNLS.2020.3001377, 2021.

[190] Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A.. Deep learning for time series classification: a review. Data Mining and Knowledge Discovery , vol. 33. , pp. 917–963. doi:10.1007/s10618-019-00619-, 2019.

[191] Wang, J., Zhao, R., Wang, D., Yan, R., Mao, K., Shen, F.. Machine health monitoring using local feature-based gated recurrent unit networks. IEEE Transactions on Industrial Electronics , vol. 65. , pp. 1539 – 1548. doi:10.1109/TIE.2017.2733438, 2017.

[192] FIDLE: Formation Introduction au Deep Learning. URL: https://gricad-gitlab.univ-grenoble-alpes.fr/talks/fidle, .

[193] Beck, M., Pöppel, K., Spanring, M., Auer, A., Prudnikova, O., Kopp, M., Klambauer, G., Brandstetter, J., Hochreiter, S.. xLSTM: Extended Long Short-Term Memory, in: Advances in Neural Information Processing Systems. doi:10.48550/arXiv.2405.04517, 2024.

[194] Yu, Y., Si, X., Hu, C., Zhang, J.. A review of recurrent neural networks: Lstm cells and network architectures. Neural Computation , vol. 31. , pp. 1235–1270. doi:10.1162/neco_a_01199, 2019.

[195] Schuster, M., Paliwal, K.K.. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing , vol. 45. , pp. 2673–2681. doi:10.1109/78.650093, 1997.

[196] Robin, C., Gérard, M., Quinaud, M., D'Arbigny, J., Bultel, Y.. Proton exchange membrane fuel cell model for aging predictions: Simulated equivalent active surface area loss and comparisons with durability tests. Journal of Power Sources , vol. 326. , pp. 417–427. doi:10.1016/j.jpowsour.2016.07.018, 2016.

[197] Jia, F., Tian, X., Liu, F., Ye, J., Yang, C.. Oxidant starvation under various operating conditions on local and transient performance of proton exchange membrane fuel cells. Applied Energy , vol. 331. doi:10.1016/j.apenergy.2022.120412, 2023.

[198] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. URL: https://www.tensorflow.org/, 2015.

[199] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V. Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research , vol. 12. , pp. 2825–2830, 2011.

[200] Van Der Maaten, L., Hinton, G.. Visualizing data using t-SNE. Journal of Machine Learning Research , vol. 9. , pp. 2579–2625, 2008.

[201] Sobie, C., Freitas, C., Nicolai, M.. Simulation-driven machine learning: Bearing fault classification. Mechanical Systems and Signal Processing , vol. 99. , pp. 403–419. doi:10.1016/j.ymssp.2017.06.025, 2018.

[202] Liu, C., Gryllias, K.. Simulation-Driven Domain Adaptation for Rolling Element Bearing Fault Diagnosis. IEEE Transactions on Industrial Informatics , vol. 18. , pp. 5760–5770. doi:10.1109/TII.2021.3103412, 2022.

[203] Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D., Erhan, D.. Domain separation networks, in: Advances in Neural Information Processing Systems, pp. 343–351. doi:10.48550/arXiv.1608.06019, 2016.

[204] Wang, Q., Taal, C., Fink, O.. Integrating Expert Knowledge with Domain Adaptation for Unsupervised Fault Diagnosis. IEEE Transactions on Instrumentation and Measurement , vol. 71. doi:10.1109/TIM.2021.3127654, 2022.

[205] Kumar, A., Fu, Z., Pathak, D., Malik, J.. RMA: Rapid Motor Adaptation for Legged Robots, in: Robotics: Science and Systems. doi:10.15607/RSS.2021.XVII.011, 2021.

[206] Du, Y., Watkins, O., Darrell, T., Abbeel, P., Pathak, D.. Auto-Tuned Sim-to-Real Transfer, in: Proceedings - IEEE International Conference on Robotics and Automation, pp. 1290–1296. doi:10.1109/ICRA48506.2021.9562091, 2021.

[207] Gao, S., Yu, Z., Zhang, Z., Feng, C., Yan, T., He, B., Zio, E.. Physical model embedding-based generative adversarial networks for unsupervised fault detection of underwater thrusters. Ocean Engineering , vol. 298. doi:10.1016/j.oceaneng.2024.117190, 2024.

[208] Xiong, J., Fink, O., Zhou, J., Ma, Y.. Controlled physics-informed data generation for deep learning-based remaining useful life prediction under unseen operation conditions. Mechanical Systems and Signal Processing , vol. 197. doi:10.1016/j.ymssp.2023.110359, 2023.

[209] Hospedales, T., Antoniou, A., Micaelli, P., Storkey, A.. Meta-Learning in Neural Networks: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence , vol. 44. , pp. 5149–5169. doi:10.1109/TPAMI.2021.3079209, 2022.

[210] Weiss, K., Khoshgoftaar, T.M., Wang, D.. A survey of transfer learning. Journal of Big Data , vol. 3. , pp. 9. doi:10.1186/s40537-016-0043-6, 2016.

[211] Weber, M., Auch, M., Doblander, C., Mandl, P., Jacobsen, H.A.. Transfer Learning with Time Series Data: A Systematic Mapping Study. IEEE Access , vol. 9. , pp. 165409 – 165432. doi:10.1109/ACCESS.2021.3134628, 2021.

[212] Storkey, A.. When training and test sets are different: Characterizing learning transfer, in: Dataset Shift in Machine Learning. The MIT Press. doi:10.7551/mitpress/7921.003.0004, 2008.

[213] Ragab, M., Eldele, E., Tan, W.L., Foo, C.S., Chen, Z., Wu, M., Kwoh, C.K., Li, X.. ADATIME: A Benchmarking Suite for Domain Adaptation on Time Series Data. ACM Transactions on Knowledge Discovery from Data , vol. 17. doi:10.1145/3587937, 2023.

[214] Wang, Q., Michau, G., Fink, O.. Domain Adaptive Transfer Learning for Fault Diagnosis, in: Proceedings - 2019 Prognostics and System Health Management Conference, PHM-Paris 2019, pp. 279–285. doi:10.1109/PHM-Paris.2019.00054, 2019.

[215] Zhou, S., Lu, Y., Bao, D.. Fault diagnosis of PEMFC systems based on an auxiliary transfer network. International Journal of Hydrogen Energy , vol. 48. , pp. 19262–19278. doi:10.1016/J.IJHYDENE.2023.01.334, 2023.

[216] Zhao, C., Zio, E., Shen, W.. Domain generalization for cross-domain fault diagnosis: An application-oriented perspective and a benchmark study. Reliability Engineering & System Safety , vol. 245. , pp. 109964. doi:10.1016/J.RESS.2024.109964, 2024.

[217] Chen, Z., He, G., Li, J., Liao, Y., Gryllias, K., Li, W.. Domain Adversarial Transfer Network for Cross-Domain Fault Diagnosis of Rotary Machinery. IEEE Transactions on Instrumentation and Measurement , vol. 69. , pp. 8702–8712. doi:10.1109/TIM.2020.2995441, 2020.

[218] Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Vaughan, J.. A theory of learning from different domains. Machine Learning , vol. 79. , pp. 151–175. doi:10.1007/s10994-009-5152-4, 2010.

[219] Wang, Q., Fink, O., Van Gool, L., Dai, D.. Continual Test-Time Domain Adaptation, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 7191 – 7201. doi:10.1109/CVPR52688.2022.00706, 2022.

[220] Zhao, Z., Zhang, Q., Yu, X., Sun, C., Wang, S., Yan, R., Chen, X.. Applications of Unsupervised Deep Transfer Learning to Intelligent Fault Diagnosis: A Survey and Comparative Study. IEEE Transactions on Instrumentation and Measurement , vol. 70. doi:10.1109/TIM.2021.3116309, 2021.

[221] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.. Domain-adversarial training of neural networks. Journal of Machine Learning Research , vol. 17. doi:10.48550/arXiv.1505.07818, 2016.

[222] Zhang, W., Li, X., Ma, H., Luo, Z., Li, X.. Universal Domain Adaptation in Fault Diagnostics With Hybrid Weighted Deep Adversarial Learning. IEEE Transactions on Industrial Informatics , vol. 17. , pp. 7957 – 7967. doi:10.1109/TII.2021.3064377, 2021.

[223] He, Q.Q., Siu, S.W.I., Si, Y.W.. Attentive recurrent adversarial domain adaptation with Top-k pseudo-labeling for time series classification. Applied Intelligence , vol. 53. , pp. 13110 – 13129. doi:10.1007/s10489-022-04176-x, 2023.

[224] Tonutti, M., Ruffaldi, E., Cattaneo, A., Avizzano, C.A.. Robust and Subject-Independent Driving Manoeuvre Anticipation through Domain-Adversarial Recurrent Neural Networks. Robotics and Autonomous Systems , vol. 115. doi:10.1016/j.robot.2019.02.007, 2019.

[225] Nejjar, I., Geissmann, F., Zhao, M., Taal, C., Fink, O.. Domain adaptation via alignment of operation profile for Remaining Useful Lifetime prediction. Reliability Engineering & System Safety , vol. 242. , pp. 109718. doi:10.1016/J.RESS.2023.109718, arXiv:2302.01704, 2024.

[226] Wang, B., Baraldi, P., Zio, E.. Deep Multiadversarial Conditional Domain Adaptation Networks for Fault Diagnostics of Industrial Equipment. IEEE Transactions on Industrial Informatics , vol. 19. , pp. 8841–8851. doi:10.1109/TII.2022.3222400, 2023.

[227] Lahiri, A., Maliakal, G., Klasky, M.L., Fessler, J.A., Ravishankar, S.. Sparse-View Cone Beam CT Reconstruction Using Data-Consistent Supervised and Adversarial Learning From Scarce Training Data. IEEE Transactions on Computational Imaging , vol. 9. , pp. 13 – 28. doi:10.1109/TCI.2022.3225680, 2023.

[228] Kirchmeyer, M., Rakotomamonjy, A., de Bézenac, E., Gallinari, P.. Mapping conditional distributions for domain adaptation under generalized target shift, in: ICLR 2022 - 10th International Conference on Learning Representations. doi:10.48550/arXiv.2110.15057, 2022.

[229] She, D., Jia, M., Pecht, M.G.. Weighted Entropy Minimization Based Deep Conditional Adversarial Diagnosis Approach under Variable Working Conditions. IEEE/ASME Transactions on Mechatronics , vol. 26. , pp. 2440 – 2450. doi:10.1109/TMECH.2020.3040175, 2021.

[230] Li, X., Zhang, W., Ma, H., Luo, Z., Li, X.. Partial transfer learning in machinery cross-domain fault diagnostics using class-weighted adversarial networks. Neural Networks , vol. 129. , pp. 313–322. doi:10.1016/J.NEUNET.2020.06.014, 2020.

[231] Rombach, K., Michau, G., Fink, O.. Controlled generation of unseen faults for Partial and Open-Partial domain adaptation. Reliability Engineering and System Safety , vol. 230. doi:10.1016/j.ress.2022.108857, 2023.

[232] Sicilia, A., Zhao, X., Hwang, S.J.. Domain adversarial neural networks for domain generalization: when it works and how to improve. Machine Learning , vol. 112. doi:10.1007/s10994-023-06324-x, 2023.

[233] Rombach, K., Michau, G., Ratnasabapathy, K., Ancu, L.S., Burzle, W., Koller, S., Fink, O.. Contrastive Feature Learning for Railway Infrastructure Fault Diagnostic, in: Proceedings of the 32nd European Safety and Reliability Conference, ESREL 2022 - Understanding and Managing Risk and Reliability for a Sustainable Future, pp. 1875–1881. doi:10.3850/978-981-18-5183-4_S02-07-645-cd, 2022.

[234] Wilson, G., Doppa, J.R., Cook, D.J.. CALDA: Improving Multi-Source Time Series Domain Adaptation With Contrastive Adversarial Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence , vol. 45. , pp. 14208 – 14221. doi:10.1109/TPAMI.2023.3298346, 2023.

[235] Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.. Optuna: A Next-generation Hyperparameter Optimization Framework, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2623–2631. doi:10.1145/3292500.3330701, 2019.

[236] Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.. Tent: Fully Test-time Adaptation by Entropy Minimization, in: ICLR 2021 - 9th International Conference on Learning Representations. doi:10.48550/arXiv.2006.10726, 2021.

[237] Crawshaw, M.. Multi-Task Learning with Deep Neural Networks: A Survey arXiv:arXiv preprint arXiv:2009.09796, 2020.

[238] Dong, H., Chatzi, E., Fink, O., 2025. Towards Multimodal Open-Set Domain Generalization and Adaptation Through Self-supervision. volume 15117 LNCS. doi:10.1007/978-3-031-73202-7_16.

[239] Kim, T., Sohn, S.. Multitask learning for health condition identification and remaining useful life prediction: deep convolutional neural network approach. Journal of Intelligent Manufacturing , vol. 32. , pp. 2169–2179. doi:10.1007/s10845-020-01630-w, 2021.

[240] Liu, P., Qiu, X., Huang, X.. Adversarial Multi-task Learning for Text Classification, in: ACL 2017 - 55th Annual Meeting of the Association for Computational Linguisticsx, pp. 1–10. doi:10.18653/v1/P17-1001, 2017.

[241] Che, Y., Forest, F., Zheng, Y., Xu, L., Teodorescu, R.. Health Prediction for Lithium-Ion Batteries under Unseen Working Conditions. IEEE Transactions on Industrial Electronics , vol. 71. , pp. 14254–14264. doi:10.1109/TIE.2024.3379664, 2024.

[242] Dersin, P., Goglio, D., Bajarunas, K., Chao, M.A.. Analytical health indices: Towards reliability-informed deep learning for phm. International Journal of Prognostics and Health Management , vol. 16. doi:10.36001/ijphm.2025.v16i2.4262, 2025.

[243] Baraldi, P., Bonfanti, G., Zio, E.. Differential evolution-based multi-objective optimization for the definition of a health indicator for fault diagnostics and prognostics. Mechanical Systems and Signal Processing , vol. 102. , pp. 382 – 400. doi:10.1016/j.ymssp.2017.09.013, 2018.

[244] Vichard, L., Harel, F., Ravey, A., Venet, P., Hissel, D.. Degradation prediction of PEM fuel cell based on artificial intelligence. International Journal of Hydrogen Energy , vol. 45. , pp. 14953–14963. doi:10.1016/j.ijhydene.2020.03.209, 2020.

[245] Nagulapati, V.M., Kumar, S.S., Annadurai, V., Lim, H.. Machine learning based fault detection and state of health estimation of proton exchange membrane fuel cells. Energy and AI , vol. 12. , pp. 100237. doi:10.1016/J.EGYAI.2023.100237, 2023.

[246] Pahon, E., Bouquain, D., Hissel, D., Rouet, A., Vacquier, C.. Performance analysis of proton exchange membrane fuel cell in automotive applications. Journal of Power Sources , vol. 510. doi:10.1016/j.jpowsour.2021.230385, 2021.

[247] Li, J., Zhao, D., Ma, X., Xie, L., Zhou, Z., Zhang, L., Chen, Q.. Health status assessment for proton exchange membrane fuel cell by polarization processes identification utilizing relaxation time analysis. International Journal of Hydrogen Energy , vol. 128. , pp. 131–145. doi:10.1016/j.ijhydene.2025.03.422, 2025.

[248] Hua, Z., Zheng, Z., Pahon, E., Pera, M.C., Gao, F.. Multi-Timescale Lifespan Prediction for PEMFC Systems under Dynamic Operating Conditions. IEEE Transactions on Transportation Electrification , vol. 8. , pp. 345–355. doi:10.1109/TTE.2021.3103921, 2022.

[249] Wang, C., Li, Z., Outbib, R., Dou, M., Zhao, D.. Symbolic deep learning based prognostics for dynamic operating proton exchange membrane fuel cells. Applied Energy , vol. 305. , pp. 117918. doi:10.1016/J.APENERGY.2021.117918, 2022.

[250] Morando, S., Jemei, S., Hissel, D., Gouriveau, R., Zerhouni, N.. Proton exchange membrane fuel cell ageing forecasting algorithm based on Echo State Network. International Journal of Hydrogen Energy , vol. 42. , pp. 1472–1480. doi:10.1016/j.ijhydene.2016.05.286, 2017.

[251] Mezzi, R., Yousfi-Steiner, N., Péra, M., Hissel, D., Larger, L.. An Echo State Network for fuel cell lifetime prediction under a dynamic micro-cogeneration load profile. Applied Energy , vol. 283. doi:10.1016/j.apenergy.2020.116297, 2021.

[252] Ma, R., Li, Z., Breaz, E., Liu, C., Bai, H., Briois, P., Gao, F.. Data-fusion prognostics of proton exchange membrane fuel cell degradation. IEEE Transactions on Industry Applications , vol. 55. , pp. 4321–4331. doi:10.1109/TIA.2019.2911846, 2019.

[253] He, K., Zhang, X., Ren, S., Sun, J.. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, in: EEE International Conference on Computer Vision (ICCV). doi:10.48550/arXiv.1502.01852, 2015.

[254] Gawlikowski, J., Tassi, C.R., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., Shahzad, M., Yang, W., Bamler, R., Zhu, X.X.. A survey of uncertainty in deep neural networks. Artif Intell Rev , vol. 56 (Suppl. , pp. 1513–1589. doi:https://doi.org/10.1007/s10462-023-10562-9, 2023.