Centre Eau Terre et Environnement

TECHNIQUES D'INTELLIGENCE ARTIFICIELLE DANS LA MODÉLISATION DU PROCESSUS ÉLECTROCHIMIQUE POUR LE TRAITEMENT DES EAUX RÉSIDUAIRES

Par

Majid Gholami Shirkoohi

Thèse présentée pour l'obtention du grade de
Philosophiae Doctor (Ph.D.)
en sciences de l'eau

Jury d'évaluation

Président du jury et examinateur interne	Erwan Gloaguen INRS-ETE
Examinateur externe	Olivier Patrick Lefebvre National University of Singapore
Examinateur externe	Nicolás Peleato University of British Columbia Okanagan
Directeur de recherche	Patrick Drogui INRS-ETE
Codirecteur de recherche	Peter Vanrolleghem Université Laval

ACKNOWLEDGEMENTS

First, I am grateful to my supervisor, Prof. Patrick Drogui and co-supervisor, Prof. Peter Vanrolleghem, for their invaluable academic guidance, support, and encouragement during my Ph.D. study. Their trust enabled me to undertake this journey and pursue my project, interests, and goals. I would like to thank them for the discussions, comments, and suggestions in each step of my doctoral program.

I thank all my colleagues, current and former group members, in Prof. Drogui's and Prof. Vanrolleghem's groups for their technical assistance and accompanying. I express my gratitude to the INRS-ETE staff for all the technical and administrative help I got throughout my studies.

I am also grateful to the Natural Science and Engineering Research Council of Canada (NSERC) and CREATE TEDGIEER program for the financial support for this study.

Also, I want to thank the members of the jury who kindly evaluated this thesis.

Finally, words cannot express my deepest gratitude to my wife, Mahsa, and my family for their tremendous help, support, understanding, and encouragement throughout my doctoral program.

ABSTRACT

Electrochemical technologies have been known and utilized for the treatment of wastewater containing recalcitrant organic pollutants because conventional treatments such as physico-chemical or biological methods are not able to completely degrade them. However, to make these technologies competitive with the conventional technologies that are in use today (e.g., coagulation or advanced oxidation processes), proper design of the processes and operating conditions through process modelling and optimization are necessary. Electrochemical processes are complicated nonlinear processes, making it difficult to describe the process behaviour using phenomenological or conventional empirical modelling methodologies. In this work, artificial intelligence techniques, including artificial neural networks (ANNs), adaptive neuro-fuzzy inference systems (ANFIS), support vector regression (SVR), genetic algorithms (GA), and particle swarm optimization (PSO) have been used as alternatives for modelling and optimization of the electrochemical processes.

In the first part of this thesis, the development of ANN models and multi-objective optimization based on a genetic algorithm was carried out for active chlorine production using the electrolysis process. In order to diagnose and prevent the over-fitting problem during the learning process of the ANN models, learning curves and the regularization factor were utilized. The results showed that the trained ANN models were able to successfully predict the active chlorine production and energy consumption of the process (R^2=0.979 and MSE=3.826 for active chlorine production and R^2=0.985 and MSE=6.952 for energy consumption). Multi-objective optimization for maximizing active chlorine production and minimizing energy consumption was carried out by a non-dominated sorting genetic algorithm (NSGA-II) using the best derived ANN models. The Pareto front obtained led to multiple non-dominated optimal points, which resulted in insights regarding the optimal operating conditions for the process.

In the second part of this study, the ANFIS modelling approach was applied as another AI technique and compared with the response surface methodology (RSM) for modelling and optimization of psychoactive pharmaceutical caffeine removal by electrochemical oxidation. Results showed that the anode type, followed by the electrolysis time are the

most important variables affecting caffeine degradation. While both RSM and ANFIS models were able to successfully predict the electrochemical process behaviour, ANFIS models performed slightly better (R^2=0.993, RMSE=2.694, MAPE=6.582 for caffeine removal efficiency, and R^2=0.976, RMSE=0.261, MAPE=9.221 for energy consumption). The optimal conditions were applied to the treatment of real municipal wastewater effluent with caffeine removal efficiency varying between 78.0±4.3% and 92.5±1.0% for different initial caffeine concentrations, showing the effectiveness of the process. Furthermore, it was demonstrated that toxicity generated by the electrooxidation process could be reduced by extending the electrolysis time or could be completely eliminated using granular activated carbon.

The third part of this thesis dealt with hyperparameter optimization of AI models by integrating metaheuristic algorithms such as GA and PSO to predict the removal efficiency of phosphate from wastewaters using the electrocoagulation process. To tackle the relatively low number of sample data available from an experimental electrochemical process and increase the reliability of data-driven models, the proposed hybrid models were built on repeated random sub-sampling validation (10 data subsets) instead of a single split approach. The performance comparison of models showed that the effectiveness of the data-driven models depends on how the data is distributed over the training, validation, and test sets. The ANFIS and hybrid SVR models were more sensitive than the hybrid ANN models to the distribution of data points. The hybrid ANN models showed greater accuracy than the ANFIS and hybrid SVR models when compared using different performance criteria. The hybrid ANN models showed less dispersed performance for the different test sub-datasets. Remarkably, PSO-ANN models illustrated exceptional generalization performance for the 10 data subsets examined.

Efforts made in each part of this thesis provide insight into the application and adaptation of AI techniques for modelling and optimization of electrochemical processes for water treatment. It was shown that AI models could be successfully applied to electrochemical processes despite the relatively low number of data points available, but the reliability and robustness of these models should be taken into account.

RÉSUMÉ

Les technologies électrochimiques sont connues et utilisées pour le traitement des eaux usées contenant des polluants organiques récalcitrants car les traitements conventionnels tels que les méthodes physico-chimiques ou biologiques ne sont pas capables de les dégrader complètement. Cependant, pour rendre ces technologies compétitives par rapport aux technologies conventionnelles utilisées aujourd'hui (par exemple, la coagulation ou les procédés d'oxydation avancée), il est nécessaire de concevoir correctement les procédés et les conditions d'exploitation par le biais de la modélisation et de l'optimisation des procédés. Les processus électrochimiques sont des processus non linéaires compliqués, ce qui rend difficile la description du comportement du processus à l'aide de méthodologies de modélisation phénoménologiques ou empiriques conventionnelles. Dans ce travail, les techniques d'intelligence artificielle, incluant les réseaux de neurones artificiels (RNA), les systèmes à inférences floues à réseaux adaptatifs (SIFRA), la régression des vecteurs de support (RVS), les algorithmes génétiques (AG) et l'optimisation par essaims particulaires (OEP) ont été utilisées comme alternatives pour la modélisation et l'optimisation des procédés électrochimiques.

Dans la première partie de cette thèse, le développement des modèles RNA et l'optimisation multi-objectifs basée sur un algorithme génétique ont été réalisés pour la production de chlore actif en utilisant le procédé d'électrolyse. Afin de diagnostiquer et de prévenir le problème de sur-ajustement pendant le processus d'apprentissage des modèles RNA, des courbes d'apprentissage et le facteur de régularisation ont été utilisés. Les résultats ont montré que les modèles RNA formés étaient capables de prédire avec succès la production de chlore actif et la consommation d'énergie du processus (R^2=0,979 et MSE=3,826 pour la production de chlore actif et R^2=0,985 et MSE=6,952 pour la consommation d'énergie). L'optimisation multi-objectifs pour maximiser la production de chlore actif et minimiser la consommation d'énergie a été réalisée par un algorithme génétique de tri non dominé (NSGA-II) en utilisant les meilleurs modèles RNA dérivés. Le front de Pareto obtenu a conduit à de multiples points optimaux non dominés, qui donnent des indications sur les conditions de fonctionnement optimales du processus.

Dans la deuxième partie de cette étude, l'approche de modélisation SIFRA, une autre technique d'intelligence artificielle, a été appliquée et comparée à la méthodologie de surfaces de réponses (MSR) pour modéliser et optimiser le processus électrochimique de dégradation de la caféine. La caféine est un produit pharmaceutique psychoactif souvent présent dans les eaux résiduaires. Les résultats ont montré que le type d'anode, suivi du temps d'électrolyse sont les variables les plus importantes affectant la dégradation de la caféine. Les modèles MSR et SIFRA ont permis de prédire avec succès le comportement du processus électrochimique, mais les modèles SIFRA sont légèrement plus performants (R^2=0,993, RMSE=2,694, MAPE=6,582 pour l'efficacité de dégradation de la caféine, et R^2=0,976, RMSE=0,261, MAPE=9,221 pour la consommation d'énergie). Les conditions optimales ont été appliquées sur des effluents d'eaux usées municipales réelles, dans lesquelles l'efficacité de dégradation de la caféine a varié entre 78,0±4,3% et 92,5±1,0% pour différentes concentrations initiales de caféine, ce qui montre l'efficacité du processus. La toxicité générée par le procédé d'électro-oxydation pouvait être réduite en prolongeant le temps d'électrolyse ou encore en utilisant du charbon actif en grains suite à l'application du traitement électrolytique.

La troisième partie de cette thèse traitait de l'optimisation des hyperparamètres des modèles d'IA en intégrant des algorithmes métaheuristiques tels que l'AG et l'OEP pour prédire l'efficacité de l'élimination du phosphate des eaux usées en utilisant le processus d'électrocoagulation. Pour faire face au nombre relativement faible de données d'échantillons disponibles dans le processus électrochimique et augmenter la fiabilité des modèles basés sur les données, les modèles hybrides proposés ont été construits sur une 'repeated random sub-sampling validation' (10 sous-ensembles de données) au lieu d'une approche à répartition unique. La comparaison des performances des modèles a montré que l'efficacité des modèles pilotés par les données dépend de la manière dont les données sont distribuées parmi les ensembles de formation, de validation et des tests. Les modèles SIFRA et RVS hybrides étaient plus sensibles que les modèles RNA hybrides à la distribution des points de données. Les modèles RNA hybrides ont montré une plus grande précision que les modèles SIFRA et RVS hybrides auxquels ils ont été comparés en utilisant différents critères de performance. Ces modèles indiquent une performance moins dispersée pour les ensembles de tests pour les différents sous-

ensembles de données. De manière remarquable, les modèles OEP-RNA ont illustré une performance de généralisation exceptionnelle pour les 10 sous-ensembles de données examinées.

Les efforts réalisés dans chaque volet de cette thèse donnent un aperçu de l'application et de l'adaptation des techniques d'IA pour la modélisation et l'optimisation des processus électrochimiques pour le traitement de l'eau et des eaux usées. Il a été démontré que les modèles d'IA pouvaient être appliqués avec succès aux processus électrochimiques malgré le nombre relativement faible de points de données disponibles, mais la fiabilité et la robustesse de ces modèles doivent être prises en compte.

SYNOPSIS

Introduction

Avec la croissance rapide de la population mondiale et l'industrialisation intense du 20ème siècle, la pollution environnementale est devenue un problème mondial avec des impacts négatifs sur le secteur de l'eau. La grande majorité de ces problèmes de pollution sont causés par des composés organiques persistants en raison de leur résistance aux traitements conventionnels tels que les méthodes physico-chimiques ou biologiques. Il en résulte la détection de polluants réfractaires tels que les pesticides, les composés phénoliques, les colorants synthétiques, les composés halogénés, les hydrocarbures aromatiques polycycliques (HAP), les polychlorobiphényles (PCB), les perturbateurs endocriniens (PE) et autres dans les rivières, les lacs, les océans et même les eaux potables du monde entier. Ils peuvent avoir des effets dangereux sur la santé des organismes vivants, y compris des êtres humains. Par conséquent, le traitement avancé de l'eau et des eaux usées est devenu une préoccupation sociale, politique et environnementale majeure (Drogui *et al.*, 2007; Moreira *et al.*, 2017; Zheng *et al.*, 2017).

Ces dernières années, les procédés électrochimiques ont suscité un intérêt croissant en tant que méthode alternative pour le traitement de l'eau et des eaux usées. Ces procédés sont considérés comme des technologies écologiques et vertes car le principal réactif impliqué, l'électron, est considéré comme un réactif propre et tire parti de la chimie de couplage (génération in situ d'oxydant) avec la science électronique (transfert d'électrons). Parmi les autres avantages intéressants, citons la polyvalence, le rendement énergétique élevé, la possibilité d'automatisation et la rentabilité. (Feng *et al.*, 2016; Rajeshwar *et al.*, 1997). Plusieurs publications portant sur différentes méthodes électrochimiques telles que l'électrooxydation, l'électrocoagulation, l'électroflottation, l'électro-Fenton et l'électrodialyse ont été publiées au cours de la dernière décennie pour améliorer les performances de traitement des eaux usées et des eaux potables (Daghrir *et al.*, 2013; Martín de Vidales *et al.*, 2012b; Olvera-Vargas *et al.*, 2015; Zhang *et al.*, 2011).

La modélisation des processus est une exigence pour l'optimisation des processus. Les approches de modélisation phénoménologique et empirique sont généralement utilisées pour les processus de traitement de l'eau et des eaux usées. Bien que la modélisation phénoménologique fournisse des indications précieuses sur le comportement du processus et ait la capacité d'extrapoler, les phénomènes de transport de chaleur et de masse ainsi qu'une connaissance détaillée de la cinétique de la réaction sont nécessaires. Dans la modélisation empirique, la structure du modèle d'adaptation aux données doit être spécifiée a priori, ce qui constitue un défi car il faut choisir le modèle approprié parmi les nombreux modèles disponibles, en particulier pour les processus non linéaires (Nandi *et al.*, 2004). Les processus électrochimiques pour le traitement de l'eau et des eaux usées sont des systèmes non linéaires très compliqués en raison des relations complexes entre les paramètres d'entrée et les sorties. Il est donc difficile d'utiliser des modèles phénoménologiques ou empiriques pour modéliser, simuler et optimiser les processus. Les techniques d'intelligence artificielle telles que les réseaux neuronaux artificiels (RNA), le système d'inférence neuro-floue adaptatif (SIFRA), la régression par vecteur de support (RVS) ainsi que les algorithmes génétiques (AG) et les méthodes d'optimisation par essaim de particules (OEP) sont apparues comme des approches alternatives attrayantes pour la modélisation et l'optimisation de ces processus non linéaires lorsque les modèles de régression phénoménologiques ou conventionnels ne sont pas pratiques (Curteanu *et al.*, 2014).

Dans ce travail, les applications des techniques d'intelligence artificielle axées sur leur fiabilité et leur validation pour la modélisation des processus électrochimiques pour les procédés de traitement de l'eau et des eaux usées seront discutées. À ce sujet, une comparaison sera faite entre les modèles d'IA et les approches de modélisation conventionnelles (par exemple, la méthode des surfaces de réponses (MSR)). En outre, les algorithmes métaheuristiques seront liés aux modèles d'intelligence artificielle, soit à des fins d'optimisation multi-objectifs, soit pour optimiser les hyperparamètres des modèles d'intelligence artificielle afin d'accroître leurs performances.

Problématique

Les procédés électrochimiques ont suscité un intérêt croissant en tant que méthode alternative pour le traitement des eaux polluées. Cependant, pour rendre ces technologies compétitives par rapport aux technologies conventionnelles utilisées aujourd'hui, une conception adéquate des procédés et des conditions de fonctionnement est nécessaire. Les améliorations nécessaires des performances et du fonctionnement des procédés électrochimiques nécessitent une meilleure modélisation en vue de la conception et de l'optimisation. La modélisation des procédés est importante pour aider à améliorer la conception et à réduire les coûts d'équipements et d'exploitation. Elle peut fournir des solutions aux défis de traitement électrochimique des eaux, et ainsi permettre de prédire les performances des procédés dans une large gamme de conditions de fonctionnement. La modélisation des processus est une condition préalable à leur optimisation. Les approches de modélisation phénoménologique et empirique sont généralement utilisées dans les processus de traitement des eaux. La modélisation phénoménologique fournit des informations précieuses sur le comportement du processus et permet l'extrapolation. D'autre part, les phénomènes de transport de chaleur et de masse ainsi qu'une connaissance détaillée de la cinétique de la réaction sont nécessaires pour les modèles phénoménologiques des procédés électrochimiques. La collecte de ces données est une tâche difficile à obtenir pour un système multi-variable, en particulier lorsqu'une connaissance limitée de celui-ci est disponible. De plus, le comportement non linéaire des procédés électrochimiques donne lieu à des modèles non linéaires complexes qui, dans la plupart des cas, ne se prêtent pas à des solutions analytiques ; leur résolution nécessite donc des méthodes numériques à forte intensité de calcul.

La modélisation empirique (régression) est l'une des approches de modélisation alternatives à la modélisation phénoménologique. Dans la plupart des cas, un modèle de régression linéaire quadratique sera choisi, mais il n'est souvent pas adéquat pour décrire les non-linéarités des systèmes.

Les processus électrochimiques dans le traitement des eaux sont des systèmes non linéaires très compliqués en raison des relations complexes entre les paramètres d'entrée

et les sorties. Par conséquent, l'utilisation de modèles phénoménologiques ou empiriques classiques pour modéliser, simuler et optimiser les processus n'est pas toujours la meilleure option. Dans ces situations, les techniques d'intelligence artificielle basées par exemple sur des réseaux neuronaux artificiels, constituent une méthode alternative intéressante. Ils pourraient surmonter les difficultés de modélisation classiques, grâce aux avantages suivants : la possibilité de les appliquer même à des processus non linéaires complexes à entrées et sorties multiples, la capacité de les construire uniquement à partir de données historiques d'entrées et de sorties de processus (ensemble de données expérimentales), et une excellente capacité de généralisation lorsqu'ils sont correctement formés. D'autre part, les inconvénients pourraient portés sur l'obtention d'un modèle fiable et robuste avec un nombre limité de données expérimentales disponibles. Par conséquent, l'objectif de cette étude est d'appliquer les techniques d'intelligence artificielle en mettant l'accent sur leur fiabilité et leur robustesse, pour modéliser et optimiser les processus électrochimiques pour le traitement des eaux résiduaires. Cela comprend la comparaison des modèles d'IA avec les approches de modélisation conventionnelles (par exemple, le MSR), l'association d'algorithmes métaheuristiques aux modèles d'IA pour les processus multi-objectifs, et l'optimisation des hyperparamètres des modèles d'IA pour augmenter leurs performances.

Hypothèses

Selon l'énoncé de la problématique, la présente étude comprend les hypothèses suivantes:

Hypothèse 1: Les RNA sont l'une des méthodes d'intelligence artificielle qui pourraient être utilisées efficacement pour modéliser les processus électrochimiques pour le traitement des eaux usées.

Les réseaux neuronaux artificiels présentent des avantages considérables par rapport aux autres méthodes empiriques, ce qui les rend pratiques pour la modélisation des processus électrochimiques. Tout d'abord, les RNA ont des connexions parallèles massives entre les nœuds. Chaque nœud fonctionne indépendamment des autres et a un léger effet sur le modèle d'entrée-sortie. Cette caractéristique de parallélisme permet aux ANN d'obtenir de meilleures performances que les modèles empiriques, en particulier

dans les processus électrochimiques non linéaires et complexes. En outre, les RNA peuvent être entraînés plusieurs fois, ce qui signifie que lorsque les performances du réseau sont insuffisantes en raison d'un changement des conditions, les RNA peuvent être entraînés à nouveau avec de nouvelles conditions pour améliorer leurs performances. Cette caractéristique adaptative n'est pas courante dans les modèles empiriques. La caractéristique MIMO (entrées et sorties multiples) des RNA est un autre aspect qui en fait une option appropriée pour la modélisation des processus électrochimiques. Les RNA peuvent mettre en correspondance de nombreuses variables indépendantes avec autant de variables dépendantes que nécessaire, ce qui constitue une caractéristique unique par rapport aux autres modèles empiriques. En outre, contrairement à la MSR, la conception expérimentale et la sélection a priori de la structure du modèle d'adaptation aux données ne sont pas une condition préalable pour les RNA. Cela est bénéfique, en particulier pour les processus non linéaires où le choix d'un modèle approprié parmi les nombreux modèles disponibles constitue un défi.

Hypothèse 2: Les algorithmes évolutionnaires liés à des modèles d'intelligence artificielle sont plus performants que d'autres approches (par exemple, la MSR) pour l'optimisation multi-objectifs.

Les algorithmes évolutionnaires liés à des modèles d'intelligence artificielle constituent une approche utile pour les processus non linéaires complexes à entrées multiples et à sorties multiples. Il n'existe pas de solution unique à un problème d'optimisation multi objectifs, mais un ensemble de solutions mathématiquement aussi bonnes, connues sous le nom de solutions optimales non dominées (ou Pareto). À cet égard, des algorithmes évolutionnaires très robustes comme les AG peuvent être associés à des modèles d'IA pour les problèmes d'optimisation multi-objectifs. Les AG peuvent surpasser les méthodes d'optimisation conventionnelles car ils n'ont pas besoin que la fonction "objectif" soit continue et/ou différentiable, ils ne nécessitent pas une formulation approfondie du problème et ils ne sont pas sensibles au point de départ.

Hypothèse 3: L'optimisation des hyperparamètres des modèles d'IA permettra

d'améliorer les performances de modélisation des processus électrochimiques pour le traitement des eaux.

Les performances des modèles d'intelligence artificielle dépendent en grande partie de l'optimisation des hyperparamètres intégrés qui contrôlent le processus d'apprentissage. Par exemple, dans le cas des modèles RNA, ces hyperparamètres comprennent le taux d'apprentissage, le nombre d'époques, l'échec de validation maximum, le facteur de régularisation, le nombre de couches cachées et de neurones cachés, et les fonctions de transfert.

Objectifs

L'objectif global de cette étude est d'appliquer et d'optimiser les modèles d'intelligence artificielle pour la modélisation et l'optimisation des processus électrochimiques pour le traitement des eaux. En d'autres termes, cette recherche vise à démontrer et à comparer la capacité et la performance des modèles d'IA optimisés avec d'autres types d'approches conventionnelles de modélisation et d'optimisation utilisées dans les procédés électrochimiques. Les objectifs spécifiques de cette étude sont les suivants :

Objectif 1: Comparaison des modèles d'IA avec d'autres approches de modélisation conventionnelles, telles que le MSR, utilisées pour la modélisation et l'optimisation des systèmes électrochimiques.

La méthodologie de la surface de réponse a été largement utilisée pour étudier l'effet des différentes variables sur la réponse. Elle permet de minimiser le nombre d'expériences tout en obtenant une réponse optimale à l'aide d'expériences bien conçues. Malgré les avantages du MSR, il arrive que l'impact de divers paramètres opérationnels dans des processus électrochimiques non linéaires complexes ne puisse pas être entièrement défini par une simple corrélation linéaire multivariée. Pour cette raison, l'un des principaux objectifs de cette étude est de comparer les performances des approches de modélisation IA et MSR.

De même, les algorithmes évolutionnaires tels que l'AG seront utilisés dans cette étude pour l'optimisation multi-objectifs des procédés électrochimiques pour le traitement des eaux usées. Pour les problèmes d'optimisation multi-objectifs, il n'y a pas de solution

unique, ce qui est typiquement obtenu par la MSR. Au lieu de cela, un ensemble de solutions mathématiquement aussi bonnes, connues sous le nom de solutions optimales de Pareto, sont obtenues. Dans cette étude, les algorithmes d'optimisation évolutionnaire seront liés aux modèles RNA afin d'améliorer les conditions opérationnelles puisque les procédés électrochimiques nécessitent l'introduction d'énergie et un temps d'expérimentation minimum pour atteindre une bonne efficacité du procédé. Cette approche peut optimiser les conditions expérimentales en améliorant les efficacités d'élimination/production et en réduisant les coûts énergétiques.

Objectif 2: Modélisation et optimisation de la dégradation de la caféine par oxydation électrochimique.

La SIFRA, une autre technique de modélisation de l'IA, est utilisée dans cette partie de la thèse pour la dégradation de la caféine par oxydation électrochimique. La caféine est choisie comme polluant modèle pharmaceutique psychoactif pour le processus d'oxydation électrochimique. La dégradation de la caféine par oxydation électrochimique sera étudiée en considérant l'effet de l'intensité du courant, de la concentration initiale de caféine, du temps d'électrolyse et du type d'anode. Le BDD et l'IrO2 ont été sélectionnés comme anodes non actives et actives à cette fin. Un plan factoriel (FD) a été utilisé pour étudier les effets principaux et les interactions des différents facteurs sur l'efficacité de l'élimination de la caféine. Par la suite, un plan central composite (CCD) utilisant des modèles polynomiaux quadratiques a été appliqué pour déterminer les conditions optimales pour la dégradation de la caféine. En outre, les résultats fournis par le CCD ont été comparés aux valeurs prédites par la SIFRA.

Objectif 3: Optimiser les hyperparamètres des modèles d'intelligence artificielle pour trouver les modèles les mieux adaptés et accroître la fiabilité et la robustesse des modèles d'intelligence artificielle développés.

En raison de l'investissement important en temps et en argent dans le travail expérimental, un nombre limité d'échantillons dans les ensembles de données est disponible pour les modèles basés sur les données. Les modèles d'IA comprennent des

hyperparamètres intégrés qui doivent être réglés avec précision pour que le modèle puisse résoudre le problème d'apprentissage automatique à son plein potentiel. Le processus d'apprentissage est contrôlé par ces hyperparamètres, qui ont un impact direct sur les performances du modèle. Dans le cas des modèles RNA, ces hyperparamètres comprennent le taux d'apprentissage, le nombre d'époques, l'échec maximal de validation, le nombre de couches cachées et de neurones cachés, et les fonctions de transfert. Pour les modèles RVS, le facteur de pénalité, la marge de tolérance aux erreurs, le type de fonction noyau et les paramètres du noyau doivent être sélectionnés de manière optimale. Les fonctions d'appartenance et le nombre de clusters affectent les performances des modèles SIFRA. D'après la revue de la littérature, la sélection des hyperparamètres des modèles RNA n'a pas fait l'objet d'une étude approfondie et implique généralement une approche par essais et erreurs, qui consomme du temps et des ressources informatiques. Il est donc hautement souhaitable de disposer d'une méthode permettant de rechercher automatiquement et efficacement les hyperparamètres optimaux. Dans cette étude, une investigation approfondie des hyperparamètres des modèles d'IA sera menée pour démontrer les effets de ces paramètres sur la performance du modèle.

Originalité

Les techniques d'intelligence artificielle, principalement les RNA, ont souvent été utilisées pour la modélisation et l'optimisation des procédés électrochimiques pour le traitement des eaux. Malgré la variété des applications, la fiabilité et la robustesse des modèles IA, en raison du nombre limité de données expérimentales disponibles, n'ont pas été étudiées en profondeur dans ce domaine. Par conséquent, dans cette étude, un effort intensif a été fait sur la validation des modèles d’IA et l'optimisation de leurs hyperparamètres inhérents. La méthode habituelle d'optimisation des hyperparamètres est la méthode d'essai et d'erreur. Dans ce travail, cela sera fait en appliquant différents algorithmes métaheuristiques liés aux modèles d’IA pour optimiser leurs hyperparamètres. Une autre partie de l'originalité de cette étude est due à l'application de techniques d'IA à des données qui ont déjà été utilisées pour d'autres approches de modélisation (par exemple, le MSR) afin de comparer leurs performances par rapport à

ces techniques. Dans ce domaine, des données issues de travaux antérieurs de notre groupe de recherche et de la littérature seront utilisées.
Enfin, une étude comparative entre MSR et SIFRA pour la modélisation et l'optimisation de l'élimination de la caféine, un produit pharmaceutique psychoactif, par oxydation électrochimique à l'aide de deux anodes actives et non actives dans des eaux usées synthétiques et réelles qui n'ont pas été étudiées auparavant sera discutée.

Méthodologie générale

Alors que de nombreux modèles dans le domaine des processus électrochimiques sont des modèles empiriques (MSR), nous prévoyons développer des modèles prédictifs basés sur des données expérimentales. Cette démarche était motivée par leur grande capacité à décrire les relations non linéaires entrée-sortie liées aux processus électrochimiques utilisés pour le traitement des eaux.

Des données issues d'expériences antérieures et celles issues de la littérature ont été utilisées. Pour le premier objectif, les données issues d'une étude antérieure de notre groupe de recherche portant sur "Statistical optimization of active chlorine production from a synthetic saline effluent by electrolysis (Zaviska *et al.*, 2012b)" ont été exploitées. La conception expérimentale du procédé d'électrochloration a été réalisée en utilisant successivement la méthodologie du plan factoriel (FD) et du plan central composite (CCD). Les expériences consistaient en 16 expériences pour le FD et 14 expériences supplémentaires pour le CCD, soit un total de 30 expériences.

Pour le second objectif, dans lequel la SIFRA est utilisée comme technique d'IA, les données expérimentales obtenues à partir de l'élimination de la caféine, un produit pharmaceutique psychoactif, par un processus d'oxydation électrochimique ont été utilisées à des fins de modélisation et d'optimisation. Sur la base du plan expérimental, un plan factoriel a été développé pour étudier les effets principaux et les interactions des différents facteurs sur l'efficacité de l'élimination de la caféine. Des modèles polynomiaux quadratiques utilisant le CCD ont été utilisés pour déterminer les conditions expérimentales optimales pour la dégradation de la caféine et la consommation d'énergie. Enfin, pour le troisième objectif portant sur l'optimisation des hyperparamètres des modèles IA à l'aide d'algorithmes métaheuristiques, des données issues de la

littérature ont été utilisées. (Ano *et al.*, 2019). Un nombre total de 62 données expérimentales pour l'élimination du phosphate des eaux usées synthétiques en utilisant le procédé d'électrocoagulation a été rassemblé. Dans leur étude, le plan factoriel et le plan composite central comme méthodologie de surface de réponse ont été utilisés pour étudier l'effet de l'intensité du courant, la concentration initiale de phosphate, le pH initial, le temps de traitement et le type d'électrode.

Après la collecte initiale des données, un prétraitement des données a été nécessaire afin de les manipuler dans un format utilisable pour l'ingestion par les modèles IA. La normalisation des caractéristiques renvoie une version normalisée de la caractéristique (entrée) X où la valeur moyenne de chaque caractéristique est égale à 0 et l'écart type à 1. La normalisation est utile car elle garantit que l'apprentissage du réseau prend en compte toutes les caractéristiques d'entrée dans une mesure similaire et qu'aucune variable unique n'oriente les performances du modèle dans une direction donnée simplement parce qu'elle est plus nombreuse. Les réseaux neuronaux MLP à action directe avec une couche d'entrée, une couche cachée et une couche de sortie sont utilisés dans ce travail pour mettre en correspondance les entrées et les sorties. La descente de gradient, la régularisation bayésienne et Levenberg-Marquardt ont été utilisés comme algorithme d'apprentissage pour former le réseau neuronal.

Afin de diagnostiquer et de prévenir le problème de surajustement pendant le processus d'apprentissage, des courbes d'apprentissage et le facteur de régularisation ont été utilisés. Les courbes d'apprentissage des performances du modèle sur les ensembles de données d'apprentissage et de validation peuvent être utilisées pour diagnostiquer un modèle sous-ajusté (biais élevé), surajusté (variance élevée) ou bien ajusté. Une courbe d'apprentissage montre comment l'erreur varie avec l'augmentation de la taille de l'ensemble d'apprentissage et indique si l'on a besoin ou non d'un modèle plus complexe pour les prédictions. Dans ce travail, les courbes d'apprentissage seront tracées pour les échantillons d'entraînement. Le paramètre de régularisation est une entrée du modèle RNA pour réduire le surajustement, ce qui réduit la variance des paramètres de régression estimés. En d'autres termes, cette technique décourage l'apprentissage d'un modèle plus complexe ou plus souple pour éviter le risque de surajustement. Elle est définie comme un terme ajouté à la fonction de coût du modèle.

Malgré la nature de boîte noire des RNA, il est possible d'effectuer une analyse de sensibilité sur les réseaux neuronaux pour indiquer l'influence de différentes variables d'entrée sur les résultats du modèle. Le niveau de participation de chaque variable d'entrée dans la simulation de la sortie souhaitée peut être obtenu par le biais de la matrice de poids des neurones. L'équation de Garson (Garson, 1991) basée sur le partitionnement des poids de connexion est appliquée à cette fin.

Les hyperparamètres d'un modèle RNA, qui définissent sa topologie et ses options d'apprentissage, influencent la précision et l'efficacité du modèle entraîné. Le nombre de couches cachées et de neurones dans chaque couche cachée, le taux d'apprentissage, le paramètre de régularisation, l'algorithme d'apprentissage et l'échec maximal de validation sont considérés comme des hyperparamètres de la RNA. Des algorithmes métaheuristiques seront utilisés pour trouver les réseaux optimaux.

La performance du modèle RVS dépend fortement de la sélection précise de ses hyperparamètres. Ceux-ci comprennent la contrainte de boîte (C), l'epsilon (ε), le type de fonction noyau et le paramètre du noyau. La contrainte de boîte (C) est un compromis entre la complexité du modèle et la capacité de généralisation. L'hyperparamètre ε influence le nombre de vecteurs de support et donc la performance du RVS en déterminant la taille de la zone insensible à ε. La fonction noyau et son paramètre pertinent transposent les données d'entrée non linéaires dans l'espace des caractéristiques de dimension supérieure pour aider le RVS à traiter les problèmes non linéaires. Dans cette étude, les algorithmes métaheuristiques sont appliqués pour trouver les valeurs optimales de ces hyperparamètres.

Les modèles SIFRA ont été construits sur la base des ensembles de données disponibles en utilisant la méthode de regroupement Fuzzy C-Means. La méthode FCM intégrée à la SIFRA permet d'obtenir un nombre relativement faible de règles, ce qui empêche le modèle d'être trop complexe et minimise le problème de surajustement. La méthode de clustering FCM divise les données d'entrée en différents clusters et est utilisée pour identifier les fonctions d'appartenance floues et la base de règles floues pour le modèle SIFRA. Dans cette étude, la méthode FCM a été utilisée pour le modèle SIFRA, et le

nombre de clusters sera sélectionné manuellement pour obtenir les meilleures performances de généralisation.

L'AG et l'OEP, des algorithmes métaheuristiques, ont été utilisés à des fins d'optimisation. Pour le premier objectif, l'algorithme génétique de tri non dominé (NSGA-II) a été utilisé pour l'optimisation multi-objectifs, qui consiste à trouver un ensemble de solutions (front de Pareto) qui correspondent aux conditions expérimentales en ce qui concerne la maximisation de la production de chlore actif et la minimisation de la consommation d'énergie. Pour le troisième objectif, les algorithmes métaheuristiques ont été utilisées pour trouver les hyperparamètres optimaux des modèles IA pour prévoir l'efficacité de l'élimination du phosphate des eaux usées en utilisant le processus d'électrocoagulation.

Afin de générer les résultats les plus valables pour les modèles et dans le but de comparer ces modèles avec d'autres, nous avions besoin de résultats à la fois précis et interprétables. Dans ce cas, nous avons utilisé des mesures de performance telles que le coefficient de détermination (R^2), l'erreur quadratique moyenne (MSE), l'erreur quadratique moyenne (RMSE) et le pourcentage d'erreur absolu moyen (MAPE).

MATLAB (version R2019a) a été utilisé dans notre étude pour développer et appliquer différents aspects des techniques d'intelligence artificielle. MATLAB est couramment utilisé pour mettre en œuvre des algorithmes pilotés par les données et des algorithmes évolutifs dans notre champ d'étude et convient à la recherche et au développement. À cet égard, tous les scripts et fonctions des techniques d'IA ont été écrits dans l'environnement MATLAB dans cette étude.

Organisation de la thèse

Cette thèse est divisée en six parties, qui correspondent à une introduction générale (chapitre 1), une analyse documentaire approfondie sous forme d'article de synthèse (chapitre 2), trois articles de revues de recherche (chapitres 3, 4 et 5), et une conclusion générale et des perspectives (chapitre 6). Les chapitres 3 et 4 sont déjà publiés dans la *Canadian Journal of Chemical Engineering* et dans *Separation and Purification Technology*, respectivement. Le chapitre 2 est en cours de révision dans le Journal of *Environmental Health Science & Engineering*, tandis que le chapitre 5 est soumis à la

revue *Digital Chemical Engineering*. Des efforts ont été faits pour réduire les répétitions, mais comme chaque article doit fonctionner indépendamment, certaines sections, comme la description des algorithmes IA, sont légèrement répétitives d'un chapitre à l'autre.

Le chapitre 1 couvre le contexte, l'énoncé du problème, les objectifs et la méthodologie générale utilisée dans cette thèse. Une analyse documentaire approfondie est fournie sous forme d'un article de synthèse dans le chapitre 2. Le chapitre 3 correspond au premier objectif de cette thèse : comparer les modèles RNA avec d'autres approches de modélisation conventionnelles telles que la MSR et l'utilisation de l'optimisation multi-objectifs à l'aide d'algorithmes évolutionnaires pour la modélisation et l'optimisation de systèmes électrochimiques. Ensuite, le chapitre 4 couvre l'application de la SIFRA comme autre approche IA pour la modélisation et l'optimisation de la dégradation de la caféine par oxydation électrochimique liée au deuxième objectif de ce travail. L'objectif des chapitres 3 et 4 était de fournir des modèles d'IA fiables avec la moindre chance de tomber dans le problème du surajustement. L'optimisation des hyperparamètres des modèles IA afin de trouver les modèles les mieux adaptés et d'augmenter la fiabilité et la robustesse des modèles IA développés, le dernier objectif de cette thèse, a été abordé dans le chapitre 5. Enfin, le chapitre 6 résume les conclusions et discute des implications du présent travail sur l'application des modèles IA à la modélisation et à l'optimisation des systèmes électrochimiques utilisés pour le traitement de l'eau et des eaux usées.

Conclusions générales et perspectives

Dans ce travail, les applications des techniques d'intelligence artificielle dans la modélisation des processus électrochimiques pour le traitement de l'eau et des eaux usées ont été évaluées. Les techniques d'IA ont été développées sur la base des données obtenues à partir d'études antérieures et d'expériences en laboratoire. En raison du peu de données disponibles dans le domaine des processus électrochimiques pour le traitement de l'eau et des eaux usées pour les techniques d'IA, des efforts ont été faits sur la fiabilité et la robustesse des modèles développés.

La première partie de la thèse a porté sur le développement d'une méthodologie RNA-AG pour la production de chlore actif par le procédé d'électrolyse. Les résultats ont

également été comparés à l'approche de modélisation classique du MSR. La MSR a été largement appliquée pour la modélisation et l'optimisation des processus électrochimiques en raison de ses caractéristiques permettant de minimiser le nombre d'expériences tout en obtenant une réponse optimale en utilisant des expériences bien conçues. D'autre part, les modèles d'intelligence artificielle tels que les RNA sont apparus comme des approches alternatives pour la modélisation et l'optimisation des processus électrochimiques non linéaires. Par conséquent, cette partie de la thèse visait à comparer la performance de l'approche de modélisation conventionnelle MSR avec une technique IA pour la modélisation et l'optimisation d'un processus d'électrolyse. L'originalité de cette partie a été d'utiliser les courbes d'apprentissage et le facteur de régularisation des modèles RNA pour mettre en lumière la formation et la validation de ces modèles boîte noire pilotés par les données. Ces techniques peuvent donner un aperçu du processus de modélisation des RNA et peuvent être utilisées à la place ou en même temps qu'une procédure d'essai et d'erreur pendant l'entraînement des réseaux neuronaux. Elles permettent d'obtenir des modèles RNA fiables, sans risque de surajustement malgré le nombre limité de données disponibles. Les modèles RNA ont réussi à décrire le comportement du processus expérimental (R^2=0,979 et MSE=3,826 pour la production de chlore actif et R^2=0,985 et MSE=6,952 pour la consommation d'énergie). En outre, le front de Pareto dérivé par l'algorithme NSGA-II pour l'optimisation multi-objectifs a conduit à la génération de points optimaux non dominés (conditions de fonctionnement) pour une production maximale de chlore actif avec une consommation d'énergie minimale. Il a été constaté que l'approche RNA-AG proposée permet d'obtenir des conditions plus économiques pour une production de chlore actif plus élevée que les conditions optimales suggérées par l'ancien MSR. La méthodologie RNA-AG proposée peut donner un aperçu de la manière de choisir efficacement les paramètres de fonctionnement du processus (variables de décision) pour atteindre les objectifs souhaités. Cette approche peut être adaptée à d'autres procédés si les données expérimentales existent déjà.

La deuxième partie de la thèse portait sur l'application de l'approche de modélisation SIFRA comme autre technique d'IA et la comparaison avec le MSR pour la modélisation et l'optimisation de l'élimination de la caféine, un produit pharmaceutique psychoactif, par le processus d'oxydation électrochimique. Les variables d'entrée ont été considérées

comme le temps d'électrolyse, l'intensité du courant, la concentration initiale de caféine et le type d'anode. Les résultats ont montré que le type d'anode suivi du temps d'électrolyse sont les variables les plus importantes affectant la dégradation de la caféine. Les modèles SIFRA ont été construits sur la base de l'ensemble des données obtenues à partir de la conception MSR en utilisant la méthode de regroupement Fuzzy C-Means (FCM) plutôt que la méthode de partitionnement de la grille. La méthode FCM intégrée à la SIFRA permet d'obtenir un nombre relativement faible de règles, ce qui empêche le modèle d'être trop complexe et minimise le problème de surajustement. Les résultats des modélisations MSR et SIFRA ont montré que toutes deux sont capables de prédire avec succès le comportement du processus électrochimique, les modèles SIFRA étant légèrement plus performants (R^2=0,993, RMSE=2,694, MAPE=6,582 pour l'efficacité d'élimination de la caféine, et R^2=0,976, RMSE=0,261, MAPE=9,221 pour la consommation d'énergie). Cependant, les modèles MSR ont pu prédire le processus d'OE en utilisant beaucoup moins de paramètres de modèle que les modèles SIFRA, ce qui pourrait réduire les incertitudes du modèle étant donné le faible nombre de données disponibles. D'un point de vue expérimental, les résultats ont montré que même si l'oxydation directe était le mécanisme dominant pendant le processus d'OE pour l'élimination de la caféine, le mécanisme d'oxydation pour la dégradation de la caféine dans cette étude semblait similaire à d'autres processus d'oxydation avancée. L'efficacité de l'élimination de la caféine dans des conditions optimales à partir d'effluents d'eaux usées municipales réelles a varié de 78,0±4,3 % à 92,5±1,0 % à différentes concentrations initiales de caféine, ce qui montre l'efficacité du processus en présence d'autres polluants. Enfin, la toxicité des eaux usées réelles a augmenté après l'OE dans des conditions optimales, ce qui pourrait être réduit en prolongeant le temps d'électrolyse ou pourrait être éliminé entièrement en utilisant la colonne CAG comme post-traitement.

La troisième partie de la recherche s'est concentrée sur la comparaison de modèles d'intelligence artificielle pour prédire l'efficacité de l'élimination des phosphates des eaux usées en utilisant le processus d'électrocoagulation. L'objectif de cette partie de la thèse était d'étudier l'optimisation des hyperparamètres des modèles d'IA en utilisant des algorithmes métaheuristiques pour trouver des modèles optimaux. Ceci a été fait en considérant la fiabilité et la robustesse des modèles IA développés pour les procédés

électrochimiques pour le traitement de l'eau et des eaux usées où une quantité limitée de données est disponible. À cet égard, les modèles hybrides proposés ont été construits sur une validation par ‘repeated random sub-sampling validation’ (10 sous-ensembles de données) au lieu d'une approche à répartition unique. Les résultats ont montré que la performance des modèles IA dépend de la façon dont les données sont distribuées dans les ensembles d'apprentissage, de validation et de test. Les modèles SIFRA et SVS hybride étaient plus vulnérables à la distribution des points d'échantillonnage des données que les modèles RNA hybride. Les modèles RNA hybrides ont surpassé les modèles SIFRA et SVS hybrides et ont montré des performances moins dispersées pour les ensembles de test des différents sous-ensembles de données. Il a été constaté que les modèles RNA de l'OEP ont une performance de généralisation exceptionnelle pour les 10 sous-ensembles de données examinés. Les valeurs moyennes de MSE, R^2 et MAPE des 10 sous-ensembles de test pour l'OEP-RNA ont été déterminées comme étant de 7,201, 0,981 et 2,022, respectivement. On peut conclure que l'approche de modélisation RNA est plus polyvalente et robuste pour le processus d'électrocoagulation étudié. Ce travail fournit une méthode pour trouver des modèles optimaux basés sur les données des processus électrochimiques pour le traitement de l'eau et des eaux usées lorsqu'un nombre limité de données est disponible.

L'adaptation et l'application de modèles IA dans les processus électrochimiques pour le traitement de l'eau et des eaux usées ont été abordées dans cette thèse. Alors que l'utilisation de modèles IA est de plus en plus répandue dans plusieurs disciplines scientifiques, y compris les processus électrochimiques, la fiabilité des modèles développés est critique en raison des données limitées disponibles. Lorsqu'il n'y a pas assez de données pour entraîner les paramètres des modèles IA, il en résulte un surajustement, ce qui signifie que la généralisation du modèle est inexacte. Davantage de données sont nécessaires pour résoudre ce problème ; cependant, dans les applications pratiques, la collecte de données supplémentaires est parfois difficile en raison de facteurs tels que les contraintes de temps et de coût des expériences. Certaines études récentes ont suggéré d'utiliser des techniques d'augmentation des données telles que l'interpolation pour résoudre ce problème. Bien que les approches d'augmentation des données aient été utilisées dans l'apprentissage automatique dans diverses

disciplines, en particulier le traitement des images et la reconnaissance vocale, elles doivent être utilisées avec prudence pour la régression des travaux expérimentaux avec des données limitées. Cela est dû au fait que le comportement des sorties dans les études expérimentales peut être beaucoup plus complexe que leur description à l'aide de fonctions d'interpolation prédéterminées, ce qui permettrait clairement au modèle IA d'anticiper facilement le comportement des nouvelles données interpolées. Une autre approche pour améliorer la fiabilité des modèles IA développés est de développer le modèle le moins complexe possible pour éviter le surajustement. Cela a été fait dans cette thèse en utilisant des courbes d'apprentissage et une régularisation pour les modèles RNA. Dans le cas de la SIFRA, la SIFRA initiale a été générée sur la base de la méthode de clustering FCM plutôt que d'autres alternatives telles que le partitionnement de grille. Le clustering FCM permet d'avoir moins de règles et donc moins de paramètres pour le modèle développé plutôt que le partitionnement de grille. Même si les résultats obtenus par le partitionnement en grille peuvent être plus satisfaisants, l'incertitude augmente avec des modèles plus complexes et de petits ensembles de données. D'autres techniques d'apprentissage automatique telles que les arbres de décision (par exemple, la forêt aléatoire), qui sont bien connues pour des tâches plus simples, peuvent être utilisées et évaluées.

Notre recherche a montré que la performance des modèles IA avec de petits ensembles de données dépend fortement de la distribution des données en ensembles de formation, de validation et de test. Pour surmonter ce problème, différentes techniques de validation, “hold-out”, “k-fold cross-validation”, et “repeated random sub-sampling validation”, ont été appliquées. Une technique de validation croisée stratifiée peut être une autre alternative pour améliorer la fiabilité des modèles. Bien que cette technique soit généralement utilisée pour la classification, elle peut être adaptée aux problèmes de régression. La distribution des valeurs cibles est garantie comme étant à peu près la même dans toutes les partitions de la validation croisée k-fold utilisant la stratification. Cela permet de s'assurer que les performances de validation et de test reflètent les performances attendues du modèle avec moins de biais et de variance. Enfin, les recherches futures peuvent utiliser des approches de modélisation d'ensemble, qui construisent plusieurs modèles et les combinent ensuite pour obtenir de meilleurs résultats. Les approches

d'ensemble donnent souvent des résultats plus précis qu'un modèle unique. Pour ce faire, on peut utiliser différents modèles dans le même ensemble de données d'entraînement, le même modèle avec différentes divisions de l'ensemble de données d'entraînement, ou toute autre méthode. Cette technique réduit souvent le surajustement et produit un modèle de régression plus lisse.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AARE	:	Average Absolute Relative Error
AI	:	Artificial Intelligence
ANFIS	:	Adaptive Neuro-Fuzzy Inference System
ANN	:	Artificial Neural Networks
APC-ECF	:	Alternating Pulse Current Electrocoagulation-Flotation
BDD	:	Boron-Doped Diamond
BOD	:	Biological Oxygen Demand
BP	:	Back-Propagation Algorithm
CCD	:	Central Composite Design
CCRD	:	Central Composite Rotatable Design
COD	:	Chemical Oxygen Demand
DB86	:	Direct Blue 86
DO	:	Dissolved Oxygen
EA	:	Evolutionary Algorithms
EC	:	Electrocoagulation
ED	:	Electrodialysis
EDCs	:	Endocrine Disrupting Chemicals
EF	:	Electro-Fenton
EO	:	Electrooxidation
FC	:	Fecal Coliform
FCM	:	Fuzzy C-mean Clustering
FD	:	Factorial Design
FFA	:	Fire Fly optimization Algorithm

GA	:	Genetic Algorithm
GRNN	:	Generalized Regression Neural Network
HA	:	Humid Acid
LM	:	Levenberg-Marquardt Algorithm
LM-BP	:	Levenberg-Marquardt Back-Propagation
MAPE	:	Mean Absolute Percentage Error
MLP	:	Multilayer Perceptrons
MSE	:	Mean Squared Error
MWWTP	:	Municipal Wastewater Treatment Plants
NSGA	:	Non-dominated Sorting Genetic Algorithm
ORP	:	Oxidation Reduction Potential
OTC	:	Oxytetracycline
PAHs	:	Polycyclic Aromatic Hydrocarbons
PCBs	:	Polychlorinated Biphenyls
PSO	:	Particle Swarm Optimization
R^2	:	Coefficient of Determination
RBF	:	Radial Basis Function
RMSE	:	Root Mean Squared Error
RSM	:	Response Surface Methodology
RY145	:	Reactive Yellow 145
SVM	:	Support Vector Machine
SVR	:	Support Vector Regression
TC	:	Total Coliform
TDS	:	Total Dissolved Solids

TS	:	Total Solids
TSS	:	Total Suspended Solids

GLOSSARY OF THE TECHNICAL TERMS

Term	Definition
Artificial intelligence	Advanced analysis and logic-based techniques, including machine learning, to interpret events, support and automate decisions, and take actions.
Machine learning	The field of study that gives computers the ability to learn without explicitly being programmed.
Artificial neural networks	Computing systems inspired by the biological neural networks that constitute animal brains.
Back-Propagation Algorithm	Widely used algorithm for training feedforward neural networks.
Training algorithm	A step-by-step procedure for adjusting the connection weights of an artificial neural network.
Hyperparameters	In machine learning, a hyperparameter is a parameter whose value is used to control the learning process.
Underfitting	When a data model is unable to capture the relationship between the input and output variables accurately, generating a high error rate on both the training set and unseen data.
Overfitting	When a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new unseen data.
Training set	A data set of examples used during the training process and is used to fit the parameters (e.g., weights).
Validation set	A data set of examples, separate from the training set, that is used to validate our model performance during training.
Test set	A secondary (or tertiary) data set of examples used to test a machine learning model after it has been trained on an initial training data set.
Metaheuristic algorithms	A class of stochastic algorithms using a combination of randomization and local search.
Genetic algorithm	A metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms.
Particle swarm optimization	A population based stochastic optimization technique developed inspired by the social behavior of birds or schools of fish.

1. SYNTHESIS

1.1. Introduction

With the world's rapid population growth and intense industrialization in the 20th century, environmental pollution has become a global problem with adverse impacts on the water sector. The vast majority of these pollution issues are caused by persistent organic compounds because of their resistance to conventional treatments such as physico-chemical or biological methods. This results in the detection of refractory pollutants such as pesticides, phenolic compounds, synthetic dyes, halogenated compounds, polycyclic aromatic hydrocarbon (PAHs), polychlorinated biphenyls (PCBs), endocrine disrupting chemicals (EDCs), and others in rivers, lakes, oceans and even drinking waters all over the world. They can cause hazardous health effects on living organisms, including human beings. Therefore, advanced water and wastewater treatment have become a primary social, political, and environmental concern (Drogui *et al.*, 2007; Moreira *et al.*, 2017; Zheng *et al.*, 2017).

In recent years, electrochemical processes have been gaining attention as an alternative method for water and wastewater treatment. This is due to their high capability to remove persistent organic pollutants. For instance, it has been shown that electrochemical advanced oxidation processes are much more efficient than biological methods in the case of pharmaceutical pollutants (Tiwari *et al.*, 2017). The removal efficiency using conventional activated sludge for Metoprolol and Ketoprofen were reported as 0% and 50%, respectively. However, removal efficiency for Metoprolol using electro-Fenton and Ketoprofen using anodic oxidation are presented as 66% and 100%, respectively (Ganzenko *et al.*, 2014; Jelic *et al.*, 2011).

Electrochemical processes are considered eco-friendly and green technologies since the leading reagent involved, the electron, is regarded as a clean reagent and takes advantage of coupling chemistry (in situ generation of oxidant) with electronic science (electron transfer). Other attractive benefits include versatility, high energy efficiency, amenability to automation, reduced demand for chemicals, less sludge production in case

of electrocoagulation, and cost-effectiveness (Feng *et al.*, 2016; Rajeshwar *et al.*, 1997). On the other hand, disadvantages include production of toxic intermediate by-products in some cases, electrode fouling due to oxidation of pollutants on the electrode surfaces, and necessity of the wastewater to be conductive. Several publications focusing on different electrochemical methods such as electrooxidation, electrocoagulation, electroflotation, electro-Fenton, and electrodialysis have been published within the last decade for improving the treatment performance of wastewaters and drinking waters (Daghrir *et al.*, 2013; Martín de Vidales *et al.*, 2012b; Olvera-Vargas *et al.*, 2015; Zhang *et al.*, 2011).

Process modelling is a requirement for process optimization. Phenomenological and empirical modelling approaches are generally used for water treatment processes. Although phenomenological modelling provides valuable insights into the behaviour of the process and has the ability of extrapolation, heat and mass transport phenomena along with detailed knowledge of the reaction kinetics are required. In empirical modelling, the structure of the data-fitting model should be specified a priori which makes it challenging as one needs to choose the suitable model from the numerous available ones, especially for nonlinear processes (Nandi *et al.*, 2004). Electrochemical processes for water treatment are highly complicated nonlinear systems due to the complex relationships between input parameters and outputs. It is thus difficult to use phenomenological or empirical models to model, simulate, and optimize the processes. Artificial intelligence techniques such as artificial neural networks (ANNs), adaptive neuro-fuzzy inference system (ANFIS), support vector regression (SVR) along with genetic algorithms (GA) and particle swarm optimization (PSO) methods have emerged as attractive alternative approaches for modelling and optimization of these nonlinear processes in case phenomenological or conventional regression models are not practical (Curteanu *et al.*, 2014).

In this work, applications of artificial intelligence techniques focusing on their reliability and validation for the modelling of electrochemical processes for water treatment processes will be discussed. On this matter, a comparison will be made between AI models and conventional modelling approaches (e.g., response surface methodology (RSM)). Furthermore, metaheuristic algorithms will be linked to AI models either for multi-

objective optimization purposes or to optimize the hyperparameters of the AI models to increase their performances.

1.2. Electrochemical processes

Electrochemical methods, which take advantage of coupling chemistry (in situ generation of oxidant) with electronic science (electron transfer), have widely proved to be a clean, flexible and powerful tool for the development of new methods for wastewater treatment. Likewise, electrochemical treatment is generally characterized by simple equipment, easy operation, short retention time and negligible equipment for adding chemicals (Drogui *et al.*, 2007). In the scope of modelling and optimization of these processes, four most applied electrochemical processes are briefly described here.

1.2.1 Electrooxidation

Municipal wastewater treatment plants (MWWTP) are not able to completely remove persistent organic pollutants, pesticides, and pharmaceuticals. Therefore, the persistence of these pollutants in the effluent is of particular importance because it can increase the risk of long-term exposure, which could be responsible for chronic toxicity and subtle effects on animals, plants and the aquatic environment (El-Hanafi *et al.*, 2014).

Electrochemical oxidation is a promising advanced oxidation technique for treating various wastewaters polluted by organic compounds (Aquino *et al.*, 2014; Guitaya *et al.*, 2017; Martín de Vidales *et al.*, 2012a; Senghor *et al.*, 2015; Zaviska *et al.*, 2012a; Zaviska *et al.*, 2013). Since it combines chemistry (generation of *in situ* oxidants) and electricity (electron transfer), it makes it an environmentally friendly technology (Jardak *et al.*, 2016). Electrochemical oxidation occurs based on two different mechanisms: (i) direct oxidation: hydroxyl radicals ($\mathrm{E^{\circ}(OH^{\circ}/H_2O)}$ = 2.80 V vs. SHE) are produced at the electrode surface by the oxidation of water molecules (Eq. (1.1)), and organic compounds can be completely mineralized or degraded by reacting with absorbed $\mathrm{OH^{\circ}}$ radicals (Eq. (1.2)). (ii) indirect oxidation: Other radical systems can be promoted by the generation of different oxidant mediators in the bulk solution, such as $\mathrm{H_2O_2}$, HClO and $\mathrm{S_2O_8^{2-}}$ (Daghrir *et al.*, 2014; Tran *et al.*, 2013).

$$\mathrm{M} + \mathrm{H_2O} \rightarrow \mathrm{M(OH^{\circ})} + \mathrm{H^+} + \mathrm{e^-} \quad (1.1)$$

$$\mathrm{M(OH^{\circ})} + \mathrm{Organics} \rightarrow \mathrm{M} + \mathrm{Oxidized\ products} \quad (1.2)$$

1.2.2 Electrocoagulation

Electrocoagulation (EC), which originates from chemical coagulation, produces in-situ coagulant agents ($\mathrm{Fe^{2+}/Fe^{3+}}$ or $\mathrm{Al^{3+}}$) to effectively remove pollutants by deposition on the cathode electrode or by flotation caused by the generation of hydrogen gas at the cathode (Asselin *et al.*, 2008a). The following equations describe the main reactions occurring in an EC cell:

At the anode: $M_{(s)} \rightarrow M_{(aq)}^{n+} + ne^-$ (1.3)

At the cathode: $2H_2O + 2e^- \rightarrow 2OH^- + H_2$ (1.4)

In the bulk solution: $M_{(aq)}^{n+} + nOH^- \rightarrow M(OH)_{n(s)}$ (1.5)

where $M_{(s)}$ is the metal, $\mathrm{M_{(aq)}^{n+}}$refers to the metallic ion (iron or aluminum ion), $\mathrm{M(OH)_{n(s)}}$ shows metallic hydroxide, and $\mathrm{ne^-}$ is the number of electrons transferred in the reaction at the electrode. It is worth mentioning that Eq. (1.5) describes a simple case of metallic hydroxide formation. In fact, depending on the pH level and the type of metal involved, the formation of different metallic complex species is possible (Dia *et al.*, 2017).

EC has several advantages over chemical coagulation, such as easy automation, low salinity of the effluent after treatment, low footprint, and reduced production of solid residuals (Drogui *et al.*, 2007). EC process has been widely studied for environmental applications to treat drinking water, urban wastewater, textile wastewater, restaurant wastewater, refractory oily wastewater, and heavy metal-containing wastewaters (Al-Shannag *et al.*, 2015; Asselin *et al.*, 2008b; Daghrir *et al.*, 2012; Elazzouzi *et al.*, 2017; Kobya *et al.*, 2014; Mólgora *et al.*, 2013).

1.2.3 Electro-Fenton

The Electro-Fenton (EF) process is an indirect electrochemical advanced oxidation process since hydroxyl radicals are not generated directly from charge transfer at the electrode level but in the solution from the well-known Fenton reaction. The electro-Fenton process has been developed to overcome the drawbacks of the classical Fenton process and to increase the efficiency of pollutant degradation and removal (Ganzenko *et al.*, 2014). In the Fenton process, homogeneous hydroxyl radicals ($^{\cdot}OH$) are generated from Fenton's reagent, a mixture of H_2O_2 and Fe^{2+}, added externally to the solution to be treated (Tran *et al.*, 2010). In contrast, for the EF process, Fenton's reagent is electrochemically produced at the cathode (Eq. (1.6)). The process is based on the continuous generation of hydrogen peroxide (H_2O_2) in an acidic medium through the electrochemical reduction of O_2 at the cathode (Eq. (1.7)). Also, ferric cations (Fe^{3+}) are reduced, and Fe^{2+} is formed (Eq. (1.8)). At the anode, by the oxidation of water, oxygen is produced (Eq. (1.9)) (Mansour *et al.*, 2015; Monteil *et al.*, 2018).

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + {}^{\cdot}OH + OH^- \quad (1.6)$$

$$O_2 + 2H^+ + 2e^- \rightarrow H_2O_2 \quad (1.7)$$

$$Fe^{3+} + e^- \rightarrow Fe^{2+} \quad (1.8)$$

$$2H_2O \rightarrow O_2 + 4H^+ + 4e^- \quad (1.9)$$

EF has been widely applied to the treatment of organic pollutants in water. These studies include pharmaceuticals (Isarain-Chávez *et al.*, 2010; Loaiza-Ambuludi *et al.*, 2013; Panizza *et al.*, 2014), dyes and textile wastewaters (Ghanbari *et al.*, 2015; Kaur *et al.*, 2019), endocrine disrupting compounds (Rosales *et al.*, 2018), pesticides (Abdessalem *et al.*, 2010), polycyclic aromatic hydrocarbons (Yap *et al.*, 2011), surfactants (Panizza *et al.*, 2013) and landfill leachates (Atmaca, 2009).

1.2.4 Electrodialysis

Electrodialysis offers an electrochemical technique that removes ionic pollutants from an aqueous solution with the aid of an electrical potential difference used as a driving force,

producing two new solutions: one concentrate of ions and another consisting of almost pure water. The first solution can be reintroduced to an industrial process, and the water can be reused. In general, the ion separation efficiency in electrodialysis varies depending on the inherent characteristics of the ion exchange membrane, operating conditions, and physicochemical properties of the metal ions (de Barros Machado *et al.*, 2014; Min *et al.*, 2019a). This process has been widely used for the treatment of industrial wastewaters, production of drinking and processed water from brackish water and seawater, recovery of valuable materials from effluents and salt production because of its high chemical stability, flexibility and high ionic conductivity due to its strong ionic characteristics (Shahi *et al.*, 2002; Wu *et al.*, 2019).

1.3. AI techniques

Fundamentals of AI techniques used in this thesis to model and optimize electrochemical processes for water treatment processes are reviewed in this section. These include ANN, SVM, ANFIS and metaheuristic algorithms.

1.3.1 Artificial neural networks

As the name implies itself, artificial neural networks, commonly referred to as “neural networks”, imitate the essential characteristics of the human brain, which itself is a highly nonlinear, complex, and parallel computer, such as self-adaptability, self-organization, and error-tolerant (Haykin, 1998; Singh *et al.*, 2009).

Aleksander et al. (Aleksander *et al.*, 1990) have one of the famous descriptions of the neural networks:

“A neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

1. Knowledge is acquired by the network from its environment through a learning process.

2. Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.”

ANNs, a form of machine learning (ML) model, are now competitive with traditional regression and statistical models in terms of utility (Abiodun *et al.*, 2018). By the mentioned definition, ANNs can explore many competing hypotheses simultaneously using a massively parallel network composed of non-linear, relatively computational elements (neurons or nodes) interconnected by links with variable weights. This mentioned interconnected set of weights contains the knowledge generated by the ANN (Adya *et al.*, 1998). Each neuron, at certain times, examines its inputs and computes an output called an activation. The new activation then is passed along those connections to other neurons. The mathematical model of an artificial neuron is represented in Figure 1.1. The weights can have the + or – sign and thus may influence the receiving element to produce a similar or a different activation. The size of the weight determines the magnitude of the influence of a sending neuron's activation upon the receiving neuron. These connections and weights are vital parameters in the ANN model since they determine the behaviour of the model (Abiodun *et al.*, 2019; Gallant, 1993).

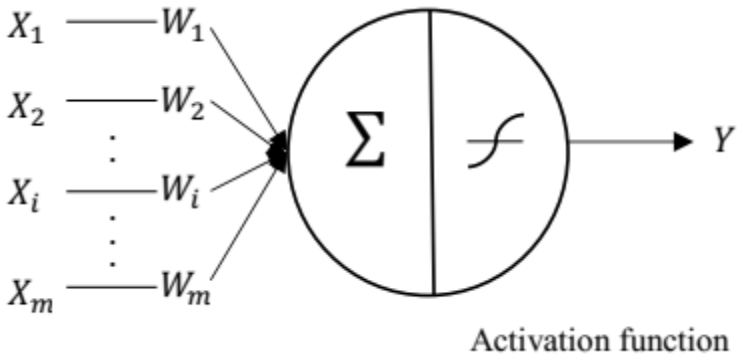


Figure 1.1 Representation of an artificial neuron

Generally, neural networks consist of independent variables (input layer), possible hidden layers, and an output layer.

- The input layer receives information from the outside world to the network for processing.

- As the name describes itself, the hidden layer has no connection with the outside world and computes the information received from the input layer and transfers it to the output layer.

- The output layer is responsible for receiving the processed information and sending it to the outside world.

Artificial neural networks have been utilized in many different areas, including pattern classification, prediction and optimization, signal processing, process forecasting, modelling, and adaptive control (Baughman *et al.*, 1995). ANNs have been used to anticipate the success or failure of banks and stock market estimates. Additionally, it is widely utilised in weather and climate forecasting, which aids in human safety and the security of assets including buildings, the environment, installations, homes, and transportation. Additionally, ANNs have been effectively used in a variety of agricultural fields such remote sensing, notably in the classification of crops and the calculation of crop yield (Abiodun *et al.*, 2018). The main architectures of the artificial neural networks, considering how the different neurons are dispositioned and connected to each other as well as the composition of layers, can be divided as follows: (i) single-layer feedforward networks, (ii) multilayer feedforward networks, and (iii) recurrent networks.

With the development of neural networks in recent years, other types of ANNs including the wavelet neural network, radial basis function, and the Elman neural network have been introduced (Elsheikh *et al.*, 2019).

Single-layer feedforward networks

This type of artificial neural network has just one input layer that projects onto an output layer of neurons; in other words, the information always flows in a single direction from the input layer to the output layer, and not vice versa. These networks are usually applied to linear filtering and pattern classification problems (Da Silva *et al.*, 2017).

Multilayer feedforward networks

The second type of artificial neural network is different from the single-layer feedforward networks in that one or more hidden layers with multiple neurons are present. Figure 1.2 represents a typical multilayer feedforward artificial neural network. These networks are usually applied to diverse problems, including function approximation, pattern classification, system identification, process control, process optimization, and so on (Carvalho *et al.*, 2011; Da Silva *et al.*, 2017). The nature and complexity of the problem

in addition to the desired accuracy and the available data, would determine the number of the hidden layers and the corresponding number of neurons in each hidden layer.

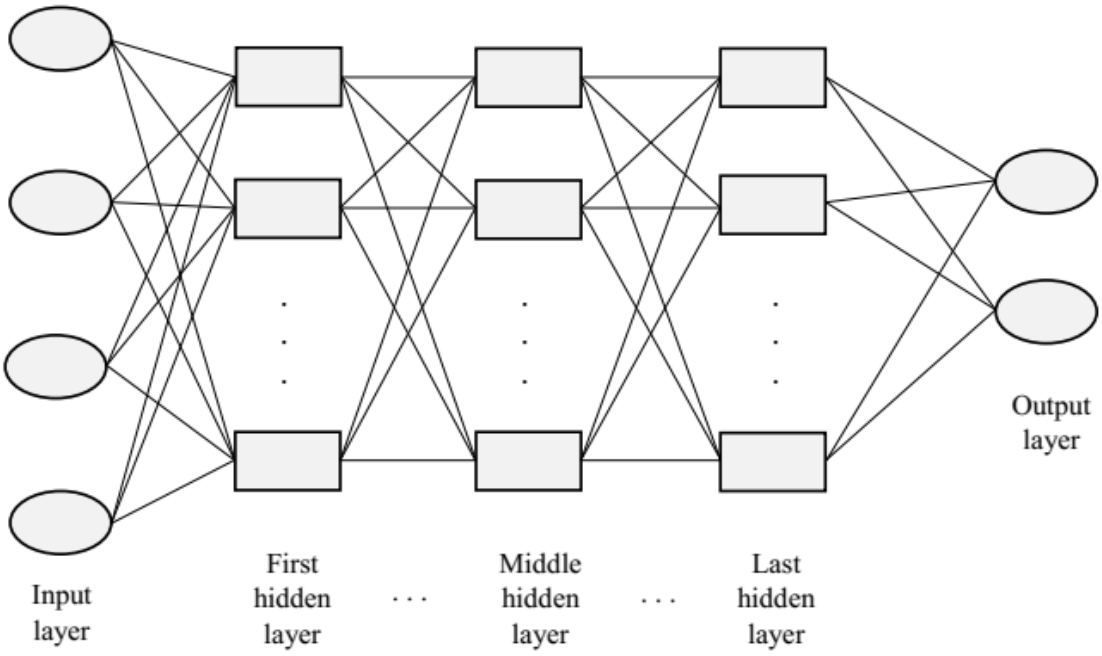


Figure 1.2 Schematic of the multi-layer feedforward ANN structure

Recurrent networks

The recurrent architecture consists of some neurons in a layer whose output signal would be used as an input signal for other neurons. This feedback feature results in a nonlinear dynamical behaviour which can be applied to time-variant systems, like time series prediction, system optimization, and process control (Jain *et al.*, 1999). This presence of feedback structure has a significant impact on the learning capability of the network, hence producing current outputs with taking into consideration of the previous outputs.

1.3.1.1 Learning process

The process in which the connection weights of the network are adjusted and learned by using examples so that the ANN can perform a particular task is called learning (training). The objective of the learning process is to minimize an error function by searching for a

set of connection strengths to produce outputs that are close to the desired targets (output of any given input) (Hamed *et al.*, 2004).

The learning process is basically divided into three categories, supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, an external teacher is involved who compares the actual output of the neural network with the desired target. In unsupervised learning, mainly used for pattern recognition or clustering, there is no performance evaluation of the system by a supervisor or any predetermined correct or incorrect answer. Thus, there is no need for additional input information or knowledge of the targets, and the neural network organizes itself with the correlations among input data to identify groups of similar input patterns. Reinforcement learning, one form of supervised learning, is the type of learning that only requires the fact that the result is good (correct) or bad (incorrect) and thus, requires less information (Baughman *et al.*, 1995; Hammerstrom, 1993).

1.3.1.2 Multilayer perceptrons and their training process

Multilayer perceptrons (MLP) neural networks, which can be used to imitate human brain activities, consist of the input layer of source nodes, one or more hidden layers of computation nodes and an output layer that can include multiple nodes (Assefi *et al.*, 2014; Feng *et al.*, 2008; Ridha *et al.*, 2008). Between these layers, just the hidden and output layers are the processing layers. Usually, the dimensions of the first and last layers are defined by the studied problem, but there is no specified rule for determining the size of the hidden layer. Although there are some indications for the hidden layer dimension, in practice, a trial-and-error procedure is usually done (Padovese, 2002). The standard learning algorithm for MLP neural networks for any pattern recognition or function fitting process is known to be the back-propagation algorithm (BP) (Carvalho *et al.*, 2011). The back-propagation algorithm can be viewed as a generalization form of the least mean square procedure (Haykin, 1998) that can be used to train multilayer neural networks. In the BP algorithm, data enters the network via the input layer, which merely transfers the data value to the hidden layer over-weighted connections. The hidden and output neurons process their inputs by multiplying each input by its weight, adding the product to a total amount, and passing through a function (transfer or activation function) to generate its

result. Each neuron is linked to all neurons in the next layer and the relationship between its inputs and output can be described as follows:

$$x_j = f\left(\sum_i y_i\, w_{ji}\right) \tag{1.10}$$

Where x_j is the output of the neuron j, y_i is the outputs of the neurons that are connected to j and are considered as inputs to neuron j, w_{ji} are the weights of these connections, and f is the linear or non-linear transfer (activation) function. The whole aim of the back-propagation algorithm is to change the values of all the network weights in response to the minimum error between the predicted output and actual targets. During the BP algorithm, first, in the forward direction, the information is passed through the network, and a response is generated at each output node. Then an error is found based on the subtraction of the predicted response and the actual output target. In the backward process, the derivatives of the output errors are passed through the network to the hidden layer using the former values of the connection weights. BP algorithm redistributes the errors to each hidden node, and weights are adjusted accordingly. To minimize the sum of the network's squared errors, the equation that changes the weights is as follows:

$$\Delta w_{ji}(m) = -\eta \times \frac{\partial E}{\partial w_{ji}} + \alpha \times \Delta w_{ji}(m-1) \tag{1.11}$$

where E is the error function being minimized, η is the learning rate or step size parameter (related to changing the correction each time), α is the momentum factor (related to oscillation and convergence), w_{ji} is a generic weight in the network, and m and $m-1$ are the two successive iterations (Ghaedi *et al.*, 2017; Pendashteh *et al.*, 2011).

The error surface can resemble a bowl shape in which the goal is to find the bottom of the bowl- that is, the best set of weights. BP is programmed to achieve the set of weights whose sum of square errors is the smallest by calculating the instantaneous slope of the error surface with respect to the current weights. By incrementally changing the weights in the direction which is the locally steepest path toward the goal (bottom of the bowl), the training process will be accomplished. This process is called gradient descent which, by

aggregating the corrections during training, improves the overall accuracy of the network (Hammerstrom, 1993).

Different transfer functions can be used as the neuron activation function to the sum of weighted inputs and biases. Generally, three transfer functions have been used mainly for MLP neural networks as homogeneous (all transfer functions are the same at the hidden and output layers) or heterogeneous (different combinations of transfer functions are used at the hidden and output layers) configurations. Log-sigmoid transfer function, which generates the outputs between 0 and 1 as the neuron's net input goes from negative to positive infinity, is often utilized and is as follows:

$$logsig(n) = \frac{1}{(1 + \exp(-n))} \tag{1.12}$$

Alternatively, the tan-sigmoid transfer function can be used as another sigmoid activation function which is often used for pattern recognition problems:

$$tansig(n) = \frac{2}{(1 + \exp(-2 * n))} - 1 \tag{1.13}$$

Also, Purelin linear transfer function is used for function fitting problems:

$$purelin(x) = x \tag{1.14}$$

While these are the most commonly used transfer functions for MLP neural networks, other differentiable transfer functions can be created and used if desired (Ghaedi *et al.*, 2017; Khataee *et al.*, 2010).

Since the traditional BP uses a gradient descent algorithm to determine the network weights, it usually slows down near minima and computes rather slowly due to linear convergence. Hence, Levenberg-Marquardt (LM) algorithm, which is the blend of the simple gradient descent and the Gauss-Newton method and is much faster due to adopting the method of the approximate second derivative, is often utilized for MLP neural

networks (Singh *et al.*, 2010). The following equation shows the algorithm for parameter updating:

$$\Delta w = -[J^T J + \mu I]^{-1} J^T E \tag{1.15}$$

where $E = \begin{bmatrix} e_1 & e_2 & \dots & e_p \end{bmatrix}^T$ is the vector of network errors, μ is learning rate (positive constant), I is the identity matrix, and J is the Jacobian matrix as follows:

$$J = \begin{bmatrix} \frac{\partial e_1}{\partial w_1} & \frac{\partial e_1}{\partial w_2} & \dots & \frac{\partial e_1}{\partial w_N} \\ \frac{\partial e_2}{\partial w_1} & \frac{\partial e_2}{\partial w_2} & \dots & \frac{\partial e_2}{\partial w_N} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial e_p}{\partial w_1} & \frac{\partial e_p}{\partial w_2} & \dots & \frac{\partial e_p}{\partial w_N} \end{bmatrix} \tag{1.16}$$

The learning rate μ is increased or decreased during training by a scale at weight updates. When μ is zero, it would be just Newton's method, and when μ is large, it will become a gradient descent method with a small step size (Singh *et al.*, 2009). LM is found to have the fastest convergence for training moderate-sized neural networks, where the training rate is 10 to 100 times faster than the usual gradient descent BP method (Pendashteh *et al.*, 2011).

Therefore, the process of modelling with ANNs can be explained in terms of the following steps:

1. Data set collection, which includes analysis and pre-processing of the data. Usually, the data will be normalized due to very large or small weights to avoid numerical overflows.

2. Constructing and training the neural network includes choosing the network architecture, training algorithm, activation functions, and network parameters.

3. Test the trained and optimized network to evaluate the network performance based on performance criteria.

4. Utilization of the best-trained network for simulation and prediction.

1.3.1.3 Over-fitting problem

The selection of an appropriate number of neurons at the hidden layer is a crucial task in the MLP neural networks since too many neurons can cause a problem, so-called over fitting. In this case, the error on the training set is very low due to the very well learning process, but the error on the new data presented to the network is very high. The network has memorized the training data but has not learned the generalization ability (Valente *et al.*, 2014). Regularly for obtaining network generalization, the method uses such a network which is large enough to provide an appropriate fit. Although it is difficult to have the perspective to know how large should be a network in each case, three generalization learning methods of cross-validation (early stopping), regularization, and pruning can be applied. Regularization is conducted by adding a penalty function to the training objective to minimize the complexity of the model and the prediction error simultaneously. Pruning physically omits some excessive neurons to generate the least size network. For the cross-validation (early stopping) method, the data set will be split into three non-overlapping subsets. The training dataset, which is utilized for learning the network parameters; the validation dataset, which is used for monitoring the training process and for approximating the generalization error; and the test dataset which is an unseen set of data by the model during training, utilized for examining the unbiased generalization error of the trained network. In the early-stopping method, when the validation error rises over a number of iterations (due to over-fitting), the training algorithm stops, and the values of the weights and biases are returned to the point where the validation error was minimum (Chan *et al.*, 2006; Zhang *et al.*, 2003).

1.3.2 Support vector regression

Support vector machines (SVM), first presented by Boser et al. (Boser *et al.*, 1992), with the basis of modern statistical machine learning techniques, have been widely applied to the classification and regression problems due to their promising generalization performance (Saradhi *et al.*, 2007). In a simple binary classification problem, the basic idea of an SVM is to find a hyperplane having the maximum distance (margin) from both sides of the hyperplane. SVM can be adopted for regression problems, and the technique

will be called support vector regression (SVR). To reach the aim of estimating the predictor function (*f*), a loss function (*L*), which assesses the quality of a prediction (*f(x)*), is introduced by *L(x, y, f(x)).* There are different popular loss functions utilized for regression problems, such as ***ε-insensitive loss, Huber's loss, logistic loss,*** and *pinball loss* (Van Messem *et al.*, 2010).

In SVR, the main goal is to obtain a predictor function ($f(x)$) that describes the relationship between inputs and output data with an error value less than ε deviation for all the training data. At the same time, the function $f(x)$ is required to be as flat as possible, meaning that the errors are not significant as long as they are less than ε, but any deviation larger than this amount is not tolerated. This function can be written as:

$$f(x) = w.\phi(x) + b \tag{1.17}$$

where *w* is a weight vector, $\phi(x)$ is a mapping function in the feature space, and *b* is a bias. The coefficients of *w* and *b* are determined by minimizing the following optimization problem:

$$Min \quad \frac{1}{2}\|w\|^2 \tag{1.18}$$

$$subject\ to \begin{cases} y_i - w.\phi(x_i) - b \leq \varepsilon \\ w.\phi(x_i) + b - y_i \leq \varepsilon \end{cases} \tag{1.19}$$

In real conditions, some data may not lie inside the ε-insensitive zone or ε-tube and to permit deviations for training data outside the ε-tube, slack variables ξ, ξ^* are introduced to the problem. Slack variables represent the difference between actual values and the corresponding boundary values of ε-tube.

Hence, the optimization problem can be rewritten as:

$$Min \quad R(w, \xi, \xi^*) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{1.20}$$

$$subject\ to \begin{cases} y_i - w.\phi(x_i) - b \le \varepsilon + \xi_i \\ w.\phi(x_i) + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, i = 1,2,\dots,n \end{cases} \tag{1.21}$$

where R is the regression risk, and box constraint or penalty parameter C is a positive value that determines the trade-off between the training error and generalization ability.

The ε-insensitive loss function can be defined as:

$$L(e_i) = \begin{cases} 0 & if\ e_i \le \varepsilon \\ |e_i| - \varepsilon & otherwise \end{cases} \tag{1.22}$$

As Figure 1.3 shows, the points outside ε-tube will be penalized in a linear fashion.

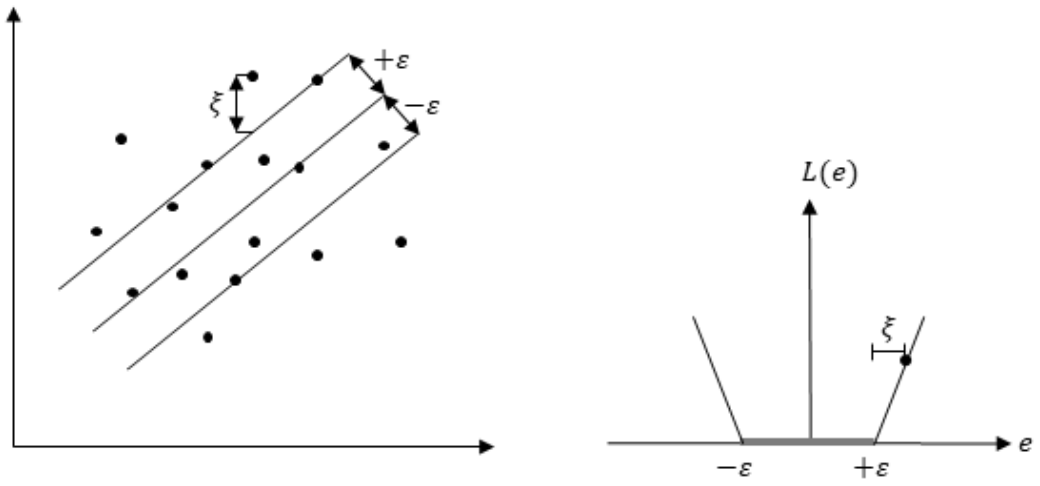


Figure 1.3 Loss function of SVM

To solve the quadratic convex optimization problem, the Lagrangian multipliers are introduced to obtain the dual Lagrangian form:

$$Min \quad \frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) + \varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) - \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) \tag{1.23}$$

$$subject\ to \begin{cases} \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0, & i = 1,2,\dots,n \\ 0 \le \alpha_i,\ \alpha_i^* \le C, & i = 1,2,\dots,n \end{cases} \tag{1.24}$$

Then the regression function can be given as:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{1.25}$$

where α_i and α_i^* are the non-zero Lagrange coefficients, and $K(x_i, x)$ is the kernel function that transforms nonlinear inputs into higher-dimensional feature space. There are several kernel functions used in SVR, such as Linear, Polynomial, and Gaussian or Radial Basis Function (RBF):

Linear: $K(x_i, x_j) = x_i x_j$ (1.26)

Polynomial: $K(x_i, x_j) = (1 + x_i x_j)^q$ (1.27)

Gaussian or RBF: $K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right)$ (1.28)

where q is the polynomial order and σ is the RBF kernel parameter. While all these kernel functions can be opted, the selection of suitable kernel functions and their parameters is essential for the SVR performance.

1.3.3 Adaptive neuro-fuzzy inference system

ANFIS was introduced by Jang (Jang, 1993), as a hybrid technique of artificial intelligence that combines a Sugeno-type Fuzzy Inference System (FIS) and an artificial neural network. Fuzzy Logic is utilized in ANFIS to produce fuzzy rules and map the inputs to an output based on a given input-output data set. ANFIS applies the neural network learning process to learn from a given set of training data (like an ANN model), the rules and membership functions for tuning the FIS parameters. Likewise, the solution mapped out into the fuzzy model is explained in linguistic terms based on IF-THEN fuzzy logic rules (Abdulshahed *et al.*, 2015).

Figure 1.4 shows the schematic of an ANFIS structure with two inputs, two rules and one output. As can be seen, the ANFIS structure includes five layers. Similar to ANN learning, ANFIS training comprises a forward pass and a backward pass. Output values are the results of the forward pass through the net. In the backward pass, the error is calculated

and is propagated back to the earlier layers in a similar manner as the backpropagation learning algorithm (Hussein, 2016).

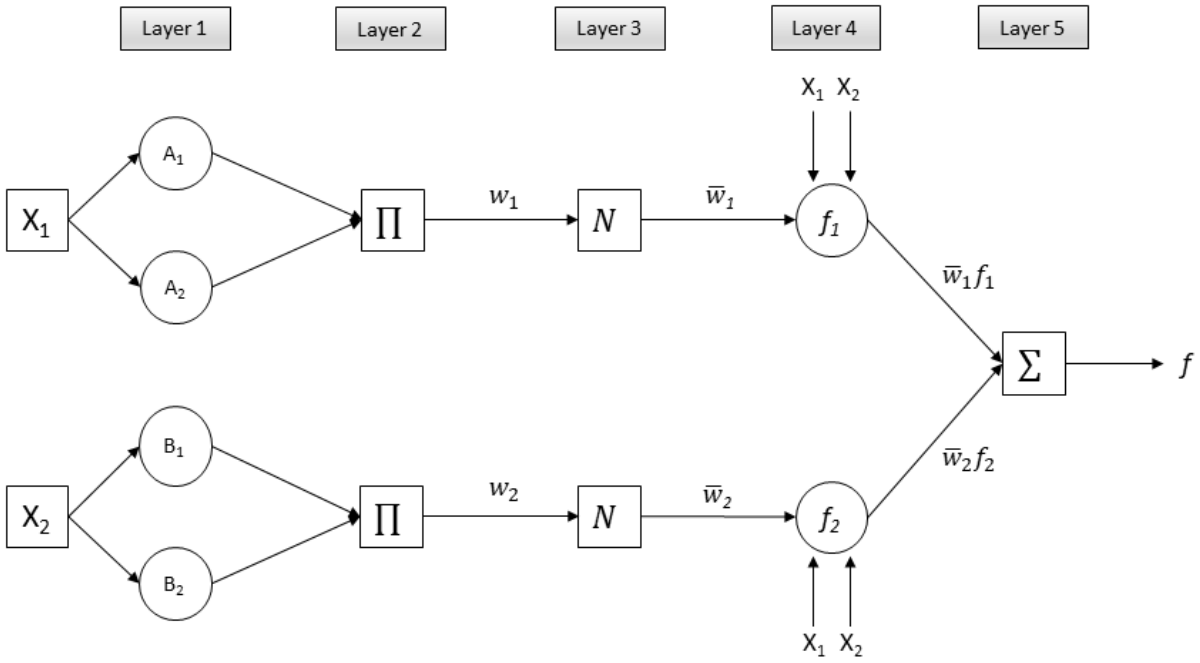


Figure 1.4 The ANFIS structure

In a simple ANFIS architecture with two inputs and one output, the following steps will be carried out:

Input layer: In the Input layer, inputs are introduced to the ANFIS network.

Layer 1: The first layer is called the fuzzification layer which generates membership functions (MF) for each of the inputs. The nodes in this layer are adaptive nodes which have adjustable premises parameters related to input membership functions.

$$O_{1,i} = \mu_{Ai}(X_1), \quad for\ i = 1,2 \tag{1.29}$$

$$O_{1,i} = \mu_{Bi-2}(X_2), \quad for\ i = 3,4 \tag{1.30}$$

where $X_1$and X_2 are the inputs to node i, A_i and B_i are the linguistic labels associated with this node, $\mu(X_1)$ and $\mu(X_2)$ are the membership functions. Fuzzy membership values can be computed by various membership functions. For instance

Gaussian: $(x;\sigma,c) = e^{\frac{-(x-c)^2}{2\sigma^2}}$; where σ is the standard deviation, and c is the mean.

Generalized bell-shaped: $(x;a,b,c) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}}$; where $a, b,$ and c are the parameters.

Triangular: $f(x;a,b,c) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases}$; where $a, b,$ and c are the parameters.

Layer 2: In this layer, the so-called rule layer, the firing strength of each rule is calculated. The nodes are fixed nodes. Fuzzy operators (e.g., AND) are involved in this layer.

$$O_{2,i} = w_i = \mu_{Ai}(X_1).\mu_{Bi}(X_2), \quad for\ i = 1,2 \tag{1.31}$$

where $O_{2,i}$ is the output of layer 2, and w_i represents the firing strength of the rule.

Layer 3: This layer is called the normalization layer, and it computes the normalized value of the firing strength of each rule with respect to the firing strength of all rules. Every node in this layer is a fixed node.

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad for\ i = 1,2 \tag{1.32}$$

Layer 4: The fourth layer or defuzzification layer is an adaptive layer which calculates the values of the consequences of the rules. Consequent parameters in a *p*-th order polynomial function of the input signals (used in the Sugeno model) are adaptive parameters in this layer that must be adjusted. The output of a first-order polynomial function is calculated as follows

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i(p_i X_1 + q_i X_2 + r_i), \quad for\ i = 1,2 \tag{1.33}$$

where $\overline{w}_i$ is the output of layer 3, and p_i, q_i, and r_i are the consequent parameters.

Layer 5: In the final layer or the sum layer, the final output is calculated by summation of all incoming signals.

$$O_{5,i} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{1.34}$$

The two IF-THEN rules for the ANFIS architecture described above would be represented as:

Rule 1: IF X_1 is A_1 AND X_2 is B_1, THEN $f_1 = p_1 X_1 + q_1 X_2 + r_1$.

Rule 2: IF X_1 is A_2 AND X_2 is B_2, THEN $f_2 = p_2 X_1 + q_2 X_2 + r_2$.

1.3.4 Metaheuristic algorithms for optimization

Metaheuristic algorithms have received growing attention in recent years among optimization techniques. Population based algorithms (e.g., genetic algorithms and particle swarm optimization) are effective metaheuristic search methods that try to capture global solutions to complex optimization problems while maintaining robustness and flexibility. These algorithms do not need any type of fitness gradient information to operate, and they can get out of local minima where deterministic optimization approaches could fail or are not applicable (Ryan, 2003). These optimization algorithms have been coupled with AI models in two ways. First, they have been applied in optimization procedures where the objective is to determine the optimum conditions for the best value of the system output. Secondly, they have been used in the training process of the AI models as an alternative optimization algorithm (besides general training algorithms such as back-propagation) or to find optimal hyperparameters of the AI models. Between various metaheuristic algorithms, genetic algorithms and particle swarm optimization have been widely used in literature (Alam *et al.*, 2020; Kulkarni *et al.*, 2015).

1.3.4.1 Genetic algorithms

GAs with good global searching ability and flexibility, ease of operation and without the need for gradient information of the objective (fitness) functions have become powerful techniques for optimization problems (Curteanu *et al.*, 2007; Ding *et al.*, 2011b). In GA, a chromosome (also sometimes called a genotype) is a set of parameters which define a

proposed solution to the problem that the GA is trying to solve by searching through the space of possible chromosome values. The major steps of a GA are as follows:

1. Initial population and chromosome representation: GA starts with the generation of a set of individuals (population) whose chromosomes are stochastically chosen. All the individuals in the initial population are evaluated by the fitness (loss) function and submitted to the genetic operators defined below.

2. Fitness/ cost function: The fitness or cost function evaluates how an individual is adapted to its environment. Any performance criteria (e.g., mean squared error) can be used as a fitness or cost function.

3. Genetic Operators:

i) Selection: Three selection operators are employed to select the most fitted individuals as the first and second parent based on their fitness evaluations to go through the crossover, namely Roulette wheel, Tournament, and Random selection (Zhong et al. 2005).

ii) Crossover: In the crossover process, genes of the two individuals (parents) selected are exchanged to form offsprings with mixed characteristics.

iii) Mutation: Just like a biological phenomenon to keep genetic diversity, the mutation is introduced in GA too. Mutation randomly changes the information in a gene as a stochastic factor to create another individual hoping that it has a better fitness evaluation.

The generation of new populations and calculation of the fitness value for each population is repeated over and over in an iterative method. When a specific termination criterion is met, e.g., when there is no more change in the population from one iteration to the next or when a satisfactory fitness value is identified, this process ends (Ansari *et al.*, 2014; Niculescu, 2003; Ridha *et al.*, 2008).

1.3.4.2 Particle swarm optimization

The particle swarm optimization algorithm, first introduced by Kennedy and Eberhart (Kennedy *et al.*, 1995), is based on social behaviour simulation of a flock of birds called ‘swarm’, searching for food. PSO is also a stochastic population-based optimization

approach in which particles, a swarm of potential solutions, fly in the problem space to find better regions and, finally, the optimal solution while cooperating and competing with other ones (Chen *et al.*, 2010). In PSO, a particle is analogous to a chromosome (population member) in GA and represents a candidate solution to the problem being studied (Eberhart *et al.*, 1998). Each particle's condition is changed by the impact of three factors: (1) its own inertia; (2) the personal most optimal position; and (3) the swarm's most optimal position (Juneja *et al.*, 2016). In the d-dimensional search space of the problem, particle *i* of the swarm can be represented by $X_i = (x_{i1}, x_{i2}, \dots, x_{id})$. The velocity of this particle and the best previous position, which is the position giving the best fitness value, are represented as $V_i = (v_{i1}, v_{i2}, \dots, v_{id})$, $P_i = (p_{i1}, p_{i2}, \dots, p_{id})$. Also, the global best position, the position of the best individual, is noted as $G = (g_1, g_2, \dots, g_d)$ (Talebi *et al.*, 2010). The velocity and position of the particles are updated as follow:

$$V_i^{j+1} = \omega * V_i^j + c_1 * rand_1 * \left(P_i^j - X_i^j\right) + c_2 * rand_2 * \left(G_i^j - X_i^j\right) \quad (1.35)$$

$$X_i^{j+1} = X_i^j + V_i^{j+1} \quad (1.36)$$

where V_i^{j+1} and X_i^{j+1} are the updated velocity and position vector of particle *i*, ω is the momentum or inertia weight factor, c_1 and c_2 are the learning factors, and $rand_1$ and $rand_2$ are random numbers between (0,1) (Moghaddam *et al.*, 2012). Detailed information about GA and PSO algorithms can be found in the literature (Juneja *et al.*, 2016; Whitley, 1994).

1.4. Problem statement, hypothesis, objectives and originality

1.4.1 Problem statement

Electrochemical processes have gained increasing interest as an alternative method for the treatment of polluted waters. These processes are considered eco-friendly and green technologies since the leading reagent involved, the electron, is regarded as a clean reagent and takes advantage of coupling chemistry (in situ generation of oxidant) with electronic science (electron transfer). Electrochemical systems offer several advantages over other approaches, such as versatility, operation at ambient temperature and

pressure, amenability to automation, and capability to adjust to variations in influent composition and flow rate. Several publications focusing on different electrochemical methods such as electrooxidation, electrocoagulation, electroflotation, electro-Fenton, and electrodialysis have been published within the last decade for improving the treatment performance of wastewaters and drinking waters.

The necessary improvements in electrochemical process performances and operation require better modelling in view of design and optimization. Process modelling is important to help improve the design and reduce both equipment and operating costs. It can provide robust and accurate solutions to electrochemical challenges and thus enable to predict process performances under a wide range of operating conditions.

Process modelling is a prerequisite for process optimization. Phenomenological and empirical modelling approaches are generally used for water treatment processes. Phenomenological modelling provides valuable insights into the behaviour of the process and has the ability of extrapolation. On the other hand, heat and mass transport phenomena, along with detailed knowledge of the reaction kinetics, are required for phenomenological models based on an electrochemical process. Collecting this is a difficult task for a multivariate system, especially when limited knowledge is available. Moreover, the nonlinear behaviour of electrochemical processes results in complex nonlinear models, which in most cases are not flexible to analytical solutions; they thus require computationally intensive numerical methods to solve.

Empirical (regression) modelling is one of the alternative modelling approaches for phenomenological modelling. In most cases, a quadratic linear regression model will be selected that is often not adequate to describe the nonlinearities of the systems.

Electrochemical processes for water treatment are highly complicated nonlinear systems due to the complex relationships between input parameters and outputs. This is due to the fact that several mechanisms usually happen at the same time in an electrochemical system. For instance, In the electrocoagulation process detailed mechanisms of charge transport, electrochemical kinetics, thermodynamics, adsorption isotherms and kinetic models, flocculation, flotation, settling, and complexation should be known. Also, in electrooxidation, the concentration of every compound in an electrochemical cell depends

on time and space, that is, their distance from the electrode surface. Describing the profile of compounds under such conditions involves a number of partial differential equations, which are often difficult to solve and involve many model parameters. The complexity of these models depends on the number of species included in the model. In a multivariable model, all the significant species in an electrochemical cell are included. This however requires further knowledge on reaction pathways to account for subsequent formations and transformations.

Therefore, it is not always the best option to use phenomenological or conventional empirical models to model, simulate, and optimize the processes. In these situations, artificial intelligence techniques based on data-driven models such as artificial neural networks are an attractive alternative method. They could overcome the classical modelling difficulties, having the following advantages: the possibility to apply them even on multiple input–multiple output complex nonlinear processes, the ability to be constructed solely from historic process input-output data (experimental dataset), and excellent generalization ability when adequately trained. On the other hand, the disadvantages seem to be obtaining a reliable and robust model with a limited number of experimental data available. Therefore, this study aims to apply artificial intelligence techniques focusing on their reliability and robustness to model and optimize electrochemical processes for water treatment. This includes comparing AI models over conventional modelling approaches (e.g. RSM), linking metaheuristic algorithms to AI models for multi-objective processes, and optimizing the hyperparameters of the AI models to increase their performance.

1.4.2 Hypothesis

According to the problem statement, the present study includes the following hypotheses:

Hypothesis 1: ANNs are one of the artificial intelligence methods that could be effectively used to model electrochemical processes for wastewater treatment.

Artificial neural networks have some considerable advantages over other empirical methods that make them practical for the modelling of electrochemical processes. First of all, ANNs have massive parallel connections between the nodes. Each node operates

independently from the other ones and has a slight effect on the input-output pattern. This parallelism feature helps ANNs have better performance than empirical models, especially in complex electrochemical nonlinear processes. Furthermore, ANNs are adaptive to be trained multiple times, meaning that whenever the network performance is inadequate due to any change in the conditions, ANNs can be trained further with new conditions to enhance their performance. This adaptive feature is not common with empirical models. Multiple-input and multiple-output (MIMO) characteristic of ANNs is another aspect that makes them a suitable option for modelling electrochemical processes. ANNs can map many independent variables to as many dependent variables as needed, which is a unique feature compared to other empirical models. Also, unlike RSM, experimental design and selecting the structure of the data-fitting model a priori is not a prerequisite for ANNs. This is beneficial, especially for nonlinear processes where choosing a suitable model from the numerous available ones makes it challenging.

Hypothesis 2: Evolutionary algorithms linked to AI models outperform other approaches (e.g. RSM) for multi-objective optimization.

Evolutionary algorithms linked to AI models are helpful for multiple input–multiple output complex nonlinear processes. There is no unique solution to a multi-objective optimization problem but a set of mathematically equally good solutions known as nondominated or Pareto optimal solutions. In this regard, highly robust metaheuristic evolutionary algorithms like GAs can be linked to AI models for multi-objective optimization problems. GAs can outperform conventional optimization methods because they do not require the objective function to be continuous and/or differentiable, they do not require extensive problem formulation, and they are not sensitive to starting point.

Hypothesis 3: Optimizing the hyperparameters of AI models will lead to a better modelling performance of electrochemical processes for water treatment.

The performance of AI models very much depends on the optimization of built-in hyperparameters which control the learning process. For instance, in the case of ANN models, these hyperparameters include learning rate, number of epochs, maximum

validation failure, regularization factor, number of hidden layers and hidden neurons, and transfer functions.

1.4.3 Objectives

The overall objective of this study is to apply and optimize artificial intelligence models for modelling and optimization of electrochemical processes for water treatment. In other words, this research aims to demonstrate and compare the capability and performance of optimized AI models with other conventional types of modelling and optimization approaches used in electrochemical processes for water treatment. The specific objectives of this study are as follows:

1.4.3.1 Specific Objectives

Objective 1: Comparing AI models with other conventional modelling approaches such as RSM used for modelling and optimization of electrochemical systems

The response surface methodology has been widely used to study the effect of different variables on the response for modelling and optimization of electrochemical systems. It helps to minimize the number of experiments while obtaining an optimal response using well-designed experiments. Despite the RSM benefits, sometimes, the impact of various operational parameters in complex nonlinear electrochemical processes cannot fully be defined by a simple linear multivariate correlation. Because of this, one of the main objectives of this study is to compare the performance of AI and RSM modelling approaches.

Also, evolutionary algorithms such as GA will be used in this study for multi-objective optimization of electrochemical processes for wastewater treatment. For multi-objective optimization problems, there is no unique solution which is what is typically obtained by RSM. Instead, a set of mathematically equally good solutions, known as Pareto optimal solutions, are obtained. In this study, evolutionary optimization algorithms will be linked to ANN models to improve operational conditions since electrochemical processes require introducing energy and minimum experimental time to achieve good process

efficiencies. This approach can optimize experimental conditions by enhancing removal/production efficiencies and reducing energy costs.

Objective 2: Modelling and optimization of caffeine degradation by electrochemical oxidation

ANFIS, as another AI modelling technique, is used in this part of the thesis for caffeine degradation by electrochemical oxidation. Caffeine is selected as a psychoactive pharmaceutical model pollutant for the electrochemical oxidation process. The degradation of caffeine by electrochemical oxidation will be studied by considering the effect of the current intensity, initial concentration of caffeine, electrolysis time, and anode type. The BDD and IrO_2 have been selected as non-active and active anodes for this purpose. Based on the experimental design, a factorial design (FD) followed was developed to investigate the main and interaction effects of different factors on caffeine removal efficiency. Quadratic polynomial models using CCD determined the optimal experimental conditions for caffeine degradation and energy consumption. Furthermore, the results provided by CCD were compared with the ANFIS predicted values.

Objective 3: Optimizing hyperparameters of AI models to find the optimally tuned models and increase the reliability and robustness of the developed AI models

With the significant investment of time and money in experimental work, a limited number of samples in datasets are available for data-driven models. AI models include built-in hyperparameters that should be fine-tuned so that the model can solve the machine learning problem to its full potential. These hyperparameters control the learning process, which directly impacts the model performance. In the case of ANN models, these hyperparameters include learning rate, number of epochs, maximum validation failure, number of hidden layers and hidden neurons, and transfer functions. For SVR models, the penalty factor, margin of error tolerance, and the type of kernel function and the kernel parameters should be optimally selected. Membership functions and the number of clusters affect the performance of ANFIS models. Based on the literature review, hyperparameter selection of the ANNs models has not been studied thoroughly and

usually involves a trial-and-error approach, which consumes time and computing resources. It is thus highly desirable to have a method to automatically search for the optimal hyperparameters efficiently. In this study, a deep investigation of AI models' hyperparameters will be conducted to demonstrate the effects of these parameters on model performance.

1.4.4 Originality

Artificial intelligence techniques, mainly ANNs, have often been utilized for modelling and optimization of electrochemical processes for water treatment. Despite the variety of applications, the reliability and robustness of AI models due to a limited number of experimental data available have not been thoroughly considered in the field. Therefore, in this study, an intensive effort has been made on AI models' validation and optimizing their inherent hyperparameters. The usual way for hyperparameter optimization is the trial-and-error method. In this work, this will be done by applying different metaheuristic algorithms linked to AI models for optimizing their hyperparameters.

Another part of the originality of this study is due to applying AI techniques to data that already have been used for other modelling approaches (e.g. RSM) to compare their performance over these techniques. Data from previous works in our research group and literature will be used in this domain.

Finally, a comparative study between RSM and ANFIS for modelling and optimization of psychoactive pharmaceutical caffeine removal by electrochemical oxidation by two active and non-active anodes in synthetic and real wastewater that have not been studied before will be discussed.

1.5. General methodology

A brief explanation of the general methodology used in this work will be explained in this section.

1.5.1 Data-driven approach

While many models in the field of electrochemical processes are conventional empirical (RSM), we attempted to develop data-driven predictive models. This was driven by their

high capability to describe the nonlinear input-output relationships related to the electrochemical processes used for water treatment.

1.5.2 Data collection

In this thesis, data from previous experiments and from the literature were used. For the first objective, the data was derived from a study in our group with the title “Statistical optimization of active chlorine production from a synthetic saline effluent by electrolysis (Zaviska *et al.*, 2012b)”. Experimental design of the electrochlorination process was carried out using successively factorial design (FD) and central composite design (CCD) methodology. The experiments consisted of 16 experiments for FD and extra 14 experiments for CCD, a total of 30 experiments. Table 1.1 shows the description and statistical parameters of the dataset used in this study.

Table 1.1 Description of the dataset used for the first objective of this work from Zaviska *et al.* (2009)

Statistical parameters	Independent variables/Inputs				Dependent variables/Outputs	
	Electrolysis time (min)	Current intensity (A)	$[H_3O^+]$ (M)	[NaCl] (M)	HClO production (mg/l)	Energy consumption (kWh/m^3)
Number of samples	30	30	30	30	30	30
Range	15-35	0.8-1.6	0.05-0.11	0.3-0.8	0.1-46.6	0.083-0.820
Average	25	1.2	0.08	0.55	14.15	0.362
Standard deviation	9	0.4	0.03	0.23	13.78	0.217

The assays were carried out in a batch electrolytic reactor with 400 ml of capacity made in PVC material with a dimension of 135 mm (L)×35 mm (l)×140 mm (h). The electrolytic cell comprised a rectangular anode of Ti/IrO_2 in the form of expanded metal and rectangular stainless steel (SS) plate used as a cathode electrode.

For the second objective, in which ANFIS is utilized as an AI technique, the experimental data obtained from the psychoactive pharmaceutical caffeine removal by electrochemical oxidation (EO) process was utilized for modelling and optimization purposes. The description and statistical parameters of the dataset are represented in Table 1.2.

Table 1.2 Description of the dataset used for the second objective of this work

Statistical parameters	Independent variables/Inputs				Dependent variables/Outputs	
	Electrolysis time (min)	Current intensity (A)	$[Caffeine]_0$ (M)	Anode type	Removal efficiency (%)	Energy consumption (Wh/mg)
Number of samples	40	40	40	40	40	40
Range	13-47	0.7-2.3	13-47	BDD/IrO_2	5.11-100	0.27-6.78
Average	30	1.5	30	-	50.72	2.32
Standard deviation	8.4	0.41	8.4	-	32.62	1.71

Based on the experimental design, a factorial design (FD) followed was developed to investigate the main and interaction effects of different factors on caffeine removal efficiency. Quadratic polynomial models using CCD determined the optimal experimental conditions for caffeine degradation and energy consumption. The electrolytic cell used was made of Plexiglas material with a dimension of 17.1 cm (depth) × 3.3 cm (width) × 11.4 cm (length). It was comprised of one anode and one cathode with an interelectrode gap of 1 cm. Lab-scale electrolysis of 460 cm^3 of caffeine solutions was carried out under galvanostatic conditions with current intensity ranging from 0.7 to 2.3 A (10.3 to 33.8 mA cm^{-2}) according to FD and CCD matrices.

Finally, for the third objective dealing with hyperparameter optimization of AI models with metaheuristic algorithms, data from literature was used (Ano *et al.*, 2019). Table 1.3 shows the description and statistical parameters of the dataset used for this part of the

study. A total number of 62 experimental data for phosphate removal from synthetic wastewaters using the electrocoagulation process was gathered. In their study, factorial design (FD) and central composite design (CCD) as response surface methodology were used to investigate the effect of current intensity, initial phosphate concentration, initial solution pH, treatment time, and electrode type.

Table 1.3 Description of the dataset used for the third objective of this work (Ano *et al.*, 2019)

Statistical parameters	Independent variables/Inputs					Dependent variable/Output
	Current Intensity (A)	$[Phosphate]_0$ (mg/L)	pH	Treatment time (min)	Electrode type	Removal efficiency (%)
Number of samples	62	62	62	62	62	62
Range	0.25-1.25	15-75	2-10	10-90	Al/Fe	29.17-100
Average	0.75	45	6	50	-	74.02
Standard deviation	0.22	13.3	1.77	17.74	-	20.59

1.5.3 Data analysis

1.5.3.1 Data preprocessing

After initial data collection, data preprocessing was necessary to manipulate the data into a usable format for ingestion by the AI models. Feature standardization has been selected as the Eq. (1.37) and returns a standardized version of feature (input) X where the mean value of each feature is 0, and the standard deviation is 1.

$$y = \frac{X_i - \mu_i}{S_i} \tag{1.37}$$

where y is the standardized value of X_i. The μ_i and the S_i are the mean and standard deviation values of X_i, respectively.

Standardization helps because it ensures that the network's learning regards all input features to a similar extent, and no single variable directs model performance in one direction just because they are greater numbers.

1.5.3.2 ANN

Feedforward MLP neural networks with one input layer, one hidden layer, and one output layer are used in this work to map the inputs to the outputs. Gradient descent and Levenberg-Marquardt have been used as the learning algorithm for training the neural network.

Learning curves

Learning curves of model performance on the training and validation datasets can diagnose an underfit (high bias), overfit (high variance), or well-fit model. A learning curve demonstrates how the error varies with an increase in the training set size and shows if one needs a more complex model for the predictions or not. In this work, learning curves will be plotted for the training samples.

Regularization factor

The regularization parameter (lambda) is an input to the ANNs model to reduce overfitting, which reduces the variance of the estimated regression parameters. In other words, this technique discourages learning a more complex or flexible model so as to avoid the risk of overfitting. It is defined as a term added to the cost function of the model (Eq. (1.38)).

$$J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda\sum_{j=1}^{n}\theta_j^2\right] \tag{1.38}$$

where $J(\theta)$ is the cost function (error), m is the number of the data points, x is the input neuron, $h_\theta\left(x^{(i)}\right)$ is the predicted value of sample i, $y^{(i)}$ is the actual value of sample i, λ regularization parameter, and θ is the network parameters. The regularization parameter λ is a control of the fitting parameters.

Relative importance

Despite the black-box nature of ANNs, It is possible to carry out sensitivity analysis on the neural networks to indicate the influence of different input variables on the model's results. The level of participation of each input variable in the simulation of the desired output can be obtained through the neural weight matrix. Garson's equation based on the partitioning of connection weights can be applied:

$$I_j = \frac{\sum_{m=1}^{m=N_h}\left(\left(\left|W_{jm}^{jh}\right| / \sum_{k=1}^{N_i}\left|W_{km}^{ih}\right|\right) \times \left|W_{mn}^{ho}\right|\right)}{\sum_{k=1}^{k=N_i}\left[\sum_{m=1}^{m=N_h}\left(\left|W_{km}^{ih}\right| / \sum_{k=1}^{N_i}\left|W_{km}^{ih}\right|\right) \times \left|W_{mn}^{ho}\right|\right]} \quad (1.39)$$

where I_j is the relative importance of the jth input variable on the output variable, N_i and N_h are the numbers of input and hidden neurons, respectively; the Ws are connection weights, the superscripts i, h and o refer to input, hidden, and output layers, respectively. Also, the subscripts k, m and n refer to input, hidden and output neurons, respectively.

1.5.3.3 Hyperparameter optimization

The hyperparameters of an ANN model, which define its topology and learning options, influence the accuracy and effectiveness of the trained model. The numbers of hidden layers and neurons in each hidden layer, learning rate, regularization parameter, learning algorithm, and maximum validation failure are considered ANN hyperparameters. Metaheuristic algorithms will be utilized to find the optimal networks.

The performance of the SVR model highly depends on the accurate selection of its hyperparameters. These include the box constraint (C), the epsilon (ε), the type of kernel function, and the kernel parameter. The box constraint C is a trade-off between model complexity and generalization ability. The ε hyperparameter influences the number of support vectors and hence the performance of the SVR by determining the size of the ε-insensitive zone. The kernel function and its relevant parameter map nonlinear input data into the higher dimensional feature space to help SVR handle nonlinear problems. In this study, metaheuristic algorithms are applied to find the optimal values of these hyperparameters.

ANFIS models were constructed based on the available datasets using the Fuzzy C-Means clustering method. FCM integrated with ANFIS helps obtain a relatively small number of rules which prevents the model from being too complex and minimizes the overfitting issue. The FCM clustering method partitions the input data into different clusters and is used to identify the fuzzy membership functions and fuzzy rule base for the ANFIS model. In this study, FCM has been used for the ANFIS model, and the number of clusters will be manually selected for the best generalization performance.

1.5.3.4 Metaheuristic algorithms

GA and PSO as population based metaheuristic algorithms have been utilized for optimization purposes. For the first objective, the non-dominated sorting genetic algorithm (NSGA-II) has been used for multi-objective optimization, which is finding a set of solutions (Pareto front) that are the experimental conditions with respect to maximization of active chlorine production and minimization of energy consumption. For the third objective, these algorithms have been used to find optimal hyperparameters of the AI models to forecast the removal efficiency of phosphate from wastewaters using the electrocoagulation process.

1.5.4 Assessment

To generate the most valuable results for the models and to compare these models with other ones, we needed both accurate and interpretable results. In this case, performance metrics such as the coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) were utilized.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1.40}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{1.41}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (1.42)$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (1.43)$$

1.5.5 Programming language and software

MATLAB (version R2019a) has been used in our study to develop and apply different aspects of artificial intelligence techniques. MATLAB is commonly used for implementing data-driven and metaheuristic algorithms in our scope of research and is suitable for research and development. In this regard, all the AI techniques' scripts and functions were written in the MATLAB environment in this study.

1.5.6 Analytical Details

The analytical details related to the experimental part of this work regarding the electrochemical degradation of caffeine in synthetic solution and real municipal wastewater effluent by the EO process are briefly mentioned in this section.

Progress of the electrochemical degradation of caffeine in solution was monitored and quantified by absorbance measurements (absorption peaks previously determined) using a Varian Cary 100 ultraviolet (UV) spectrophotometer. The absorption peak measured at the wavelength of 273 nm was chosen to evaluate the residual caffeine concentration. The concentrations of caffeine in real wastewater and identification of intermediates were monitored and quantified by LC/MS/MS (Thermo TSQ Quantum Access). Chromatographic separation was achieved using a Hypersil Gold C18 column (Thermo Hypersil Ltd., Runcorn, UK) with a particle size of 3.0 mm and a 100 mm length × 2.1mm inner diameter. Total organic carbon (TOC) was measured by the high-temperature catalytic combustion method and infrared detection using a Shimadzu TOC VCPH analyzer (Shimadzu Scientific Instruments, Kyoto, Japan). The sample was acidified to remove inorganic carbon. Anionic species (Cl^-, NO_3^-, SO_4^{2-}) were measured by ion

chromatography using a Thermo Integrion High-Pressure Ion Chromatography (HPIC). The ammonia concentration was determined according to the analytical method proposed by LACHAT Instrument (QuikChem® Method 10-107-06-2-B). To study the indirect effect of electrochemical process, the oxidant production was evaluated by absorbance measurements at λ=353 nm using a UV-spectrophotometer based on the reaction of electro-generated oxidants with iodide to form triiodide. To evaluate the acute toxicity of the untreated and treated real municipal wastewater effluent, a daphnia (Daphnia magna) test was used.

1.6. Thesis organization

This thesis is divided into six parts, which correspond to a general introduction (Chapter 1), an extensive literature review as a review article (Chapter 2), three research journal articles (Chapters 3, 4, and 5), and a general conclusion and perspectives (Chapter 6). Chapters 3 and 4 are already published in *The Canadian Journal of Chemical Engineering* and *Separation and Purification Technology*, respectively. Chapter 2 is under review in the *Journal of Environmental Health Science & Engineering,* whereas chapter 5 is submitted to the journal of *Digital Chemical Engineering.* Efforts were taken to reduce duplication, but because each journal article must function independently, certain sections, such as the description of the AI algorithms, are slightly repetitive across chapters.

Chapter 1 covers the background, problem statement, objectives, and general methodology used in this thesis. An extensive literature review is provided as a review article in Chapter 2. Chapter 3 corresponds to the first objective of this thesis to compare ANN models with other conventional modelling approaches such as RSM and the use of multi-objective optimization utilizing evolutionary algorithms for modelling and optimization of electrochemical systems. Then, Chapter 4 covers the application of ANFIS as another AI approach for modelling and optimization of caffeine degradation by electrochemical oxidation related to the second objective of this work. Chapters 3 and 4 focus on providing reliable AI models with the slightest chance of getting trapped into the overfitting issue. Optimizing hyperparameters of AI models to find the optimally tuned models and increasing the reliability and robustness of the developed AI models as the

last objective of this thesis has been discussed in Chapter 5. Finally, Chapter 6 summarizes the conclusions and discusses the implications of the present work on the application of AI models to the modelling and optimization of electrochemical systems used for water treatment. Also, perspectives for the future work are provided in this chapter.

2. REVIEW OF ARTIFICIAL INTELLIGENCE TECHNIQUES IN MODELLING OF ELECTROCHEMICAL PROCESSES FOR WATER AND WASTEWATER TREATMENT

ÉTUDE DES TECHNIQUES D'INTELLIGENCE ARTIFICIELLE DANS LA MODÉLISATION DES PROCÉDÉS ÉLECTROCHIMIQUES POUR LE TRAITEMENT DE L'EAU ET DES EAUX USÉES

Majid Gholami Shirkoohi[a,d], Rajeshwar Tyagi[b], Peter A. Vanrolleghem[c,d], Patrick Drogui[a,d]

[a] Institut National de la Recherche Scientifique (INRS), Centre-Eau Terre Environnement, Université du Québec, 490 Rue de la Couronne, Québec (QC) G1K 9A9, Canada

[b] BOSK Bioproducts, 399 Rue Jacquard, suite 100, Québec (QC) G1N 4J6, Canada

[c] modelEAU, Département de génie civil et de génie des eaux, Université Laval, 1065 av. de la Médecine, Québec (QC) G1V 0A6, Canada

[d] CentrEau, Centre de recherche sur l'eau, Université Laval, Québec (QC), Canada

Title of the Journal: Journal of Environmental Health Science and Engineering

Date of submission: 02 March 2021

Abstract

Artificial intelligence techniques have been useful alternatives for modelling and optimization of electrochemical processes when phenomenological models cannot be applied to these multivariate systems. In this work, artificial intelligence techniques such as artificial neural networks (ANNs), support vector machines (SVM), adaptive neuro-fuzzy inference system (ANFIS), genetic algorithms (GA), and particle swarm optimization (PSO), used in water and wastewater treatment processes, are reviewed. ANN and SVM are two popular machine learning approaches for supervised learning and appear to be good alternatives for modelling complex nonlinear processes. GA and PSO algorithms have been utilized for either optimizing the output (outputs) or as a tool for estimating ANN parameters during the training process. This paper describes applications of the mentioned artificial intelligence techniques for the modelling and optimization of electrochemical processes for water and wastewater treatment processes. Most research in the mentioned scope of study consists of electrooxidation, electrocoagulation, electro-Fenton, and electrodialysis.

Keywords: Artificial neural networks (ANNs), Electrocoagulation, Electrooxidation, Genetic algorithms (GA), Mathematical modelling, Process optimization

2.1. INTRODUCTION

With the world's rapid population growth and intense industrialization in the 20th century, environmental pollution has become a global problem with adverse impacts on the water sector. The vast majority of the remaining pollution issues is caused by heavy metals and persistent organic compounds because of their resistance to conventional treatments such as physico-chemical or biological methods. This results in detection of refractory pollutants such as pesticides, phenolic compounds, synthetic dyes, halogenated compounds, polycyclic aromatic hydrocarbon (PAHs), polychlorinated biphenyls (PCBs), endocrine disrupting chemicals (EDCs), and others in rivers, lakes, oceans and even drinking waters all over the world. They can cause hazardous health effects on living organisms including human beings. Therefore, advanced water and wastewater treatment have become a primary social, political, and environmental concern (Drogui *et al.*, 2007; Moreira *et al.*, 2017; Zheng *et al.*, 2017).

In recent years, electrochemical processes have been gaining attention as an alternative method for water and wastewater treatment. These processes are considered as eco-friendly and green technologies since the leading reagent involved, the electron, is considered a clean reagent and takes advantage of coupling chemistry (*in situ* generation of oxidant) with electronic science (electron transfer). Other attractive advantages include: versatility, high energy efficiency, amenability to automation, and cost-effectiveness (Feng *et al.*, 2016; Rajeshwar *et al.*, 1997). Several publications focusing on different electrochemical methods such as electrooxidation, electrocoagulation, electroflotation, electro-Fenton, and electrodialysis have been published within the last decade for improving the treatment performance of wastewaters and drinking waters (Daghrir *et al.*, 2013; Martín de Vidales *et al.*, 2012b; Olvera-Vargas *et al.*, 2015; Zhang *et al.*, 2011).

Process modelling is a requirement for process optimization. Phenomenological and empirical modelling approaches are generally used for water and wastewater treatment processes (Cañizares *et al.*, 2004a; Cañizares *et al.*, 2004b; Zaviska *et al.*, 2013). Although phenomenological modelling provides valuable insights on the behavior of the process and has the ability of extrapolation, heat and mass transport phenomena along with detailed knowledge of the reaction kinetics are required. In empirical modelling the

structure of the data-fitting model should be specified a priori which makes it challenging as one needs to choose the suitable model from the numerous available ones, especially for nonlinear processes (Nandi *et al.*, 2004). Electrochemical processes for water and wastewater treatment are highly complicated nonlinear systems due to the complex relationships between input parameters and outputs. It is thus difficult to use phenomenological or empirical models to model, simulate, and optimize the processes. Artificial intelligence methods such as artificial neural networks (ANNs) and support vector machines (SVM) along with genetic algorithms (GA) and particle swarm optimization (PSO) methods have emerged as attractive alternative approaches for modelling and optimization of these nonlinear processes in case phenomenological or conventional regression models are not practical (Curteanu *et al.*, 2014).

Evolutionary algorithms (EA), and in particular genetic algorithms (GA) and particle swarm optimization (PSO), have received growing attention in recent years among available optimization techniques. EA, with good global searching ability and flexibility, ease of operation and without the need for gradient information of the objective (fitness) functions, have become powerful techniques for optimization problems (Curteanu *et al.*, 2007; Ding *et al.*, 2011b). EA have been utilized in two ways with ANNs. First, they have been applied in optimization studies where the objective is to determine the optimum conditions for the best value of the system output. The trained neural network is used here as the objective (fitness) function of the EA. Secondly, whereas the BP is the most widely used training algorithm, it can get trapped in suboptimal solutions (local optima) for systems containing complex nonlinear relationships. In these cases, EA can be used in the training process for hyperparameter optimization of neural networks to avoid local optima by searching in several regions simultaneously. Detailed information about GA and PSO algorithms can be found in the literature (Juneja *et al.*, 2016; Whitley, 1994).

In this work, applications of artificial intelligence techniques in modelling of electrochemical processes for water and wastewater treatment processes are discussed. To make AI modelling approach performance competitive to other conventional modelling approaches usually used (e.g., response surface methodology), it is important to build robust and reliable AI models. While the trend to use AI models is increasing in different fields of science, including electrochemical processes, the lack of attention to reliability

and robustness of these models can have a negative impact on the progression of this field. This has been the main motivation of the authors for this paper since, to the knowledge of the authors, there is no specific review for this particular subject. Therefore, the purpose of this review paper is to provide an overview of the current knowledge and to present future perspectives in this field.

2.2. DATA SETS

2.2.1 Electrochemical processes

Most of the data sets in published articles derive from four electrochemical processes: electrooxidation, electrocoagulation, electro-Fenton, and electrodialysis. In this section, a brief explanation of these electrochemical processes along with their AI modelling applications for water and wastewater treatment processes has been reviewed.

2.2.1.1 Electrooxidation

Municipal wastewater treatment plants (MWWTP) are not able to completely remove persistent organic pollutants, pesticides, and pharmaceuticals. Hence, their persistence in the effluent is of particular importance because it can increase the risk of long-term exposure, responsible for chronic toxicity and subtle effects in animals, plants and the aquatic environment (El-Hanafi *et al.*, 2014; Särkkä *et al.*, 2015).

Electrochemical oxidation is a promising advanced oxidation technique for treating various wastewaters polluted by persistent organic compounds (Aquino *et al.*, 2014; Martín de Vidales *et al.*, 2012a; Polcaro *et al.*, 2004; Senghor *et al.*, 2015; Zaviska *et al.*, 2012a; Zaviska *et al.*, 2013). Since it combines chemistry (generation of *in situ* oxidants) and electricity (electron transfer), it is an environmentally friendly technology (Jardak *et al.*, 2016). Electrochemical oxidation occurs based on two different mechanisms:

(i) direct oxidation: hydroxyl radicals ($E^{\circ}(OH^{\circ}/H_2O)$ = 2.80 V vs. SHE) are produced at the electrode surface by the oxidation of water molecules (Eq. (2.1)), and organic compounds can be completely mineralized (electrochemical combustion) or degraded

(electrochemical conversion) by reacting with absorbed OH° radicals (Grimm *et al.*, 1998) (Eq. (2.2)).

$$M + H_2O \rightarrow M(OH^{\circ}) + H^{+} + e^{-} \quad (2.1)$$

$$M(OH^{\circ}) + Organics \rightarrow M + Oxidized\ products \quad (2.2)$$

(ii) indirect oxidation: other radical systems can be promoted by the generation of different oxidant mediators in the bulk solution, such as $H_2O_2, HClO \text{ and } S_2O_8^{2-}$ (Anglada *et al.*, 2009; Daghrir *et al.*, 2014).

Table 2.1 summarizes the application of artificial intelligence (AI) modelling approaches of electrochemical oxidation for water and wastewater treatment processes.

Table 2.1 Application of AI modelling of electrochemical oxidation for water and wastewater treatment processes

Type of pollutant to treat / process	Input variables	Output variables	Modelling approach	Performance	Ref.
COD, specialty chemical manufacturer effluent	Current density, time, salt concentration	COD Removal, energy Consumption	ANN; 3:7:1; One hidden layer	R=0.9977	(Ahmed Basha *et al.*, 2010)
Malachite green dye	Current density, time, salt concentration	Dye removal	ANN; 3:9:1; One hidden layer	R=0.9987	(Soloman *et al.*, 2010)
CBSOL LE red wool dye	pH, current, reaction time	Dye degradation, color removal, energy consumption	ANN; 3:8:3; One hidden layer	R=0.995	(Sangal *et al.*, 2015)
Oxytetracycline	Nature and concentration of the supporting electrolyte, initial pH, current intensity, reaction time	Removal efficiency	ANN; 5:14:1; One hidden layer	R=0.99	(Belkacem *et al.*, 2017)

2-Chlorophenol	Current density, pH, electrolysis time, supporting electrolyte concentration, oxidation–reduction potential	COD removal efficiency and total energy consumption	ANN; 5:10:2; One hidden layer	$MSE_{COD\ removal}$= 0.0015526 $MSE_{Energy\ consumption}$=0.023456 $R^2_{COD\ removal}$= 0.9900 $R^2_{Energy\ consumption}$=0.9944	(Mei *et al.*, 2019)
Rhodamine-B dye	Current density, electrolyte concentration, initial pH, electrolysis time	Decolorization and energy consumption	ANN; 4:12:2; One hidden layer	R=0.9865	(Kothari *et al.*, 2020)
Synthetic saline effluent	Electrolysis time, current intensity, hydrochloric acid concentration, chloride ion concentration	Active chlorine production and energy consumption	ANN; 4:5:1 and 4 :4:1; One hidden layer	$MSE_{Chlorine\ production}$= 3.826 $MSE_{Energy\ consumption}$=6.952 $R^2_{Chlorine\ production}$= 0.979 $R^2_{Energy\ consumption}$=0.985	(Gholami Shirkoohi *et al.*, 2021)

COD, distillery effluent	Flow rate, current density, supporting electrolyte concentration	COD removal	ANN; 3:3:3:1; Two hidden layers	R=0.9987	(Manokaran *et al.*, 2014)
Sulfate wastewaters with Bromophenol blue dye	Electrolysis time, flow, current density, pH, and initial dye concentration	Discoloration efficiency	ANN; 5:5:3:1; Two hidden layers	RMSE= 10.73 MAPE=8.81 R=0.946	(Picos-Benitez *et al.*, 2020)
Phenolic compounds	Pollutant concentration, pH, temperature, current density, current charge	COD	ANN; Stacked neural networks	Average relative error of 4.92%	(Piuleac *et al.*, 2010)
Phenol compounds	Temperature, initial COD, pH, current density, charge, type of chlorine phenol compounds, type of nitrophenol compounds	COD	Hybrid and stacked neural networks	R^2=0.998	(Piuleac *et al.*, 2012)
Substituted phenols	Quantum chemical descriptors of the phenols	Degradation of substituted phenols	SVM	RMSE=0.202 R^2=0.892	(Yuan *et al.*, 2006)

Organic compounds and the available micro-organisms in activated sludge of the sewage	Current density, time, electrode type, pH, COD, TC, FC, EC, TDS	COD, total coli form (TC), fical coli form (FC), electroconductivity (EC), total dissolved solids (TDS)	Multiple ANN topologies and SVM	R^2>0.977	(Curteanu *et al.*, 2014)
Ciprofloxacin	Current density, electrolyte concentration, pH, electrolysis time	Removal efficiency	ANN, SVM, ANFIS	Multiple performance criteria	(Farzin *et al.*, 2020)

2.2.1.2 Electrocoagulation

Electrocoagulation (EC), developed from chemical coagulation, produces coagulant agents (Fe^{2+}/Fe^{3+} or Al^{3+}) in-situ to effectively remove pollutants by deposition on the cathode or by floatation caused by the generation of hydrogen gas at the cathode (Asselin *et al.*, 2008a). The following equations describe the main reactions occurring in an EC cell:

At the anode: $$M_{(s)} \rightarrow M_{(aq)}^{n+} + ne^{-} \quad (2.3)$$

At the cathode: $$2H_2O + 2e^{-} \rightarrow 2OH^{-} + H_2 \quad (2.4)$$

In the bulk solution: $$M_{(aq)}^{n+} + nOH^{-} \rightarrow M(OH)_{n(s)} \quad (2.5)$$

where M(s) is the metal, $M_{(aq)}^{n+}$ refers to the metallic ion (iron or aluminum ion), $M(OH)_{n(s)}$ represents the metallic hydroxide, and ne^{-} is the number of electrons transferred in the reaction at the electrode. It is worth mentioning that Eq. (2.5) describes a simple case of metallic hydroxide formation. In fact, depending on the pH and the type of metal involved, the formation of different metallic complex species is possible (Dia *et al.*, 2017).

EC has several advantages over chemical coagulation, such as easy automation, low salinity of the effluent after treatment, low footprint, and reduced production of solid residuals (Drogui *et al.*, 2007). The EC process has been widely studied for environmental applications to treat drinking water, urban wastewater, textile wastewater, restaurant wastewater, refractory oily wastewater, and heavy metal containing wastewaters (Al-Shannag *et al.*, 2015; Asselin *et al.*, 2008b; Daghrir *et al.*, 2012; Elazzouzi *et al.*, 2017; Kobya *et al.*, 2014; Mólgora *et al.*, 2013).

There are a number of studies regarding the application of artificial neural networks for modelling of wastewater treatment by electrocoagulation processes (see Table 2.2).

Table 2.2 Application of ANNs for modelling of wastewater treatment by electrocoagulation

Type of pollutant to treat / process	Input variables	Output variables	ANN architecture	Performance	Ref.
Color removal, textile dye solution	Current intensity, electrolysis time, initial pH, initial dye solution, conductivity, retention time of sludge, distance between electrodes	Color removal percentage	7:10:1; One hidden layer	R=0.974	(Daneshvar *et al.*, 2006)
COD, dairy industry effluent	Total Solids, Total Suspends Solids, Total Dissolved Solids, turbidity and initial COD, initial pH, electrolysis time, distance between electrodes and current density	Final COD	9:10:1; One hidden layer	MSE=0.00406 R^2=0.9560	(Valente *et al.*, 2014)
Cr(VI)	Current density, time of electrolysis, initial	Residual Cr(VI) concentration	4:10:1; One hidden layer	R^2=0.976	(Aber *et al.*, 2009)

	concentration of Cr(VI) and salt concentration				
Cr(VI)	Voltage and time of electrolysis	Cr(VI) removal efficiency and energy consumption	2:4:2; One hidden layer	MSE=0.0242 $R^2_{Cr(VI)removal}$=0.975 $R^2_{Energy\ consumption}$=0.99	(Bhatti *et al.*, 2011b)
Direct Blue 86 (DB86) and Reactive Yellow 145 (RY145)	Current density, time of electrolysis, initial dye concentration, conductivity and pH	Removal efficiency of dyes	5:20:3; One hidden layer	R^2=0.976	(Keskin *et al.*, 2011)
Reactive Black 5 dye	Time, current, conductivity and flocculant dosage	Percentage of dye removal and the operating cost	4:10:2; One hidden layer	R^2= 0.9764	(Nourouzi *et al.*, 2011)
Endosulfan	Current density, time of electrolysis, initial dye concentration, conductivity and pH	Removal of Endosulfan	5:8:1; One hidden layer	R^2=0.976	(Mirsoleimani-Azizi *et al.*, 2015)

Sunfix Red S3B dye	Initial concentration of dye, initial pH, agitation speed, gum dosage and electrolysis time	Color and COD removal	5:7:2; One hidden layer	R^2=0.836 RMSE=9.844% MAPE=13.776%.	(Bui, 2016)
Grey water	Current density, electrolysis time, and gap between electrodes	Removal efficiency of turbidity	3:6:1; One hidden layer	R=0.89	(Nasr *et al.*, 2016)
Boron, mining wastewater	Current intensity, pH, and treatment time	Percentage of boron removal	3:10:1; One hidden layer	R^2=0.973 SSE=0.616	(da Silva Ribeiro *et al.*, 2019)
Distillery spent wash	Current intensity, pH, mixing speed, electrolysis time, and the inter-electrode space	Colour removal efciency	5:10:1; One hidden layer	R^2=0.987	(David *et al.*, 2020)
Acid Blue 113 textile effluent	Effluent concentration, electrolyte pH, current density, electrolysis time	Percentage of COD removal	Multiple ANN topologies	R^2=0.983	(Murugan *et al.*, 2009)

Humic acid	Initial humic acid (HA) concentration, initial pH, electrical conductivity, current density and number of pulses	HA removal	Multiple ANN topologies	R^2=0.999 MSE=0.00006	(Hasani *et al.*, 2018)
Chlorophyll, final effluent of aerated lagoons	Electric power, temperature, time, electrode distance, electrode type and initial concentrations of chlorophyll a, TSS, COD	Final concentrations of chlorophyll a, TSS, and COD	Stacked neural networks	Relative errors (E_r): $E_{r,\ TSS}$=5.89%, $E_{r,\ chlorophyll\ a}$=14.86% $E_{r,\ COD}$=9.55%	(Curteanu *et al.*, 2011)

2.2.1.3 Electro-Fenton

The Electro-Fenton (EF) process is an indirect electrochemical advanced oxidation process since hydroxyl radicals are not generated directly from charge transfer at the electrode level but in the solution from the well-known Fenton reaction. The electro-Fenton process has been developed to overcome the drawbacks of the classical Fenton process and to increase the efficiency of pollutant degradation and removal (Ganzenko *et al.*, 2014). In the Fenton process, homogeneous hydroxyl radicals (˙OH) are generated from Fenton's reagent, a mixture of H_2O_2 and Fe^{2+}, added externally to the solution to be treated (Tran *et al.*, 2010). In contrast, for the EF process, Fenton's reagent is electrochemically produced in situ at the cathode (Eq. (2.6)). The process is based on the continuous generation of hydrogen peroxide (H_2O_2) in an acidic medium through the electrochemical reduction of O_2 at the cathode (Eq. (2.7)). Also, ferric cations (Fe^{3+}) are reduced, and Fe^{2+} is formed (Eq. (2.8)). At the anode, by the oxidation of water, oxygen is produced (Eq. (2.9)) (Mansour *et al.*, 2015; Monteil *et al.*, 2018).

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + {}^{\cdot}OH + OH^- \quad (2.6)$$

$$O_2 + 2H^+ + 2e^- \rightarrow H_2O_2 \quad (2.7)$$

$$Fe^{3+} + e^- \rightarrow Fe^{2+} \quad (2.8)$$

$$2H_2O \rightarrow O_2 + 4H^+ + 4e^- \quad (2.9)$$

EF has been widely applied to the treatment of organic pollutants in water and wastewater. These studies include pharmaceuticals (Isarain-Chávez *et al.*, 2010; Loaiza-Ambuludi *et al.*, 2013; Panizza *et al.*, 2014), dyes and textile wastewaters (Ghanbari *et al.*, 2015; Kaur *et al.*, 2019), endocrine disrupting compounds (Rosales *et al.*, 2018), pesticides (Abdessalem *et al.*, 2010), polycyclic aromatic hydrocarbons (Yap *et al.*, 2011), surfactants (Panizza *et al.*, 2013) and landfill leachates (Atmaca, 2009).

Applications of ANNs for water and wastewater treatment using the EF process are presented in Table 2.3.

Table 2.3 Applications of ANNs for the water and wastewater treatment using the EF process

Type of pollutant to treat / process	Input variables	Output variables	ANN architecture	Performance	Ref.
Decolorization, BR46 dye	Time, initial pollutant concentration, applied current, $Fe^{2+/3+}$, pH	Percentage of dye removal	5:16:1; One hidden layer	R^2=0.986	(Zarei *et al.*, 2010)
Naphtol Blue Black	Initial Fe^{3+} concentration, initial pH, concentration of Na_2SO_4, temperature, applied current, and initial dye concentration	Degradation efficiency and the rate constant	6:32:2; One hidden layer	MSE=10^{-5} R^2>0.99	(Bouasla *et al.*, 2014)
Phenolic wastewater	Time, initial pollutant concentration, applied current, $Fe^{2+/3+}$	Phenol degradation efficiency	4:20:1; One hidden layer	R^2=0.9742	(Radwan *et al.*, 2018)
Composting plant leachate	Time, pH, TDS, current density, H_2O_2 concentration	Removal efficiency	5:8:1; One hidden layer	R^2=0.9907	(Alavi *et al.*, 2019)

				MSE=8.77	
Textile wastewater	Dissolved oxygen and oxidation reduction potential related inputs	Fe^{2+} dose and COD removal efficiency	4:8:1 and 4:12:1; One hidden layer	$R^2_{ANN,\ Fe^{2+}}$=0.9944 $R^2_{ANN,\ COD}$=0.9952	(Yu *et al.*, 2013)

2.2.1.4 Electrodialysis

Electrodialysis offers an electrochemical technique that removes ionic pollutants from an aqueous solution with the aid of an electrical potential difference used as a driving force, producing two new solutions: one concentrate of ions and another consisting of almost pure water. The first solution can be reintroduced to an industrial process, and the water can be reused. In general, the ion separation efficiency in electrodialysis varies depending on the inherent characteristics of the ion exchange membrane, operating conditions, and physicochemical properties of the metal ions (de Barros Machado *et al.*, 2014; Min *et al.*, 2019b). This process has been widely used for the treatment of industrial wastewaters, production of drinking and processed water from brackish water and seawater, recovery of useful materials from effluents and salt production because of its high chemical stability, flexibility and high ionic conductivity due to its strong ionic characteristics (Lu *et al.*, 2016; Mohammadi *et al.*, 2004; Sadrzadeh *et al.*, 2007b; Shahi *et al.*, 2002; Wu *et al.*, 2019).

Table 2.4 summarizes the applications of ANNs for water and wastewater treatment with the ED process.

Table 2.4 Applications of ANNs for water and wastewater treatment with ED process

Type of pollutant to treat / process	Input variables	Output variables	ANN architecture	Performance	Ref.
Lead ions	Lead ions concentration, flow rate, temperature and cell voltage	Separation percent of lead ions	4:5:4:1; two hidden layers	Mean absolute error below 1%	(Sadrzadeh *et al.*, 2007a)
Lead ions	Lead ions concentration, flow rate, temperature and cell voltage	Separation percent of lead ions	4:5:4:1; two hidden layers	MSE=0.102 R^2=0.999	(Sadrzadeh *et al.*, 2008)
Lead ions	Lead ions concentration, flow rate, temperature and cell voltage	Separation percent of lead ions	4:6:2:1; two hidden layers	Standard deviation not more than 1%	(Sadrzadeh *et al.*, 2009)
Saline wastewater	Time, concentrations of NaCl, Fe^{2+}, and H_2O_2	TOC/TOC_0	Multiple ANN topologies	R^2=0.960	(Borges *et al.*, 2009)
NaCl separation	Feed concentration, flow rate, temperature and cell voltage	Separation percent	Multiple ANN topologies	MSE<0.3 R^2=0.99	(Jing *et al.*, 2012)

2.2.2 Size of data sets

Data-driven AI techniques highly depend on the quantity and quality of the data sets fed into them. In other words, it is required to have enough reliable data to reasonably capture the relationships both between input variables and between input and output variables. It should be mentioned that the size of data sets required for machine learning approaches depends on the complexity of the problem and complexity of the learning algorithm, and there are no advanced certainties about the amount of data required for these approaches. Since data used for modelling and optimization of electrochemical processes for water and wastewater treatment processes are derived mainly from experimental studies, acquiring sufficient large data sets requires a huge amount of time and resources. Figure 2.1 shows the distribution of the number of samples in data sets used in the field in literature. As can be seen, most of the studies have implemented AI techniques with a relatively low number of samples (<150) in data sets. Hence, it can be concluded that considering the amount of data available, most of the effort should be focused on the reliability and robustness of the AI models derived from these data sets.

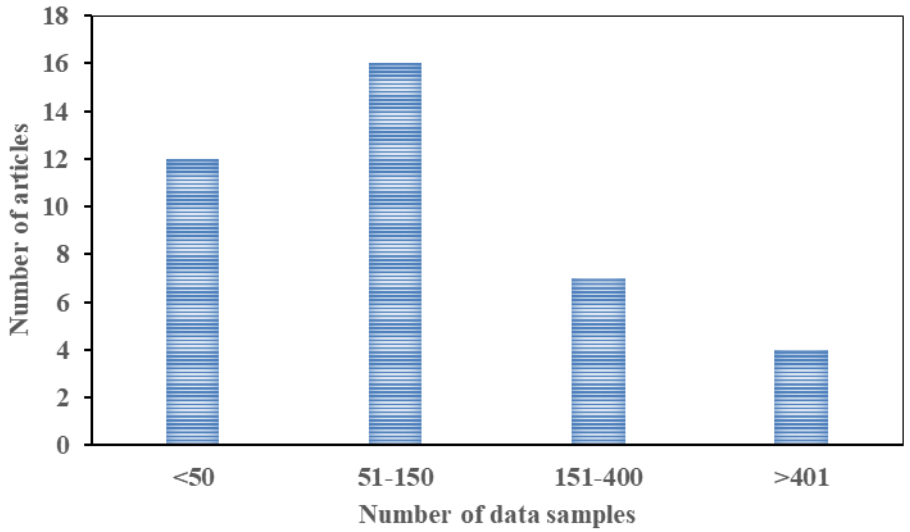


Figure 2.1 Frequencies of articles in literature regarding the size of the data sets

However, to overcome the limitation of a low number of data in AI modelling based on the experimental results, some authors have proposed using data augmentation techniques

such as interpolation (Farzin *et al.*, 2020). When insufficient data sets are not enough to learn many of the parameters of learning algorithms, it would cause overfitting meaning that the generalization of the model is unreliable. In order to solve this problem, more data needs to be collected, but in actual applications, additional data collection is often difficult for various reasons such as time and cost limitations. Data augmentation is a solution to address this (Oh *et al.*, 2020). Although data augmentation techniques have been applied to machine learning in different fields in literature, especially image processing and speech recognition (Abayomi-Alli *et al.*, 2020; Li *et al.*, 2020c; Qian *et al.*, 2019), one should be cautious about using these techniques for the goal of regression of experimental work with limited data. This is because the behaviour of outputs in experimental studies can be much more complicated than describing them with predefined interpolation functions, which obviously would not be too hard for AI model to predict the behavior of the new interpolated data.

2.2.3 Data preprocessing

Experimental data obtained in electrochemical processes are used by AI models as inputs and outputs. Those independent and dependent experimental variables will be used as inputs and outputs, respectively. While various independent variables have been used in electrochemical processes, Figure 2.2 shows the most common ones specified as inputs. As can be seen, electrolysis time and applied current have been the most frequent input variables for AI modelling of electrochemical processes. Other independent variables with a low number of frequencies used as inputs were feed flowrate, temperature, mixing speed, nature of the electrolyte and type of the pollutant.

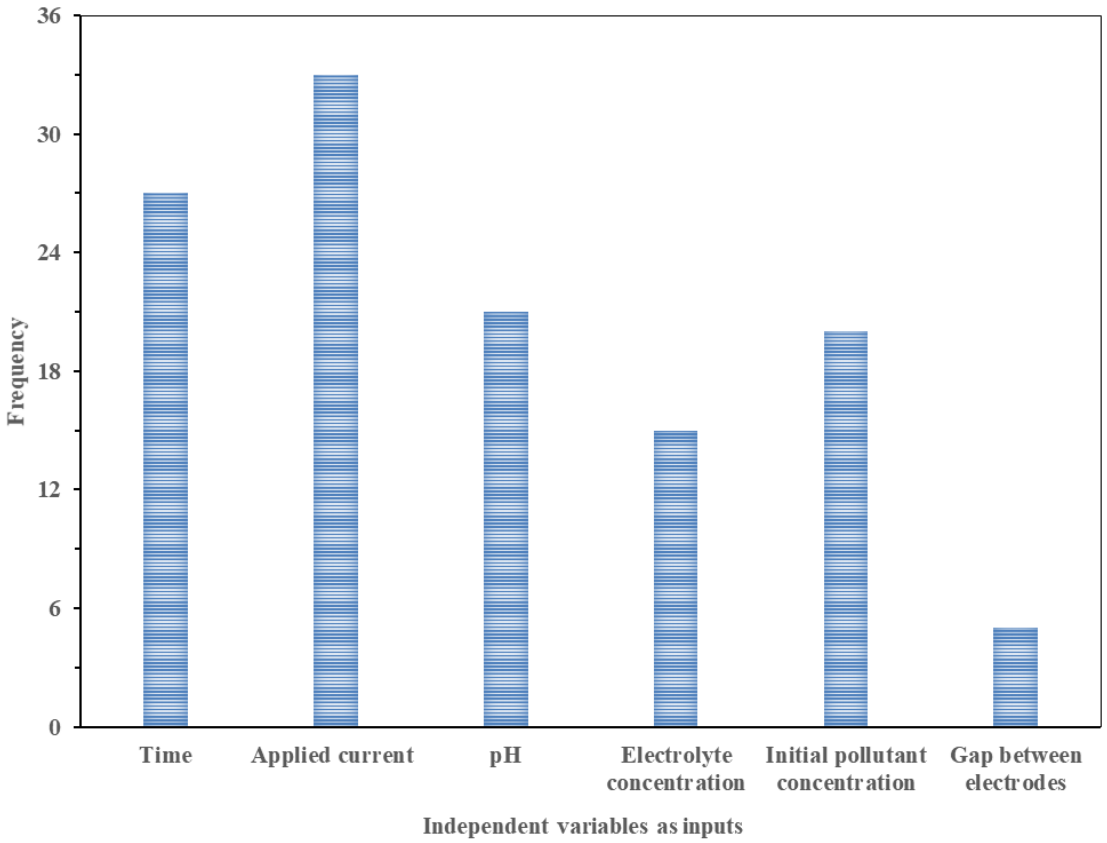


Figure 2.2 Frequencies of different independent variables used as AI model inputs in literature

Feature scaling has often been used to scale the variables in the dataset. If the input and output variables are not of the same order of magnitude, some variables may appear to have more significance than they actually do. The training algorithm has to compensate for order-of-magnitude differences by adjusting the network weights which is not very effective in many of the training algorithms (i.e., the backpropagation algorithm in ANN) (Baughman *et al.*, 1995).

Normalization and standardization have been utilized as feature scaling techniques in the reviewed studies. In the common normalization technique, so-called Min-Max scaling, values are shifted and rescaled so that they end up ranging between 0 and 1 (Ghaedi *et al.*, 2017). In standardization, values will be centred around the mean with a unit standard deviation meaning that the mean of the feature becomes zero while the resultant distribution has a unit standard deviation (Gholami Shirkoohi *et al.*, 2021).

2.3. Performance Evaluation

To evaluate the performance of the ANNs, there are different prediction accuracy criteria used in the literature. The most utilized criteria in the publications studied in this review for the performance evaluation of the networks are listed in Table 2.5.

Table 2.5 Performance evaluation criteria mostly used in the literature

Performance evaluation criterion	Equation
R (coefficient of correlation)	$\frac{\sum_{i=1}^{N}(P_i - P_m)(E_i - E_m)}{\sqrt{\sum_{i=1}^{N}(P_i - P_m)^2}\sqrt{\sum_{i=1}^{N}(E_i - E_m)^2}}$
R^2 (coefficient of determination)	$1 - \frac{\sum_{i=1}^{N}(E_i - P_i)^2}{\sum_{i=1}^{N}(E_i - E_m)^2}$
MSE (mean squared error)	$\frac{\sum_{i=1}^{N}(P_i - E_i)^2}{N}$
RMSE (root mean squared error)	$\sqrt{\frac{\sum_{i=1}^{N}(P_i - E_i)^2}{N}}$
SSE (sum of squared errors)	$\sum_{i=1}^{N}(E_i - P_i)^2$
MAPE (mean absolute percentage error)	$\frac{100}{N}\sum_{i=1}^{N}\left\|\frac{(E_i - P_i)}{E_i}\right\|$
AARE (average absolute relative error)	$\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{(E_i - P_i)}{E_i}\right\|$
Relative error	$\frac{E_i - P_i}{E_i}.100$

E_i is the ith experimental value, and P_i is the ith predicted value obtained from the neural network model, E_m is the mean value of all experimental values, P_m is the mean value of all predicted values, and N is the number of data samples.

2.4. AI TECHNIQUES

AI techniques in literature applied to electrochemical processes for water and wastewater treatment processes are reviewed in this section. These include ANNs, SVM, ANFIS and evolutionary algorithms.

2.4.1 ANNs

As the name itself implies, artificial neural networks, commonly referred to as “neural networks”, imitate the essential characteristics of the human brain, which itself is a highly nonlinear, complex, and parallel computer, such as self-adaptability, self-organization, and error tolerance (Haykin, 1998; Singh *et al.*, 2009). Multilayer perceptrons (MLP) feedforward neural networks are the type of ANNs that have been utilized frequently for modelling electrochemical processes. These networks that can imitate human brain activities consist of an input layer of source nodes, one or more hidden layers of computation nodes and an output layer that can include multiple nodes (Assefi *et al.*, 2014; Elsheikh *et al.*, 2019; Feng *et al.*, 2008). These networks are usually applied to diverse problems, including function approximation, pattern classification, system identification, process control, process optimization, and so on (Carvalho *et al.*, 2011; Da Silva *et al.*, 2017; Zarra *et al.*, 2019).

The nature and complexity of the problem in addition to the desired accuracy and the available data, determine the number of hidden layers and the corresponding number of neurons in each hidden layer. The selection of an appropriate number of neurons in the hidden layer is a crucial task for MLP neural networks since too many neurons can cause the so-called over-fitting problem (Madhiarasan *et al.*, 2017). In this case, the network has memorized the training data but has not exploited its generalization ability (Valente

et al., 2014). Although there are some indications for the hidden layer dimension, in practice, a trial-and-error procedure is usually applied (Padovese, 2002).

Single hidden layer MLP networks have been considered sufficient enough to correlate inputs to outputs in most of the electrochemical processes studied for water and wastewater treatment (e.g., (Alavi *et al.*, 2019; David *et al.*, 2020; Gholami Shirkoohi *et al.*, 2021; Mei *et al.*, 2019; Sangal *et al.*, 2015; Valente *et al.*, 2014). Soloman *et al.* (2010) developed an ANN model to predict the electro-oxidation of malachite green, a triphenyl methane dye, based on experimental data collected in a batch electrochemical reactor. A three layer back-propagation network with a 3:9:1 configuration of was found adequate to predict the COD removal efficiency with R=0.9987, and RMSE=1.1428 (mean experimental value=61.25). Daneshvar *et al.* (2006) investigated the use of artificial neural networks for modelling the colour removal from a textile dye solution containing C. I. BY28 by the electrocoagulation method. Comparison between experimental results and three layer 7:10:1 neural network model predicted outputs showed a correlation coefficient of 0.974, which proved that an ANN model can successfully describe the behaviour of such an electrocoagulation system containing complex reactions.

Multiple hidden layers instead of a single hidden layer were also considered for correlating inputs to outputs (Manokaran *et al.*, 2014; Picos-Benitez *et al.*, 2020; Sadrzadeh *et al.*, 2008; Sadrzadeh *et al.*, 2009). Manokaran *et al.* (2014) also used a feedforward back-propagation ANN model to predict the degradation of a distillery effluent by electro-oxidation. In total, 200 data sets were used for training (150 samples), validation (30), and testing (20) using the scaled conjugate gradient (trainscg) algorithm for training. The performance of four different configurations of three layer networks (1 hidden layer) and four different configurations of four layer networks (2 hidden layers) were compared. They showed that the four layer 3:3:3:1 BP neural network had the best performance for COD removal: RMSE=0.8633, AARE=3.4613, R=0.9987. Comparing regression and ANN models, Radwan *et al.* (2018) showed that ANN model performs slightly better ($R^2_{regression}$=0.9525, R^2_{ANN}=0.9742) for modelling an EF process for the treatment of phenolic wastewater.

While the previous studies examined a single optimum network for predicting the process outputs, some authors suggested using multiple networks or so-called stacked neural networks. Stacked neural networks as an ensemble machine learning technique that have been used in other supervised methods such as SVM, k-nearest neighbours algorithm or decision trees (Wolpert, 1992), are based on the premise that the use of multiple networks, instead of simply just one single network, can be an optimal network and improved predictions can be obtained (Torres-Sospedra *et al.*, 2006). Thus, combining the outputs of different neural network models which each capture certain aspects of the process and aggregating their information, can provide more accurate predictions (Figure 2.3).

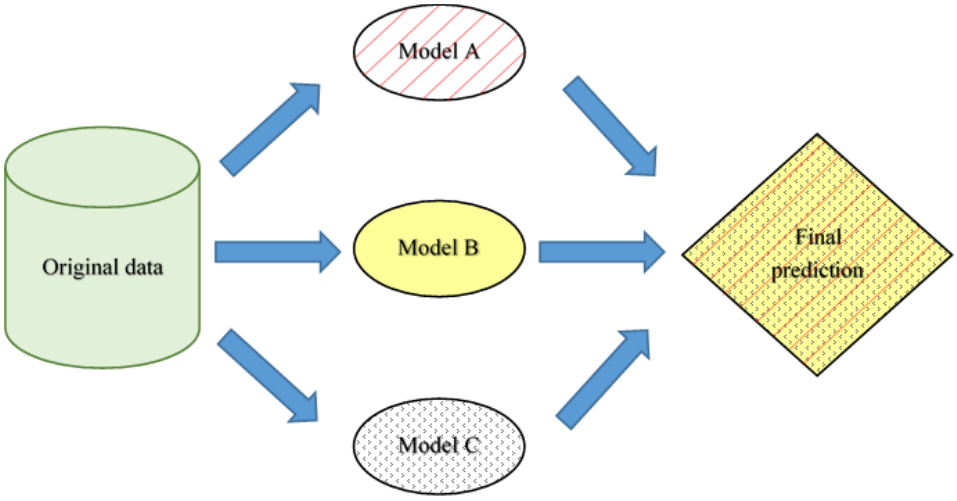


Figure 2.3 Scheme of stacked neural networks as an ensemble machine learning technique

Piuleac *et al.* (2010) applied stacked neural network modelling to the electrolysis of wastes polluted with phenolic compounds, including phenol, 4-chlorophenol, 2,4-dichlorophenol, 2,4,6-trichlorophenol, 4-nitrophenol, and 2,4-dinitrophenol. In their work, various ANN types of artificial neural networks were aggregated in a stack whose output response was a weighted sum of the individual networks. A comparison between the tested methodologies indicated that utilizing stacked neural networks or the assembly of neural networks could obtain smaller validation errors of 5.8% and 4%, respectively, rather than a single optimal MLP neural network. Curteanu *et al.* (2011) studied the idea of stacked neural network modelling for the removal of chlorophyll a from the final effluent of aerated lagoons by an electrolysis process which consisted of both mechanisms of

electro-flotation and electro-coagulation. Three different approaches were studied, including: using a single neural network for the whole data set, three separate neural networks for different sets of data for each output, and a stacked neural network combining the outputs of three selected individual networks. The best result among the three approaches was attributed to the stacked neural network with relative errors for each of the three outputs: $\mathrm{E_{r,TSS}} = 5.89\%$, $\mathrm{E_{r,chlorophyll\ a}} = 14.86\%$ and $\mathrm{E_{r,COD}} = 9.55\%$. With their good results, they concluded that their modelling methodology could be adapted and applied for other treatment processes and systems.

The response surface methodology (RSM) and ANN models were compared in terms of their performance in modelling of electrocoagulation processes (Bhatti *et al.*, 2011a; Gholami Shirkoohi *et al.*, 2021; Nourouzi *et al.*, 2011). Nourouzi *et al.* (2011) employed a three layer ANN model to predict the removal of Reactive Black 5 dye by a sequential electrocoagulation-flocculation process. The results obtained using the ANN model were compared with the RSM and showed that both models are able to predict the process, the ANN having a slightly better performance than the RSM model (R^2= 0.9764 and 0.9446, respectively).

Within the scope of process optimization, by process control, Pinto *et al.* (2018) applied an ANN feedforward controller to a hybrid system of electrocoagulation and organic coagulation for removing Reactive Blue 5G dye from textile effluent. Their ANN was trained by the LM method using data sets created by the central composite rotatable design (CCRD) model. The ANN based controller could manipulate the current intensity and organic coagulant dosage to act upon a disturbance in the influent dye load. Although the controller was not able to keep the controlled variable perfectly at the set point, its action was positive based on the experimental data, which demonstrated the controller to be able to either reduce the oscillation or keep the controlled variable in regions closer to the set point, when comparing to the system without the controller.

In the domain of controlling electro-Fenton processes using artificial neural networks, Yu *et al.* (2013) studied textile wastewater treatment using on-line monitoring of dissolved oxygen (DO) and oxidation-reduction potential (ORP). Their research was in line with their previous efforts on using artificial neural networks to control the Fenton process, both in

batch and continuous operation mode, for textile wastewater treatment (Yu *et al.*, 2009; Yu *et al.*, 2010). In their study, two feedforward back-propagation ANNs were used to predict the Fe^{2+} dosage requirement and COD removal efficiency. One ANN predicted the Fe^{2+} dose based on the following inputs: reaction time to reach the ORP valley (min), the time for DO rising point (min), the ORP value at the ORP valley (mV), and the desired COD removal efficiency (%), with a 4:8:1 configuration. Their efforts to demonstrate the ANN's capability for EF process control was a step forward in the application of ANNs in wastewater treatment.

In the scope of utilizing artificial neural networks for process integration, Borges *et al.* (2009) used the ANN approach to model an integrated electrodialysis and photochemical process for saline wastewater treatment. Two three layer feedforward artificial neural networks were put in series to model the photo-Fenton process. The first (4:4:1) neural network was responsible for modelling the output values of TOC/TOC_0 as a function of the input parameters time, concentrations of NaCl, Fe^{2+}, and H_2O_2. The output value of the first ANN was sent to the second neural network to calculate the reaction rate with input parameters TOC/TOC_0, concentrations of NaCl, Fe^{2+}, and H_2O_2. This model was used to design a plug flow reactor and to determine its volume (V), for different process conditions and TOC reaction rates. Their work using neural networks showed an essential step in understanding the behaviour of the integrated process.

2.4.2 SVR

Support vector machines (SVM), first presented by Vapnik (1963), based on modern statistical machine learning techniques, have been widely applied to classification and regression problems thanks to their promising generalization performance (Saradhi *et al.*, 2007). In a simple binary classification problem, the basic idea of an SVM is to find a hyperplane that has the maximum distance (margin) from both sides of the hyperplane. SVM can be adopted for regression problems, thus called support vector regression (SVR). To reach the aim of estimating the predictor function (*f*), a loss function (*L*), which assesses the quality of a prediction (*f(x)*), is introduced as *L(x, y, f(x)).* There are different popular loss functions utilized for regression problems, such as *ε-insensitive loss, Huber's loss, logistic loss,* and *pinball loss* (Van Messem *et al.*, 2010). In SVR, the goal is to obtain

a function that has the maximum ε (difference) among all targets and at the same time is as flat as possible, meaning that the errors are not significant as long as they are less than ε, but any deviation larger than this amount is not accepted (Smola *et al.*, 2004). A detailed description of the SVM algorithm and its parameters can be found in the literature (Çevik *et al.*, 2015; Ding *et al.*, 2011a).

Curteanu *et al.* (2014) applied two machine learning techniques (artificial neural networks and support vector machines) for the prediction of the performance of an electro-oxidation method to decrease the organic compounds and remove micro-organisms from activated sludge effluent. Although very good results were obtained by different combinations of ANNs, a minimum R^2 of 0.90 and a maximum MSE of 0.04 in the testing phase, the SVM obtained more accurate results after many trial and error examinations. It was reported that overall, the SVM outperformed the ANN models when comparing correlation coefficients.

Farzin *et al.* (2020) applied different approaches to data mining, including least square support vector machine (LSSVM) used for electrochemical removal of Ciprofloxacin (CIP) as a model pollutant. LSSVM needs to solve quadratic programming with only equality constraints, or equivalently a linear system of equations, which makes it simpler and faster than SVM (Tian *et al.*, 2012). They showed that their tuned LSSVM model has superiority over other investigated algorithms for their problem. SVM was also used by Yuan *et al.* (2006) for predicting the electrochemical degradation of substituted phenols by developing a quantitative structure-property relationship model. Their SVM model had a good predictive ability for the quantitative relationship between rate constants and the structure of substituted phenols with a performance of RMSE=0.202 and R^2=0.892.

2.4.3 ANFIS

Adaptive neuro-fuzzy inference system (ANFIS) (Jang, 1993) combines advantages of both ANN (e.g., learning, self-adapt, optimization and connectionist structures) and fuzzy logic (e.g., humanlike “if–then” rules thinking, reasoning, and incorporating expert knowledge). ANFIS is a kind of feed-forward artificial neural network that each layer is a neuro-fuzzy system component based on a fuzzy inference system. By employing a

hybrid-learning method, the ANFIS architecture can build an input-output model based on constructing a set of fuzzy if-then rules with appropriate membership functions (according to human knowledge) to produce the conditional input-output pairs (Ghaedi *et al.*, 2017). Details of the ANFIS can be found in the literature (Jang, 1993; Jang *et al.*, 1993).

ANFIS and RSM models were applied for optimization of Reactive Blue 19 removal using a combined electrocoagulation/coagulation process. The effects of five independent parameters, including applied current, reaction time, initial dye concentration, initial pH and dosage of Poly Aluminum Chloride, were studied by RSM. According to the results, all the independent parameters were equally important in dye removal efficiency. In addition, ANFIS was applied for dye removal efficiency and operating costs modelling. High R^2 values (85%) indicate that the predictions of RSM and ANFIS models are acceptable for both responses.

A comparison of ANFIS and other AI techniques was carried out by Farzin *et al.* (2020) for the modelling of Ciprofloxacin electrochemical removal from wastewater. While the interpolation method was used for increasing the number of data, ANFIS performed better than ANN for in interpolated data which was in accordance with some other studies (Sahoo *et al.*, 2020; Souza *et al.*, 2018). ANFIS, along with ANN, has also been studied for the treatment of greywater using electrocoagulation by Nasr et al. (Nasr *et al.*, 2016). Their three layer (3:6:1) ANN model successfully simulated the experimental results with R-values of 0.99 (training), 0.84 (validation) and 0.89 (testing) to predict the removal efficiency of turbidity. Their ANFIS application performed an exhaustive search within the available inputs to determine the most influential input attribute in predicting the turbidity removal. It was indicated that current density is the most influential input on turbidity removal.

2.4.4 Evolutionary algorithms

Evolutionary algorithms, mainly GA and PSO, have been utilized for the optimization of electrochemical processes for water and wastewater treatment. However, recently other nature-inspired algorithms like the fire fly optimization algorithm (FFA) have been utilized

by researchers (Farzin *et al.*, 2020). EA as optimization techniques have been applied for process output optimization and hyperparameter selection of AI models, especially in the case of ANN. Figure 2.4 represents ANN-EA approaches for finding optimal process conditions and optimization of hyperparameters of ANN models.

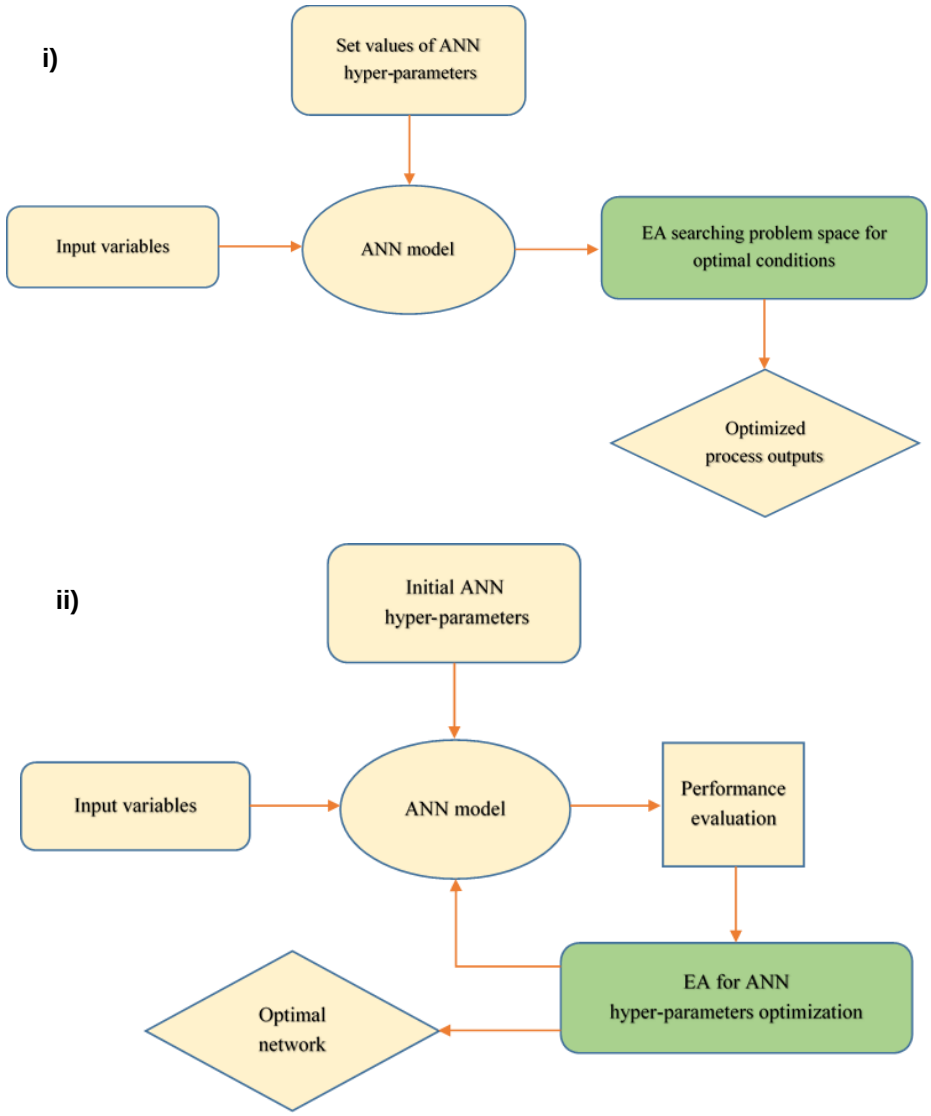


Figure 2.4 ANN-EA approaches for: i) finding optimal process conditions, ii) optimization of hyperparameters of ANN models

The ANN-GA approach for electrooxidation process output optimization to find optimal conditions has been applied by some authors (Maleki *et al.*, 2015; Picos-Benitez *et al.*, 2020; Picos *et al.*, 2018). Picos *et al.* (2018) used this approach for the prediction of discoloration of a dye by an electro-oxidation process in a press type reactor. A total of 52 experiments (70% for the training and 30% for the testing) were used for the ANN

modelling with a Levenberg-Marquardt back-propagation (LM-BP) algorithm resulting in an ANN configuration of 4:9:5:1 and sigmoid transfer functions (logsig at the hidden layer and tansig at the output layer). The ANN with performance MAPE=8.3868% and RMSE=7.5537% values was linked to GA optimization to find the best operational conditions, where the EO can reach a maximum discoloration at the lowest current density, flow rate, experimental time and at the highest dye concentration. They experimentally validated the ANN-GA result that about 95% discoloration can be obtained in an experimental time of 110 min, a flowrate of 12 Lps, a current density of 27.34 mA/cm^2 and a dye concentration of about 230 mg/L. The Same group studied the ANN-GA approach for the prediction of discoloration of Bromophenol blue dye for an electro-oxidation process (Picos-Benitez *et al.*, 2020). Mean discoloration efficiency of 88.8 %, compared to 95.5 % predicted by the model, could be obtained at the optimal conditions. Similar discoloration efficiencies were obtained, which proved that this AI model could be used as a helpful tool in the design, control and operation of similar EO processes to wastewaters with similar dyes.

In the scope of electrocoagulation process optimization, Taheri *et al.* (2015) used ANN modelling and a GA algorithm to improve the Taguchi design optimization for the degradation of three different dyes, including Acid Orange 7, Acid Brown 14, and Acid Red 18 azo dyes by electrocoagulation. In their study, a multilayer feedforward back-propagation ANN with 10 neurons in the hidden layer and using tansig and purelin transfer functions at hidden and output layers, was used. A total of 241 data sets were utilized for ANN modelling, and the model showed excellent reliability to predict the EC system with R^2=0.9804. A GA was used for techno-economical optimization of the Taguchi design for dye removal. Their GA used the ANN model to search for the best conditions for removal efficiencies between the minimum and maximum levels of the Taguchi design. The operating cost for the optimal conditions was calculated by the regression of the Taguchi method. Their GA optimization results showed removal efficiencies of 96.79% and 76.74% for Acid Orange 7 and Acid Red 18, respectively, at nearly the same operating conditions. At a lower operating cost, a 98.12% removal efficiency was reported for Acid Brown 14. Their work illustrated the ANN and GA approach to be a powerful tool for techno-economical optimization of selected dye removal using the EC process.

When there are multiple responses to consider, the problem shifts to a multi-objective optimization problem. There is no unique solution to a multi-objective optimization problem but a set of mathematically equally good solutions known as nondominated or Pareto optimal solutions. Bhatti *et al.* (2011a) used multi-objective optimization by genetic algorithms for electrocoagulation of copper from simulated wastewater. Their system was modelled by both RSM and ANN modelling approaches. Despite the limited experimental data, the 4:5:2 ANN model performed as well as the RSM (R^2=0.993 for copper removal efficiency and R^2=0.870 for energy consumption) to describe the nonlinearities of the electrocoagulation process, with MSE=0.571 and combined regression coefficient of 0.982 for copper removal efficiency and energy consumption. A genetic algorithm linked to the ANN model was utilized to derive the Pareto front, which defined a set of optimum operating points with respect to removal efficiency and energy consumption. Their multi-objective optimization linked to the ANN model resulted in an insight into the optimal operating conditions of the process. Maleki *et al.* (2014) also developed a combination of ANN models and multi-objective genetic algorithm optimization for the electrocoagulation of orange 25 dye (DO25) in a batch process. Based on their results, the optimum value for initial dye concentration in the Pareto front was almost the minimum value since higher values of initial dye concentration would contribute to lower colour removal and also higher energy consumption.

Multi-objective PSO algorithm has also been used for techno-economical optimization of combined electrocoagulation/coagulation's performance in the removal of RB 19 from simulated wastewater using the ANFIS model (Taheri *et al.*, 2013). Minimum and maximum values of 58.27% and 99.67% for RB 19 removal efficiencies were reported by the selected ANFIS model, respectively. The difference between the minimum and maximum dye removal efficiency levels for operating costs was 0.39 US$/m3.

2.5. INSIDE THE BLACK-BOX MODELS

2.5.1 Tuning AI model parameters

AI models have inherent hyperparameters that should be tuned so that the model can optimally solve the machine learning problem. These hyperparameters control the

learning process and have a direct effect on the model performance. As most of the studies have focused on ANN modelling approach, Figure 2.5 shows the hyperparameters of an ANN model.

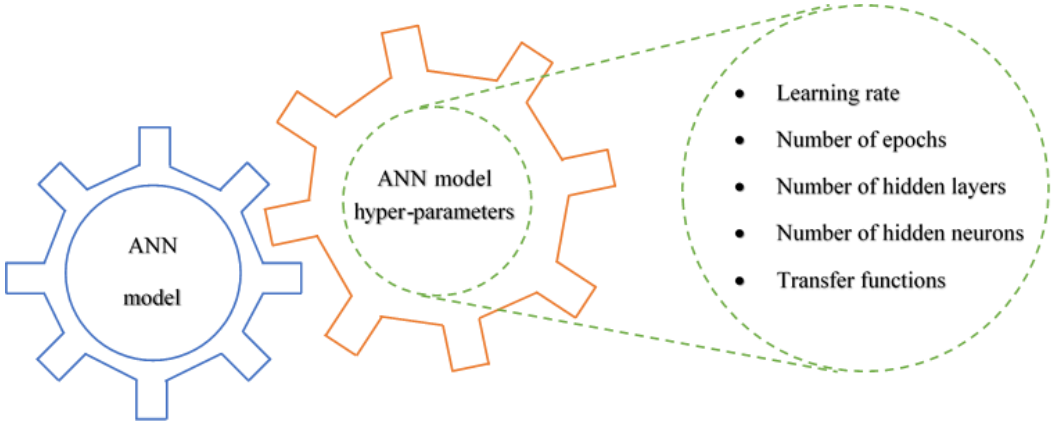


Figure 2.5 Hyperparameters of an ANN model

The network configuration, i.e. the number of hidden layers and hidden neurons, has received the most attention (Aber *et al.*, 2009; Bui, 2016; da Silva Ribeiro *et al.*, 2019; Hasani *et al.*, 2018; Keskin *et al.*, 2011; Mirsoleimani-Azizi *et al.*, 2015; Nourouzi *et al.*, 2011; Valente *et al.*, 2014). In most of the studies, the coefficient of determination and MSE were chosen as criteria for network performance.

Valente *et al.* (2014) studied the prediction of COD concentration in dairy industry effluent treated by electrocoagulation using artificial neural networks. In order to select an appropriate number of neurons in the hidden layer to prevent overfitting and loss of the network's generalization ability, several ANN architectures were evaluated using MSE and correlation coefficient as performance parameters. A neural network with 9:10:1 configuration was selected with MSE=0.00406 and R^2=0.9560 for the test set. According to their results based on ANN simulation, the efficiency of the COD removal can be described as a function of time, pH, current density and distance between electrodes.

Single hidden layer networks with a trial and error procedure on the network configuration were utilized for correlating inputs to outputs (Ahmed Basha *et al.*, 2010; Soloman *et al.*, 2010). Ahmed Basha *et al.* (2010) used ANNs for modelling of the electro-oxidation process applied to an effluent of a specialty chemical manufacturer, which was highly loaded with organic matter (COD: 48,000 mgL^{-1} and BOD_5: 1100 mgL^{-1}). In their work, a

single hidden layer network with 3:7:1 configuration led to a reasonable prediction of the COD removal efficiency, with R=0.9977, and RMSE=0.8378 (mean experimental value=53.59). It was shown that an increase in number of hidden neurons can enhance the performance of the three layered network but can have an adverse effect on the performance of the four layered network. The importance of the number of hidden layers and hidden neurons were also investigated in other studies (Soloman *et al.*, 2010).

The trial and error procedure was also applied by other authors to determine the optimum number of hidden layer neurons based on different error functions (Belkacem *et al.*, 2017; Kothari *et al.*, 2020; Manokaran *et al.*, 2014; Sangal *et al.*, 2015). Sangal *et al.* (2015) developed a three-layer ANN model to predict the removal of CBSOL LE red wool dye from wastewater by electro-oxidation. The optimal 3:8:3 ANN architecture could estimate the outputs with a correlation coefficient of 0.995, 0.996, 0.992, and 0.995 for training, validation, testing, and all data sets, respectively. It was reported that the proposed ANN could accurately simulate the outputs from given inputs.

Other than the number of hidden layers and hidden neurons in each layer which have been widely considered in ANN modelling, the initial weights are another important factor that affects the performance of the network. Choosing an improper set of initial weights can lead to local minima, which results in bad performance of the network. This effect has been rarely considered in ANN modelling studies for water and wastewater treatment using electrochemical processes. In their two studies, Sadrzadeh et al. (2008; 2009) took this point into account by performing 20 runs using different random values of initial weights for each of their different structured networks based on hidden layer and hidden neuron numbers. This approach can lead to reducing the uncertainties related to the neural networks.

They also studied the effect of different transfer functions of hidden and output layers on the performance of the network. Transfer functions used as the neuron activation function to the sum of weighted inputs and biases are one of the neural network hyperparameters that can affect the network performance. Generally, three transfer functions have been used for MLP neural networks: log-sigmoid, tan-sigmoid, and purelin (equations 4-6)

$$logsig(n) = \frac{1}{(1 + \exp(-n))} \quad (2.10)$$

$$tansig(n) = \frac{2}{(1 + \exp(-2 * n))} - 1 \quad (2.11)$$

$$purelin(n) = n \quad (2.12)$$

Piuleac *et al.* (2010) illustrated that a transfer function combination for hidden and output layers performed better than the single transfer function for all layers. Their optimal network was then tested with a real wastewater of a fine-chemicals plant and showed an average error around 4.92% between experimental and predicted COD concentrations, which gave a very good illustration of using neural networks in the case of wastewater treatment.

The effect of transfer functions in hidden layers on the neural network performance was investigated again by the same team for the electrolysis treatment of wastewater polluted by phenol compounds (Piuleac *et al.*, 2012). Based on trial and error, different configurations and architectures of neural networks were examined and based on validation stage results, the MLP (7:25:20:1) with tansig transfer function for all hidden and output layers, obtained the best relative error of about 8.93% and a correlation value of 0.998.

Still aiming to find optimal ANN structures da Silva Ribeiro *et al.* (2019) studied an artificial neural network for the prediction of boron removal from mining wastewaters by electrocoagulation. Different types of transfer functions and network structures were examined in their study to observe their performance. The 3:10:1 network with a logsig transfer function in the hidden layer and a purelin transfer function in the output layer, showed the best performance based on the correlation coefficient (R^2) and the sum of squared error (SSE) with values of 0.973 and 0.616, respectively.

One of the most thorough studies on the effect of various network architectures and parameters on the modelling performance was performed by Hasani *et al.* (2018) for the modelling of alternating pulse current electrocoagulation-flotation (APC-ECF) for humic acid (HA) removal. Their study focused on the effect of various network architectures and

parameters (e.g., two different ANN architectures as MLP and generalized feedforward-GFF, number of hidden neurons, transfer functions, and learning parameters) on the modelling performance. Their extensive comparisons between different networks revealed that the single hidden-layer GFF NN (5:6:1), using sigmoid transfer function at both hidden and output layers and LM training algorithm, had the best performance with R^2=0.999 and MSE=0.00006. Their computational analysis proved that ANN-based modelling can effectively simulate the experimental data and predict the optimum conditions of the electrocoagulation/flotation process for the removal of HA from aqueous solutions.

As mentioned before, EA optimization techniques can be utilized to find the optimal configured network by searching in the hyperparameter space of the neural network (Farzin *et al.*, 2020; Mei *et al.*, 2019; Viana *et al.*, 2018). Viana *et al.* (2018) presented artificial neural networks and statistical analysis to predict and optimize the electrochemical degradation of the textile dye Reactive Black 5 using a $\mathrm{Ti}/(\mathrm{RuO_2})_{0.8} - (\mathrm{Sb_2O_3})_{0.2}$ in a batch treatment system. By using the PSO algorithm, they optimized their neural network model parameters, including hidden neuron number, transfer function, and learning rate. Their 4:8:3 neural network could successfully predict colour removal, COD removal, and energy consumption for the textile dye Reactive Black 5 degradation with the performance of $\mathrm{R^2_{test}}$=0.982, $\mathrm{MSE_{test}}$=0.0146.

In the scope of EA, it is worth noting that different values of the GA control parameters can have significant impacts on the optimal results obtained. Piuleac *et al.* (2013) studied an ANN-based optimization methodology in detail, including the impacts of the genetic algorithm parameters, to optimize an electro-coagulation process involving three different pollutants of: kaolin, Eriochrome Black T solutions, and an oil/water emulsion. Time, current density and initial pH were considered as decision variables for the GA optimization alongside the size of the initial population, the number of generations, crossover rate, and mutation rate as GA control parameters. To observe the impacts of these GA control parameters, they conducted different series of optimizations with different values for these control parameters. Various scenarios with different sets of GA control parameters were developed in order to select the most convenient working

conditions regarding the decision variables. The ANN-GA approach was found to be an efficient optimization method for their EC process and could predict the optimal conditions for maximum removal efficiency of the three pollutants with a maximum relative error of 11.46% and an average relative error of 6.61%.

2.5.2 Regularization techniques to prevent overfitting

The selection of an appropriate number of neurons in the hidden layer is a crucial task for MLP neural networks since too many neurons can cause the so-called overfitting problem. In this case, the fitting error on the training set will be very low due to the very successful learning process, but the error on new data presented to the network is very high. The network has memorized the training data but has not exploited its generalization ability (Valente *et al.*, 2014). Regularly, to obtain good network generalization, the method is to propose a network which is large enough to provide an appropriate fit. Although it is difficult to have the perspective to know how large a network should be in each case, three generalization learning methods of cross-validation (early stopping), regularization, and pruning can be applied. Regularization is conducted by adding a penalty function to the training objective to minimize the complexity of the model and the prediction error at the same time; while pruning physically omits some excessive neurons to generate the least size network. For the cross-validation (early stopping) method, the data set will be split into three non-overlapping subsets. The training dataset is utilized for learning the network parameters, the validation dataset is utilized for monitoring the training process and for approximating the generalization error, and the test dataset, a set of data not seen by the model during training, is utilized for examining the unbiased generalization error of the trained network. In the early-stopping method, when the validation error rises over a number of iterations (due to over-fitting), the training algorithm stops, and the values of the weights and biases are returned to the point where the validation error was minimal (Chan *et al.*, 2006; Zhang *et al.*, 2003).

While early-stopping method has been used in most of the studies in the domain, recently Gholami Shirkoohi *et al.* (2021) applied regularization method to their problem of modelling and optimization methodology for active chlorine production using the electrolysis process. Learning curves were used to diagnose whether there is a high bias

(underfit) problem or a high variance (overfit) issue. In the presence of a high variance problem, using the regularization factor can help. Regularization makes slight modifications to the learning algorithm such that the model generalizes better and the model's performance on unseen data is improved. They showed that utilizing learning curves along with regularization factor analysis can help to obtain reliable ANN models to predict the production of active chlorine and energy consumption using an electrolysis process.

2.5.3 Sensitivity analysis

The weights obtained from ANN training are coefficients between artificial neurons that are analogous to synaptic strengths between the axon and dendrites in a biological neuron in the brain. As in real life, the proportion of the incoming signal to be transmitted to the neuron's body is decided by these weights (Khataee *et al.*, 2010). Despite the black-box nature of ANNs, to estimate the influence of different independent variables on the output, it is possible to conduct a sensitivity analysis on the ANNs. The relative importance of each input independent variable on the desired output can be obtained through the neural connection weight matrix. First, Garson (Garson, 1991) and then Goh (Goh, 1995) proposed a procedure for partitioning the connection weights to determine the relative importance of the various inputs. This method basically involves partitioning the hidden-output connection weights of each hidden neuron into components associated with each input neuron (Gevrey *et al.*, 2003).

Belkacem *et al.* (2017) reported that neural network modelling can effectively forecast the electro-oxidation of oxytetracycline (OTC) in a batch process using a platinized titanium anode. A 5:14:1 neural network was selected as the best network with performance MSE=0.0002 and R=0.99. Furthermore, they studied the effect of input parameters based on the Garson algorithm. They showed that the reaction time has the most influence on the process output with a relative importance of 50.70% followed by the current intensity and the nature of the electrolyte, 15.24%, and 14%, respectively.

For an electrocoagulation process, Aber *et al.* (2009) modelled the removal of Cr(VI) from polluted solutions using artificial neural networks. With a coefficient of determination

R^2=0.976, their 4:10:1 neural network could successfully predict the present system. Also, the Garson algorithm was employed to evaluate the relative importance of the input variables and the results showed that all input variables have significant effects on the removal of Cr(VI). In further work, Bui (2016), applied artificial neural networks to predict dye removal efficiency (colour and COD) of electrocoagulation for a Sunfix Red S3B aqueous solution. The proposed 5-7-2 ANN model optimized by the Bayesian regulation algorithm (trainbr) could successfully predict the electrocoagulation process with a correlation coefficient of R^2=0.836, RMSE=9.844% and MAPE=13.776%. A sensitivity analysis showed that the efficiency of the EC process is highly dependent on current density, electrolysis time and initial pH for colour removal whereas it is highly dependent on initial dye concentration, sulphate concentration, and electrolysis time combined with the initial pH for COD removal.

For EF processes, one study showed that while all of the independent variables have a strong influence on the output, the initial pH is slightly more influential for the PEF/TiO_2 process (Zarei *et al.*, 2010). Conversely, time and current intensity were the two most important parameters for phenol removal using the EF process (Radwan *et al.*, 2018). These two parameters were also shown to be the most influential factors in an EF process for the treatment of composting plant leachate (Alavi *et al.*, 2019).

The relative importance of each input independent variable on the desired output obtained by the Garson algorithm can help ANN modelling approach to provide meaningful insights from the process, usually driven by a well-known RSM approach for experimental studies. Gholami Shirkoohi *et al.* (2021) showed that electrolysis time and current intensity have about 81.5% influence on active chlorine production compared to an 82.8% influence in the factorial design analysis using RSM. The H_3O^+ and NaCl concentration represented the remaining 18.5% of the investigated response. They reported that their findings are similar to the RSM outcomes showing the compatibility and reliability results of the ANN model.

2.6. CONCLUSIONS AND FUTURE PERSPECTIVES

Based on the extensive literature reviewed, it can be observed that artificial intelligence techniques have demonstrated their potential for modelling, performance prediction and optimization of electrochemical processes used for water and wastewater treatment processes. With the limitation of the number of samples in data sets regarding the requirement of huge investment in time and resources, most of the effort should be focused on the reliability and robustness of the AI models derived from these data sets. Although some attempts have been made to use interpolation techniques to solve the data set's limitation problem, cautions should be considered about using these techniques for the goal of regression of experimental work with limited data. This is because the behaviour of outputs in experimental studies can be much more complicated than describing them with predefined interpolation functions. Therefore, it seems tuning AI model hyperparameters and use of regularization techniques to prevent overfitting problem would be the principal part to focus.

The most utilized modelling approach has been the artificial neural networks, thanks to their strong ability to simulate complex, nonlinear input-output systems. Multilayer feedforward neural networks with back-propagation training were widely used in treatment applications. Of course, one of the major objectives of process modelling is to pursue process optimization. GA and PSO were linked mostly to ANNs for optimizing the ANN outputs and have been proved as useful hybrid approaches. Though AI techniques are indicated to be a promising alternative to traditional linear and parametric, and phenomenological methods for modelling and optimization of the electrochemical processes used in water and wastewater treatment processes, there are still some areas requiring further research:

(1) Tuning AI model parameters which control the learning process and have a direct effect on the model performance is a crucial aspect. In the case of ANNs, the selection of optimum network parameters such as number of hidden layers, number of neurons in hidden layers, learning rate, momentum factor, transfer functions, and learning algorithms are still major tasks in ANN modelling and the usual way to overcome these difficulties is

the trial and error method. There have just been a few studies so far to use optimization algorithms such as PSO to optimize the ANN model structure and parameters.

(2) Most of the studies reviewed considered single neural networks for modelling and predicting the performance of their systems. The downside of this approach is that as neural networks are sensitive to the training data, they would find different sets of weights each time they are trained. This will lead to different predictions each time and to high variance. Ensemble modelling, which consists in training multiple models instead of a single model and combining them to find the predictions, is one of the proposed approaches to overcome this challenge. It can be conducted by:

- single learning algorithm, different data sets;

- single learning algorithm, different configurations options;

- different algorithms.

In the reviewed papers, only a few studies applied ensemble modelling approaches like stacked neural networks, but they showed promises.

(3) So far, most of the relevant studies have been performed by a conventional feedforward ANN with the BP algorithm. However, with the advances in machine learning, the BP-MLP NN with regular activation functions and long training time are not the best option. Further research is still required to apply different machine learning algorithms (e.g., SVM) or neural networks (e.g., GRNN, ANFIS) for the modelling and optimization of electrochemical processes used in water and wastewater treatment processes.

DECLARATIONS

Funding

Sincere thanks are extended to the National Sciences and Engineering Research Council of Canada and the CREATE TEDGIEER program for their financial contribution to this study. Peter A. Vanrolleghem holds the Canada Research Chair on Water Quality Modelling.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

3. ARTIFICIAL NEURAL NETWORKS AND GENETIC ALGORITHMS: AN EFFICIENT MODELLING AND OPTIMIZATION METHODOLOGY FOR ACTIVE CHLORINE PRODUCTION USING THE ELECTROLYSIS PROCESS

RÉSEAUX DE NEURONES ARTIFICIELS ET ALGORITHMES GÉNÉTIQUES : UNE MÉTHODOLOGIE EFFICACE DE MODÉLISATION ET D'OPTIMISATION DE LA PRODUCTION DE CHLORE ACTIF PAR LE PROCÉDÉ D'ÉLECTROLYSE

Majid Gholami Shirkoohi[a,c], Rajeshwar Tyagi[a,c], Peter A. Vanrolleghem[b,c], Patrick Drogui[a,c]

[a] Institut National de la Recherche Scientifique (INRS), Centre-Eau Terre Environnement, Université du Québec, Québec (QC), Canada

[b] modelEAU, Département de génie civil et de génie des eaux, Université Laval, Québec (QC), Canada

[c] CentrEau, Centre de recherche sur l'eau, Université Laval, Québec (QC), Canada

Title of the Journal: The Canadian Journal of Chemical Engineering

Date of published online: 16 January 2021

DOI: 10.1002/cjce.24036

Abstract

This study evaluates the effectiveness of a modelling and optimization methodology based on artificial neural networks and genetic algorithms in the prediction of the behaviour of an electrolysis process of active chlorine production from a synthetic saline effluent. Multilayer perceptrons feedforward neural networks were developed for the active chlorine production and energy consumption based on the following inputs: electrolysis time, current intensity, hydrochloric acid concentration and chloride ion concentration. In order to diagnose and prevent the over-fitting problem during the learning process, learning curves and the regularization factor were utilized. The trained ANN models were able to successfully predict the active chlorine production and energy consumption of the process (R^2=0.979 and MSE=3.826 for active chlorine production and R^2=0.985 and MSE=6.952 for energy consumption). Multi-objective optimization for maximizing active chlorine production and minimizing energy consumption was carried out by a genetic algorithm using the best derived ANN models. The Pareto front obtained led to multiple non-dominated optimal points, which result in insights regarding the optimal operating conditions for the process.

Keywords: ANN-GA, Electrochemical processes, Learning curves, Multi-objective optimization, Response surface methodology

3.1. Introduction

In recent years, electrochemical processes have been gaining attention as an alternative method for water and wastewater treatment. These processes are considered as eco-friendly and green technologies since the leading reagent involved, the electron, is considered a clean reagent and takes advantage of coupling chemistry (in situ generation of oxidant) with electronic science (electron transfer). Other attractive advantages include: versatility, high energy efficiency, amenability to automation, and cost-effectiveness (Feng *et al.*, 2016; Rajeshwar *et al.*, 1997). Several studies focusing on the use of electrolysis with different electrochemical methods such as electrooxidation, electrocoagulation, electroflotation, electro-Fenton reaction, and electrodialysis have been published over the last decade for improving the treatment performance of wastewaters and drinking waters (Daghrir *et al.*, 2012; Martín de Vidales *et al.*, 2012b; Olvera-Vargas *et al.*, 2015; Zhang *et al.*, 2011).

In literature, phenomenological and empirical modelling approaches are generally used for mathematical modelling of electrochemical water and wastewater treatment processes. Although phenomenological (white-box) modelling provides valuable insights into the behaviour of the process and has the ability of extrapolation, heat and mass transport phenomena along with detailed knowledge of the reaction kinetics are required. First principles related to the underlying science and engineering laws lead to governing equations that ultimately arrange these models (Zendehboudi *et al.*, 2018). In empirical modelling the structure of the data-fitting model should be specified a priori which makes it challenging as one needs to choose a suitable model structure among the many available ones, especially for non-linear processes (Nandi *et al.*, 2004). Electrochemical processes for water and wastewater treatment are generally complicated non-linear systems and dependent on many factors such as the influent concentration of contaminants (Jardak *et al.*, 2017), the applied current density and electrical potential (Tran *et al.*, 2015a), the types of electrodes (Dia *et al.*, 2017), the electrolyte type and concentration (Jardak *et al.*, 2016), and chemical interactions between contaminants (Giwa *et al.*, 2016; Moreno-Casillas *et al.*, 2007). It is thus difficult to use phenomenological or empirical models to model, simulate, and optimize the processes.

In this context, artificial intelligence methods such as artificial neural networks (ANNs) along with genetic algorithms (GAs) have emerged as attractive alternative approaches for modelling and optimization of these non-linear processes in case phenomenological or conventional regression models are not practical (Curteanu *et al.*, 2014). These black-box (data-driven) models are based on empirical data and relationships among input and output variables of the process. Artificial intelligence methods, such as ANNs, have the role of discovering relationships in which patterns of input data can be linked to the associated output data. These data-driven tools model the system behavior solely from mapping the input-output data rather than from process knowledge. As the complexity of engineering problems increases, the development of faster computers along with more advanced computational algorithms and availability of cost-efficient sensors results in a noticeable paradigm shift from white-box to black-box modelling (Kamari *et al.*, 2014; Nejatian *et al.*, 2014; Zendehboudi *et al.*, 2018). Various types of problems in science can be cast in the form of such pattern-matching, and among the methods within the machine learning tools, ANNs are one of the most effective methods(Boucheikhchoukh *et al.*, 2020; Karimi *et al.*, 2014; Piuleac *et al.*, 2013). Some recent publications illustrate successful application ANN models in various electrochemical processes (Chindapan *et al.*, 2013; da Silva Ribeiro *et al.*, 2019; Radwan *et al.*, 2018; Sangal *et al.*, 2015; Wang *et al.*, 2016).

Genetic algorithms (GAs) belong to the category of evolutionary algorithms that are used for the optimization of objective (fitness) functions by means of parameter space coding. Through the algorithm, a GA can obtain acceptable results by using three stochastic operators; selection, crossover, and mutation (Ghaedi *et al.*, 2014). Detailed information about the theory of GAs and the combination of ANNs with GAs can be found in the literature (Schaffer *et al.*, 1992; Whitley, 1994).

The present study is focused on investigation and analyses of the effectiveness of AI methods for modelling and optimization of an electrolysis process. The database used in this paper was taken from a series of experiments for active chlorine production from a synthetic saline effluent by electrolysis, from the authors' lab and previously published in the literature (Zaviska *et al.*, 2012b). Chlorine is one of the most commonly synthetically produced chemicals worldwide and due to its oxidizing power, has been used as a

disinfectant for potable water, wastewater, and swimming pools (Elmas *et al.*, 2016). Other uses of chlorine products by electrolysis have been reported for the treatment of dye-containing effluents (Zaviska *et al.*, 2009), and as electrolyzed oxidizing water in the food industry (Huang *et al.*, 2008). It is worth mentioning that studies are still being conducted about the production of undesirable active chlorine species during electrolysis (Brito *et al.*, 2015).

While the production of chlorine is commercially dependent on the electrolysis of highly concentrated solutions of sodium chloride (NaCl) (Khouzam, 2008; Oliveira *et al.*, 2007), other alternative approaches are being introduced. These approaches include seawater (Hsu *et al.*, 2015a) and deep ocean water electrolysis (Hsu *et al.*, 2015b; Hsu *et al.*, 2016). In addition, desalination plants produce brine effluents, which are highly concentrated in salts. One of the techniques for managing this saline concentrate can be to use it as a saline resource for chlorine production. This would lead to a reduction in chemical costs for the process of chlorine production (Abdul-Wahab *et al.*, 2009; Pillai *et al.*, 2009).

The ANN-GA approach for modelling and optimization of electrochemical processes has been applied before. Picos and Peralta-Hernández (Picos *et al.*, 2018) utilized ANN models to predict the behavior of an electro-oxidation pilot press-type reactor, which treats synthetic wastewater prepared with a synthetic Violet 54-B dye. Single-objective GA optimization was linked to their ANN model to find the best operational conditions for discoloration efficiency. Tuning ANN models, falling into the domain of hyperparameter (e.g., number of hidden neurons) optimization, is a crucial task to obtain neural networks with the best performance possible and a strong ability of generalization. In this regard, usually, a trial and error procedure is used to derive the best configured network (Ghiasi *et al.*, 2014).

In this work, modelling and optimization of active chlorine production by combining artificial neural networks and genetic algorithms will be studied. This method includes feedforward neural networks and considers the impacts of learning curves and the regularization factor to improve the training process. It is followed by a multi-objective genetic algorithm for the optimization process regarding active chlorine production and energy consumption. Learning curves help to acquire an insight throughout the modelling

problem in order to diagnose the problem as high-variance or high-bias, which can then help to optimally select the most suitable configuration of the network. Regularization is also utilized to prevent over-fitting which can occur with a too complex model. These techniques can give an insight into the ANN modelling process and can be used instead of or along with a trial and error procedure during training of neural networks. To the best of our knowledge, learning curves and the impact of the regularization factor in the cost function of ANNs have not been studied before for modelling of electrochemical processes. Further, Pareto optimal solutions obtained by multi-objective optimization using a genetic algorithm can help to identify optimal operating conditions regarding the production of active chlorine and energy consumption of the process.

3.2. Experimental procedure

The database used in our work was acquired from the experiments of a published study of our group entitled “Statistical optimization of active chlorine production from a synthetic saline effluent by electrolysis” (Zaviska *et al.*, 2012b). To prepare the synthetic saline effluent (SSE) used in these experiments, sodium chloride (NaCl, Fisher Scientific, ACS reagent) was added to distilled water to produce solutions at different concentrations from 0.05 mol/L to 0.105 mol/L. It was observed that produced chlorine gas could be converted to hypochlorous acid (HClO) and hypochlorite ion (ClO^-) after a value of pH = 2.0. Therefore, the initial pH of solution was adjusted by hydrochloric acid (from 0.02 mol/L to 0.14 mol/L) in the range of 0.9 to 1.3. A batch electrolytic cell was designed for conducting the assays using a power supply, an air diffuser, a 4-liter glass tank and a peristaltic pump. An expanded metal Ti/IrO_2 anode and a stainless steel cathode in the form of plates were utilized as electrodes. A 400 ml (135 mm ×35 mm ×140 mm) PVC electrolytic reactor was used to carry out the experiments.

The Wessler reaction was used to estimate the hypochlorous acid production which is based on the oxidation of iodide ions (I^-) to iodine (I_2) in the presence of active chlorine. Then tri-iodide (I_3^-) can be formed by the reaction of surplus iodide ions with iodine (Entezari *et al.*, 1994). A Carry UV 50 spectrophotometer (Varian, Canada) was used to

analyze the tri-iodide ion by measuring the absorbance at 353 nm. The electrical intensity and voltage were applied via an Enduro 250 V power supply.

Response surface methodology approach was utilized to design the experimental assays using a factorial design (FD) followed by a central composite design (CCD). The experiments consisted of 16 experiments for FD and an extra 14 experiments for CCD, a total of 30 experiments. Table 3.1 represents the experimental region for gas chlorine production.

Table 3.1 Experimental operating conditions range

Parameter	Min. value	Max. value
Electrolysis time (min)	15	35
Current intensity (A)	0.8	1.6
$[H_3O^+]$ (mol/L)	0.05	0.11
[NaCl] (mol/L)	0.3	0.8

3.3. Process modelling and optimization

3.3.1 ANNs modelling

As the name implies, artificial neural networks, commonly referred to as “neural networks”, imitate the essential characteristics of the human brain, which itself is a highly non-linear, complex, and parallel computer, such as self-adaptability, self-organization, and error tolerant (Haykin, 1998; Singh *et al.*, 2009). ANNs can explore many competing hypotheses simultaneously using a massively parallel network composed of non-linear computational elements (neurons or nodes) that are interconnected by links with variable weights. The mentioned interconnected set of weights contains the knowledge generated by the network (Adya *et al.*, 1998). Each neuron at certain times examines its inputs and computes an output called an activation. The new activation is then passed along those connections to other neurons.

One of the most common architectures of artificial neural networks, considering how the different neurons are positioned and connected to each other as well as the composition of layers, is the multilayer percepterons (MLP) feedforward network. These networks are usually applied to diverse problems, including function approximation, pattern classification, system identification, process control, process optimization, and so on (Carvalho *et al.*, 2011; Da Silva *et al.*, 2017). The nature and complexity of the problem in addition to the desired accuracy and the available data determine the number of hidden layers and the corresponding number of neurons in each hidden layer. In addition, the configuration of the MLP network including the number of hidden layers and hidden neurons, can be derived by a trial and error procedure (Ghiasi *et al.*, 2015).

The standard learning algorithm for MLP neural networks for any pattern recognition or function fitting process is known as the back-propagation algorithm (BP) (Carvalho *et al.*, 2011). The back-propagation algorithm can be viewed as a generalization of the least mean square procedure that can be used for the training of multilayer neural networks. In the BP algorithm data enters the network via the input layer which merely transfers the data value to the hidden layer over weighted connections. The hidden and output neurons process their inputs by multiplying each input by its weight, adding the product to a total amount, and then passing it through a (transfer or activation) function to generate its result. The whole aim of the back-propagation algorithm is to change the values of the network weights to achieve the minimum error between the predicted output and actual targets.

Figure 3.1 shows the MLP neural network used for the modelling and optimization of the active chlorine production from a synthetic saline effluent by electrolysis. In continuation of the work of Zaviska et al. (2012), current intensity, electrolysis time, chloride ion concentration and hydrochloric acid concentration were selected as the input neurons, whereas the output layer contains the active chlorine production or the energy consumption.

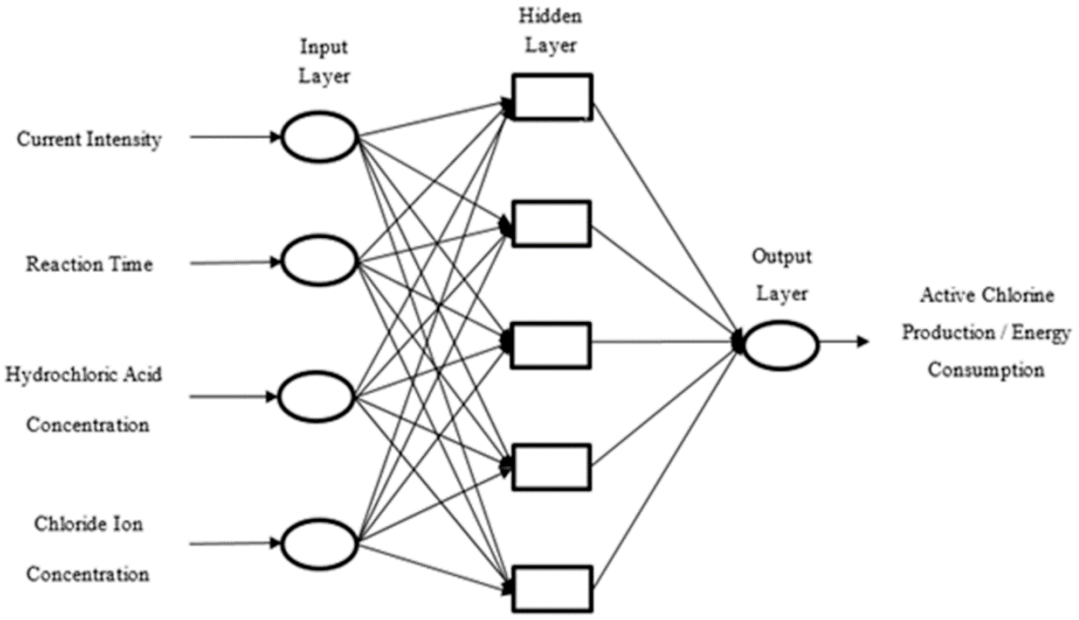


Figure 3.1 Schematic of the MLP neural network used for the modelling and optimization of the active chlorine production

The gradient descent algorithm has been selected as the learning algorithm for training the neural network with ‘tansig’ and ‘purelin’ transfer function in the hidden and output layer, respectively.

3.3.2 Learning curves

Learning curves of model performance on the training and validation datasets can be used to diagnose an underfit (high bias), overfit (high variance), or well-fit model.

At first, the data was split into two sets: training and validation. One single example from the training set was taken and used to fit a model. The error related to the model on the validation set and that single training example were measured. The error related to this training instance would be 0, since it is not too overwhelming to fit a single data point perfectly. Since the model is built around a single instance, the error related to the validation set will be quite large. This is due to the lack of generalization ability to the data that it has not seen before. Then the number of training samples is gradually increased until the entire training set is used. As the training set changes, the error values will vary more or less. Thus, two error values have to be monitored: one for the validation set, and

one for the training set. If the evolution of the two error values is plotted as the training sample sets change, two curves (so-called learning curves) are obtained. In brief, a learning curve demonstrates how the error varies with an increase in the training set size and demonstrates if one needs a more complex model for the predictions or not.

In this work, learning curves will be plotted for the training samples. To avoid the uncertainty related to the selection of the training sample, each training sample selection was replicated 50 times and the overall mean value has been calculated for that training sample batch.

3.3.3 Regularization factor

The regularization parameter (lambda) is an input to the objective function to reduce overfitting. This reduces the variance of the estimated regression parameters. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting. It is defined as a term added to the cost function of the model (Eq. (3.1)).

$$J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda\sum_{j=1}^{n}\theta_j^2\right] \tag{3.1}$$

where $J(\theta)$ is the cost function (error), m is the number of data points used for training, x are the input neurons, $h_\theta\left(x^{(i)}\right)$ is the predicted value of sample i, $y^{(i)}$ is the actual value of sample i, λ is the regularization parameter, and θ are the network parameters (weights). In fact, to have control of the fitting parameters, regularization parameter is used. With any increase in the mgnitudes of the network parameters, an increasing penalty will be applied on the cost function. As can be seen, this penalty is relevant on the magnitude of lamda and the squares of the weights. Any increase in lambda can be advantageous up to a certain point, since it reduces the variance which avoids overfitting. But after this point, important properties of the model starts to be lost introducing more bias into the model (underfitting problem). This implies the importance of the selection of the lambda value. In this study, different lambda values have been tested each time to obtain the best training and cross-validation errors and to present these errors vs. the lambda value. The

optimum lambda value was selected from this graph by considering the cross-validation and training errors.

3.3.4 Relative importance of input variables

The weights obtained from ANN training are coefficients between artificial neurons that are analogous to synaptic strengths between the axon and dendrites in a biological neuron in the brain. As in real life, the proportion of the incoming signal to be transmitted to the neuron's body is decided by these weights (Khataee *et al.*, 2010). Despite the black-box nature of ANNs, to estimate the influence of different independent variables on the output, it is possible to conduct a sensitivity analysis on the ANNs. The relative importance of each input independent variable on the desired output can be obtained through the neural connection weight matrix. First, Garson (Garson, 1991) and then Goh (Goh, 1995) proposed a procedure for partitioning the connection weights to determine the relative importance of the various inputs. This method basically involves partitioning the hidden-output connection weights of each hidden neuron into components associated with each input neuron (Gevrey *et al.*, 2003).

Garson's equation based on the partitioning of connection weights can be applied:

$$R_{\mathrm{j}} = \frac{\sum_{\mathrm{m}=1}^{\mathrm{m}=\mathrm{U_h}} \left(\left(\left| \mathrm{W}_{\mathrm{jm}}^{\mathrm{jh}} \right| / \sum_{\mathrm{k}=1}^{\mathrm{U_i}} \left| \mathrm{W}_{\mathrm{km}}^{\mathrm{ih}} \right| \right) \times \left| \mathrm{W}_{\mathrm{mn}}^{\mathrm{ho}} \right| \right)}{\sum_{\mathrm{k}=1}^{\mathrm{k}=\mathrm{U_i}} \left[\sum_{\mathrm{m}=1}^{\mathrm{m}=\mathrm{U_h}} \left(\left| \mathrm{W}_{\mathrm{km}}^{\mathrm{ih}} \right| / \sum_{\mathrm{k}=1}^{\mathrm{U_i}} \left| \mathrm{W}_{\mathrm{km}}^{\mathrm{ih}} \right| \right) \times \left| \mathrm{W}_{\mathrm{mn}}^{\mathrm{ho}} \right| \right]} \tag{3.2}$$

where R_{j} is the relative importance of the jth independent variable on the output variable, $\mathrm{U_i}$ and $\mathrm{U_h}$ denote the number of input and hidden neurons, respectively; the W is the connection weight value, the superscripts i, h and o refer to input, hidden and output layers, respectively. Also, the subscripts k, m and n refer to input, hidden and output neurons, respectively.

3.3.5 Genetic algorithm and multi-objective optimization

In recent years, evolutionary algorithms, and in particular genetic algorithms, have received growing attention among optimization techniques. GAs with their good global searching ability and flexibility, ease of operation and without the need for gradient

information on the objective (fitness) functions, have become powerful techniques for optimization problems (Curteanu *et al.*, 2007; Ding *et al.*, 2011b). A GA starts with a primary population of candidate solutions and a fitness value is calculated for each solution. Through the algorithm, three stochastic operators are applied to each population which are analogous to chromosomes in a biological context. Selection is choosing the solutions with the highest fitness value to create an intermediate population. The next population is the result of crossover or mutation. By crossover, the selected members are mated in pairs and recombined through genetic manipulation of chromosomes to generate two new solutions (offsprings). Mutation acts as an assurance against lost genetic material and consists of replacing some of the chromosome's genes with new genes. The generation of new populations and calculation of the fitness value for each population is repeated over and over in an iterative method. When a specific termination criterion is met, e.g., when there is no more change in the population from one iteration to the next or when a satisfactory fitness value is achieved, this process ends (Ansari *et al.*, 2014; Niculescu, 2003; Ridha *et al.*, 2008).

When multiple objectives are specified to a problem, selecting a single solution with specific decision variables could not satisfy all the objectives in a single manner. In fact, objective functions could have non-linear and opposite behaviour to each other. Therefore, a trade-off between all these conflicting objective functions should be made to find the decision variables. This trade-off can be illustrated as Pareto front, which is based on the domination concept. Best solutions in the problem space will be represented in this front which are the solutions for which there would be no other solution having better values regarding the objective functions. Having Pareto front helps obtain a clear insight throughout the trade-off between different objective function. This would help to find and focus on promising solutions from a possibly large population of solutions and choose the decision variables suitable regarding the objective functions (Hu *et al.*, 2014).

In this regard, the well-known non-dominated sorting genetic algorithm (NSGA-II) (Deb *et al.*, 2002) has been utilized for multi-objective optimization, leading to a set of solutions (Pareto front), that are the experimental conditions, with respect to maximization of active chlorine production and minimization of energy consumption. The flowchart of the adopted ANN-GA approach in this study is shown in Figure 3.2.

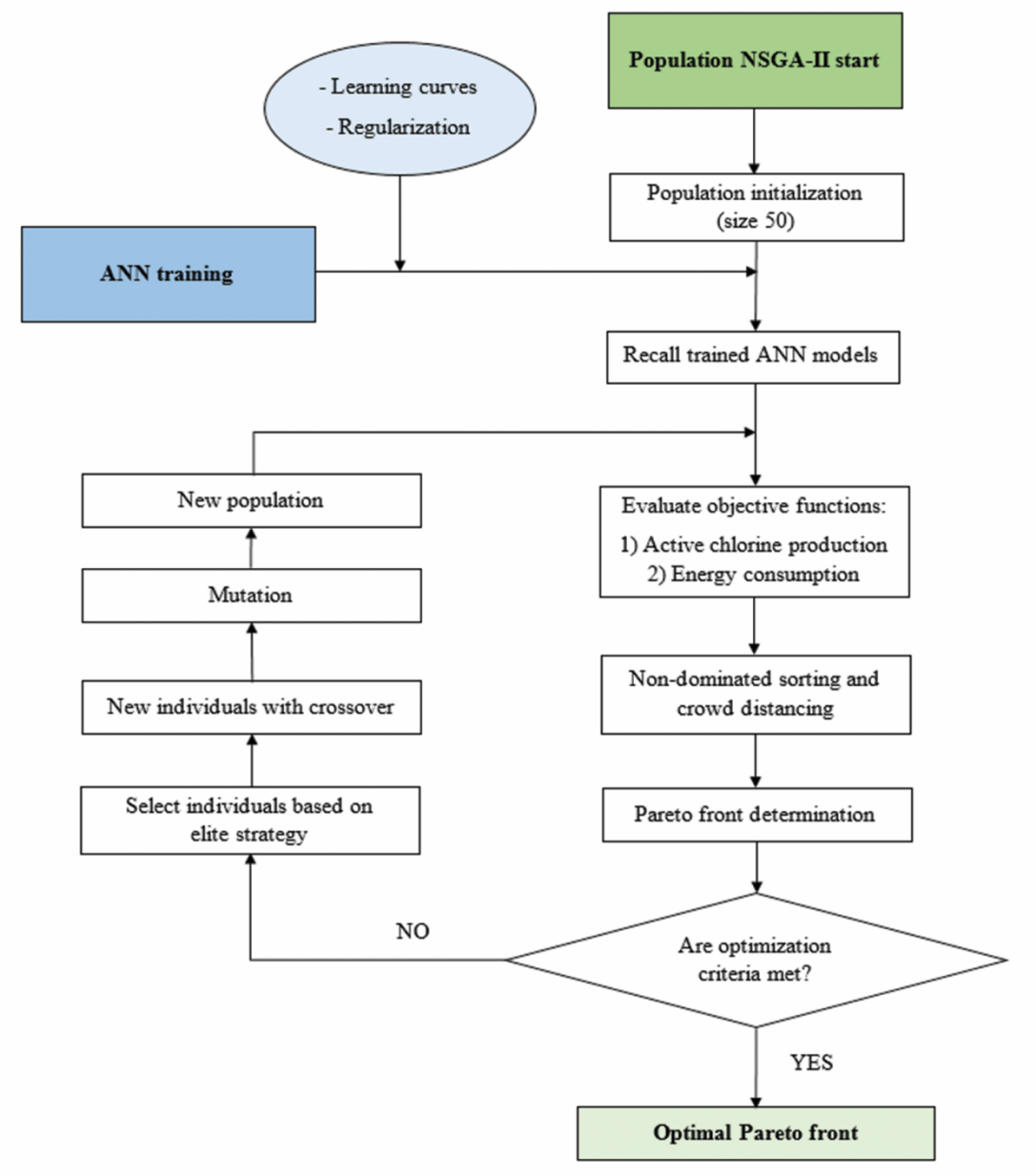


Figure 3.2 Flowchart of ANN-GA methodology used for multi-objective optimization

3.4. Results and discussion

3.4.1 ANN modelling

After initial data collection, data preprocessing was necessary to manipulate the data into a usable format for processing by the artificial neural networks. Feature normalization (Eq. (3.3)) has been selected and returns a normalized version of feature (input) X where the mean value of each feature is 0 and the standard deviation is 1.

$$y=\frac{X_i-\mu_i}{S_i} \tag{3.3}$$

where y is the normalized value of X_i. The μ_i and the S_i are the mean and standard deviation values of X_i, respectively. Normalization helps because it ensures (i) that the network's learning regards all input features to a similar extent and (ii) that there are both positive and negative values used as inputs for the next layer, which makes learning more flexible.

At the first step of the ANN modelling, a three layer network was configured with the 16 FD experiments. The ANN model was constructed with 5 neurons in the hidden layer with 'tansig' and 'purelin' transfer functions in the hidden and output layer, respectively, and trained by the gradient descent algorithm. While the coefficient of determination for the FD was reported as R^2=0.982, this value increased to R^2=0.999 with the ANN model.

Finding optimal conditions to produce active chlorine is a multi-objective optimization by taking into account of energy consumption of the process.This has been done by conducting 14 more experiments using a Central Composite Design (CCD) (Zaviska *et al.*, 2012b). Total 30 experimental data points have been used for ANN modelling for the purpose of training and validation, including sets of 24 and six samples for each, respectively.

3.4.1.1 Learning curves and impact of regularization

For an ANN model, it is necessary to have an overview of the state of the model in order to check whether there is a high bias (underfit) problem or high variance (overfit) issue.

This helps decide whether a more complex model (with more hidden layers and neurons) is required or not. Figure 3.3 shows the mean learning curve obtained for different numbers of training examples. As explained in section 3.2, learning curves show how the error changes as the training set size increases and demonstrate whether one needs a more complex model for the predictions or not.

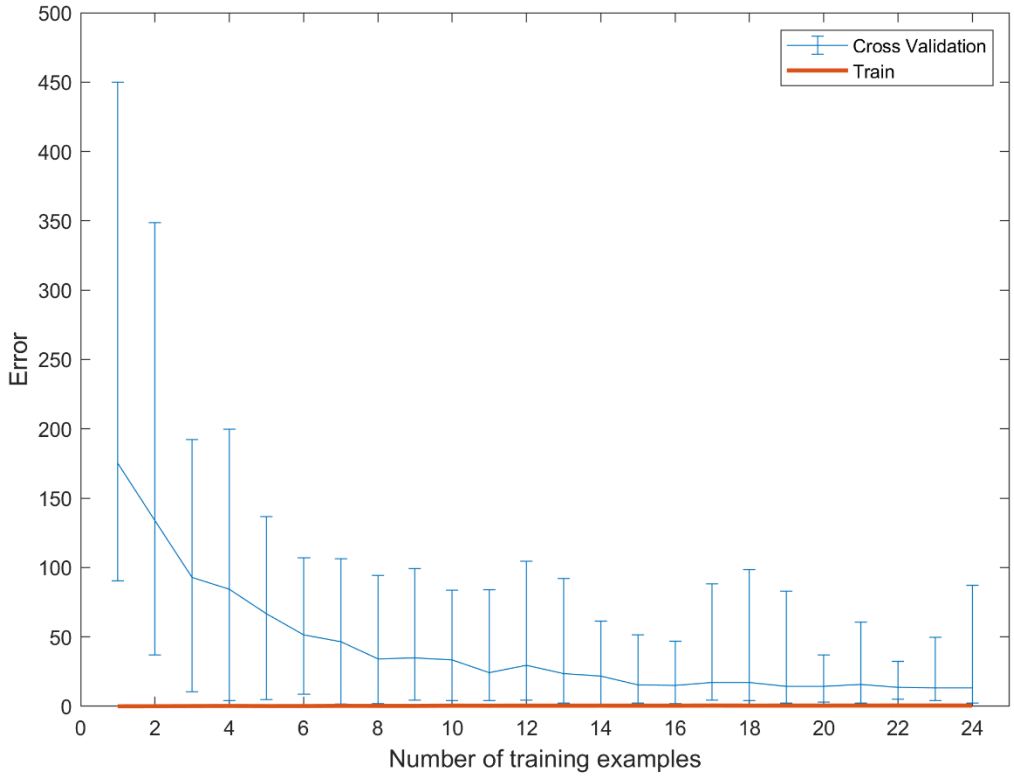


Figure 3.3 Mean learning curves for different numbers of training examples (Error bars are generated with 50 time training for each training set)

In this figure, the training and cross-validation error have been plotted versus the number of training examples in the training set. This figure helps to have an overview of the type of problem dealt with. In case of a high bias model (underfit), there would be high errors for both training and cross validation data sets. For a high variance model (overfit), the training error would be low and the cross validation error would be much higher. Also, in case of a high variance problem, having more data along with not having a more complex model (e.g. adding more hidden layers or hidden neurons) would help the modelling process. As Figure 3.3 shows, there is a gap between the training and cross-validation

error, with a very small error for the training set and a much higher error for the cross-validation set. Also, as the number of samples in the training set increases the cross-validation error decreases, which proves that increasing the number of samples is a good solution for a high variance problems as diagnosed for this case. The ANN model for each point in the learning curve has been trained 50 times with random sampling from the available data and the mean, minimum and maximum error values have been represented. Decision making based on the mean value of 50 times iterations for each training with random sampling helps to decrease the risk of uncertainty related to stochastic behaviour of ANN modelling.

Figure 3.3 shows the mean learning curve for the validation samples with the minimum and maximum values obtained for the 50 iterations of training. As said before, for a high variance problem, having a more complex model does not help. This is shown in Table 3.2 where three different neural network configurations are presented with their correlation coefficients for the training, validation, and all data sets. It can be concluded that for this high variance problem there is no need for a more complex model that inherently would increase the overfitting issue.

Table 3.2 Feedforward backpropagation networks

Configuration	R^2 training	R^2 validation	R^2 All
# Samples	24	6	30
5 hidden neurons	0.9803	0.9701	0.9791
6 hidden neurons	0.9404	0.9592	0.9453
7 hidden neurons	0.9445	0.9562	0.9481

In the presence of the high variance problem, using the regularization factor can help. Regularization makes slight modifications to the learning algorithm such that the model generalizes better and improves the model's performance on unseen data. Therefore, a graph of Error vs regularization factor (lambda) helps to optimally select the best lambda value. In our case, the best lambda value is 3 (Figure 3.4).

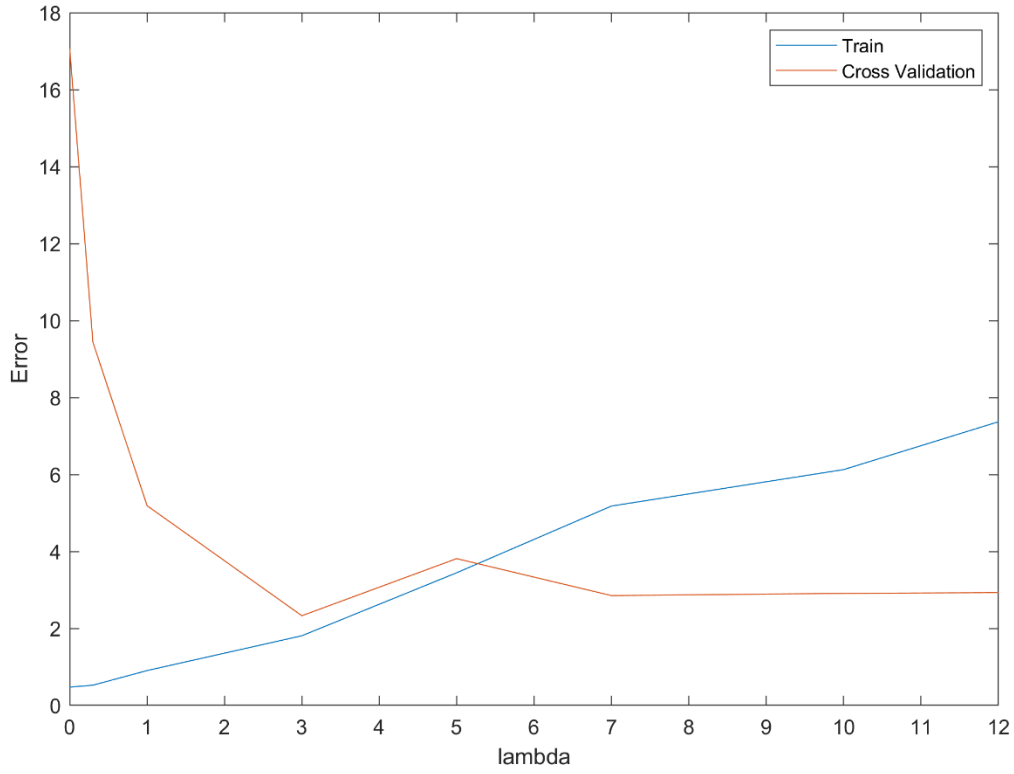


Figure 3.4 Impact of regularization factor on model performance

Thus, a 3-layer feedforward back-propagation network with 5 hidden neurons and a regularization factor value of 3 is selected for the optimization.

The selected network for active chlorine production has a coefficient of determination R^2=0.979 while this value for the RSM with the central composite design was reported as R^2=0.964.

Also, for the multi-objective optimization, the selected neural network for predicting the energy consumption is configured with 4 hidden layers and a sigmoid transfer function at the hidden layer. This network has a performance of R^2_{train}=0.997, $R^2_{validation}$=0.951, R^2_{All}=0.985 which compares favorably to the RSM regression performance R^2=0.990. Parity plots for ANN and RSM models regarding active chlorine production and energy consumption are represented in Figure 3.5.

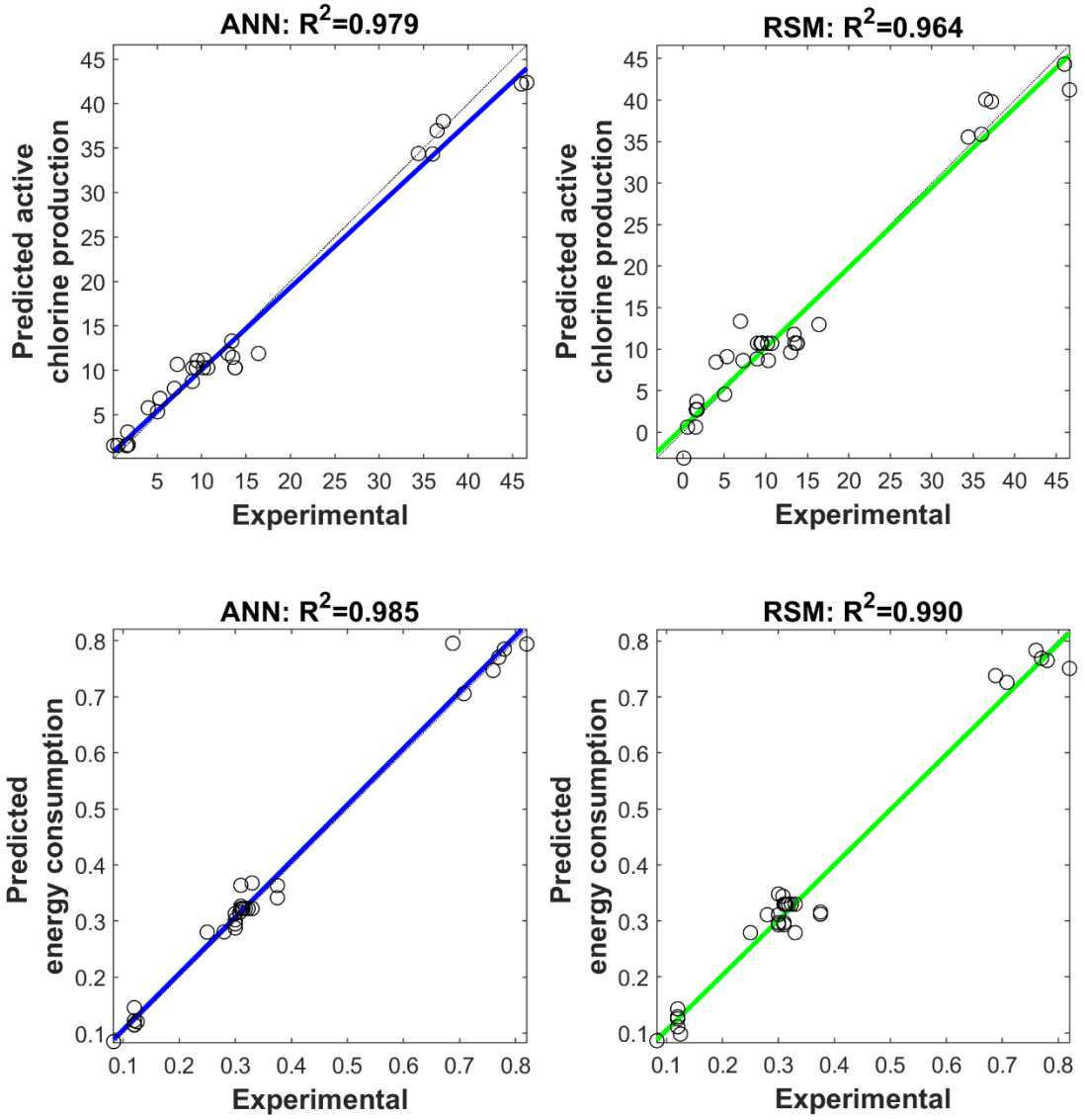


Figure 3.5 Parity plots of predicted versus experimental values of active chlorine production and energy consumption for ANN and RSM models

It should be mentioned that in RSM, all available data was used for the linear regression method for curve fitting. In ANN modelling, however the data is divided into training and validation sets. Table 3 represents the CCD experimental plan, actual and predicted values of the ANN and RSM models for active chlorine production and energy consumption. Also, performance criteria (R^2 and Mean Squared Error (MSE)) for each model and dependent variable are reported in Table 3.3. As can be seen, the ANN

method performs slightly better than RSM for predicting active chlorine prediction and energy consumption of the electrolysis process.

Table 3.3 Actual and predicted values of central composite designed experiments.

Experimental assays				Chlorine production (mg/l)			Energy consumption (kWh/m^3)		
Time (min)	Current (A)	$[H_3O^+]$ (mol/L)	$[Cl^-]$ (mol/L)	Actual	ANN predicted	RSM predicted	Actual	ANN predicted	RSM predicted
15	1.6	0.05	0.3	5.04	5.34	4.57	0.250	0.280	0.279
25	1.2	0.08	0.55	10.7	10.28	10.7	0.312	0.322	0.330
15	1.6	0.11	0.3	5.32	6.84	9.09	0.300	0.302	0.297
15	0.8	0.05	0.3	1.6	1.52	2.71	0.120	0.122	0.111
25	1.2	0.08	0.55	9	10.28	10.7	0.310	0.322	0.330
35	1.6	0.11	0.8	46	42.23	44.33	0.760	0.747	0.783
25	2	0.08	0.55	46.6	42.39	41.24	0.708	0.706	0.726
15	1.6	0.11	0.8	6.94	7.96	13.35	0.280	0.281	0.311
15	0.8	0.11	0.8	0.56	1.57	0.61	0.120	0.114	0.143
25	1.2	0.08	0.55	13.8	10.28	10.7	0.330	0.322	0.330
25	1.2	0.08	0.55	13.76	10.28	10.7	0.323	0.322	0.330
25	1.2	0.08	0.55	10.2	10.28	10.7	0.314	0.322	0.330
15	0.8	0.05	0.8	1.54	1.52	0.61	0.120	0.115	0.125
35	0.8	0.11	0.8	7.26	10.67	8.63	0.300	0.295	0.311
25	1.2	0.08	0.05	13	11.88	9.62	0.375	0.342	0.316
35	1.6	0.11	0.3	36.5	36.96	40.07	0.770	0.770	0.769
25	1.2	0.08	0.55	9.4	10.28	10.7	0.318	0.322	0.330
15	1.6	0.05	0.8	8.96	8.75	8.83	0.300	0.288	0.293
25	1.2	0.08	1.05	13.4	13.28	11.78	0.308	0.315	0.344
25	0.4	0.08	0.55	1.68	3.06	3.68	0.083	0.085	0.086
25	1.2	0.02	0.55	4	5.77	8.44	0.375	0.363	0.312
35	0.8	0.11	0.3	9.52	11.09	10.73	0.310	0.326	0.297
5	1.2	0.08	0.55	0.1	1.52	-3.12	0.125	0.120	0.098
35	1.6	0.05	0.8	37.2	38.01	39.81	0.780	0.785	0.765
45	1.2	0.08	0.55	36	34.34	35.88	0.688	0.795	0.738
25	1.2	0.14	0.55	16.4	11.90	12.96	0.300	0.313	0.348
35	1.6	0.05	0.3	34.4	34.39	35.55	0.820	0.794	0.751
35	0.8	0.05	0.8	10.3	11.17	8.63	0.310	0.364	0.293
15	0.8	0.11	0.3	1.74	1.62	2.71	0.120	0.145	0.129
35	0.8	0.05	0.3	13.5	11.46	10.73	0.330	0.368	0.279
R-squared (R^2)				-	**0.979**	**0.964**	**-**	**0.985**	**0.990**
Mean squared error (MSE)				-	**3.826**	**6.952**	**-**	**6.903e-04**	**9.043e-04**

3.4.1.2 Relative importance of input variables

Using the factorial design method, the influence of four main experimental factors was investigated. Based on the sensitivity analysis results, electrolysis time and current intensity with 82.8% contribution on the active chlorine production were the two most influential factors. In order to assess the relative importance of the input variables for the ANN model, the neural net weight matrix can be used. The relative importance of the various variables, calculated by Eq. (3.2), is shown in Figure 3.6.

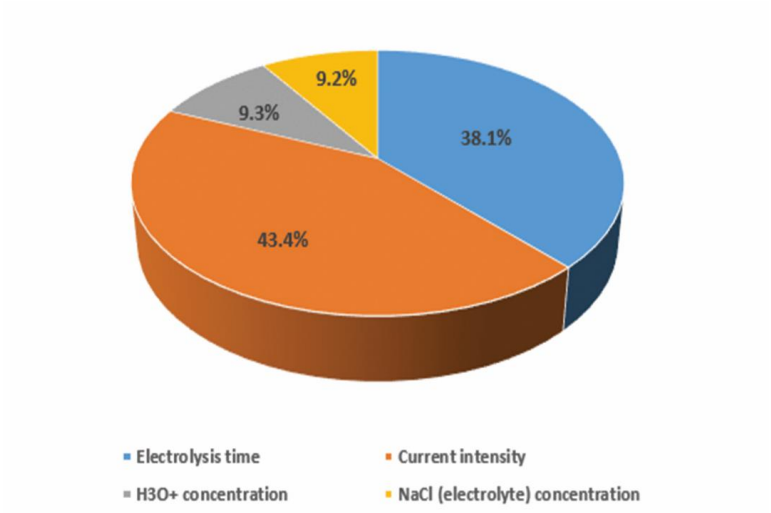


Figure 3.6 Importance (%) of the input variables on the electrochemical active chlorine production

Like the FD method, ANN weight analysis derived by Garson's algorithm described in section 3.4 illustrates that electrolysis time and current intensity are the most important factors for predicting the production of active chlorine. Garson's algorithm investigation on neural network weights shows about 81.5% influence on active chlorine production for these two main independent variables (compared to an 82.8% influence in the FD analysis). The H_3O^+ and NaCl concentration represent the remaining 18.5% of the investigated response (active chlorine production). Figure 6 shows the compatible and reliable results of the ANN model, similar to the RSM outcomes.

3.4.1.3 Response surfaces of the RSM and ANN

The effect of electrolysis time and current intensity on the production of active chlorine is illustrated in Figure 3.7. Note that the concentrations of acid and chloride are kept

constant at the center of the investigated experimental ranges (0.08 mol/L and 0.55 mol/L, respectively). In these conditions, by increasing the electrolysis time, the active chlorine concentration rises for all current intensities studied. As can be seen, the RSM response surface is a quadratic model that has to fit the predicted values on this surface, whereas the ANN model with its high ability for nonlinearity can fit the data in a much finer way. It can be concluded that active chlorine can be produced up to more than 33 mg/L at electrolysis times longer than 30 min and for current intensities at the higher values of 1.4 A.

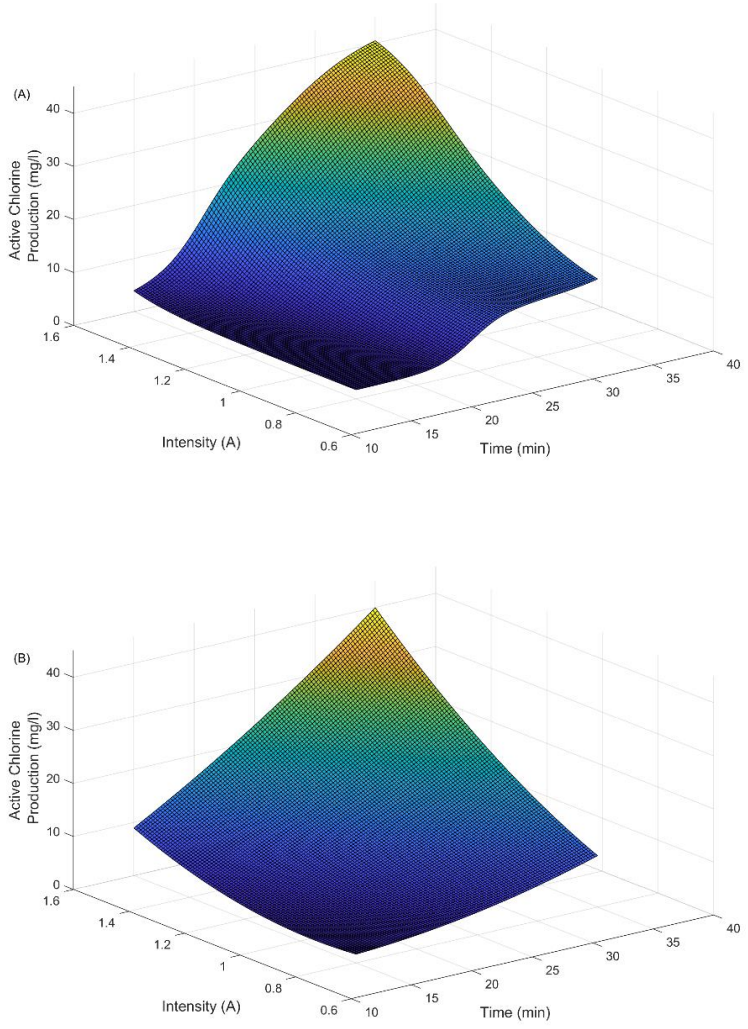


Figure 3.7 Response surface graph of active chlorine production versus electrolysis time and current intensity, (A) ANN, (B) RSM

3.4.2 Multi-objective optimisation with GA

Simultaneous optimization of hypochlorous acid production and the energy consumption is defined in the category of multi-objective optimization. No unique solution can be derived for a multi-objective optimization problem, except for Pareto front solutions which are inherently non-dominated. A MATLAB script using two ANN models developed for the hypochlorous acid production and energy consumption was written to create a cost (fitness) function. The multi-objective optimization is conducted by aiming for both maximizations of the hypochlorous acid production and minimization of energy consumption. The bounds of the four independent variables were chosen by the ranges of the experiments. The following NSGA-II algorithm options were set:

Population size: 50

Maximum number of iterations: 150

Selection function: Tournament selection

Crossover strength: 0.7

Mutation strength: 0.3

Distance measure function: distance crowding

The maximum number of iterations was used as stopping criterion. For the purpose of comparison, Pareto fronts have also been generated using RSM models with the same NSGA-II algorithm. After 150 iterations the Pareto front of Figure 3.8 is obtained. The decision variables of the electrolysis process corresponding to each of the ANN-GA Pareto front solutions are tabulated in Table 3.4.

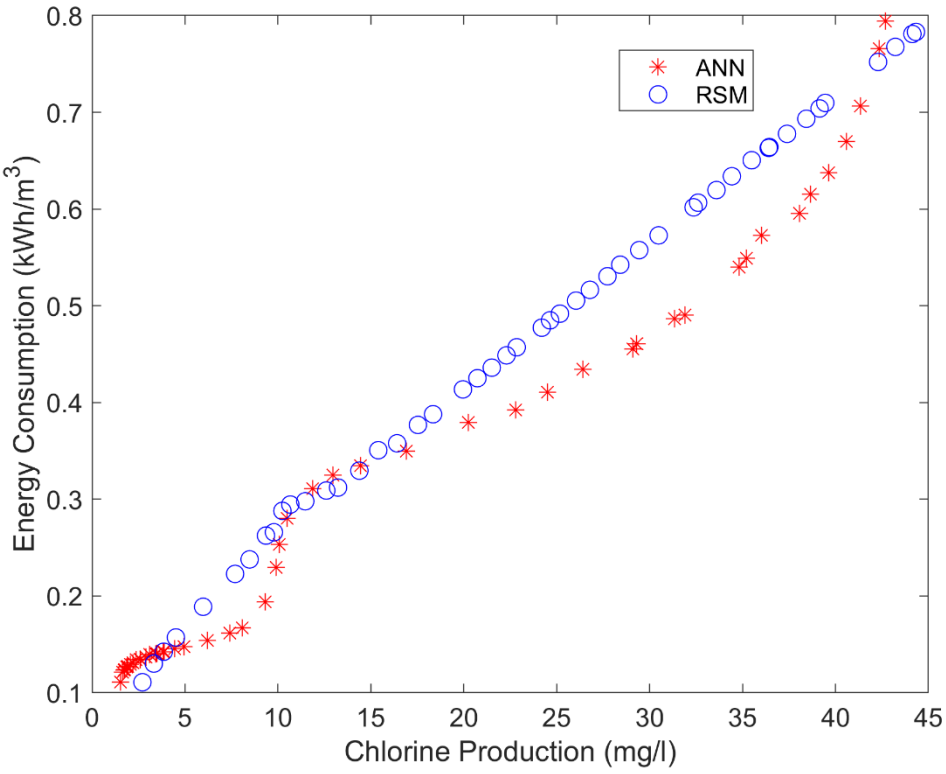


Figure 3.8 Pareto fronts for multi-objective optimization of active chlorine production and energy consumption

The general method used for RSM optimization is single optimization with multiple responses using desirability functions and weighting factors representing the importance of each response. In this approach, usually, just a single optimal point is reported based on the desirability value. No Pareto front will be provided. Since the RSM models are generated by linear regression method, the Pareto front provided by NSGA-II for this approach is linear. Each point on the Pareto front indicates that there is no other process decision variables that can have the same active chlorine production with lower cost or, in other words, with the same cost there are no other process decision variables that can produce higher active chlorine. As reported in the paper on RSM (Zaviska *et al.*, 2012b), the optimal conditions for the electrolytic reactor were obtained by 27 min of electrolysis time with the concentrations of hydrochloric acid and chloride sodium of 0.11 and 0.8 mol/L, respectively. Under these conditions, it was mentioned that production of 30.60 mg/l of active chlorine could be possible with 0.54 kWh/m^3 energy consumption.

Table 3.4 Decision variables of the electrolysis process corresponding to each of the Pareto front solutions presented in Figure 8.

S. No.	Time (min)	Current (A)	$[H_3O^+]$ (mol/L)	$[Cl^-]$ (mol/L)	HClO production (mg/l)	Energy consumption (kWh/m^3)
1	35.00	1.60	0.11	0.80	42.23	0.75
2	15.00	0.80	0.08	0.80	1.53	0.11
3	22.75	1.57	0.11	0.80	25.45	0.42
4	20.70	1.45	0.11	0.80	13.69	0.33
5	17.92	0.80	0.11	0.80	1.87	0.12
6	18.81	0.80	0.11	0.80	2.16	0.13
7	21.43	1.36	0.11	0.80	11.40	0.30
8	33.27	1.60	0.11	0.80	41.49	0.70
9	24.13	0.80	0.11	0.80	7.33	0.16
10	20.33	1.53	0.11	0.80	16.89	0.35
11	33.48	1.53	0.11	0.80	40.13	0.67
12	27.36	1.05	0.11	0.80	10.62	0.28
13	20.38	1.59	0.11	0.80	21.11	0.37
14	25.29	0.80	0.11	0.80	8.34	0.17
15	23.04	0.80	0.11	0.80	6.03	0.15
16	23.81	1.59	0.11	0.80	28.84	0.45
17	21.47	0.80	0.11	0.80	4.08	0.14
18	26.53	0.96	0.11	0.80	9.72	0.24
19	28.25	0.81	0.11	0.80	9.60	0.20
20	28.75	1.59	0.11	0.80	37.21	0.57
21	29.71	1.59	0.11	0.80	38.50	0.60
22	28.56	0.93	0.11	0.80	10.12	0.25
23	27.03	0.81	0.11	0.80	9.26	0.19
24	19.61	0.80	0.11	0.80	2.48	0.13
25	20.12	0.80	0.11	0.80	2.84	0.13
26	20.64	0.80	0.11	0.80	3.13	0.14
27	19.92	0.80	0.11	0.80	2.71	0.13
28	26.19	1.59	0.11	0.80	33.72	0.51
29	21.90	0.80	0.11	0.80	4.52	0.14
30	25.41	1.59	0.11	0.80	32.03	0.49
31	21.12	0.81	0.11	0.80	3.71	0.14
32	28.41	1.56	0.11	0.80	35.47	0.55
33	30.77	1.60	0.11	0.80	39.76	0.63
34	24.65	1.59	0.11	0.80	30.47	0.47
35	20.97	1.55	0.11	0.80	19.53	0.37
36	27.16	1.59	0.11	0.80	34.92	0.53
37	22.15	0.80	0.11	0.80	4.82	0.15
38	20.95	0.80	0.11	0.80	3.51	0.14
39	17.36	0.80	0.11	0.80	1.77	0.12
40	15.78	0.80	0.11	0.80	1.60	0.12
41	20.80	0.80	0.11	0.80	3.27	0.14
42	22.46	0.80	0.11	0.80	5.28	0.15
43	19.31	0.80	0.11	0.80	2.29	0.13
44	30.55	1.59	0.11	0.80	39.47	0.63
45	22.57	0.81	0.11	0.80	5.50	0.15
46	16.98	0.80	0.11	0.80	1.71	0.12
47	16.45	0.80	0.11	0.80	1.65	0.12
48	19.31	0.80	0.11	0.80	2.35	0.13

Conversely, in the ANN-GA approach of this paper, one of the solutions implies that with 0.53 kWh/m^3 of energy consumption 34.92 mg/l of active chlorine can be produced under different operating conditions. Also, production of 30.47 mg/l of active chlorine is achievable with 0.47 kWh/m^3 energy consumption. These imply more economic

conditions for higher active chlorine production. Unfortunately, our results could not be verified by the experiments since the experimental set up was no longer available. However, by comparing Table 3 and 4, at least one Pareto optimal point appears similar to an experimental assay (Solution number 1: Time=35 min, Current=1.6 A, $[H_3O^+]$=0.11 mol/L and $[Cl^-]$=0.8 mol/L). Under these conditions, 42.23 mg/l of active chlorine can be produced with 0.75 kWh/m^3 energy consumption compared to experimental values of 46 mg/l and 0.76 kWh/m^3 for active chlorine and energy consumption, respectively. The relative error for this optimal point is 0.082 and 0.013 for active chlorine production and energy consumption, respectively.

The ANN-GA approach introduced in this study provides optimal operational conditions based on active chlorine production and energy cost. The advantage of having a Pareto front for industrial process designers and operators is that different operational conditions (decision variables) can be selected based on preference for each objective. This gives an insight on the trade-off between the different objective functions involved in this industrial process.

Although some effort has been made in this study to obtain the best possible ANN models for describing the process, there remain some ANN hyperparameters that can be optimized in further studies. These hyperparameters, including transfer functions and learning rate, may have direct impact on the ANN modelling performance.

3.5. Conclusions

The artificial neural network-genetic algorithm (ANN-GA) methodology was successfully applied to an electrolysis process for active chlorine production. MLP feedforward neural networks were developed for actives chlorine production and energy consumption. To diagnose whether there is a danger for high-variance or high-bias error and to prevent over-fitting of the model, learning curves along with regularization factor analysis were utilized during the training of the neural network models. Figure 3.7 indicated that the ANN model was able to describe the nonlinearities related to the experimental process better than the previously proposed RSM model with a coefficient of determination of 0.979 and 0.985 for production of active chlorine and energy consumption, respectively.

Analysis of the relative importance of the variables indicated that electrolysis time and current intensity are the two most influential parameters with a total effect of 81.5% on active chlorine production.

To obtain a clear insight throughout the trade-off between different objective functions involved in the electrolysis process, the NSGA-II algorithm was used for multi-objective optimization of the process regarding active chlorine production and energy consumption. The Pareto front derived by GA led to the generation of non-dominated optimal points (operating conditions) for maximum active chlorine production at minimum energy consumption. The proposed ANN-GA methodology can give insight in how to efficiently choose the process operation parameters (decision variables) for the desired objectives. This approach can be adapted to other processes if the experimental data already exist.

Acknowledgements

Sincere thanks are extended to the National Sciences and Engineering Research Council of Canada and CREATE TEDGIEER program for their financial contribution to this study. Peter A. Vanrolleghem holds the Canada Research Chair on Water Quality Modelling.

4. MODELLING AND OPTIMIZATION OF PSYCHOACTIVE PHARMACEUTICAL CAFFEINE REMOVAL BY ELECTROCHEMICAL OXIDATION PROCESS: A COMPARATIVE STUDY BETWEEN RESPONSE SURFACE METHODOLOGY (RSM) AND ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (ANFIS)

MODÉLISATION ET OPTIMISATION DE L'ÉLIMINATION DE LA CAFÉINE, UN PRODUIT PHARMACEUTIQUE PSYCHOACTIF, PAR UN PROCESSUS D'OXYDATION ÉLECTROCHIMIQUE : UNE ETUDE COMPARATIVE ENTRE LA METHODE DE SURFACE DE REPONSE (MSR) ET DES SYSTÈMES À INFÉRENCES FLOUES À RÉSEAUX ADAPTATIFS (SIFRA)

Majid Gholami Shirkoohi[a,d], Rajeshwar D. Tyagi[b], Peter A. Vanrolleghem[c,d], Patrick Drogui[a,d]

[a] Institut National de la Recherche Scientifique (INRS), Centre-Eau Terre Environnement, Université du Québec, 490, Rue de la Couronne, Québec (QC) G1K 9A9, Canada

[b] BOSK Bioproducts, 399 Rue Jacquard, suite 100, Québec (QC) G1N 4J6, Canada

[c] modelEAU, Département de génie civil et de génie des eaux, Université Laval, 1065 av. de la Médecine, Québec, QC G1V 0A6, Canada

[d] CentrEau, Centre de recherche sur l'eau, Université Laval, Québec (QC), Canada

Title of the Journal: Separation and Purification Technology

Date of published online: 23 March 2022

DOI: 10.1016/j.seppur.2022.120902

Abstract

In this study, the modelling and optimization of psychoactive pharmaceutical caffeine removal in synthetic solution and real municipal wastewater effluent by electrochemical oxidation (EO) process was investigated using central composite design (CCD) based on response surface methodology (RSM) and adaptive neuro fuzzy inference system (ANFIS). The influence of independent variables including electrolysis time, current intensity, initial concentration of caffeine, and type of anode were studied. Results showed that anode type followed by electrolysis time are the most important variables affecting caffeine degradation. Both CCD and ANFIS models were able to successfully predict the electrochemical process behaviour, while ANFIS models performed slightly better (R^2=0.993, RMSE=2.694 for caffeine removal efficiency, and R^2=0.976, RMSE=0.261 for energy consumption). Identification of intermediates by liquid chromatography-tandem mass spectrometry (LC-MS/MS) suggests that the degradation of caffeine by the EO process follows an oxidation pathway similar to the mechanism proposed for other advanced oxidation processes. The optimal conditions determined using CCD were applied on real municipal wastewater effluent in which caffeine removal efficiency varied between 78.0±4.3% and 92.5±1.0% for different initial caffeine concentrations showing the effectiveness of the process. Finally, toxicity assessment with Daphnia magna showed that the EO of real municipal wastewater effluent in optimal conditions may increase the toxicity levels of the samples. Toxicity could be reduced by extending the electrolysis time or could be completely eliminated using granular activated carbon.

Keywords:

Artificial intelligence, Electrooxidation, Endocrine disruptor, Toxicity, Central composite design

4.1. Introduction

With the world's rapid population growth and intense industrialization in the 20th century, environmental pollution has become a global problem with adverse impacts on the water sector. The vast majority of these pollution issues is caused by persistent organic compounds because of their resistance to conventional treatments such as physico-chemical or biological methods. This results in detection of refractory pollutants such as pesticides, pharmaceutically active compounds (PhAC), phenolic compounds, synthetic dyes, halogenated compounds, polycyclic aromatic hydrocarbon (PAHs), polychlorinated biphenyls (PCBs), endocrine disrupting chemicals (EDCs), and others in rivers, lakes, oceans and even drinking waters all over the world. They can cause hazardous health effects on living organisms including human beings. Therefore, advanced water and wastewater treatment have become a primary social, political, and environmental concern (Drogui *et al.*, 2007; Rahman *et al.*, 2009; Zheng *et al.*, 2017).

Caffeine, an alkaloid compound of the methylxanthine family, is a psychoactive drug widely used legally in foods, personal care products, beverages and medicines. Caffeine is considered safe at usual moderate intakes and is consumed for its benefits such as improved physical endurance, weight loss, cognitive alertness, and reduction of perceived fatigue (Beauchamp *et al.*, 2017; Nardi *et al.*, 2009). However, in excessive doses caffeine may exacerbate underlying psychiatric disorders, increase the risk of significant toxicity, cardiotoxicity resulting in arrhythmia, anxiety and panic disorders (Elhalil *et al.*, 2018; Wikoff *et al.*, 2017). Caffeine is commonly used as a key ingredient in tea, coffee, soft drinks, and chocolates. Based on global population and average consumption of 70 mg caffeine/person/day, about 460 tons of caffeine are being consumed each day by humans (Buerge *et al.*, 2003). An unidentified amount of this consumption is introduced into the wastewater treatment plants globally, mainly as the excretory residues of consumed caffeine, disposal of unconsumed caffeinated drinks, pharmaceutical products, manufacturing plant wastes, hospital wastes, etc (Li *et al.*, 2020a). Caffeine has been detected in wastewater treatment plant influents and effluents in concentrations ranging from ng/L to mg/L (Kahl *et al.*, 2017; Tran *et al.*, 2014). The highest concentrations of caffeine in influents and effluents of wastewater treatment plants have been reported

as 3.6 mg/L in Singapore and 66 µg/L in Spain, respectively (Li *et al.*, 2020b). Once discharged to the aquatic environment, caffeine is extremely stable due its high solubility in water (37.5 ± 1.0 g/L) and reported half-life of 100-240 days (Hillebrand *et al.*, 2012; Sriamornsak *et al.*, 2007). This has resulted in considering caffeine as one of the most abundant PhAC contaminants frequently detected in aquatic environments gaining attention for its adverse impacts in aquatic biota. Recent studies have demonstrated that caffeine residues can lead to multiple detrimental effects on aquatic organisms and terrestrial insects such as lethality, decreasing general stress, inducing oxidative stress and lipid peroxidation, affecting energy reserves and metabolic activity, neurotoxic effects, affecting reproduction and development, etc (Bruton *et al.*, 2010; Li *et al.*, 2020a).

Since municipal wastewater treatment plants cannot efficiently remove caffeine, alternative methods such as advanced oxidation processes (AOPs) have been applied for the degradation of caffeine (Arfanis *et al.*, 2017; Elhalil *et al.*, 2018; Ganzenko *et al.*, 2015; Rosal *et al.*, 2009; Trovó *et al.*, 2013). AOPs have emerged as a potent technique for degradation of various pollutants (Carabin *et al.*, 2016; Ferre-Aracil *et al.*, 2016; Tran *et al.*, 2010). In recent years, new AOPs based on electrochemical technology, the so-called electrochemical advanced oxidation processes (EAOPs), have been developed for the treatment of various wastewaters containing refractory contaminants (Daghrir *et al.*, 2014; Komtchou *et al.*, 2017; Olvera-Vargas *et al.*, 2015; Tran *et al.*, 2015b). Electrochemical oxidation (EO) as one of the promising EAOPs has gained attention for its great performance regarding the degradation of non-biodegradable pollutants (Garcia-Segura *et al.*, 2018; Martínez-Huitle *et al.*, 2018; Moreira *et al.*, 2017; Yang, 2020). Electrochemical oxidation can be carried out based on two different mechanisms of direct oxidation and indirect oxidation. In direct oxidation hydroxyl radicals are produced at the electrode surface by the oxidation of water molecules whereas in indirect oxidation different oxidant mediators in the bulk solution such as H_2O_2, HClO and $S_2O_8^{2-}$ are generated (Panizza, 2010). The nature of the anode material affects the nature and quantity of the hydroxyl radical generated. In non-active anodes such as lead (IV) oxide (PbO_2), tin (IV) oxide (SnO_2), and boron doped diamond (BDD) the electro-generated hydroxyl radicals remain physisorbed on the anode surface. The BDD anode, having a high oxygen evolution overpotential, generally produces greater quantities of highly

reactive and weakly adsorbed hydroxyl radicals to mineralize organic pollutants. Because of its characteristics, the BDD anode has become the state-of-the-art EAOP catalyst. However, recent studies have developed novel non-active anodes such as Graphdiyne (GDY) to overcome the challenges related to the BDD including its inherent high cost and manageability (Liu *et al.*, 2021). In the case of active anodes such as iridium (IV) oxide (IrO_2), ruthenium (IV) oxide (RuO_2), and other mixtures of metal oxides, there is a strong electrode-hydroxyl radical interaction forming higher metal oxides which can convert organic substrates to more oxidized intermediates (Durán *et al.*, 2018).

To study the effect of different independent variables on the process and calculating the response variables, response surface methodology (RSM) based on design of experiments (DoE) is often used instead of the classical approach in which one is changing one factor at a time while keeping all other factors fixed at a specified condition method. RSM helps identifying the single and combined effects (interactions) of process variables, find the optimal response of the process using a limited number of experiments saving time and cost by reducing the number of trials (Hakizimana *et al.*, 2017). RSM has been successfully applied to EO processes for degradation of various organic pollutants in wastewaters (Jardak *et al.*, 2017; Karimifard *et al.*, 2018; Zaviska *et al.*, 2012a).

Artificial intelligence (AI) methods have become an interesting alternative option in modelling and optimization of electrochemical processes in case phenomenological or conventional regression models (e.g., RSM) are not practical (Curteanu *et al.*, 2014). These data-driven tools model the system behaviour solely from mapping the input-output data rather than from process knowledge. Various types of problems in science can be cast in the form of such pattern-matching, and among the AI techniques, artificial neural networks (ANN) are one of the most effective methods (Gholami Shirkoohi *et al.*, 2021; Shirkoohi *et al.*, 2021). Hyperparameters of the ANN model such as the number of hidden layers and hidden neurons in each layer should be tuned for an efficient proposed model. This requires a black-box trial-and-error process. The adaptive neuro fuzzy inference system (ANFIS) combining the well-established learning technique of ANN and the linguistic characteristics of fuzzy logic theory has become a powerful and attractive AI modelling approach. The parameters of the Tagaki-Sugeno-Kang (TSK) inference model are updated from training data by employing the learning algorithm of ANN. In this

manner, the hidden layers and hidden neurons are determined accurately by a FIS in the ANFIS network. Hence, the difficulty of determining the ANN model configuration is eliminated (Abdulshahed *et al.*, 2015; Naghibi *et al.*, 2021). Taheri et al. (Taheri *et al.*, 2013) utilized ANFIS and RSM models for the optimization of Reactive Blue 19 removal using a combined electrocoagulation/coagulation process. High R^2 values (≥85%) indicated that the predictions of RSM and ANFIS models are acceptable for both responses. A comparison of ANFIS and other AI techniques was carried out by Farzin et al. (Farzin *et al.*, 2020) for the modelling of Ciprofloxacin removal from wastewater by electrochemical oxidation. While the interpolation method was used for increasing number of data, ANFIS performed better than ANN for in interpolated data which was in accordance with some other studies (Sahoo *et al.*, 2020; Souza *et al.*, 2018).

To the best of the authors' knowledge, this is the first comparative study between central composite design (CCD) based on RSM and ANFIS for modelling and optimization of psychoactive pharmaceutical caffeine removal by electrochemical oxidation by two active and non-active anodes. The main objective of this study was investigation of caffeine degradation in synthetic and real municipal wastewater effluents by an EO process in different experimental conditions. Based on experimental design, a factorial design (FD) followed was developed to investigate the main and interaction effects of different factors on caffeine removal efficiency. Quadratic polynomial models using CCD were used to determine the optimal experimental conditions for caffeine degradation and energy consumption. Furthermore, the results provided by CCD were compared with the ANFIS predicted values. Contribution of direct versus indirect effects while oxidizing caffeine using the electrooxidation process were also investigated. Another objective of this study was to determine the oxidation mechanism of the EO process. Finally, real municipal wastewater effluent was used to test the effectiveness of the optimal EO process for caffeine degradation in the presence of other pollutants. Toxicity was evaluated before and after treatment.

4.2. Materials and methods

4.2.1 Preparation of the synthetic solution

Caffeine (1,3,7-Trimethylpurine-2,6-dione, $C_8H_{10}N_4O_2$) was supplied by Thermo Fisher Scientific (≥99.7%). Some important characteristics of caffeine are shown in Table AI. 1.

Anhydrous sodium sulphate used as supporting electrolyte was analytical grade purchased from Thermo Fisher Scientific. The solution of caffeine was prepared in a volumetric flask containing 1 L of distilled water at room temperature in which different concentrations of caffeine were mixed with 1 g of electrolyte to increase electrical conductivity. Conditioning was carried out with a magnetic stirrer at high speed (700 rpm) for a 40-min period.

4.2.2 Real municipal wastewater treatment effluent

The municipal wastewater effluent used throughout this study was sampled from the Quebec Urban Community (CUQ) wastewater treatment plant (WTP, Beauport, Quebec City, Quebec, Canada). It is a conventional WTP with a short sludge age having a physicochemical pre-treatment followed by a bio-filtration process. Characteristics of the effluent samples are represented in Table AI. 2. These samples were used to test the electrooxidation removal of caffeine in the presence of other types of pollutants and to perform toxicity assessment. The EO treatment of municipal wastewater contaminated by caffeine was conducted without the addition of Na_2SO_4. This was because the effluent was conductive enough and Na_2SO_4 had no indirect effect on caffeine removal (as will be explained in section 3.3).

4.2.3 Electrolytic reactor setup

The electrolytic cell used was made of Plexiglas material with a dimension of 17.1 cm (depth) × 3.3 cm (width) × 11.4 cm (length). It was comprised of one anode and one cathode with an interelectrode gap of 1 cm. BDD and IrO_2 were selected as anodes because of their relatively high overvoltage for oxygen evolution. The utilized rectangular anode electrodes (10 cm × 11 cm of 0.1 cm thick) had a solid surface area of 68 cm^2 and

a void surface area of 45 cm^2. The cathode was a stainless steel grid having a surface area of 110 cm^2 (solid surface area of 55 cm^2 and a void surface area of 55 cm^2). The electrodes were vertically installed on a perforated Plexiglas plate at 1.8 cm from the bottom of the cell. All experiments were carried out in a batch reactor as shown in Figure AI. 1. The current intensity was delivered by means of an EXTECH 382275 DC power supply (1-30 V, 0-20 A).

4.2.4 Analytical details

Progress of the electrochemical degradation of caffeine in solution was monitored and quantified by absorbance measurements (absorption peaks previously determined) using a Varian Cary 100 ultraviolet (UV) spectrophotometer. The absorption peak measured at the wavelength of 273 nm was chosen to evaluate the residual caffeine concentration. The spectrum is characterized by a main peak located in the ultraviolet region at 273 nm. Calibration curves (absorbance versus concentration) were created to calculate the residual caffeine concentration in the solution. Figure AI. 2 represents examples of UV absorbance spectra for the two utilized anodes at different time intervals. The concentrations of caffeine in real wastewater and identification of intermediates were monitored and quantified by LC/MS/MS (Thermo TSQ Quantum Access). Chromatographic separation was achieved using a Hypersil Gold C18 column (Thermo Hypersil Ltd., Runcorn, UK) with a particle size of 3.0 mm and a 100 mm length × 2.1mm inner diameter. Total organic carbon (TOC) was measured by the high temperature catalytic combustion method and infrared detection using a Shimadzu TOC VCPH analyzer (Shimadzu Scientific Instruments, Kyoto, Japan). The sample was acidified to remove inorganic carbon.

Anionic species (Cl^-, NO_3^-, SO_4^{2-}) were measured by ion chromatography using a Thermo Integrion High-Pressure Ion Chromatography (HPIC). The ammonia concentration was determined according to the analytical method proposed by LACHAT Instrument (QuikChem® Method 10-107-06-2-B).

4.2.5 Toxicity assessment

In order to evaluate the acute toxicity of the untreated and treated real municipal wastewater effluent, a daphnia (Daphnia magna) test is used. Daphnia is a freshwater microcrustacean of the order of cladocerans and has been used for the determination of toxicity in different studies (Gimeno *et al.*, 2016; Gireli *et al.*, 2019; Ouarda *et al.*, 2018). This species is sensitive to a wide range of contaminants and is relatively easy to store in the laboratory. The neonate stage of the microcrustacean Daphnia magna Strauss (≤ 24 h) is used for testing. This toxicity test consists of determining the concentration of the sample which causes 50% mortality after 48 hours of exposure (48 hour LC50) in a static system and under controlled conditions. The photoperiod is 16h of light and 8h of dark cycle with a light intensity of 500 to 1000 lux. The sample is homogenized before the start of the test and kept at 20.0 ± 2.0 °C. Also, temperature, pH, dissolved oxygen, conductivity and hardness are measured and noted. If the hardness is less than 50 mg/l, it is adjusted to 50 mg/l with a concentrated solution of calcium chloride and magnesium chloride. If, and only if, the dissolved oxygen concentration is less than 40% or greater than 100% saturation, the sample is pre-aerated for a period not exceeding 30 minutes at a rate of 25 to 50 mL/min.L. No other changes are made to the sample. However, if the pH is extreme, an additional test can be performed with a pH-adjusted sample.

To determine the 48-hour LC50, a series of sample dilutions is performed and the percentage mortality is determined for each of the concentrations after 48 hours of exposure. The dilution water can be dechlorinated, ground or bottled water. This water must have a hardness between 160 and 180 mg/L, a pH between 6.5 and 8.5 (recommended between 7.0 and 8.0) and a dissolved oxygen content between 90 and 100% saturation. A dilution factor of 0.5 to 0.7 is generally appropriate (USEPA, 2002). Five concentrations of 6.25, 12.5, 25, 50, 100 (% v/v) plus a control (0 % v/v, dilution water only) have been used in this study. Four replicates for each concentration containing five neonates in 10 mL of each sample were subject to a photo period of 48 hours. The test results are acceptable if the percentage of mortality in the control groups is equal to or less than 5%. The concentration of the sample that causes 50% mortality in the population tested (LC50) will be calculated. Mortality is determined as the lack of

movement of the antennae, appendages, and the absence of a heartbeat as observed through a dissecting microscope.

4.2.6 Experimental design

The experimental design of the electrooxidation process for caffeine degradation was carried out using the response surface methodology. RSM is a collection of mathematical and statistical techniques for modelling, optimization and analysis of processes when several variables influence the response (Anderson *et al.*, 2016). In this study, an FD and a CCD were used successively for modelling and optimization of caffeine degradation by the electrochemical oxidation process. The FD was employed to investigate the main and interaction effects of the different factors on caffeine degradation and energy consumption. Subsequently, CCD was used to optimize the electrochemical oxidation process in terms of caffeine degradation and energy consumption. Four independent variables including electrolysis time (X_1), current intensity (X_2), initial concentration of caffeine (X_3) and anode type (X_4) were selected to model the electrooxidation process. The caffeine removal efficiency (Y_1) and energy consumption (Y_2) were considered as dependent factors (responses) and calculated using Eq. (4.1) and (4.2):

$$\text{Removal efficiency } (\%) = \frac{[\text{CAF}]_0 - [\text{CAF}]}{[\text{CAF}]_0} \times 100 \tag{4.1}$$

$$\text{Energy consumption } (\text{Wh/mg}) = \frac{\text{VI}\Delta\text{T}}{\Delta[\text{CAF}]} \tag{4.2}$$

where $[\text{CAF}]_0$ is the initial concentration of caffeine (mg/L), [CAF] is concentration of caffeine after different treatment times (mg/L); V is the applied voltage in volt (V), I is the current in Ampere (A), ΔT is the electrolysis time in hours and Δ[CAF] is the amount of caffeine removed in mg.

It should be noted that process variables and their ranges were determined by conducting preliminary experiments. Table AI. 3 shows the experimental independent factors, their ranges and coded values. Lab-scale electrolysis of 460 cm^3 of caffeine solutions were

carried out under galvanostatic conditions with current intensity ranging from 0.7 to 2.3 A (10.3 to 33.8 mA cm^{-2}) according to FD and CCD matrices.

4.2.7 ANFIS modelling

ANFIS introduced by Jang (Jang, 1993), a hybrid technique of artificial intelligence, combines a Sugeno type Fuzzy Inference System (FIS) and an artificial neural network (ANN). Fuzzy Logic is utilized in ANFIS to produce fuzzy rules and map the inputs to an output based on a given input-output data set. ANFIS applies the neural network learning process to learn from a given set of training data (like an ANN model), the rules and membership functions for tuning the FIS parameters. Likewise, the solution mapped out into the fuzzy model is explained in linguistic terms based on if-then fuzzy logic rules (Abdulshahed *et al.*, 2015).

Figure 4.1 shows the schematic of an ANFIS structure with two inputs, two rules and one output. As can be seen, the ANFIS structure includes five layers. Similar to ANN learning, ANFIS training comprises of a forward pass and a backward pass. Output values are the results of the forward pass through the net. In the backward pass, the error is calculated and is propagated back to the earlier layers in a similar manner as the backpropagation learning algorithm (Hussein, 2016). In the Input layer, inputs are introduced to the ANFIS network. Layer 1 is called the fuzzification layer that generates membership functions (MF) for each of the inputs. The nodes in this layer are adaptive nodes which have adjustable premises parameters related to input membership functions. In layer 2 or rule layer, the firing strength of each rule is calculated. Layer 3 or the normalisation layer computes the normalized value of the firing strength of each rule with respect to the firing strength of all rules. The fourth layer or defuzzification layer is an adaptive layer which calculates the values of the consequences of the rules. Consequent parameters in a *p*-th order polynomial function of the input signals (used in the Sugeno model) are adaptive parameters in this layer that must be adjusted. In layer 5 or the sum layer, the final output is calculated by summation of all incoming signals.

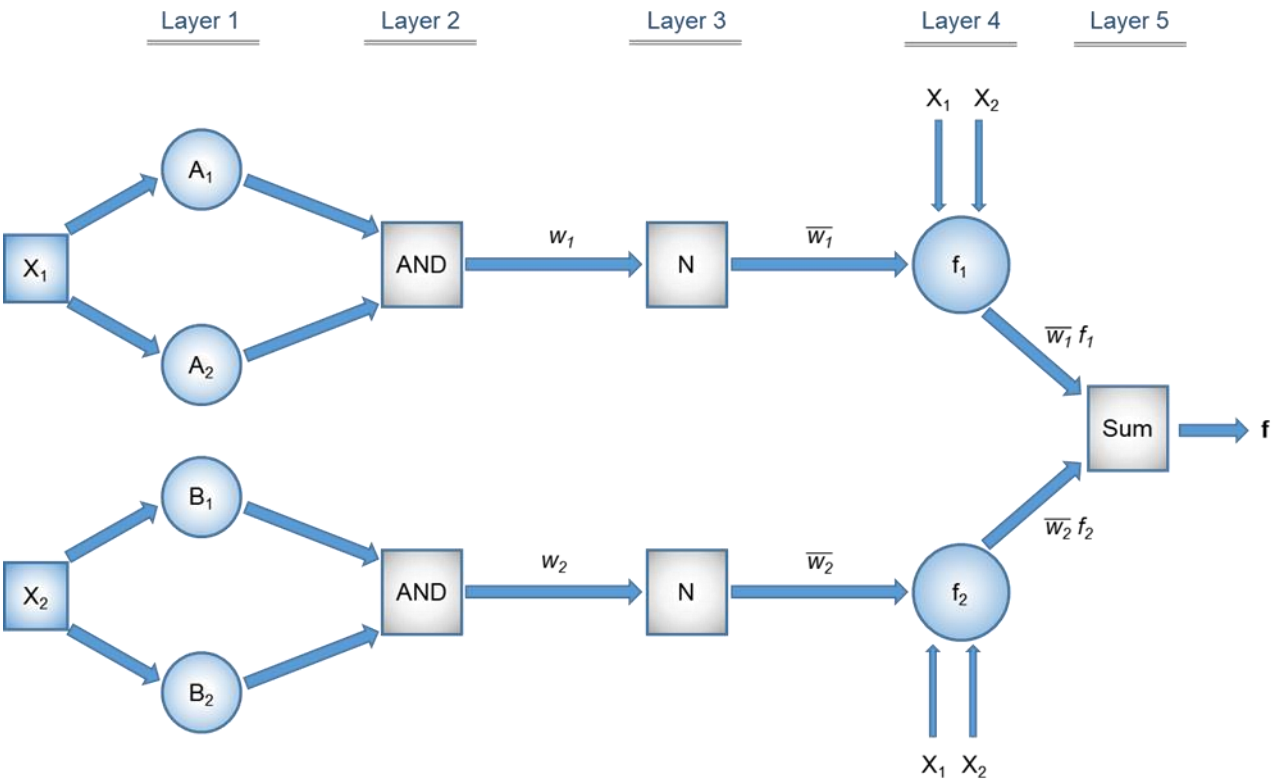


Figure 4.1 Schematic of the ANFIS structure (explanation in text)

In this study, ANFIS models were built in MATLAB ver. R2019a with the experimental data collected according the DoE. For each of the responses (removal efficiency and energy consumption) one model with four inputs and one output was developed.

To evaluate the performance of the CCD and ANFIS models, the correlation coefficient (R^2), the root-mean-square error (RMSE), and the mean absolute percentage error (MAPE) were used as comparison criteria. To train the ANFIS models RMSE was used as the error function. These functions were calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4.3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4.4}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (4.5)$$

where y_i is the experimental value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the average value of the experimental data.

4.3. Results and discussion

4.3.1 RSM modelling

4.3.1.1 Factorial design

A factorial design matrix (2^k) with k variables was used to investigate the influence of four main factors- electrolysis time (X_1), current intensity (X_2), initial concentration of caffeine (X_3), and type of anode (X_4)- on the efficiency of caffeine degradation and energy consumption. In this design, each factor is fixed to two levels normalized as -1 and +1, respectively. The experimental response associated to the 2^4 factorial design (four factors) could be represented by a linear polynomial model with interaction as follows:

$$Y=+50.79+10.85\ X_1+4.66\ X_2-3.97\ X_3-28.47\ X_4-2.93\ X_1X_4+4.14\ X_3X_4 \quad (4.6)$$

The percentage contribution of each factor and their interactions on the response can be calculated by Pareto analysis (Carabin *et al.*, 2016):

$$P_i = \left(\frac{b_i^2}{\sum_{i=1}^{k} b_i^2}\right) \times 100\ , (i \neq 0) \quad (4.7)$$

where b_i represents the estimation of the principal effects on the factor *i* and *k* is the number of factors. The percentage contribution of each of the factors is illustrated in Figure 4.2. It can be seen that caffeine degradation is mainly influenced by the anode type followed by electrolysis time; their effects are 81.0% (for anode type) and 11.8% (for electrolysis time) on the response. The current intensity, initial concentration and other interaction effects represent only 8% on the caffeine degradation. The anode electrode material is a major factor that influenced the rate of caffeine oxidation. In particular, “active” electrodes, such as IrO_2, are considered as a suitable catalyst for oxygen

evolution reactions that allow only partial oxidation of the organic compounds (electrochemical conversion). On the other hand, “non-active” electrodes, such as BDD, having high oxygen evolution potential, can lead to complete mineralization of organic pollutants (electrochemical combustion) which makes them more efficient as EO process (Durán *et al.*, 2018; Jardak *et al.*, 2016). As suggested by Comninellis (Comninellis, 1994), considering the very low oxygen overpotential of IrO_2 (E° = 0.25 V) secondary reactions may take place which decreases the degradation efficiency of organic pollutants. A BDD anode has shown its good performance for electrochemical caffeine degradation in other studies (Cotillas *et al.*, 2016; Indermuhle *et al.*, 2013; Martin de Vidales *et al.*, 2015). Moreover, different types of anodes such as graphite, graphite-PVC, Pt, and PbO_2 anodes, have also been used for electrooxidation of caffeine (Al-Qaim *et al.*, 2015; Chen, 2015; Periyasamy *et al.*, 2018). Chen et al. (Chen, 2015) utilized BDD, PbO_2, and Pt anodes for EO of caffeine under different operating conditions, and it was found that the performance of the tested anodes in terms of caffeine degradation efficiency and TOC reduction was in the order BDD > PbO_2 > Pt.

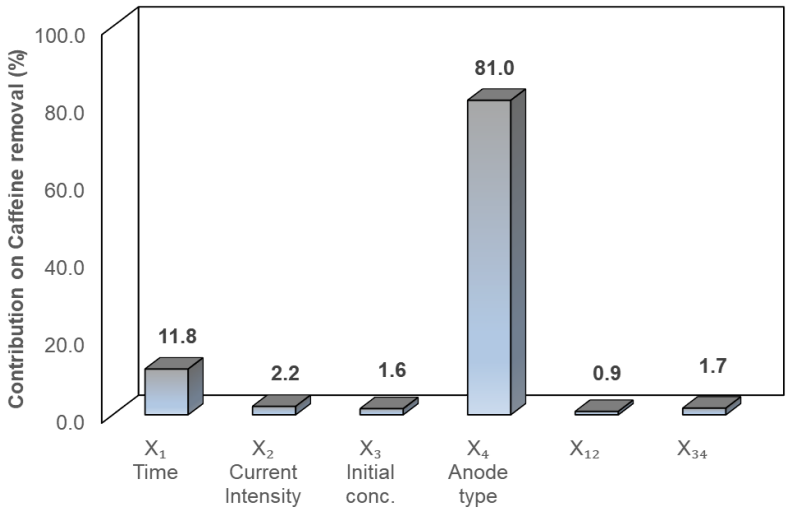


Figure 4.2 Contributions of different factors on caffeine removal

4.3.1.2 Central composite design

A CCD methodology based on RSM was employed to obtain a mathematical model for prediction of the behaviour of the electrochemical process and to find the optimal conditions for caffeine removal efficiency and energy consumption. The CCD matrix is

comprised of three sets of experiments: a factorial matrix at cubic points (16 assays), a star matrix at axial points at α=1.68 (12 assays), and a set of points at the center of the experimental region at α=0 (12 assays). The 4-factors five-level CCD matrix, and a comparison of experimental and predicted values of caffeine removal efficiency and energy consumption are listed in Table 1. The predicted responses are calculated based on the second-order quadratic polynomial equation given below:

$$Y = b_0 + \sum_{i=1}^{k} b_i \times X_i + \sum_{i=1}^{k} b_{ii} \times X_i^2 + \sum\sum b_{ij} \times X_i X_j + e_i \quad (4.8)$$

where Y is the response, X_i and X_j are the independent variables, b_0 corresponds to the average value of the responses, and b_i, b_{ii}, and b_{ij} are the linear, quadratic, and interaction effects between factors i and j for response Y, and e_i represents the random error on the response Y from the observed results. The coefficients of this model are calculated using the least-squares method:

$$B = (X^T X)^{-1} X^T Y \quad (4.9)$$

where B is the vector of estimates of the coefficients; X represents the model matrix; and Y corresponds to the vector of experiment results. The coded variables (X_i) were calculated by the following equation:

$$X_i = \frac{U_i - U_{i,0}}{\Delta U_i} \quad (4.10)$$

where $U_{i,0} = \frac{(U_{i,max} + U_{i,min})}{2}$ represents the value of U_i at the center of the experimental range; and $\Delta U_i = \frac{(U_{i,max} - U_{i,min})}{2}$ represents the step of variation; and $U_{i,max}$ $U_{i,min}$ are maximum and minimum values of the effective variable U_i, respectively.

Table 4.1 The CCD matrix with observed and predicted responses

Assay	U_1	U_2	U_3	U_4	Removal (%)			Energy Consumption (Wh/mg)		
					Observed	Predicted		Observed	Predicted	
	(min)	(A)	(mg/L)			CCD	ANFIS		CCD	ANFIS
1	30	0.7	30	BDD	78.00	74.08	78.00	0.27	0.11	0.27
2	47	1.5	30	BDD	99.69	99.59	99.69	1.11	1.63	1.11
3	30	1.5	30	BDD	80.64	80.87	83.90	0.86	0.87	0.87
4	40	1	40	IrO_2	19.35	22.54	19.35	1.82	1.60	1.71
5	30	1.5	30	IrO_2	20.14	20.66	17.24	3.35	3.54	3.94
6	30	1.5	30	BDD	86.95	80.87	83.90	0.88	0.87	0.87
7	40	2	40	IrO_2	38.79	31.02	38.79	2.77	2.99	2.77
8	30	0.7	30	IrO_2	12.46	13.87	12.46	1.56	1.99	1.56
9	40	1	20	IrO_2	23.78	24.79	23.78	3.00	3.13	3.00
10	30	1.5	47	IrO_2	24.15	24.11	24.15	1.84	1.78	1.84
11	40	2	20	IrO_2	39.01	33.27	39.01	5.41	5.61	5.41
12	13	1.5	30	IrO_2	5.11	2.73	5.11	5.83	4.70	5.83
13	20	1	40	IrO_2	11.26	7.94	11.26	1.59	2.37	1.59
14	30	1.5	30	IrO_2	16.73	20.66	17.24	3.88	3.54	3.94
15	40	2	20	BDD	100.00	105.04	100.00	2.35	2.25	2.35
16	20	2	40	IrO_2	20.56	16.42	20.56	2.73	3.76	2.73
17	20	2	40	BDD	58.88	65.08	58.88	1.05	0.77	1.05
18	47	1.5	30	IrO_2	26.73	27.56	26.73	3.84	3.39	3.89
19	30	1.5	30	BDD	85.93	80.87	83.90	0.83	0.87	0.87
20	30	1.5	30	BDD	81.29	80.87	83.90	0.91	0.87	0.87
21	30	1.5	13	IrO_2	34.11	27.93	34.11	4.28	5.31	4.28
22	13	1.5	30	BDD	49.50	51.12	49.50	0.65	1.10	0.60
23	30	1.5	30	IrO_2	17.69	20.66	17.24	3.93	3.54	3.94
24	20	1	20	BDD	71.63	68.04	71.63	0.53	0.45	1.03
25	20	2	20	BDD	79.06	76.52	79.06	1.54	1.94	1.54
26	40	1	40	BDD	82.36	85.11	82.36	0.46	0.69	0.46
27	20	1	40	BDD	52.33	56.59	52.33	0.37	0.38	0.25
28	30	1.5	30	IrO_2	15.94	20.66	17.24	4.14	3.54	3.94
29	30	1.5	30	IrO_2	15.71	20.66	17.24	4.24	3.54	3.94
30	30	1.5	30	BDD	84.70	80.87	83.90	0.85	0.87	0.87
31	30	1.5	30	BDD	81.90	80.87	83.90	0.86	0.87	0.87
32	30	1.5	47	BDD	78.72	76.50	77.01	0.58	0.34	0.58
33	40	2	40	BDD	91.06	93.59	97.63	1.27	1.08	1.27
34	30	2.3	30	IrO_2	26.40	27.44	26.94	4.99	5.09	4.99
35	30	1.5	13	BDD	86.63	95.96	93.59	1.86	1.39	1.86
36	30	1.5	30	IrO_2	18.76	20.66	17.24	3.46	3.54	3.94
37	30	2.3	30	BDD	88.60	87.65	88.75	1.68	1.62	1.68
38	20	1	20	IrO_2	9.53	10.19	14.70	3.82	3.90	3.82

39	40	1	20	BDD	98.78	96.56	97.98	0.76	0.76	0.76
40	20	2	20	IrO_2	16.24	18.67	18.50	6.78	6.39	7.37

Table 4.2 Optimal values of the process parameters for the maximum caffeine removal efficiency (%)

Solution	U_1	U_2	U_3	U_4	Removal (%)			Energy Consumption (Wh/mg)		
					Observed	Predicted		Observed	Predicted	
	(min)	(A)	(mg/L)	-		CCD	ANFIS		CCD	ANFIS
	39	0.7	13	BDD	93.82±0.80	100.46	95.24	0.70±0.02	0.26	0.91

Based on the experimental results, quadratic polynomial equations between the independent variables (current intensity, electrolysis time, initial concentration of caffeine, and type of anode) and responses (caffeine removal efficiency and energy consumption) in coded forms were obtained using the Design-Expert® Program:

$$Y_1 = 50.76+10.78X_1+4.24X_2-3.42X_3-30.10X_4-3.48X_1X_4+2.30X_3X_4-1.91X_1^2+1.86X_3^2 \quad (4.11)$$

$$Y_2 = 2.21-0.12X_1+0.72X_2-0.67X_3+1.34X_4-0.27X_1X_4-0.27X_2X_3+0.25X_2X_4 -0.37X_3X_4+0.17X_1^2 \quad (4.12)$$

According to Table 4.1, it can be observed that there is good agreement between experimental and CCD predicted values for caffeine removal efficiency and energy consumption. An analysis of variance (ANOVA) was applied to evaluate the fitted models (Table 4.3). The F-values of the developed models were calculated as 277.72 (removal efficiency) and 52.02 (energy consumption), indicating that there is only a 0.01% chance that a "Model F-Value" this large could occur due to noise. The low p-value of the models (< 0.0001) demonstrates that the models are statistically significant. In the case of caffeine removal efficiency, the model terms X_1, X_2, X_3, X_4, X_1X_4, X_3X_4, X_1^2, X_3^2 were found to be significant while in case of energy consumption the model terms X_1, X_2, X_3, X_4, X_1X_4, X_2X_3, X_2X_4, X_3X_4, X_1^2 were. Model terms with p-values larger than 0.10 were considered insignificant (Kumari *et al.*, 2019) and eliminated from the developed quadratic models. Surface plots of the CCD models as function of electrolysis time and current intensity at the center point of the initial caffeine concentration for the two anodes are illustrated in Figure 4.3. The main effects of each variable on the caffeine removal efficiency and energy consumption are represented in Figure 4.4. As can be seen, the caffeine removal efficiency is improved by increasing electrolysis time, and current intensity and is decreased by an increase in initial caffeine concentration. For the energy consumption, it can be said that the BDD anode is more energy efficient than the IrO_2 one in caffeine removal since the BDD curves are below the IrO_2 curves in all plots. Also, the energy consumption rises by increasing current intensity and initial caffeine concentration. These findings for the behaviour of the caffeine removal efficiency and energy consumption are

compatible with other studies in literature for electrochemical oxidation of caffeine in aqueous solutions (Al-Qaim *et al.*, 2015; Periyasamy *et al.*, 2018).

Table 4.3 ANOVA results of the quadratic models for caffeine removal efficiency and energy consumption

	Removal efficiency			Energy consumption		
	Coefficient	F-value	p-value	Coefficient	F-value	p-value
Model	-	277.72	< 0.0001	-	52.02	< 0.0001
Intercept	50.76	-	-	2.21	-	-
X_1	10.78	172.22	< 0.0001	-0.12	1.59	0.2175
X_2	4.24	26.66	< 0.0001	0.72	61.38	< 0.0001
X_3	-3.42	17.38	0.0002	-0.67	53.72	< 0.0001
X_4	-30.1	1967.43	< 0.0001	1.34	310.72	< 0.0001
X_1X_4	-3.48	17.93	0.0002	-0.27	8.64	0.0063
X_2X_3	-	-	-	-0.27	5.20	0.0299
X_2X_4	-	-	-	0.25	7.30	0.0112
X_3X_4	2.3	7.84	0.0087	-0.37	15.85	0.0004
X_1^2	-1.91	5.74	0.0228	0.17	3.82	0.0602
X_3^2	1.86	5.45	0.0263	-	-	-

Optimal conditions were obtained by Design-Expert 7.0 regarding maximum removal efficiency (5/5 weighting factor) and minimum energy consumption (3/5 weighting factor) (Table 4.2). Experiments showed that with an electrolysis time of 39 min, an imposed current intensity of 0.7 A, an initial caffeine concentration of 13 mg/L, and a BDD anode, 93.82±0.80% caffeine removal efficiency can be achieved with energy consumption of 0.70±0.02 Wh/mg. Under these conditions, as Figure AI. 3 shows, a TOC removal efficiency of 50.1±0.1% was measured. At least, half of the organic carbon was mineralized (transformed into carbon dioxide and water). The remaining may be transformed to smaller molecules and by-products.

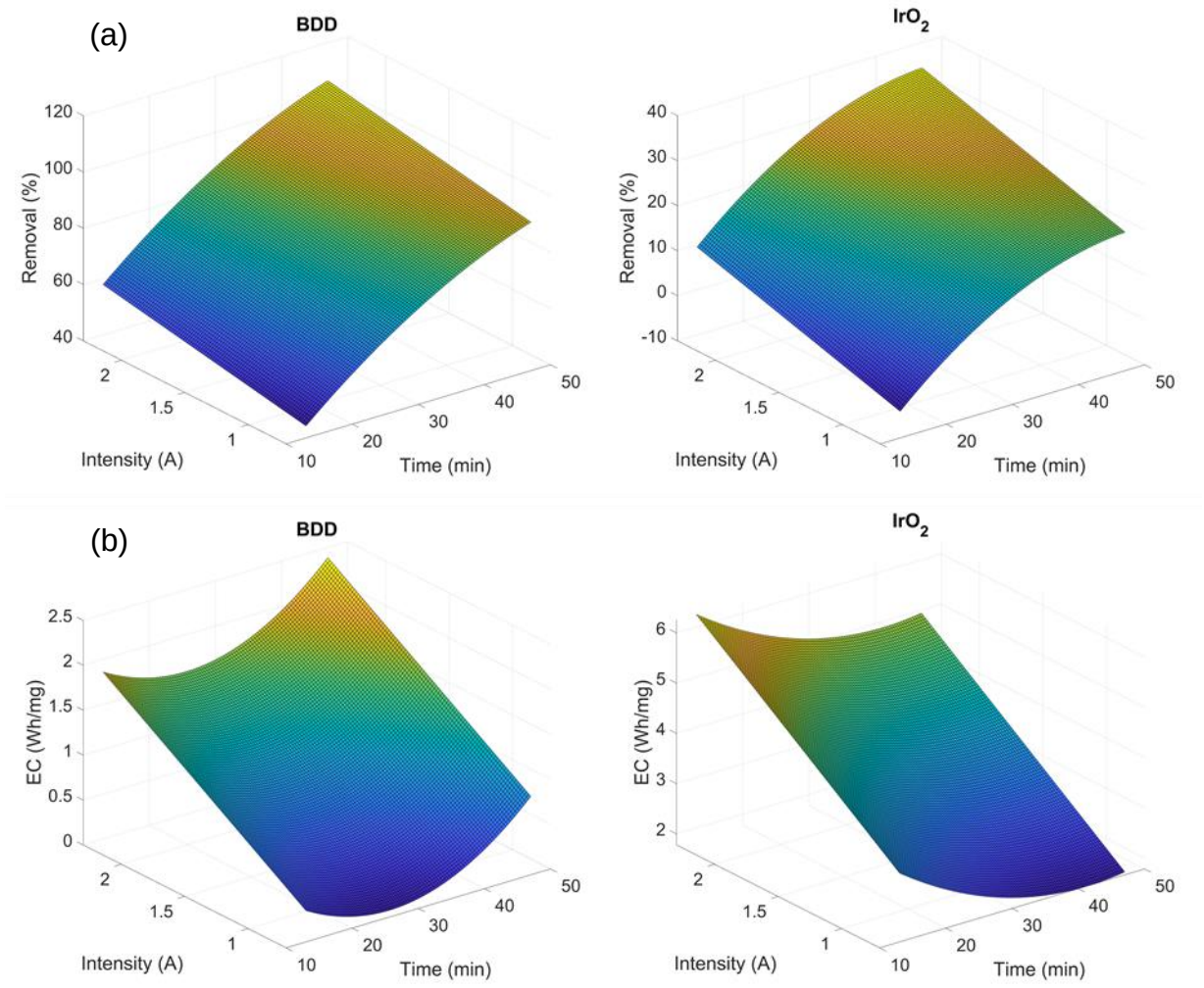


Figure 4.3 Surface plots of the CCD models as a function of electrolysis time and current intensity at an initial caffeine concentration of 30 mg/L for the two anodes; (a) caffeine removal efficiency (%), and (b) energy consumption (Wh/mg)

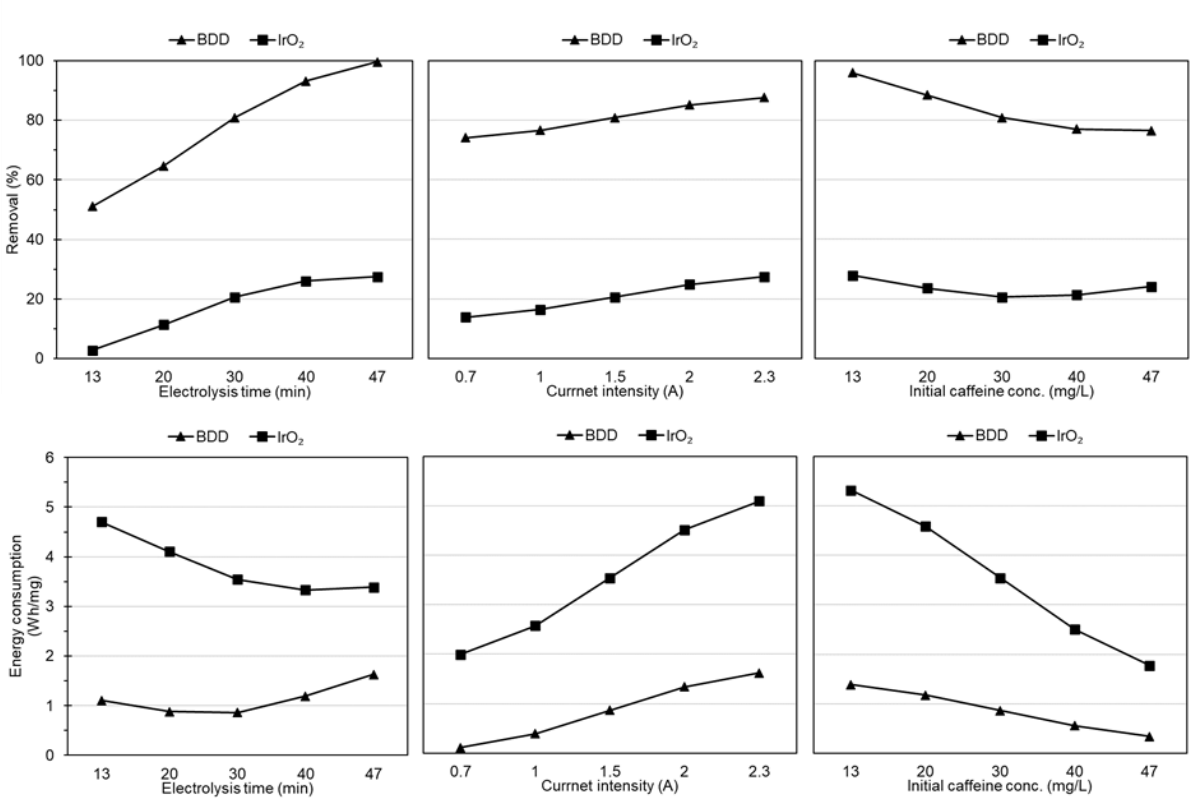


Figure 4.4 Main effect plot of each independent variable on caffeine removal efficiency and energy consumption

4.3.2 ANFIS modelling

The experimental data of the CCD design for caffeine removal efficiency and energy consumption (each consisting of 40 experiments) were divided into training (75%), validation (12.5%), and testing (12.5%) data sets. The training data set was used to train the ANFIS, whereas the validation data set was used to prevent overfitting and improve the generalization performance of the trained model. The optimal model was then applied on the test data set in order to evaluate the ANFIS model on the unseen data.

ANFIS models were constructed based on the obtained dataset from the CCD design using the Fuzzy C-Means (FCM) clustering method. FCM integrated with ANFIS helps to obtain a relatively small number of rules which prevents the model to be too complex and to minimize the overfitting issue. The FCM clustering method partitions the input data into different clusters and is used to identify the fuzzy membership functions and fuzzy rule base for the ANFIS model (Abdulshahed *et al.*, 2015).

The architecture of the optimal ANFIS models obtained after tuning the parameters used for caffeine removal efficiency and energy consumption are given in Table 4.4 and the schematic of the model structure is represented in Figure AI. 4. The prediction error analysis for ANFIS models in training, validation, and testing phases for caffeine removal efficiency and energy consumption are presented in Table 4.5. It is common to see that better solutions are obtained in the training and validation stages as compared to the testing stage where unseen data is presented to the models (Khude *et al.*, 2020; Manu *et al.*, 2017). The number of clusters is one of the key parameters of the FCM-ANFIS model. It plays an important role in the architecture of the network and simplicity or complexity of the trained model and has been considered as hyperparameter for ANFIS model tuning (Naghibi *et al.*, 2021). Table 4.6 shows the impact of increasing the number of clusters on the architecture and performance of the trained models in training, validation and test. As can be seen, setting a higher number of clusters for ANFIS models results in having more model parameters which make the model more complex and may lead to overfitting. The overall performance of the two models for removal efficiency and energy consumption increases for the training samples (lower RMSE values) but the performance drops for the validation and test sets (higher RMSE values) as the number of clusters builds up. For both models, the training RMSE reaches a minimum value after setting the number of clusters to 4 or higher but the RMSE increases for validation and test sets which defines that models have probably memorized the training set and noise but lost their ability to generalize well to new data. Hence, the number of clusters has been set to two for caffeine removal efficiency and energy consumption. The ANFIS predicted outputs are tabulated along with the CCD predicted outputs and the experimental values in Table 4.1.

Table 4.4 The ANFIS model characteristics

Parameters	Value/description
Fuzzy structure	Takagi-Sugeno
Initial FIS generated	Fuzzy C-Means clustering
Input membership function type	Gaussian ('gaussmf')
Output membership function type	Linear
No. of clusters	2
Optimization method	Hybrid (least-squares and backpropagation gradient descent method)
Number of linear parameters	10
Number of nonlinear parameters	16
Total number of parameters	26
No. of fuzzy rules	2
No. of maximum epoch	30
Number of inputs	4
Number of outputs	1

Table 4.5 Error analysis for the ANFIS models

Parameters	Caffeine removal efficiency			Energy consumption		
	Train	Validation	Test	Train	Validation	Test
R^2	0.996	0.997	0.954	0.986	0.949	0.917
RMSE	1.996	1.727	5.584	0.203	0.082	0.540
MAPE	5.853	5.891	11.647	9.959	5.226	8.788

Table 4.6 Error analysis as function of the number of clusters for the FCM method

No. of clusters	No. of fuzzy rules	Total no. of parameters	Caffeine removal efficiency			Energy consumption		
			Train	Validation	Test	Train	Validation	Test
2	2	26	1.996	1.727	5.584	0.203	0.082	0.540
3	3	39	1.834	1.619	8.570	0.284	0.195	0.431
4	4	52	1.125	4.056	14.542	0.188	0.122	1.046
5	5	65	1.122	4.972	18.087	0.147	0.264	1.224
6	6	78	1.122	5.123	23.358	0.147	0.782	1.854
7	7	91	1.122	5.675	25.245	0.147	1.391	2.442

The obtained ANFIS and CCD models for caffeine removal efficiency and energy consumption were also compared and performance results are presented in Table 4.7. As can be seen, while both models could successfully predict the electrochemical system behaviour, ANFIS models performed slightly better. The calculated R^2 of the CCD models

were 0.986 and 0.939 for caffeine removal efficiency and energy consumption, respectively. The corresponding R^2 values of the ANFIS models were obtained as 0.993 and 0.976 for caffeine removal efficiency and energy consumption, respectively. Surface plots of the ANFIS models as a function of electrolysis time and current intensity at the center point of the initial caffeine concentration for the two anodes are illustrated in Figure 4.5. Also, the ANFIS models show smaller deviations compared to the CCD models illustrating their ability to be better fitted to experimental data especially for the case of energy consumption. The superiority of ANFIS over RSM models is in accordance with studies of Noorani Khomeyrani et al. (Noorani Khomeyrani *et al.*, 2021) and Lenin Sundar et al. (Lenin Sundar *et al.*, 2021). However, it should be mentioned that with the limited number of data available, the difference in number of parameters for the two modelling approaches should be an important criterion. Indeed, CCD models are able to fit the available data with a lower number of parameters compared to ANFIS models. Figure 4.6 illustrates the comparative parity plot of the CCD and ANFIS models for the two responses. The blue dashed line in the two plots represents the line of equation y=x. Ideally, points on this line mean that the predicted value is equal to the experimental value. For a good fit, points should be close to the dashed line. As can be seen, both models perform well for the caffeine removal efficiency. For the energy consumption, points are more scattered from the dashed line and the CCD models show larger deviations than the ANFIS models.

Table 4.7 Performance Comparison between CCD and ANFIS models

Parameters	Caffeine removal efficiency		Energy consumption	
	CCD	ANFIS	CCD	ANFIS
R^2	0.986	0.993	0.939	0.976
RMSE	3.778	2.694	0.418	0.261
MAPE	10.173	6.582	16.690	9.221
Total number of parameters	9	26	10	26

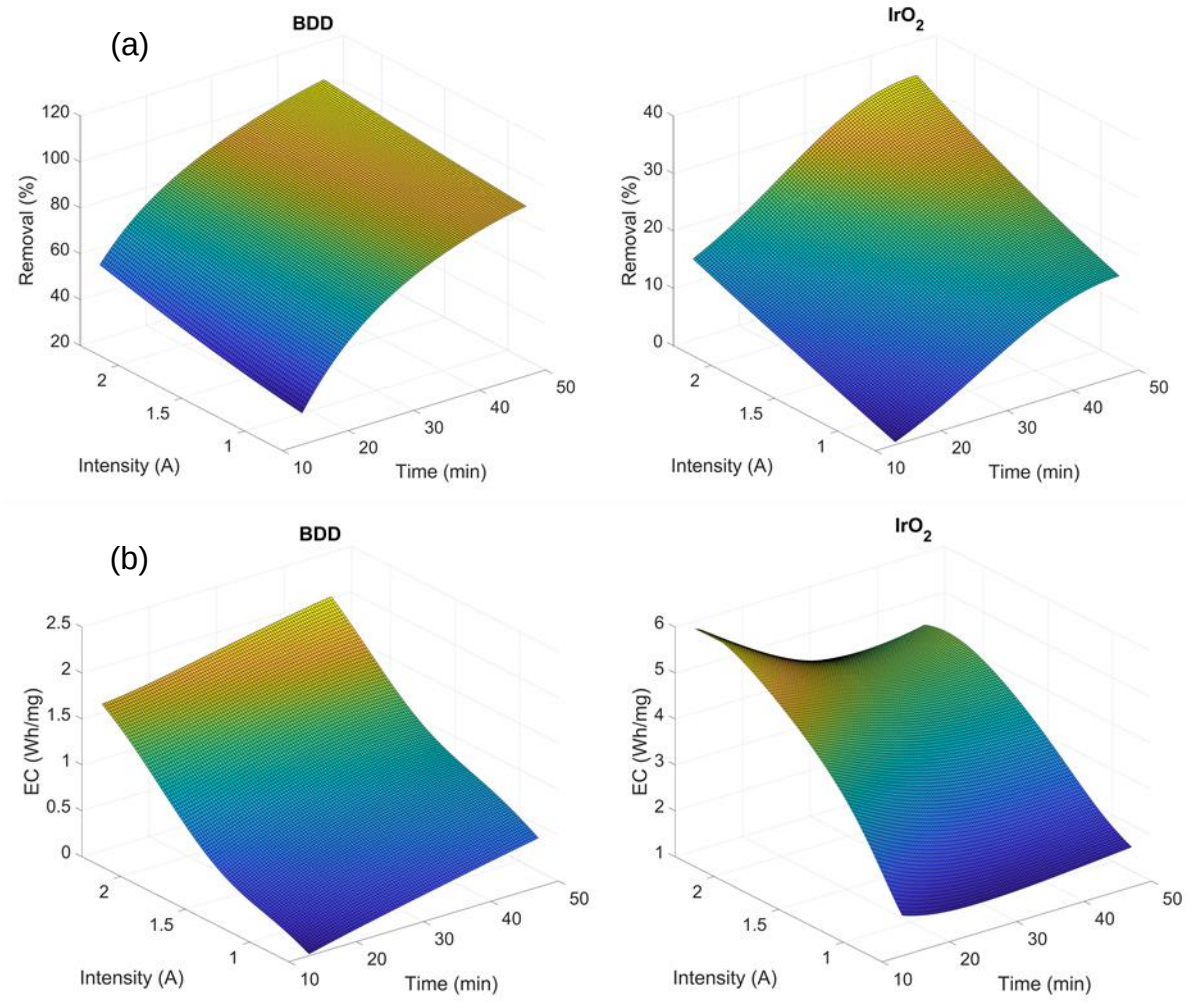


Figure 4.5 Surface plots of the ANFIS models as a function of electrolysis time and current intensity at the initial caffeine concentration of 30 mg/L for the two anodes; (a) caffeine removal efficiency (%), and (b) energy consumption (Wh/mg)

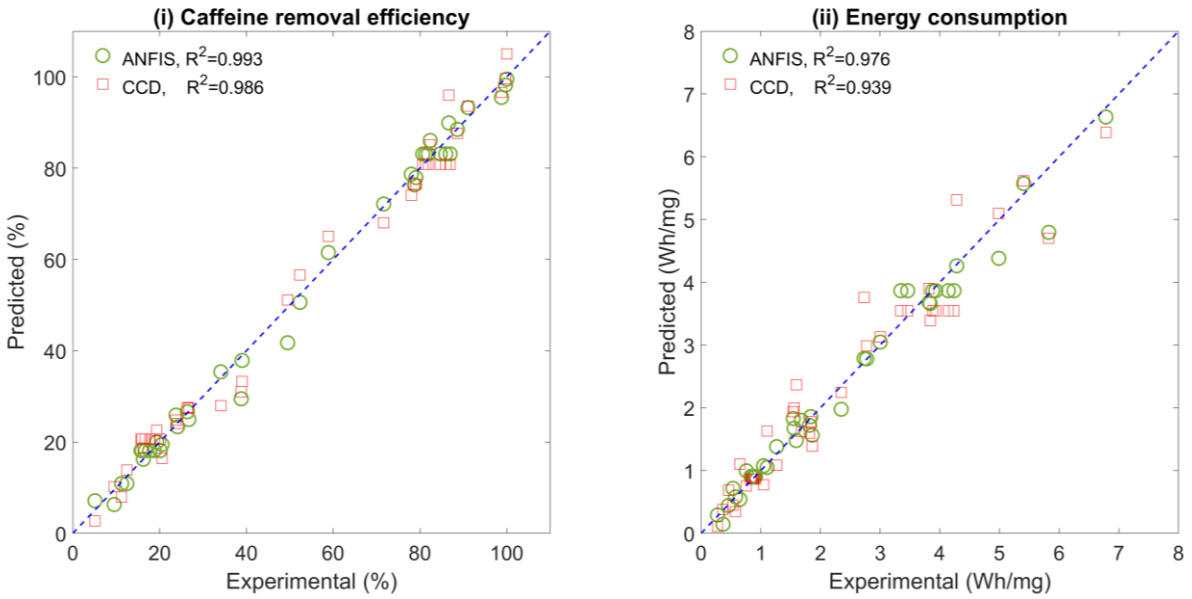


Figure 4.6 Parity plots of the experimental and predicted values for ANFIS and CCD models, (i) caffeine removal efficiency (%), and (ii) energy consumption (Wh/mg)

4.3.3 Contribution of direct-indirect oxidation in the removal of caffeine

There are two types of reaction in EO: the direct and indirect EO reaction. In direct EO, electron exchange happens at the anode surface without the contribution of other species. In this way, hydroxyl radicals ($E^{\circ}(OH^{\circ}/H_2O)$ = 2.80 V vs. SHE) are produced at the electrode surface by the oxidation of water molecules, and organic compounds can be degraded by reacting with absorbed OH° radicals. In indirect EO, other oxidant species can be promoted by the generation of different oxidant mediators in the bulk solution such as H_2O_2, HClO and $S_2O_8^{2-}$ (Tran *et al.*, 2013).

The objective of this part of study was to examine whether electrochemical oxidation of caffeine is obtained from direct electrochemical oxidation or indirect electrochemical oxidation. Upon utilization of sodium sulfate as electrolyte, powerful oxidants such as persulfate ($S_2O_8^{2-}$) can be produced in the bulk solution capable of oxidizing organic structures (Jardak *et al.*, 2016) based on the following equation:

$$2SO_4^{2-} \rightarrow S_2O_8^{2-} + 2e^- \quad (4.13)$$

To understand the mechanism behind caffeine removal, additional experiments were carried out to firstly evaluate the production of oxidants by the electrolytic cell and secondly, whether these oxidants contribute to the caffeine degradation or not. For the measurement of the oxidants produced, a synthetic solution containing 1 g/L of sodium sulfate in distilled water was electrolysed for 60 minutes with a current intensity of 0.7 A. At each time step a volume of 10 mL of sample was taken from the electrolytic cell and added to a beaker with 500 mg of potassium iodide. The mixture was then agitated for 15 min to let the electro-generated oxidant react with iodide and form triiodide. Then the oxidant production was evaluated by absorbance measurements at λ=353 nm using a UV-spectrophotometer.

The absorbance measurements with the sodium sulfate solution showed a constant increase up to 40 min after which it remained almost constant up to 60 min (Figure 4.7). It is assumed that this increase of absorbance is due to the persulfate production at the BDD anode. To evaluate the direct effect of the electrooxidation of caffeine, sodium nitrate (with the same molarity as sodium sulfate (7.04 mM)) was used as supporting electrolyte since no oxidant agent can be produced with this type of electrolyte. The conductivity of the solutions with sodium sulfate and sodium nitrate electrolytes were measured as 1501±5 and 837±6 µS/cm, respectively. The absorbance measurement for $NaNO_3$ in Figure 4.7 illustrates this.

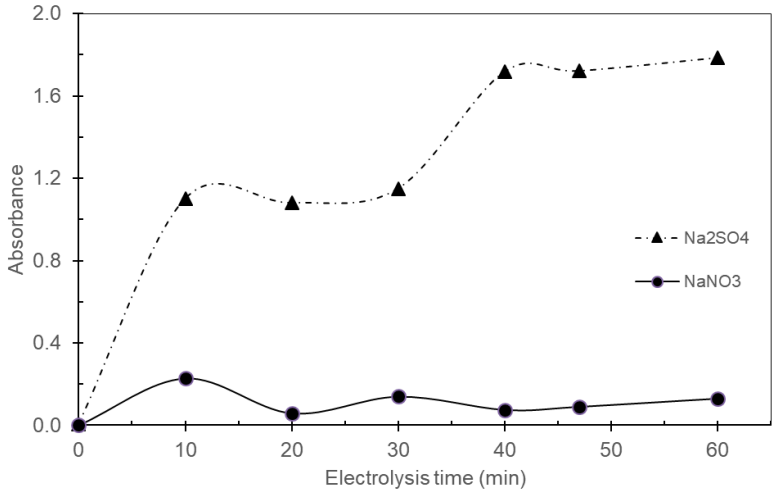


Figure 4.7 Oxidation capacity measurement of two supporting electrolytes versus time (current intensity=0.7 A; Electrolyte conc.= 7 mmol/L; anode= BDD)

While it was shown that the electrolytic reactor used in this study was capable of producing strong oxidants, additional experiments were conducted to evaluate whether the oxidant produced can oxidize caffeine by indirect electrochemical oxidation or not. Samples were taken at each time step during electrolysis of a synthetic solution of sodium sulfate and then contaminated artificially with 13 mg/L of caffeine and agitated for 30 min. Also, for comparison, another experiment was conducted with sodium nitrate as electrolyte to study direct electrochemical oxidation and one other experiment with sodium sulfate to evaluate the combination of direct and indirect electrochemical oxidation. In these two experiments, an initial caffeine concentration of 13 mg/L and a current intensity of 0.7 A were used in the electrolysis process. As Figure 4.8 shows indirect oxidation with persulfate has no contribution to caffeine removal during electrolysis. The absence of significant indirect oxidation is also represented in Figure 4.8 by the contribution of direct oxidation with sodium nitrate as electrolyte. As can be seen, the direct oxidation with sodium nitrate shows a similar behaviour as electrolysis using sodium sulfate meaning that the caffeine removal in the electrolytic reactor was mainly due to direct oxidation. Indirect oxidation was negligible.

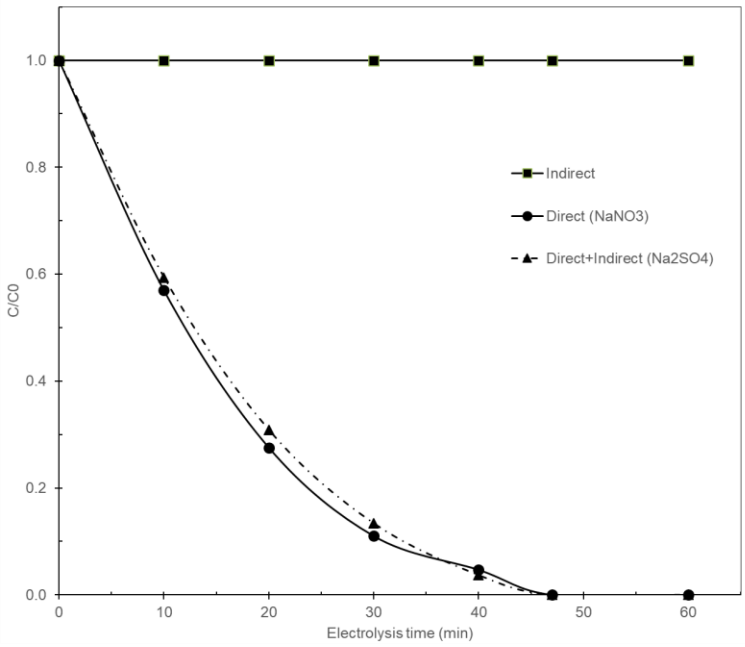


Figure 4.8 Contribution of direct oxidation, indirect oxidation, and their combination in the removal of caffeine

4.3.4 Identification of caffeine by-products

The objective of the work reported in this section was to monitor by-products formed by electrooxidation of caffeine in synthetic solution and in this way know more about the progress of the oxidation. The electrolytic cell was operated at the optimal conditions defined above but with a higher initial caffeine concentration of 47 mg/L instead of 13 mg/L to be able to analyze the intermediates easier using LC-MS/MS. Under this experimental condition, known by-products of caffeine oxidation were identified, namely dimethylparabanic acid (P_1, $C_5H_6N_2O_3$, M_w 142), di(hydroxymethyl) parabanic acid (P_2, $C_5H_7N_2O_5$, M_w 174). The oxidation of caffeine starts by a fast attack of hydroxyl radicals to the C=C double bond. Dalmázio et al. (Dalmázio *et al.*, 2005) reported that after successive hydroxylations and oxidations in different advanced oxidation processes such as UV/TiO_2, H_2O_2/UV and Fenton, dimethylparabanic acid and di(hydroxymethyl) parabanic acid can be formed where di(hydroxymethyl) parabanic acid is slowly mineralized to CO_2, NH_3, and CH_3NH_2. The same by-products were also characterized in other advanced oxidation processes with caffeine (Chuang *et al.*, 2011; Kolonko *et al.*, 1979; Li *et al.*, 2021). Similar to other studies 1,3,7-trimethyluric acid (P_3, $C_8H_{10}N_4O_3$, M_w 210), 1,3-dimethyluric acid (P_4, $C_7H_8N_4O_3$, M_w 196) and theophylline (P_5, $C_7H_8N_4O_2$, M_w 180) were identified as intermediates (P. Telo *et al.*, 1997; Stadler *et al.*, 1996). This shows that electrochemical oxidation of caffeine with the BDD anode may follow the similar reaction pathways to other advanced oxidation processes. In the present study, analysis of the EO of caffeine suggests similar multistep and interlinked pathways (Dalmázio *et al.*, 2005; Kolonko *et al.*, 1979) as presented in Figure 4.9.

Furthermore, TOC removal analysis under the same operating conditions was carried out and showed that caffeine was partially mineralized and oxidized into water and carbon dioxide. After 47 min of electrolysis the maximum TOC removal reached 54% (Figure 4.10). It should be mentioned that at the end of the experiment, more than 95% of the caffeine had been removed. This indicates that about 46% of the TOC that remained in the solution originates from caffeine by-products showing formation of smaller molecules in the electrolytic cell.

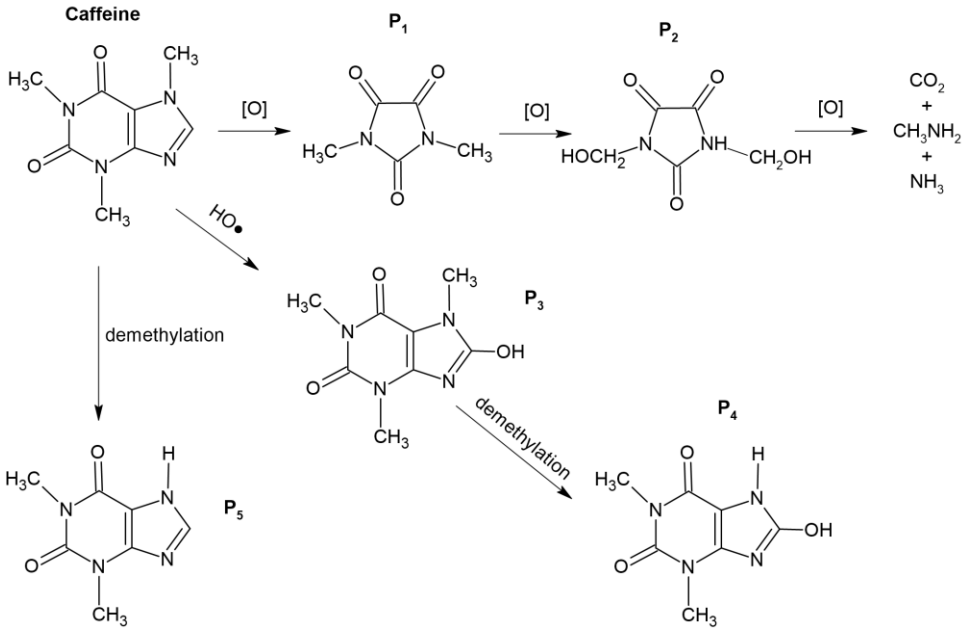


Figure 4.9 Proposed reaction pathways for the EO degradation of caffeine

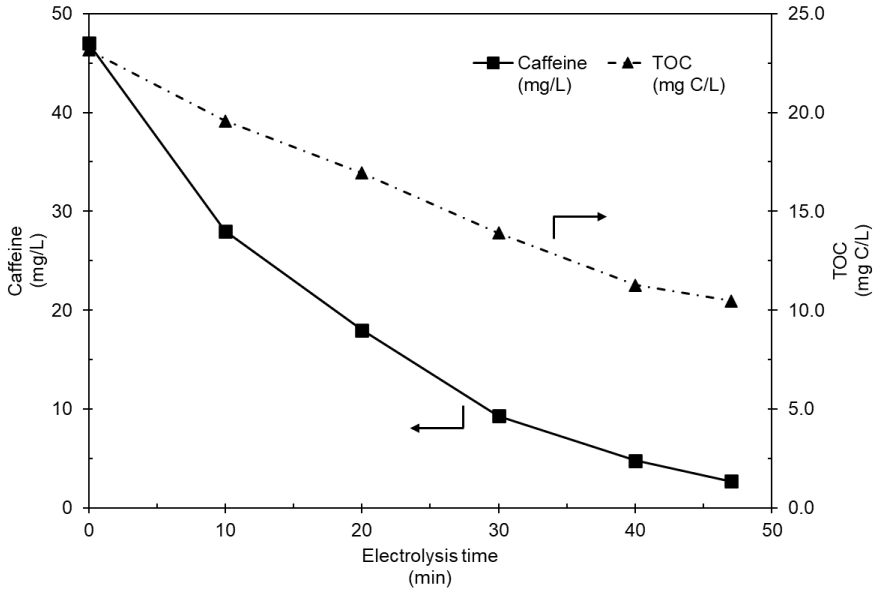


Figure 4.10 Caffeine and TOC removal efficiency at: [initial caffeine]=47 ppm, current intensity= 0.7 A, and BDD anode

4.3.5 Application of electro-oxidation on real municipal wastewater effluent

The optimal conditions obtained by the CCD methodology (current intensity of 0.7 A, electrolysis time of 39 min, and BDD electrode) without the addition of electrolyte (Na_2SO_4) was used to determine the effectiveness of the EO process for caffeine

degradation of real wastewater effluent and effluent contaminated by caffeine. The objective was to evaluate the effectiveness of the electrolytic cell as a tertiary treatment to remove emerging pollutants such as caffeine in the presence of other types of pollutants such as inorganic, organic, and microbial contaminants. In order to have reproducible results, experiments were conducted in replicates to provide reliable outcomes of the electrochemical process for treating municipal wastewater treatment effluent contaminated by caffeine. The liquid chromatography of the municipal wastewater treatment effluent showed an initial caffeine concentration of 20 µg/L. In addition to the caffeine present in municipal wastewater treatment effluent, the effluent was artificially contaminated with caffeine to reach 100 µg/L and 13 mg/L. It should be mentioned that the municipal wastewater effluent contained 116±12.7 mg/L of chloride ions that help the conductivity of the solution for electrical current passage. The formation of hypochlorous acid (HClO) as indirect electrochemical oxidation with the real wastewater effluents has been reported in other studies for degradation of refractory pollutants such as bisphenol-A and ibuprofen (Tran *et al.*, 2015b; Zaviska *et al.*, 2012a). Moreover, studies have shown that the presence of Cl^- accelerates the caffeine degradation in different AOPs (Martín de Vidales *et al.*, 2015; Rao *et al.*, 2021). This was confirmed by our findings for the EO of caffeine with different types of electrolyte in synthetic solutions (Figure AI. 5). However, as Indermuhle et al. (Indermuhle *et al.*, 2013) reported, more reaction intermediates such as organo-chlorinated compounds are formed in the electrooxidation of caffeine in the presence of NaCl as a supporting electrolyte compared to Na_2SO_4. These compounds can cause toxicity to the EO-treated solution. Since the municipal wastewater effluent considered in this study contains chloride ions, the toxicity evaluation will be studied.

The LC-MS/MS analysis showed that the caffeine removal efficiency varied between 78.0±4.3% and 92.5±1.0% for real municipal wastewater effluent compared to the removal efficiency of 93.8% recorded in the synthetic effluent (Table 4.8). These results confirm the ability of the EO process to remove caffeine even in the presence of other organic contaminants. The initial TOC concentration of the samples was 8.7±0.9 mg/L, 9.5±0.4 mg/L, and 16.2±0.4 mg/L for raw effluent, contaminated effluent of 100 µg/L, and contaminated effluent of 13 mg/L, respectively. The residual concentrations of TOC after electrolysis for the three samples were measured in the range between 6.2±0.6 and

8.4±0.1. The TOC removal efficiency of the samples was obtained as 21.3±5.3%, 34.4±3.5%, 48.1±0.5% for raw effluent ([CAF]=20 µg/L), contaminated effluent ([CAF]=100 µg/L), and contaminated effluent ([CAF]=13 mg/L), respectively (Figure 4.11). The TOC removal of the real wastewater effluent especially in the low to medium concentrations shows lower efficiencies compared to the TOC removal efficiency of 50.1±0.1% in the optimal condition of synthetic solution. The presence of other organic contaminants than caffeine in the real effluent makes the TOC removal a competitive process. Also, the low efficiency of TOC removal in real wastewater effluent reflects that only a small fraction of organic compounds consisted of caffeine and other refractory organic matters are mineralized (transformed into carbon dioxide and water), and the remaining may be transformed into the by-products. These by-products may be equally or even more toxic than the parent compounds and their formation can cause toxicity to the treated effluent. In order to investigate the toxicity of the raw and treated municipal wastewater effluent, toxicity was assessed.

Table 4.8 Application of EO process for the removal of caffeine in municipal wastewater effluent

	Raw effluent (20 µg/L)		Contaminated (100 µg/L)		Contaminated (13 mg/L)	
	Initial	Residual	Initial	Residual	Initial	Residual
CAF (µg/L)	19.0±1.4	4.2±1.1	95.5±3.5	7.1±1.2	12100±283	1350±212
CAF Removal rate (%)	78.0±4.3		92.5±1.0		88.9±1.5	
TOC (mg/L)	8.7±0.9	6.9±0.7	9.5±0.4	6.2±0.6	16.2±0.4	8.4±0.1
TOC Removal rate (%)	21.3±5.3		34.4±3.5		48.1±0.5	

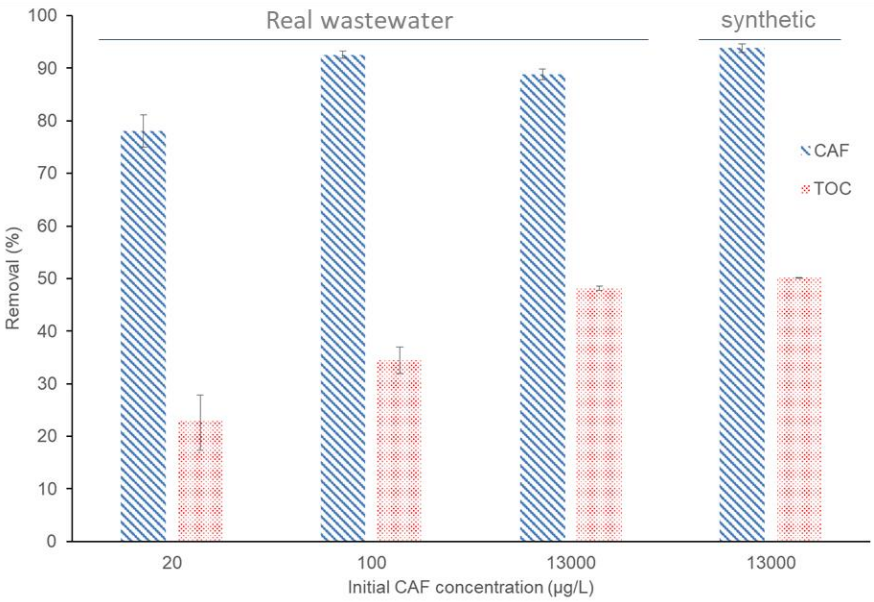


Figure 4.11 Caffeine and TOC removal efficiencies by an EO process under optimal conditions using real municipal wastewater effluent and a synthetic solution

4.3.6 Toxicity evaluation

Toxicity tests were carried out to evaluate the toxicity of a real municipal wastewater effluent before and after electrochemical oxidation treatment for caffeine degradation. Real municipal wastewater treatment effluent with initial caffeine concentration of 20 µg/L was used for the electrochemical oxidation in optimal conditions (current intensity of 0.7 A, electrolysis time of 39 min, and BDD anode). Two samples of wastewater before and after EO treatment for 39 min were evaluated for toxicity with Daphnia magna. Also, two complementary tests were conducted including one with extended electrolysis time of 2 hours and one using a granular activated carbon column after 39 min of electrolysis.

Figure 4.12 shows the result of the toxicity evaluation as mortality percentage of Daphnia magna neonates versus concentration before and after EO. As can be seen, no significant toxicity was measured for the raw effluent (LC50 > 100% v/v). On the other hand, after EO for 39 min (EO-39 min) the sample shows toxicity even at low concentrations (LC50<6.25% v/v). Several studies have suggested that by-product formation during oxidation treatment can increase the toxicity of the samples toward Daphnia magna and Vibrio fischeri (Ganzenko *et al.*, 2020; Ouarda *et al.*, 2018; Starling *et al.*, 2019).

Ganzenko et al. (Ganzenko *et al.*, 2020) reported that Vibrio Fisheri luminescence inhibition increased to 100% after 30 min of treatment of a pharmaceutical mixture (including caffeine) during the electro-Fenton process. Furthermore, luminescence inhibition started to decrease from 4 h of treatment onwards till the end of a 6 h treatment owing to degradation of toxic by-products into harmless by-products. In this regard, a toxicity test with an extended electrolysis time of 2 h (EO-2hrs) was carried out. Toxicity results show that increasing the electrolysis time decreased the toxicity and improved the LC50 from 6.25% v/v to 33% v/v. Still, acute toxicity remained. Residual reactive chemical oxidants such as chlorine produced during electrolysis could also cause the toxicity of the samples (Ouarda *et al.*, 2018). Zeng (Zeng *et al.*, 2012) suggested that residual chlorine concentration higher than 0.32 mg/L in water would significantly affect the behavioural responses of Daphnia magna. Since the electrolytic reactor used in this study was able to generate different oxidants in the presence of Br^-, Cl^-, SO_4^{2-} ions (Figure AI. 6), the increased toxicity could be related to the mentioned issue. A granular activated carbon (GAC) column which is well known for its capability to remove chlorine and its by-products (Gonce *et al.*, 1994; Sorlini *et al.*, 2005; Suidan *et al.*, 1977) was used as post treatment after the EO process. As Figure 4.12 shows, the treated effluent with combination of EO for 39 min and GAC (EO-GAC) has nearly the same behaviour as the raw untreated effluent (LC50 > 100% v/v) which proves that the GAC column had successfully eliminated the toxic compounds generated after 39 min of electrolysis substantially reducing the mortality rate of Daphnia magna. The immobility rate of Daphnia magna neonates versus concentration also showed this (Figure AI. 7).

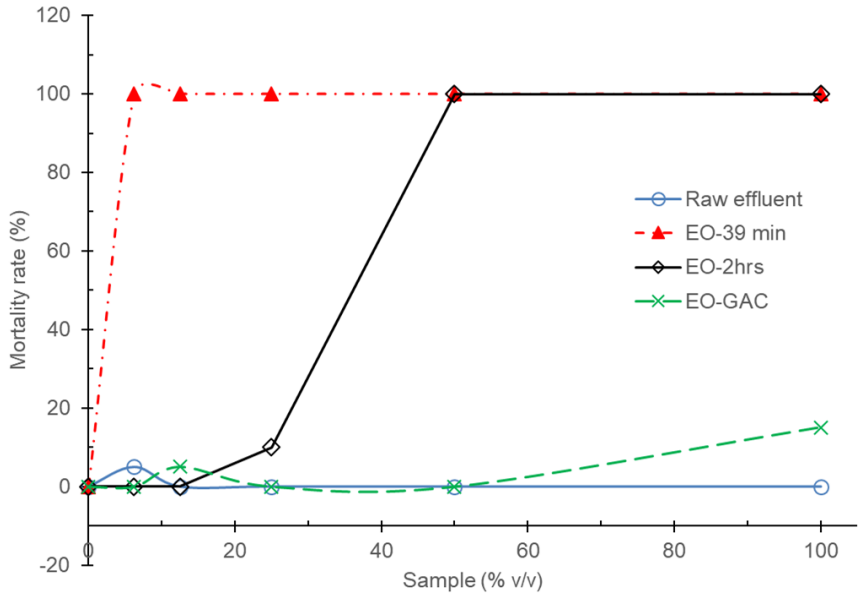


Figure 4.12 Evaluation of mortality rate of Daphnia magna for the EO process

Table 4.9 summarizes the toxicity of the different samples studied. It can be concluded that the EO process under optimal conditions for caffeine degradation in municipal wastewater treatment effluent can lead to generation of toxic compounds which could be removed by using a GAC column. Toxic unit comparison of the different samples shows that the toxicity level of the samples can be ordered as: EO-39 min > EO-2hrs > raw effluent and EO-GAC where EO-GAC treated sample demonstrates no acute toxicity as the raw municipal wastewater effluent.

Table 4.9 Summary of the toxicity assessment results of the samples

	Raw effluent	EO-39 min	EO-2hrs	EO-GAC
LC50-48h (%)	>100	<6.25	33.0	>100
Toxic unit	<1.0	>16	3.0	<1.0
Acute toxicity	No	Yes	Yes	No

4.4. Conclusion

The capability of caffeine removal in synthetic solution and real municipal wastewater effluent by the EO process in a batch electrolytic reactor was investigated using CCD and

ANFIS. The results of CCD and ANFIS modelling showed that both are able to successfully predict the electrochemical process behaviour, ANFIS models performing slightly better (R^2=0.993, RMSE=2.694, MAPE=6.582 for caffeine removal efficiency, and R^2=0.976, RMSE=0.261, MAPE=9.221 for energy consumption). However, CCD models could predict the EO process using much fewer model parameters compared to the ANFIS models which could lower the uncertainties of the model given the low number of data available. Between the selected independent variables (type of anode, electrolysis time, current intensity, and initial concentration of caffeine), type of anode and electrolysis time were the most influential factors on caffeine removal efficiency. While direct oxidation was the dominant mechanism during the EO process for caffeine removal, the oxidation mechanism for caffeine degradation in this study seemed similar to other advanced oxidation processes based on results obtained from the identification of intermediates by LC-MS/MS.

The caffeine removal efficiency in optimal conditions from real municipal wastewater effluent varied from 78.0±4.3% to 92.5±1.0% at different initial caffeine concentrations showing the effectiveness of the process in the presence of other pollutants. However, TOC removal efficiency was found to be lower due to the presence of other organic contaminants than caffeine in the effluent which makes the TOC removal a competitive process in the municipal wastewater effluent.

Toxicity assessment with Daphnia magna showed that the EO process in optimal conditions for caffeine degradation in municipal wastewater treatment effluent can lead to generation of toxic compounds but using a GAC column as post treatment can eliminate the toxicity of the EO treated effluent.

Acknowledgements

Sincere thanks are extended to the National Sciences and Engineering Research Council of Canada and the CREATE TEDGIEER program for their financial contribution to this study. Peter A. Vanrolleghem holds the Canada Research Chair on Water Quality Modelling.

5. A COMPARISON OF ARTIFICIAL INTELLIGENCE MODELS FOR PREDICTING PHOSPHATE REMOVAL EFFICIENCY FROM WASTEWATER USING THE ELECTROCOAGULATION PROCESS

COMPARAISON DE MODÈLES D'INTELLIGENCE ARTIFICIELLE POUR PRÉDIRE L'EFFICACITÉ DE L'ÉLIMINATION DES PHOSPHATES DES EAUX USÉES PAR LE PROCÉDÉ D'ÉLECTROCOAGULATION

Majid Gholami Shirkoohi[a,d], Rajeshwar D. Tyagi[b], Peter A. Vanrolleghem[c,d], Patrick Drogui*[a,d]

[a] Institut National de la Recherche Scientifique (INRS), Centre-Eau Terre Environnement, Université du Québec, 490, Rue de la Couronne, Québec (QC) G1K 9A9, Canada

[b] BOSK Bioproducts, 399 Rue Jacquard, suite 100, Québec (QC) G1N 4J6, Canada

[c] modelEAU, Département de génie civil et de génie des eaux, Université Laval, 1065 av. de la Médecine, Québec, QC G1V 0A6, Canada

[d] CentrEau, Centre de recherche sur l'eau, Université Laval, Québec (QC), Canada

Title of the Journal: Digital Chemical Engineering

Date of submission: 22 April 2022

Abstract

In this study, artificial intelligence (AI) models including adaptive neuro-fuzzy inference systems (ANFIS), artificial neural networks (ANN), and support vector regression (SVR) were applied to predict the removal efficiency of phosphate from wastewaters using the electrocoagulation process. The five input variables used in this study were current intensity, initial phosphate concentration, initial pH, treatment time, and electrode type. The optimal hyperparameters of the ANN and SVR models were found by integrating evolutionary algorithms such as genetic algorithms (GA) and particle swarm optimization (PSO) to these models. To increase the reliability and robustness of the developed AI models, a search for optimal hyperparameters was conducted based on repeated random sub-sampling validation instead of a single split approach. ANFIS models with two clusters were selected based on their generalization capability. The performance of the AI models was compared using statistical analyses based on the ten sub-datasets and the results demonstrated that the effectiveness of the data-driven model depends on how the data is distributed to the training, validation, and test sets. However, hybrid ANN models outperformed other models and PSO-ANN models showed exceptional generalization performance for the different sub-datasets. The average MSE, R^2, and MAPE values of the 10 test subsets for PSO-ANN were determined as 7.201, 0.981, and 2.022, respectively.

Keywords:

Data-driven model, Electrochemical process, Hyperparameters, Metaheuristic algorithm, Modelling, Phosphorus removal

5.1. Introduction

Human activities have significantly increased the soluble forms of phosphorus compounds in aquatic environments over the last 50 years, owing to the widespread use of detergents and chemical fertilizers, animal manure, wastewater effluents, and plant residues. Environmental concerns such as eutrophication can be caused by soluble forms of phosphorus in water, compromising the quality and sustainability of water bodies. This phenomenon can deplete oxygen levels in water due to algae breakdown, which can harm fish and other aquatic life, resulting in decreased biodiversity (Ano *et al.*, 2019; Li *et al.*, 2022; Tran *et al.*, 2012). Chemical precipitation (Lavanya *et al.*, 2021), adsorption (Gizaw *et al.*, 2021), ion exchange (Bektaş *et al.*, 2021), and biological processes (Zhang *et al.*, 2022) have all been developed to lower phosphate levels in wastewater before it is released into the environment. An alternative method for removing phosphates is electrochemical treatment, such as electrocoagulation (EC) (Kobya *et al.*, 2021). The main advantages of the (EC) process are the ease of use of the equipment, the ease of automation, and the process efficiency in the treatment of a wide range of pollutants. Furthermore, because this process does not necessitate the use of chemicals, treatment costs are reduced. Metal cations are released in situ by electrodissolution of an Al or Fe anode immersed in the effluent, as opposed to chemical precipitation (Jing *et al.*, 2021).

Modelling and optimization of the electrochemical process are seen as a key part of the study in order to examine the efficacy of the process. The concentration of pollutants, the applied current density and electrical potential, the types of electrodes, the electrolyte type and concentration, and chemical interactions between contaminants are all important aspects in electrochemical processes for water and wastewater treatment (Drogui *et al.*, 2007). Given the diversity of aspects, modelling, simulation, and optimization of these processes are challenging. Artificial intelligence methods such as artificial neural networks (ANN), adaptive neuro-fuzzy inference systems (ANFIS), support vector regression (SVR), and evolutionary algorithms such as genetic algorithms (GA) and particle swarm optimization (PSO) have emerged as appealing approaches for modelling and optimizing these nonlinear processes. These data-driven models are based on empirical data and linkages between process input and output variables rather than

process knowledge. Some recent studies have shown the application of AI models in electrochemical processes for water and wastewater treatment with reasonable accuracy (Farzin *et al.*, 2020; Gholami Shirkoohi *et al.*, 2021; Taheri *et al.*, 2013).

AI models include built-in hyperparameters that should be fine-tuned so that the model can solve the machine learning problem to its full potential. The learning process is controlled by these hyperparameters, which directly impact the model performance. In the case of ANN models, these hyperparameters include training algorithms, number of epochs, maximum validation failure, number of hidden layers and hidden neurons, and transfer functions (Viana *et al.*, 2018). For SVR models, the penalty factor, margin of tolerance for errors, and the type of kernel function and the kernel parameters should be optimally selected (Rui *et al.*, 2019). Membership functions and the number of clusters affect the performance of ANFIS models (Abdulshahed *et al.*, 2015). Grid search and random search are often used to search for the optimal values within the space of hyperparameters of the AI models. In the case of a large space to investigate, grid search would be too computationally intensive and slow, and with random search, there is a chance not to be able to find the optimal hyperparameters (Menapace *et al.*, 2021). Recently, metaheuristic algorithms such as genetic algorithm (Gu *et al.*, 2011), particle swarm optimization (Huang *et al.*, 2021), firefly algorithm (FA) (Zhang *et al.*, 2019), ant colony optimization algorithm (ACO) (Jiang *et al.*, 2020), and bat algorithm (BA) (Hafezi *et al.*, 2015) have been used to efficiently tune and optimize the AI models' parameters. For instance, Viana *et al.* (2018) used the PSO algorithm to optimize neural network model hyperparameters, including the hidden neuron number, the transfer function, and the learning rate.

With the significant investment of time and money in experimental work, only a limited number of samples is available in datasets for data-driven models. As the use of data-driven models in the field of electrochemical processes for water and wastewater treatment expands, model reliability and robustness become increasingly important. Apart from hyperparameters, the allocation of the data to training and testing sets influences the performance of the AI model generated with relatively small sample sizes. The use of single split training and test sets methodologies is a standard research

technique, but it is problematic given the limited sample sizes of experimental datasets accessible from electrochemical processes (Singh *et al.*, 2021). However, the reliability of AI models used in electrochemical processes for water and wastewater treatment in the context of the mentioned issue has not been considered thoroughly in the literature.

In this paper, different AI models including ANFIS, ANN, and SVR are developed to predict the removal efficiency of phosphate from wastewaters using the electrocoagulation process. To optimize the hyperparameters of the SVR and ANN models, GA and PSO have been integrated as the proposed approach. To increase the robustness of the AI models with the optimal hyperparameters with respect to the division of the data between training and testing sets, repeated random sub-sampling validation has been utilized for the hybrid models. In order to illustrate the predicting performance of the proposed models, results were compared based on statistical indices.

5.2. Development of the AI models

5.2.1 Data acquisition

To develop the AI models, a total number of 62 experimental data for the removal of phosphate from synthetic wastewaters using an electrocoagulation process was gathered from Ano *et al.* (2019). In their study, factorial design (FD) and central composite design (CCD) were used as response surface methodology (RSM) to investigate the effect of current intensity, initial phosphate concentration, initial pH, treatment time, and electrode type. Table 5.1 shows the description and statistical parameters of the dataset used in this study.

Table 5.1 Description of the dataset from Ano *et al.* (2019) used in this study

Statistical parameters	Independent variables/Inputs					Dependent variable/Output
	Current Intensity (A)	Initial phosphate concentration (mg/L)	pH	Treatment time (min)	Electrode type	Removal efficiency (%)
Number of samples	62	62	62	62	62	62
Range	0.25-1.25	15-75	2-10	10-90	Al/Fe	29.2-100
Average	0.75	45	6	50	-	74.0
Standard deviation	0.22	13.3	1.77	17.7	-	20.6

5.2.2 Adaptive neuro fuzzy inference system

ANFIS, introduced by Jang (Jang, 1993) is a hybrid technique of artificial intelligence in which a Sugeno-type Fuzzy Inference System (FIS) and an artificial neural network (ANN) are combined. Fuzzy Logic produces fuzzy rules that map the inputs to an output based on a given input-output data set. Using the ANFIS hybrid approach, an initial fuzzy model based on fuzzy logic from the input-output data of the system is derived. Then, the neural network learning process is utilized to fine-tune the rules of the initial fuzzy model to generate the optimal ANFIS model of the system. Therefore, ANFIS benefits from the advantages of Fuzzy Logic and neural networks in a single integrated system (Buragohain, 2009).

The FIS structure can be generated by different strategies, including grid partitioning, Fuzzy C-mean clustering (FCM) and the subtractive clustering method (SCM). FCM integrated with ANFIS helps obtaining a relatively small number of rules which prevents the model from being too complex and reduces the risk of overfitting. The FCM clustering method partitions the input data into different clusters and is used to identify the fuzzy membership functions and fuzzy rule base for the ANFIS model (Melin *et al.*, 2014). In this study, FCM has been used for the ANFIS model and the number of clusters will be manually selected for the best generalization performance.

5.2.3 Support vector regression

Support vector regression (SVR) is an extension of support vector machines (SVM), first presented by Vapnik (1963), used for prediction and regression problems. Due to its promising generalization performance, SVR has been widely applied to regression prediction problems (Saradhi *et al.*, 2007). In SVR, the main goal is to obtain a predictor function $f(x)$ that describes the relationship between input and output data with an error value less than ε for all the training data. At the same time, the function $f(x)$ is required to be as flat as possible, meaning that the errors are not significant as long as they are less than ε, but any deviation larger than this amount is not tolerated. This function can be written as:

$$f(x) = w.\phi(x) + b \tag{5.1}$$

where *w* is a weight vector, $\phi(x)$ is a mapping function in the feature space, and *b* is a bias. The coefficients w and *b* are determined by minimizing the following optimization problem:

$$Min \quad \frac{1}{2}\|w\|^2$$
$$subject\ to \begin{cases} y_i - w.\phi(x_i) - b \leq \varepsilon \\ w.\phi(x_i) + b - y_i \leq \varepsilon \end{cases} \tag{5.2}$$

In real conditions, to permit some deviations for training data outside the ε-insensitive zone, slack variables ξ, ξ^* are introduced to the problem. Hence, the optimization problem can be rewritten as:

$$Min \quad R(w, \xi, \xi^*) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$
$$subject\ to \begin{cases} y_i - w.\phi(x_i) - b \leq \varepsilon + \xi_i \\ w.\phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1,2, \dots, n \end{cases} \tag{5.3}$$

where R is the regression risk, and the box constraint or penalty parameter C is a positive value that determines the trade-off between the training error and generalization ability. To solve this quadratic optimization problem, the dual Lagrangian is used to obtain the regression function as follows:

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + b \tag{5.4}$$

where α_i and α_i^* are the non-zero Lagrange coefficients, and $K(x_i, x)$ is the kernel function that transforms the nonlinear inputs into a higher-dimensional feature space. There are several kernel functions used in SVR such as Linear, Polynomial, and Gaussian or Radial Basis Function (RBF):

Linear: $K(x_i, x_j) = x_i x_j$

Polynomial: $K(x_i, x_j) = (1 + x_i x_j)^q$

Gaussian or RBF: $K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right)$

where q is the polynomial order and σ is the RBF kernel parameter. While all these kernel functions can be applied, the selection of suitable kernel function and its parameters is essential for the SVR performance. In this study, all three kernel functions were evaluated for the best SVR performance.

SVR parameters optimization

The performance of the SVR model highly depends on the accurate selection of its hyperparameters. These include the box constraint (C), the acceptable error epsilon (ε), the type of kernel function, and the kernel parameter. The box constraint C is a trade-off between model complexity and generalization ability. The magnitude of the penalty for samples with losses greater than ε is determined by the C-value. With too small C-values, the model will be simpler (less complex), but the training error will increase. On the other hand, when C is too large, the empirical risk (the second term in the regression risk R) will be the dominant term for the minimization objective, which results in the overfitting issue (Rui *et al.*, 2019). The ε hyperparameter influences the number of support vectors and, hence, the performance of the SVR by determining the size of the ε-insensitive zone. The kernel function and its relevant parameter maps nonlinear input data into the higher dimensional feature space to help SVR handle nonlinear problems (Alade *et al.*, 2019).

Therefore, the three hyperparameters of C, ε, and kernel parameter (σ for RBF and q for polynomial kernel function) should be selected carefully in view of the effectiveness of the SVR model. In this study, PSO and GA are applied to find the optimal values of these hyperparameters. It should be noted that in the case of a linear kernel function, only C and ε will be optimized by the optimization algorithms.

5.2.4 Artificial neural networks

An ANN imitates the essential characteristics of the human brain (which itself is a highly nonlinear, complex, and parallel computer), such as self-adaptability, self-organization,

and error tolerance (Haykin, 1998). A massive parallel network composed of nonlinear computational elements (neurons) helps ANNs to explore many competing hypotheses simultaneously. Considering how the different neurons are positioned and connected to each other as well as the composition of layers, various ANN architectures can be generated. The multilayer perceptrons (MLP) feedforward network is one of the most common ANN architectures. These networks are applied to diverse problems, including function approximation, pattern classification, system identification, process control, process optimization, and so on (Gholami Shirkoohi *et al.*, 2021). The weights of the connections between the neurons are adjusted in the training process. Multiple optimization algorithms can be used for the training process, such as gradient descent, Levenberg–Marquardt, and Bayesian Regularization backpropagation.

ANN parameters optimization

The hyperparameters of an ANN model, which define its topology and learning options, influence the accuracy and effectiveness of the trained model. The numbers of hidden layers and neurons in each hidden layer, training algorithm, transfer functions, the regularization parameter, the learning algorithm, and the maximum validation failure are considered as ANN hyperparameters (Shirkoohi *et al.*, 2021; Sinha *et al.*, 2021; Valencia *et al.*, 2021; Viana *et al.*, 2018). In this study, the selection of the number of hidden neurons, the training algorithm, the type of transfer function in the hidden layer, and the number of maximum validation failures are considered for tuning with PSO and GA. As training algorithms Gradient descent with momentum and adaptive learning rate backpropagation (traingdx), Levenberg-Marquardt backpropagation (trainlm), and Bayesian Regularization backpropagation (trainbr) have been selected for the optimization process. Also, transfer functions including Log-sigmoid (logsig), hyperbolic tangent sigmoid (tansig), and positive linear (poslin) were chosen as options.

5.2.5 Genetic algorithm

Genetic Algorithms, first proposed by Holland (Holland, 1992) are evolutionary search and optimization algorithms based on natural selection. GAs, thanks with good global searching ability, flexibility, no need for gradient information of the objective (fitness)

functions, and ease of operation, have been a powerful technique for optimization problems (Curteanu *et al.*, 2007; Ding *et al.*, 2011b). GAs start with a primary population of candidate solutions and a fitness value is calculated for each solution. The algorithm applies three stochastic operators to each population, which are analogous to chromosomes in a biological connotation. Selection is choosing the solutions with the highest fitness value to create an intermediate population. The next population is the result of crossover or mutation. By crossover, the selected members are mated in pairs and recombined through genetic manipulation of chromosomes to generate two new solutions (offsprings). Mutation acts as an assurance against lost genetic material and consists of replacing some of the chromosome's genes with new genes. Generation of new populations and calculation of fitness value for each population repeat over and over in an iterative method. When a specific termination criterion is met, e.g., when there is no change in the population from one iteration to the next or when a satisfactory fitness value is identified, this process ends (Whitley, 1994).

5.2.6 Particle swarm optimization

PSO, first introduced by Kennedy *et al.* (1995), is based on the social behaviour simulation of a flock of birds, called 'swarm', searching for food. PSO is a stochastic population-based optimization approach in which particles, a swarm of potential solutions, fly in the problem space to find better regions and finally the optimal solution, while cooperating and competing with other ones (Chen *et al.*, 2010). In PSO, a particle is analogous to a chromosome (population member) in GA and represents a candidate solution to the problem being studied (Eberhart *et al.*, 1998). Each particle's condition is changed by the impact of three factors: (1) its own inertia; (2) its most optimal position; (3) the swarm's most optimal position. In the d-dimensional search space of the problem, particle *i* of the swarm can be represented by $X_i = (x_{i1}, x_{i2}, \dots, x_{id})$. The velocity of this particle and the best previous position, which is the position giving the best fitness value, are represented as $V_i = (v_{i1}, v_{i2}, \dots, v_{id})$, $P_i = (p_{i1}, p_{i2}, \dots, p_{id})$. Also, the global best position, the position of the best individual, is noted as $G = (g_1, g_2, \dots, g_d)$. The velocity and position of the particles are updated as follows:

$$V_i^{j+1} = \omega * V_i^j + c_1 * rand_1 * \left(P_i^j - X_i^j\right) + c_2 * rand_2 * \left(G_i^j - X_i^j\right) \quad (5.5)$$

$$X_i^{j+1} = X_i^j + V_i^{j+1} \quad (5.6)$$

where V_i^{j+1} and X_i^{j+1} are the updated velocity and position vector of particle *i*, ω is the momentum or inertia weight factor, c_1 and c_2 are the acceleration coefficients, and $rand_1$ and $rand_2$ are random numbers between (0,1) (Juneja *et al.*, 2016).

In general, in population-based optimization approaches, significant diversity is required during the early phase of the search to cover the whole search space. However, during the latter phase of the search, when the algorithm is converging to the optimal solution, fine-tuning of the solutions is required to effectively locate the global optimum. Thus, time-varying inertia weights and acceleration factors have been introduced and widely used with the PSO algorithm as follows:

$$\omega = \left(\frac{It_{max} - It}{It_{max}}\right) * \left(\omega_i - \omega_f\right) + \omega_f \quad (5.7)$$

$$c_1 = \left(\frac{It}{It_{max}}\right) * \left(c_{1f} - c_{1i}\right) + c_{1i} \quad (5.8)$$

$$c_2 = \left(\frac{It}{It_{max}}\right) * \left(c_{2f} - c_{2i}\right) + c_{2i} \quad (5.9)$$

where ω_i and ω_f are the initial and final values of the inertia weights, c_{1i}, c_{1f}, c_{2i}, and c_{2f} are constants, It is the current iteration number and It_{max} is the maximum number of iterations. The best reported results have been achieved in literature when $\omega_i = 0.9, \omega_f = 0.4, c_{1i} = c_{2f} = 2.5$, and $c_{1f} = c_{2i} = 0.5$ (Ratnaweera *et al.*, 2004; Shi *et al.*, 1999). These settings enable particles to travel across the whole search space rather than gravitating toward the population's best in the early stages. The time-varying parameters allow the particles to converge into the global optimum in the latter stages.

As mentioned before, GA and PSO will be used in this study to search for optimal hyperparameters of the developed SVR and ANN models. Figure 5.1 shows the flowchart of the integrated GA and PSO to find the optimal hyperparameters of the models.

Repeated random sub-sampling validation or Monte Carlo cross-validation (Picard *et al.*, 1984) is used in this study as applied before in literature (Altaf *et al.*, 2016; Cao *et al.*, 2020; Severeyn *et al.*, 2019). To have an equal distribution of data for the different AI models, at first, 10 sub-datasets are generated randomly, each containing 62 data points. From these 10 sub-datasets, 42 data points will be selected for training, 10 data points for validation, and 10 data points for testing. The search for the optimal hyperparameters will be conducted using all 10 sub-datasets. This helps to find hyperparameters that result in the best performance available for the 10 sub-datasets and to overcome the uncertainty related to the use of single split training, validation and test sets method with limited data points available.

To evaluate the performance of the AI models, the mean square error (MSE), the correlation coefficient (R^2), and the mean absolute percentage error (MAPE) were used as comparison criteria. To train and validate the hybrid models, MSE was used as the error function. These functions were calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5.10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5.11}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{5.12}$$

where y_i is the experimental value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the average value of the experimental data.

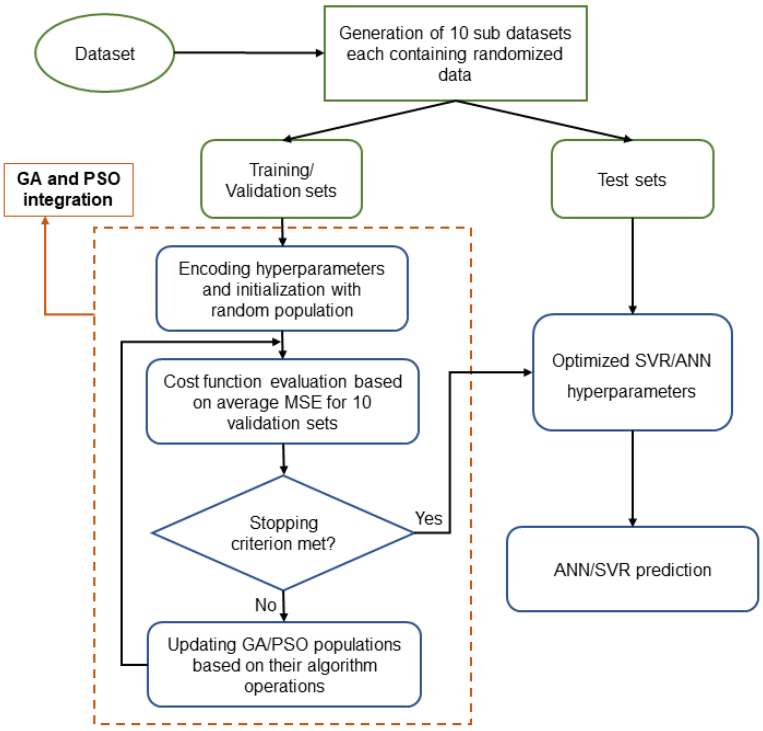


Figure 5.1 Flowchart of the proposed hybrid models

5.3. Results

5.3.1 ANFIS model

The number of clusters in FCM-FIS generation affects the performance of the ANFIS model. Table 5.2 shows the impact of increasing the number of clusters on the performance of trained ANFIS models in training, validation, and test. As can be seen, setting a higher number of clusters for an ANFIS model results in having better performance on the training set by making the model more complex. However, this causes the model to lose generalization capability and to risk overfitting. This is evident from the performance of the models on the validation and test sets when higher number of clusters are adopted. Although both models with two and three clusters perform well on the validation and test stage, the model with two clusters would have a lower total number of parameters than a model with three clusters (38 parameters compared to 57) and hence leads to a simpler model. Therefore, ANFIS models with two clusters were selected.

Table 5.2 Error analysis as a function of the number of clusters in ANFIS

No. of clusters	MSE			R^2			MAPE		
	Train	Validation	Test	Train	Validation	Test	Train	Validation	Test
2	**31.642**	**68.733**	**60.763**	**0.924**	**0.794**	**0.835**	**6.550**	**9.322**	**9.275**
3	17.647	58.422	96.994	0.958	0.832	0.757	5.066	8.176	10.823
4	16.695	103.391	120.240	0.961	0.709	0.676	4.506	11.294	12.706
5	8.306	151.575	179.181	0.981	0.557	0.504	2.690	12.961	14.650
6	1.558	932.835	1030.522	0.996	-0.986	-2.050	0.757	19.982	22.588

5.3.2 PSO-SVR and GA-SVR

As mentioned before, the selection of kernel function and its inherent parameter influence the performance of the SVR model. Therefore, it was necessary to select the kernel function of the SVR model before fine-tuning the hyperparameters. For this purpose, the three kernel functions (linear, polynomial, and RBF) were tested on the data using the 5-fold cross-validation method. This method applies the training process on 4 folds of observations and (each time) leaves one fold of observations out to calculate the generalization error of the models. Table 5.3 shows the results for the different kernel functions used. The generalization error calculated is the out-of-sample MSE. It should be mentioned that the MSE obtained in Table 5.3 is based on the default values of the hyperparameters and the optimal hyperparameters of the SVR models with the three different kernel functions on the validation sets. As can be seen, the polynomial function leads to the best results on the 5-fold cross-validation method with the default hyperparameters. The kernel parameter, C, and ε should be optimally selected by the hyperparameter optimization algorithm. The three kernel functions were also tested for hyperparameter optimization by PSO-SVR and GA-SVR on the 10 data subsets. It can be seen that optimal SVR models with polynomial kernel function obtain the lowest MSE on the validation sets of PSO and GA algorithms with a population size of 50.

Table 5.3 Effect of the kernel function on the SVR performance

Kernel functions	MSE	
	Default hyperparameters	PSO and GA optimal hyperparameters
Linear function	119.96	72.89
Polynomial function	105.48	65.43
RBF	361.61	66.09

5.3.3 GA-ANN and PSO-ANN

GA and PSO were used to find the optimal hyperparameters of the ANN models. A population size of 50 has been considered for the algorithms. Although population size is problem-dependent and no universal value can be proposed, there are studies suggesting that population sizes between 20 and 50 would be appropriate for solving optimization problems (Lobo *et al.*, 2007; Poli *et al.*, 2007; Wang *et al.*, 2018; Zhang *et al.*, 2005). Other control parameters of the GA and PSO are the same as suggested in the literature as described in section 5.2.5. The number of maximum iterations was set to 200 and 100 for GA and PSO, respectively, to have an equal number of function calls for both algorithms, allowing for an unbiased comparison. Indeed, an equal number of iterations for each test, and not an equal number of function calls, may result in a better performance of the optimization algorithm that is attributed to a larger number of function calls (Piotrowski *et al.*, 2020).

Table 5.4 represents the optimal hyperparameters of the ANN models found by GA and PSO. The prediction accuracy of the hybrid models is given as the average MSE based on the 10 validation sets. It can be seen that both GA and PSO perform reasonably on the validation subsets. The best performances, in terms of average MSE over 10 validation sets, are calculated as 7.686 and 7.830 for GA and PSO, respectively. However, their performance will be evaluated in the next section for the test subsets as well. The Levenberg-Marquardt training algorithm and log-sigmoid transfer function were found to be optimal for both GA and PSO techniques. The Levenberg-Marquardt method has shown its good performance including its ability to converge 10–100 times faster than

the conventional gradient descent backpropagation algorithm (Da Silva *et al.*, 2017; Kamosi *et al.*, 2010).

Table 5.4 Optimal hyperparameters and prediction accuracy results for the hybrid GA and PSO models

Technique	Hyperparameters				MSE
	No. of hidden neurons	Training algorithm	Transfer function	No. of maximum validation failures	
GA-ANN	7	trainlm	logsig	5	7.686
PSO-ANN	9	trainlm	logsig	7	7.830

5.3.4 Performance comparison of the proposed models

The objective of the proposed AI models including ANFIS, PSO-SVR, GA-SVR, PSO-ANN, GA-ANN was to predict the removal efficiency of phosphate from wastewaters using the electrocoagulation process. The performance of each data-driven model was evaluated by its correspondence with experimental data on training, validation, and test sets. According to the mentioned analyses, it can be concluded that the proposed hybrid ANN models have great performance in the prediction of the removal efficiency of phosphate using the electrocoagulation process. The results confirmed that PSO-ANN models have exceptional generalization performance for the different data subsets.

Table 5.5 represents the statistical analyses of the proposed models with the optimal parameters. Results are provided for training, validation, and test sets as the average of the 10 sub-datasets.

The results in According to the mentioned analyses, it can be concluded that the proposed hybrid ANN models have great performance in the prediction of the removal efficiency of phosphate using the electrocoagulation process. The results confirmed that PSO-ANN models have exceptional generalization performance for the different data subsets.

Table 5.5 show that hybrid ANN models perform better than ANFIS and hybrid SVR models in all training, validation, and test stages. Both PSO-SVR and GA-SVR models showed similar performance results on the datasets. This comes from the fact that both evolutionary algorithms found the same SVR parameters as the optimal parameters. The

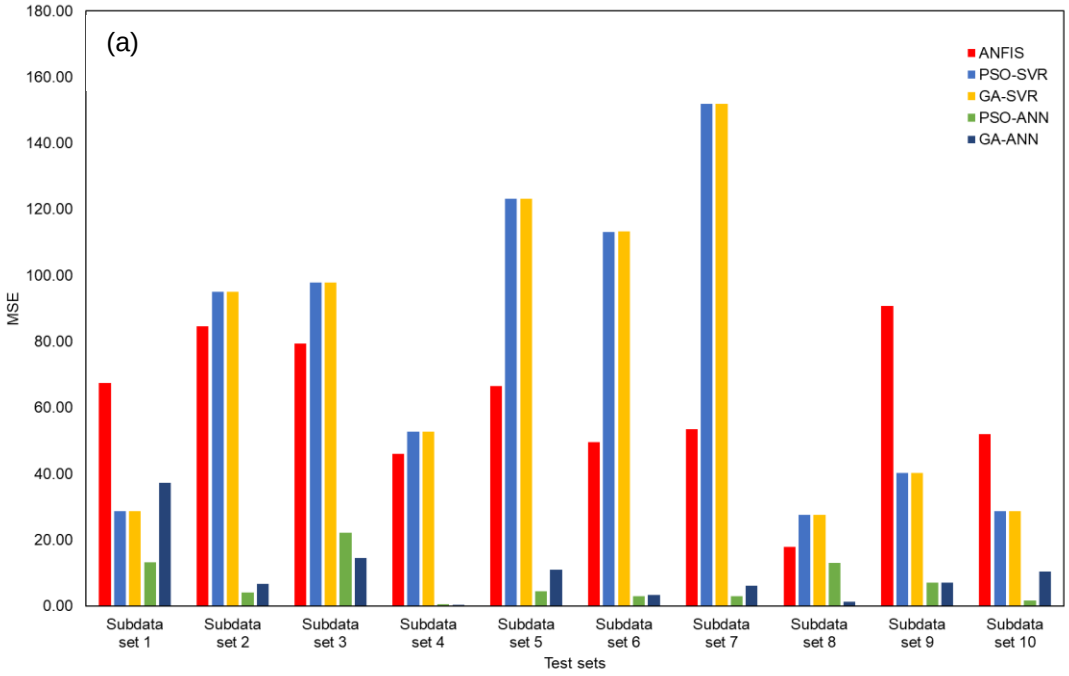
optimal parameters (C, q, ε) of PSO-SVR and GA-SVR were identified as (1, 3, 3.171). While the ANFIS models performed better in terms of generalization for test sets, it was outperformed by PSO-SVR and GA-SVR in the training and validation steps. The best performance both in training and generalization was obtained by the PSO-ANN models, while the GA-ANN models also showed their effectiveness. The average MSE, R^2, and MAPE values of the 10 sub-datasets for PSO-ANN are determined as 7.201, 0.981, and 2.022 for the test sets, respectively.

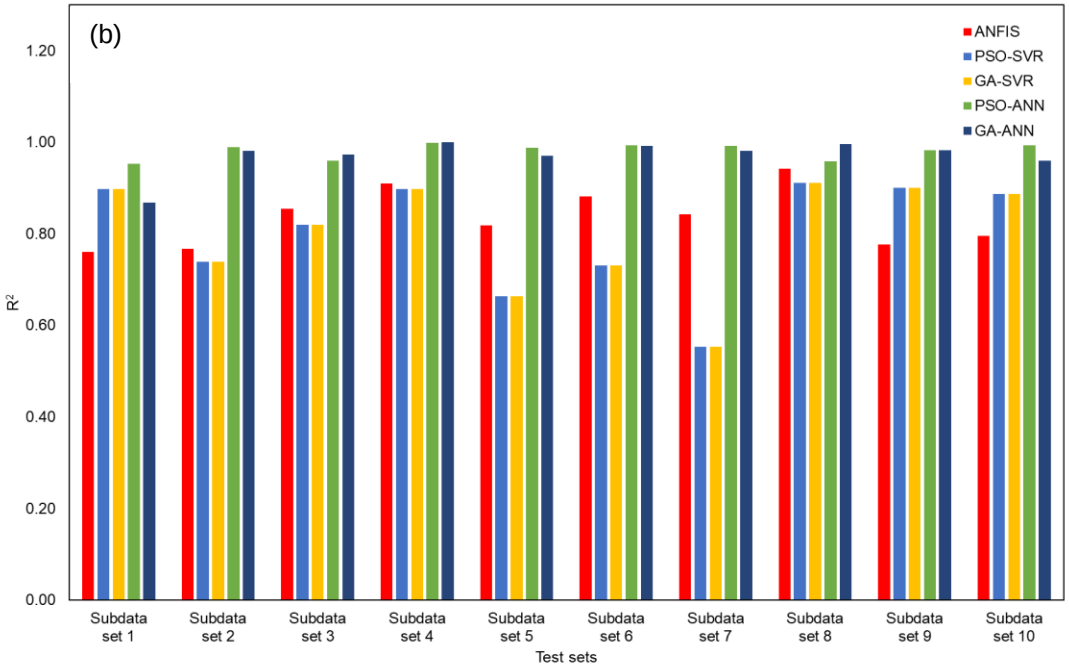
Figure 5.2 shows the performance comparison of the AI models on the test sets of each of the 10 data subsets. As can be observed, the effectiveness of the data-driven model depends on how the data is distributed over the training, validation, and test sets. For instance, the hybrid SVR models have low MSE, MAPE, and high R^2 for sub-dataset 1. Their performance is comparable to the hybrid ANN models. However, the hybrid ANN models outperform ANFIS and hybrid SVR models for the test sets of other data subsets. Figure 5.3 presents the box plots of the performance indices of the AI models for the test sets of 10 data subsets. It shows that there is a greater variability for the hybrid SVR models than for the hybrid ANN models. It can be concluded that generally, hybrid SVR models are more sensitive than hybrid ANN models to the distribution of the data points among the training, validation, and test sets. Nevertheless, hybrid ANN models show less dispersed performance for the test sets of the different data subsets.

According to the mentioned analyses, it can be concluded that the proposed hybrid ANN models have great performance in the prediction of the removal efficiency of phosphate using the electrocoagulation process. The results confirmed that PSO-ANN models have exceptional generalization performance for the different data subsets.

Table 5.5 Performance evaluation of the developed AI models

Model	MSE			R^2			MAPE		
	Train	Validation	Test	Train	Validation	Test	Train	Validation	Test
ANFIS	31.642	68.732	60.763	0.924	0.793	0.835	6.550	9.322	9.275
PSO-SVR	9.374	65.433	75.893	0.978	0.808	0.800	4.377	8.998	10.779
GA-SVR	9.374	65.433	75.894	0.978	0.808	0.800	4.377	8.998	10.779
PSO-ANN	7.259	7.830	7.201	0.983	0.978	0.981	1.958	2.286	2.022
GA-ANN	8.765	7.686	9.759	0.979	0.978	0.970	2.520	2.747	2.774


(a)
ANFIS
PSO-SVR
GA-SVR
PSO-ANN
GA-ANN
MSE
180.00
160.00
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
Subdata set 1
Subdata set 2
Subdata set 3
Subdata set 4
Subdata set 5
Subdata set 6
Subdata set 7
Subdata set 8
Subdata set 9
Subdata set 10
Test sets


(b)
ANFIS
PSO-SVR
GA-SVR
PSO-ANN
GA-ANN
R²
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Subdata set 1
Subdata set 2
Subdata set 3
Subdata set 4
Subdata set 5
Subdata set 6
Subdata set 7
Subdata set 8
Subdata set 9
Subdata set 10
Test sets

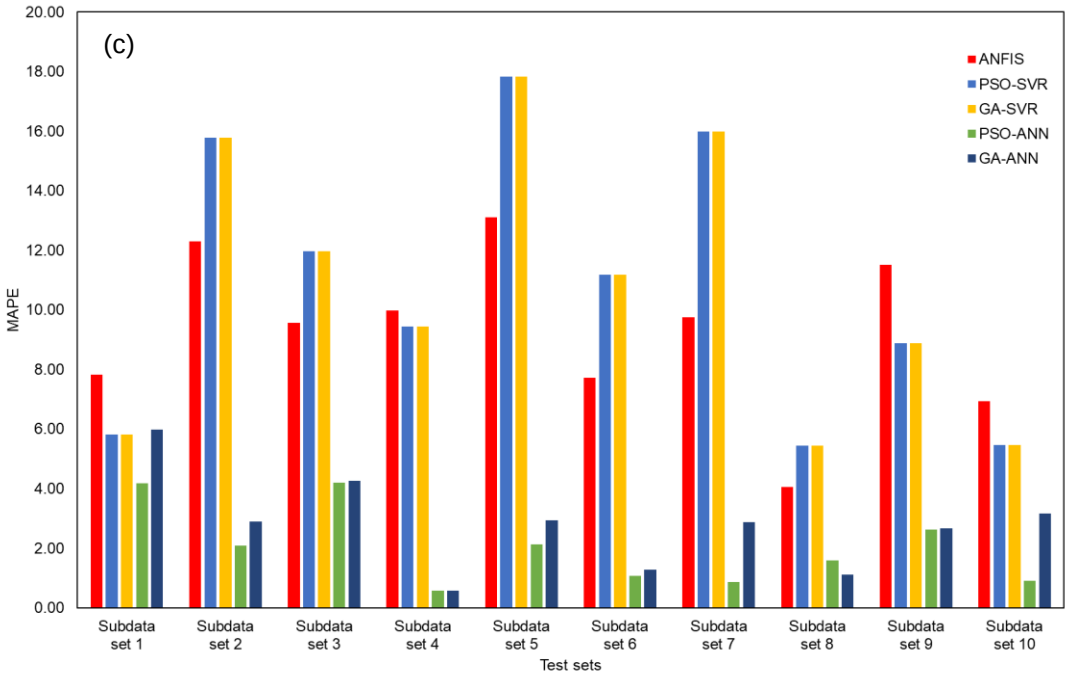


Figure 5.2 Performance evaluation of the developed AI models on the test sets of the 10 subsets; (a) MSE, (b) R^2, (c) MAPE

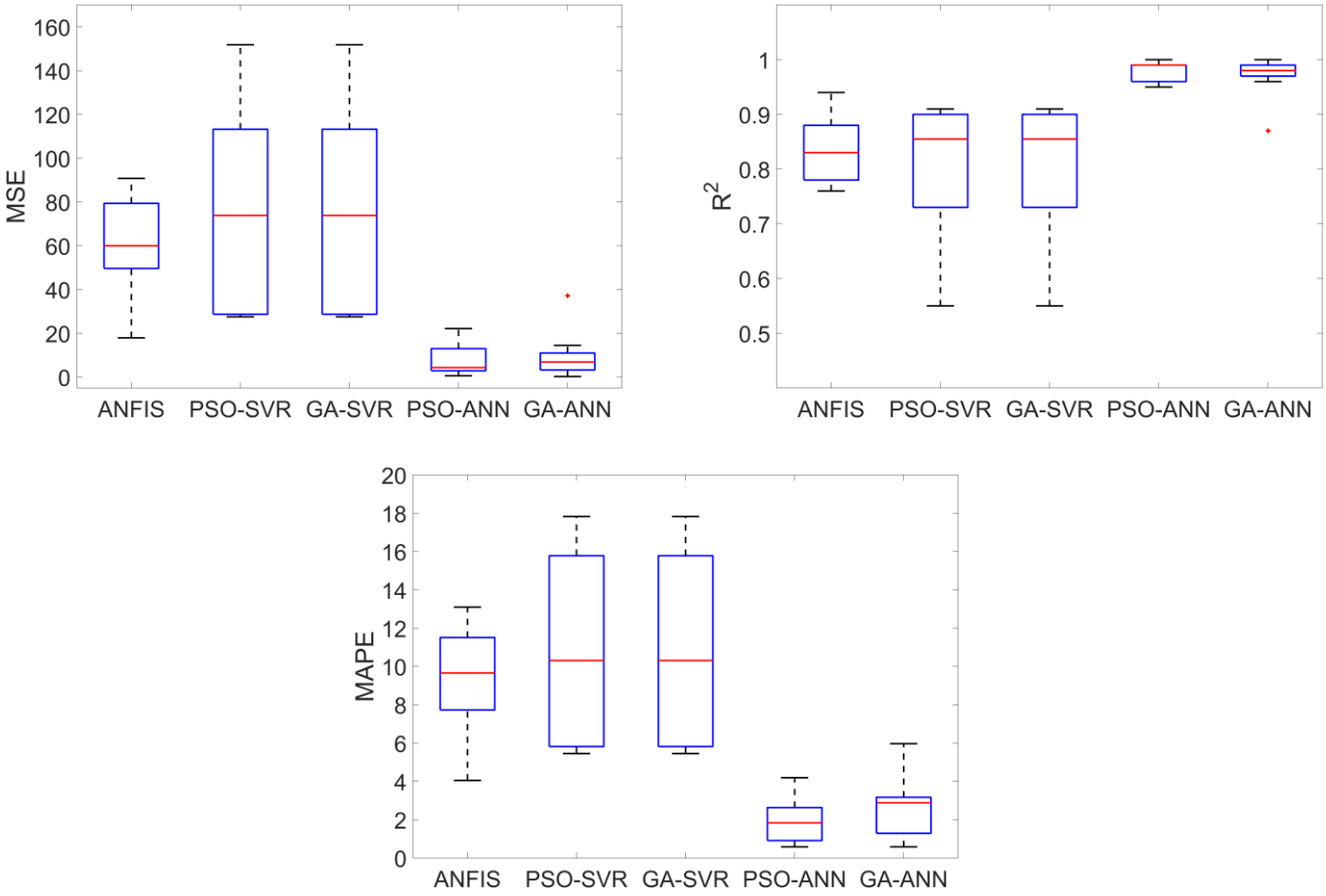


Figure 5.3 Boxplot comparison of the performance of the different AI models for the test sets using three performance criteria

5.4. Conclusion

In this study, GA and PSO algorithms were used to optimize the hyperparameters of the SVR and ANN models to forecast the removal efficiency of phosphate from wastewaters using the electrocoagulation process. The current intensity, initial phosphate concentration, initial pH, treatment time, and electrode type were considered as models' inputs. To tackle the relatively low number of sample data available from the experimental electrochemical process and increase the reliability of data-driven models, the proposed hybrid models were built on repeated random sub-sampling validation (10 data subsets) instead of a single split approach. The performance criteria (MSE, R^2, MAPE) comparison of models showed that the effectiveness of the data-driven models depends on how the

data is distributed to the training, validation, and test sets. The ANFIS and hybrid SVR models were more sensitive than hybrid ANN models to the distribution of data points. The hybrid ANN models showed greater accuracy than the ANFIS and hybrid SVR models that they were compared to using different performance criteria and indicated less dispersed performance for the test sets of the different sub-datasets. Remarkably, PSO-ANN models illustrated exceptional generalization performance for the 10 data subsets examined. Further research in the application of hybrid evolutionary algorithms and AI models may be carried out in electrochemical processes for water and wastewater treatment with respect to the reliability and robustness of the models.

Acknowledgements

Sincere thanks are extended to the National Sciences and Engineering Research Council of Canada and the CREATE TEDGIEER program for their financial contribution to this study. Peter A. Vanrolleghem holds the Canada Research Chair on Water Quality Modelling.

6. CONCLUSIONS AND PERSPECTIVES

In this work, different types of artificial intelligence techniques including ANNs, SVR, ANFIS, GA, and PSO were applied on different electrochemical processes for wastewater treatment. The first part of the thesis focused on developing an ANN-GA methodology for production of active chlorine by electrolysis for disinfection. The second part of the thesis involved the application of the ANFIS modelling approach and a comparison with the RSM approach for modelling and optimization of electrochemical oxidation of caffeine as an emerging organic pollutant. And the final part focused on the removal of phosphates as an inorganic pollutant from wastewater. The third part of the research focused on comparing different artificial intelligence models to predict the phosphate (as an inorganic pollutant) removal efficiency from wastewater using the electrocoagulation process.

Efforts were made in this study to obtain not only models with good prediction performance but also with high reliability. This was done in the first part of study by using learning curves and the regularization technique. These techniques were shown to give an insight into the ANN modelling process and can be used instead of or along with a trial and error procedure during the training of neural networks. This results in obtaining reliable ANN models and prevents from overfitting even with a limited number of data available. The ANN models were successful in describing the behaviour of the experimental process as well as the conventional RSM modelling approach, showing their capability to be adapted on an electrolysis process while being well-trained.

In the case of ANFIS, the initial FIS was generated based on the FCM clustering method rather than other alternatives such as grid partitioning. FCM clustering helps to have fewer rules and thus fewer parameters for the developed model than grid partitioning would lead to and minimizes the overfitting issue. Even though the results obtained by grid partitioning could be more satisfying, the uncertainty on model outputs increases with more complex models developed with small datasets

Also, the Pareto front derived by the NSGA-II algorithm for multi-objective optimization allowed identifying non-dominated optimal points (operating conditions) for maximum active chlorine production at minimum energy consumption. The proposed ANN-GA

methodology in the first part of the study can give insight into how to efficiently choose the process operation parameters (decision variables) for the desired objectives. This approach can be adapted to other processes if the experimental data already exist.

The hyperparameters of the AI models affect the overall performance of the developed models. Searching for the optimal hyperparameters of AI models utilizing metaheuristic algorithms were studied in the third part of this work. This optimization was done considering both reliability and robustness of the AI models developed for electrochemical processes for water treatment in case only a limited amount of data is available. In this regard, the proposed hybrid models were built on repeated random sub-sampling validation (10 data subsets) instead of a single split approach. The results demonstrated that the performance of the AI models depends on how the data is distributed over training, validation, and test sets. The ANFIS and hybrid SVR models were more vulnerable to the particular distribution of sample points of data than the hybrid ANN models. The hybrid ANN models outperformed the ANFIS and hybrid SVR models and showed less scattered performance for the prediction of the phosphate removal efficiencies of different test subsets. It was found that PSO-ANN models have exceptional generalization performance for the 10 data subsets examined. This part of the work provides a method for finding optimal data-driven models of electrochemical processes for wastewater treatment where a limited number of data is available.

The adaptation and application of AI models in electrochemical processes for water treatment were discussed in this thesis. While the usage of AI models is becoming more prevalent in several scientific disciplines, including electrochemical processes, the reliability of the developed models is still critical owing to the limited data available. When there is insufficient data to train the parameters of AI models, it results in overfitting, which means that the model's generalization is problematic. More data is required to address this problem; however, in practical applications, additional data gathering is sometimes difficult due to factors such as time and cost constraints of the experiments. Some recent studies suggested to use data augmentation techniques such as interpolation to address this issue. Although data augmentation approaches have been used in machine learning in various disciplines, particularly image processing and speech recognition, they should

be used with caution for the purpose of regression of experimental work with limited data. This is due to the fact that the behaviour of outputs in experimental studies might be far more complex than describing them using predetermined interpolation functions, which would clearly make it easy for the AI model to anticipate the behaviour of the new interpolated data. Another approach to improve the reliability of the developed AI models and avoid overfitting is to develop the least complex model possible. This was done in this thesis by using learning curves and regularization for the ANN models. In the case of ANFIS, the initial FIS was generated based on the FCM clustering method rather than other alternatives such as grid partitioning. Other machine learning techniques such as decision trees (e.g., random forest), which are well-known for simpler tasks, can be utilized and evaluated.

It was shown in our research that the performance of the AI models with small datasets is highly dependent on the distribution of the data into training, validation, and test sets. To overcome this issue, different validation techniques, including hold-out, k-fold cross-validation, and repeated random sub-sampling validation, were applied. A stratified cross-validation technique can be another alternative for improving the reliability of the models. Although this technique is typically used for classification, it can be adapted for regression problems. The distribution of target values is guaranteed to be about the same across all partitions of the k-fold cross-validation using stratification. This is helpful in ensuring that validation and test performance portray the expected performance of the model with less bias and variance. Finally, future research can employ ensemble modelling approaches, which construct multiple models and then combine them to provide better findings. Ensemble approaches often yield more accurate results than a single model. This can be done using various models in the same training dataset, the same model with different training dataset splits, or any other method. This technique often reduces overfitting and produces a smoother regression model.

APPENDIX I. SUPPLEMENTARY MATERIAL FOR CHAPTER 4

Table AI. 1 Characteristics of caffeine

Parameters	Unit	Caffeine
Formula	-	$C_8H_{10}N_4O_2$
Molecular weight	(g/mol)	194.19
Density	(g/ml)	1.23
Solubility in water	(mg/ml)	20

Table AI. 2 Characteristics of the Quebec City municipal wastewater treatment effluent

Parameter	Value
Cl^- (mg/L)	116.0±12.7
NO_3^- (mg/L)	2.8±1.5
SO_4^{2-} (mg/L)	64.9±8.4
NH_4^+ (mg N/L)	20.4±6.9
Turbidity (NTU)	3.5±0.2
TSS (mg/L)	7.1±1.4
TOC (mg C/L)	8.7±0.9
COD (mg O_2/L)	<80
Conductivity (µS/cm)	962±9
pH	7.5±0.3

Table AI. 3 Experimental ranges and levels of the independent variables

Independent variable	Factor code	Range and level				
		-1.68	-1	0	+1	+1.68
Electrolysis time	X_1	13	20	30	40	47
Current intensity	X_2	0.7	1	1.5	2	2.3
Initial concentration of caffeine	X_3	13	20	30	40	47
Anode type	X_4	-	BDD	-	IrO_2	-

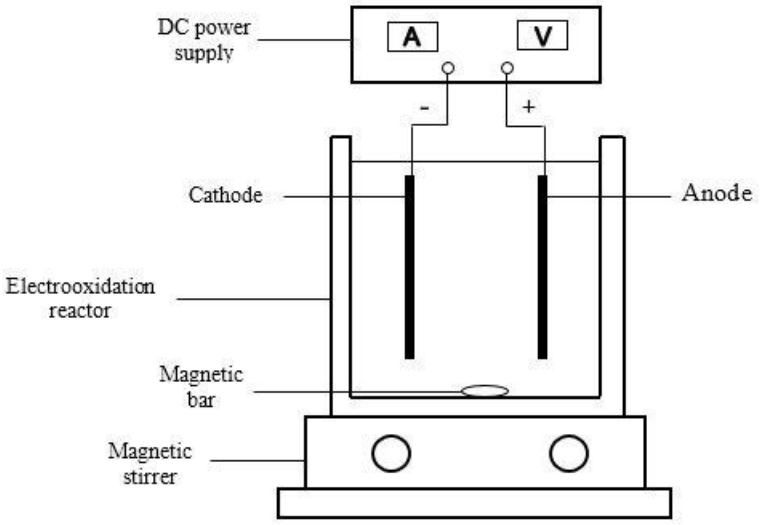
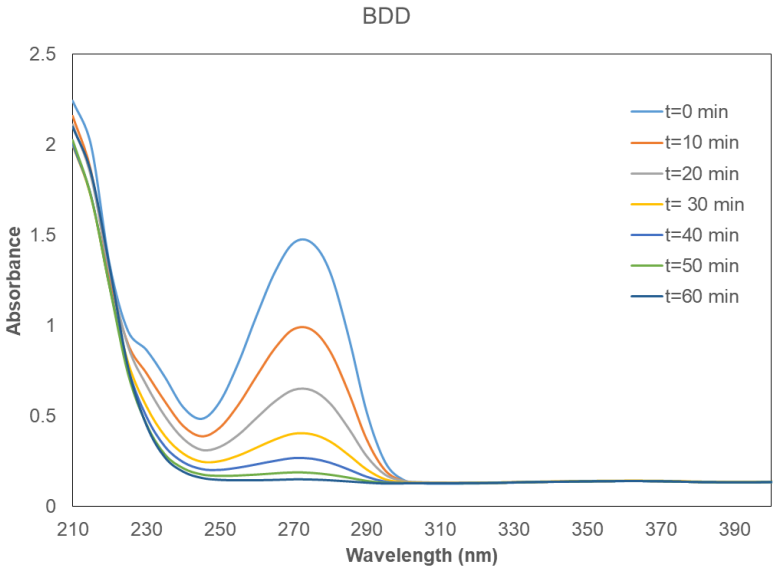


Figure AI. 1 Schematic diagram of the electrooxidation reactor for caffeine degradation



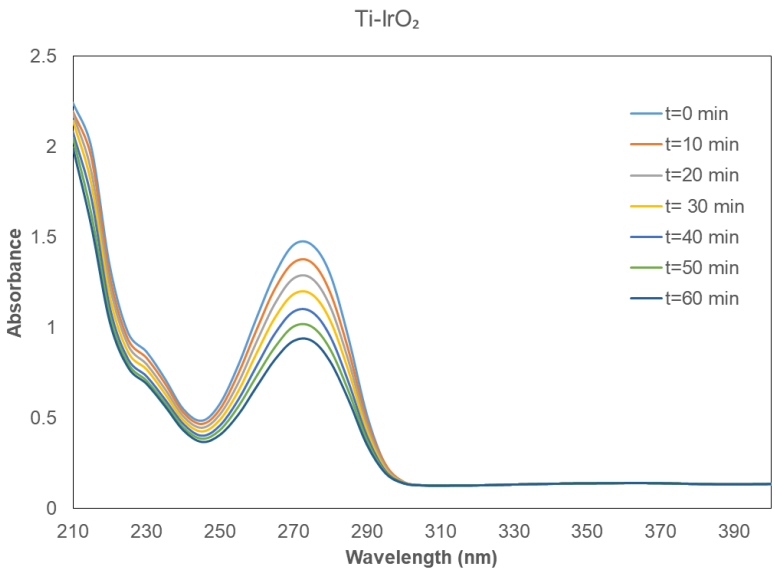


Figure AI. 2 UV absorption spectra of caffeine during the EO process using BDD and Ti-IrO_2 electrodes (I=1.5 A, Na_2SO_4=1 g/L).

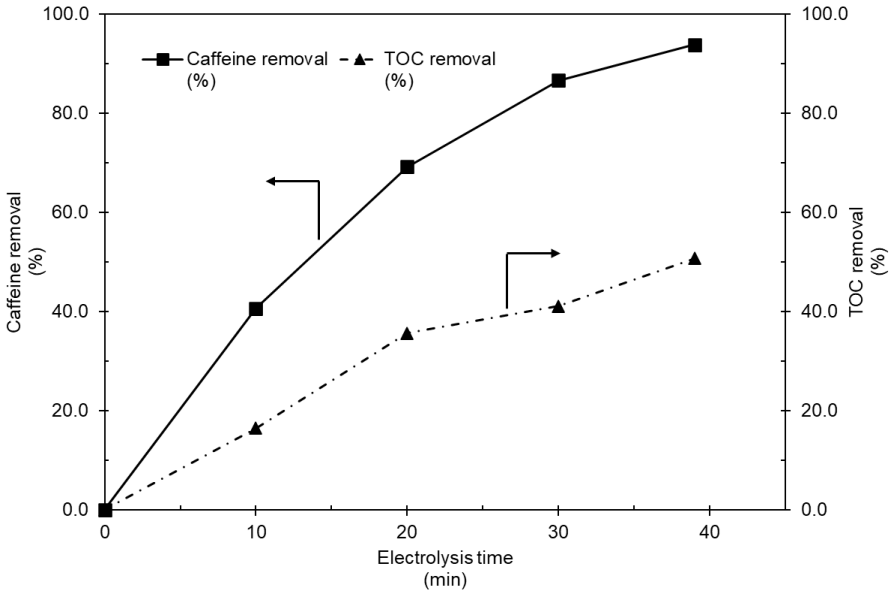


Figure AI. 3 Caffeine and TOC removal efficiency at optimal conditions: [initial caffeine]=13 ppm, current intensity= 0.7 A, and BDD anode

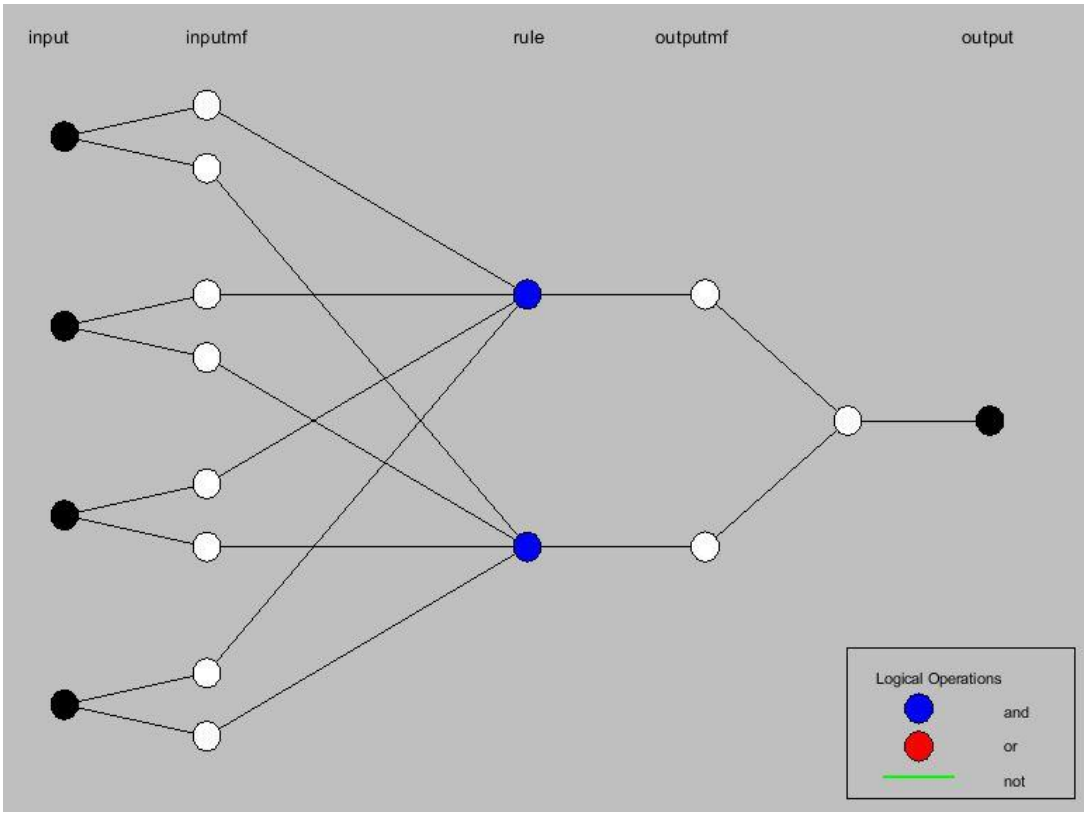


Figure AI. 4 Schematic of the ANFIS model structure

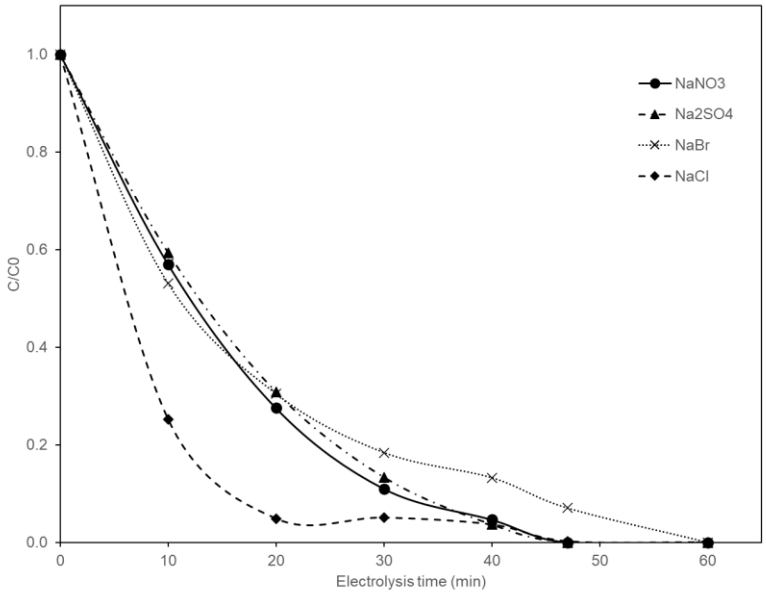


Figure AI. 5 Effect of supporting electrolyte on caffeine degradation in synthetic solution at I=0.7 A, and $[CAF]_0$=13 ppm

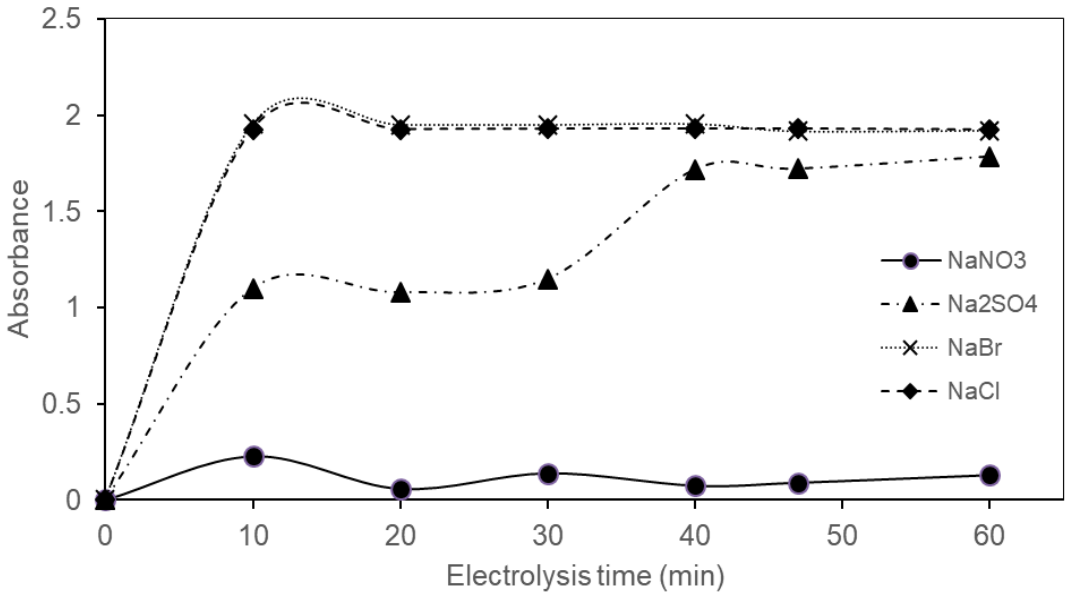


Figure AI. 6 Oxidation capacity measurement for different supporting electrolytes versus electrolysis time (current intensity=0.7 A; Electrolyte conc.= 7 mmol/L; anode= BDD)

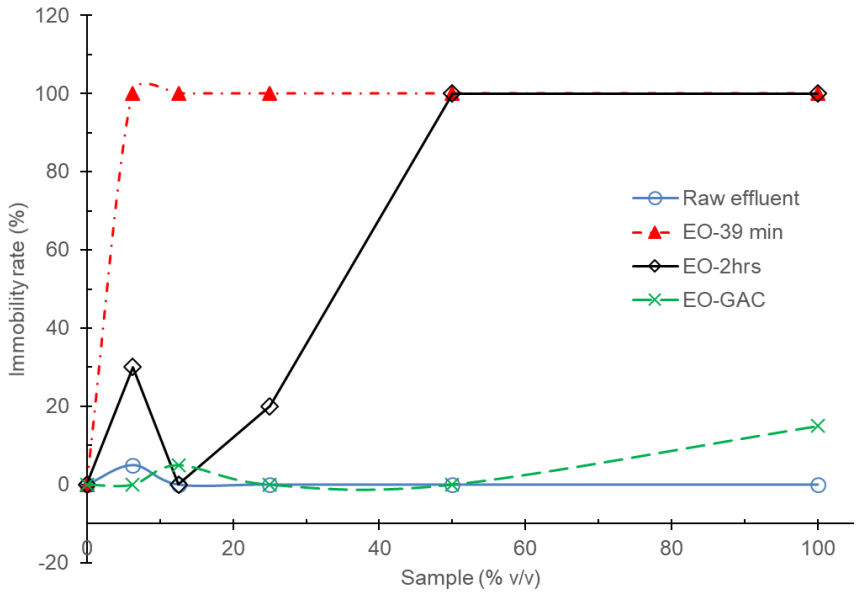


Figure AI. 7 Evaluation of the immobility rate of Daphnia magna for the EO process

References

Abayomi-Alli OO, Damaševičius R, Maskeliūnas R, Abayomi-Alli A (2020) BiLSTM with data augmentation using interpolation methods to improve early detection of parkinson disease. *15th Conference on Computer Science and Information Systems (FedCSIS).* (Sofia, Bulgaria, 6-9 Sept), p 371-380.

Abdessalem AK, Bellakhal N, Oturan N, Dachraoui M, Oturan MA (2010) Treatment of a mixture of three pesticides by photo- and electro-Fenton processes. *Desalination* 250(1):450-455.

Abdul-Wahab SA, Al-Weshahi MA (2009) Brine management: substituting chlorine with nn-site produced sodium hypochlorite for environmentally improved desalination processes. *Water Resources Management* 23(12):2437-2454.

Abdulshahed AM, Longstaff AP, Fletcher S (2015) The application of ANFIS prediction models for thermal error compensation on CNC machine tools. *Applied Soft Computing* 27:158-168.

Aber S, Amani-Ghadim AR, Mirzajani V (2009) Removal of Cr(VI) from polluted solutions by electrocoagulation: Modeling of experimental results using artificial neural network. *Journal of Hazardous materials* 171(1-3):484-490.

Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H (2018) State-of-the-art in artificial neural network applications: A survey. *Heliyon* 4(11):e00938.

Abiodun OI, Jantan A, Omolara AE, Dada KV, Umar AM, Linus OU, Arshad H, Kazaure AA, Gana U, Kiru MU (2019) Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition. *IEEE Access* 7:158820-158846.

Adya M, Collopy F (1998) How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting* 17(5-6):481-495.

Ahmed Basha C, Soloman PA, Velan M, Miranda LR, Balasubramanian N, Siva R (2010) Electrochemical degradation of specialty chemical industry effluent. *Journal of Hazardous materials* 176(1):154-164.

Al-Qaim FF, Mussa ZH, Othman MR, Abdullah MP (2015) Removal of caffeine from aqueous solution by indirect electrochemical oxidation using a graphite-PVC composite electrode: A role of hypochlorite ion as an oxidising agent. *Journal of Hazardous materials* 300:387-397.

Al-Shannag M, Al-Qodah Z, Bani-Melhem K, Qtaishat MR, Alkasrawi M (2015) Heavy metal ions removal from metal plating wastewater using electrocoagulation: Kinetic study and process performance. *Chemical Engineering Journal* 260:749-756.

Alade I, Saleh T, Abd Rahman MA (2019) Modeling and prediction of the specific heat capacity of Al2O3/water nanofluids using hybrid genetic algorithm/support vector regression model. *Nano-Structures & Nano-Objects* 17:103-111.

Alam T, Qamar S, Dixit A, Benaida M (2020) Genetic Algorithm: Reviews, Implementations, and Applications. *International Journal of Engineering Pedagogy (iJEP)* 10(6):pp. 57-77.

Alavi N, Dehvari M, Alekhamis G, Goudarzi G, Neisi A, Babaei AA (2019) Application of electro-Fenton process for treatment of composting plant leachate: kinetics, operational parameters and modeling. *Journal of Environmental Health Science and Engineering* 17(1):417-431.

Aleksander I, Morton H (1990) *An introduction to neural computing.* Van Nostrand Reinhold Co., USA. 240 p

Altaf MAB, Zhang C, Radakovic L, Yoo J (2016) Design of energy-efficient on-chip EEG classification and recording processors for wearable environments. *2016 IEEE International Symposium on Circuits and Systems (ISCAS).* (Montreal, Canada, 22-25 May), p 1126-1129.

Anderson MJ, Whitcomb PJ (2016) *RSM simplified: optimizing processes using response surface methods for design of experiments.* Productivity Press, USA

Anglada Á, Urtiaga A, Ortiz I (2009) Contributions of electrochemical oxidation to waste-water treatment: fundamentals and review of applications. *Journal of Chemical Technology and Biotechnology* 84(12):1747-1755.

Ano J, Assémian AS, Yobouet YA, Adouby K, Drogui P (2019) Electrochemical removal of phosphate from synthetic effluent: A comparative study between iron and aluminum by using experimental design methodology. *Process Safety and Environmental Protection* 129:184-195.

Ansari A, Bakar AA (2014) A comparative study of three artificial intelligence techniques: Genetic Algorithm, Neural Network, and Fuzzy Logic, on scheduling problem. *4th International Conference on Artificial Intelligence with Applications in Engineering and Technology.* (Kota kinabalu, Sabah, Malaysia, 3-5 Dec.), p 31-36.

Aquino JM, Rocha-Filho RC, Saez C, Canizares P, Rodrigo MA (2014) High efficiencies in the electrochemical oxidation of an anthraquinonic dye with conductive-diamond anodes. *Environmental Science and Pollution Research* 21(14):8442-8450.

Arfanis MK, Adamou P, Moustakas NG, Triantis TM, Kontos AG, Falaras P (2017) Photocatalytic degradation of salicylic acid and caffeine emerging contaminants using titania nanotubes. *Chemical Engineering Journal* 310:525-536.

Assefi P, Ghaedi M, Ansari A, Habibi MH, Momeni MS (2014) Artificial neural network optimization for removal of hazardous dye Eosin Y from aqueous solution using CO2O3-NP-AC: Isotherm and kinetics study. *Journal of Industrial and Engineering Chemistry* 20(5):2905-2913.

Asselin M, Drogui P, Benmoussa H, Blais JF (2008a) Effectiveness of electrocoagulation process in removing organic compounds from slaughterhouse wastewater using monopolar and bipolar electrolytic cells. *Chemosphere* 72(11):1727-1733.

Asselin M, Drogui P, Brar SK, Benmoussa H, Blais JF (2008b) Organics removal in oily bilgewater by electrocoagulation process. *Journal of Hazardous materials* 151(2-3):446-455.

Atmaca E (2009) Treatment of landfill leachate by using electro-Fenton method. *Journal of Hazardous materials* 163(1):109-114.

Baughman DR, Liu YA (1995) *Neural Networks in Bioprocessing and Chemical Engineering.* Academic Press, Boston

Beauchamp G, Amaducci A, Cook M (2017) Caffeine toxicity: A brief review and update. *Clinical Pediatric Emergency Medicine* 18(3):197-202.

Bektaş TE, Uğurluoğlu BK, Tan B (2021) Phosphate removal by Ion exchange in batch mode. *Water Practice and Technology* 16(4):1343-1354.

Belkacem S, Bouafia S, Chabani M (2017) Study of oxytetracycline degradation by means of anodic oxidation process using platinized titanium (Ti/Pt) anode and modeling by artificial neural networks. *Process Safety and Environmental Protection* 111:170-179.

Bhatti MS, Kapoor D, Kalia RK, Reddy AS, Thukral AK (2011a) RSM and ANN modeling for electrocoagulation of copper from simulated wastewater: Multi objective optimization using genetic algorithm approach. *Desalination* 274(1-3):74-80.

Bhatti MS, Reddy AS, Kalia RK, Thukral AK (2011b) Modeling and optimization of voltage and treatment time for electrocoagulation removal of hexavalent chromium. *Desalination* 269(1-3):157-162.

Borges F, Roux-de Balmann H, Guardani R (2009) Modeling electrodialysis and photochemical process for their integration in saline wastewater treatment. *Brazilian Journal of Chemical Engineering* 27(3):473 - 482.

Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. in *Proceedings of the fifth annual workshop on Computational learning theory* (ACM, Pittsburgh, Pennsylvania, USA), p 144-152.

Bouasla C, Samar ME-H, Bendjama H (2014) Kinetics study and neural network modeling of degradation of Naphtol Blue Black by electro-Fenton process: effects of anions, metal ions, and organic compound. *Desalination and Water Treatment* 52(34-36):6733-6744.

Boucheikhchoukh A, Thibault J, Fauteux-Lefebvre C (2020) Catalyst design using artificial intelligence: SO2 to SO3 case study. *The Canadian Journal of Chemical Engineering* 98(9):2016-2031.

Brito CdN, de Araújo DM, Martínez-Huitle CA, Rodrigo MA (2015) Understanding active chlorine species production using boron doped diamond films with lower and higher sp3/sp2 ratio. *Electrochemistry Communications* 55:34-38.

Bruton T, Alboloushi A, de la Garza B, Kim B-O, Halden RU (2010) Fate of Caffeine in the Environment and Ecotoxicological Considerations. *In: Contaminants of Emerging Concern in the Environment: Ecological and Human Health Considerations,* (ACS Symposium Series: 1048) American Chemical Society, Vol 1048. p 257-273.

Buerge IJ, Poiger T, Müller MD, Buser H-R (2003) Caffeine, an anthropogenic marker for wastewater contamination of surface waters. *Environmental Science & Technology* 37(4):691-700.

Bui H (2016) Modeling the removal of Sunfix Red S3B from aqueous solution by electrocoagulation process using artificial neural network. *Journal of the Serbian Chemical Society* 81:32-32.

Buragohain M (2009) *Adaptive Network based Fuzzy Inference System (ANFIS) as a Tool for System Identification with Special Emphasis on Training Data Minimization*. DOCTOR OF PHILOSOPHY (Indian Institute of Technology Guwahati, India).

Cañizares P, Garcia-Gomez J, Lobato J, Rodrigo MA (2004a) Modeling of wastewater electro-oxidation processes Part I. General description and application to inactive electrodes. *Industrial and Engineering Chemistry Research* 43(9):1915-1922.

Cañizares P, Garcia-Gomez J, Lobato J, Rodrigo MA (2004b) Modeling of Wastewater Electro-oxidation Processes Part II. Application to Active Electrodes. *Industrial and Engineering Chemistry Research* 43(9):1923-1931.

Cao W, Liu X, Ni J (2020) Parameter optimization of support vector regression using Henry Gas Solubility optimization algorithm. *IEEE Access* 8:88633-88642.

Carabin A, Drogui P, Robert D (2016) Photocatalytic oxidation of carbamazepine: Application of an experimental design methodology. *Water, Air, & Soil Pollution* 227(4).

Carvalho AR, Ramos FM, Chaves AA (2011) Metaheuristics for the feedforward artificial neural network (ANN) architecture optimization problem. *Neural Computing and Applications* 20(8):1273-1284.

Çevik A, Kurtoğlu AE, Bilgehan M, Gülşan ME, Albegmprli HM (2015) Support vector machines in structural engineering: a review. *Journal of Civil Engineering and Management* 21(3):261-281.

Chan ZSH, Ngan HW, Rad AB, David AK, Kasabov N (2006) Short-term ANN load forecasting from limited data using generalization learning strategies. *Neurocomputing* 70(1):409-419.

Chen T-S (2015) Anodic degradation of caffeine under different operating conditions. *Fresenius Environmental Bulletin* 24:800.

Chen W, Yu X, Wang H, Wen B-g (2010) PSO-GA on Endmember extraction for hyperspectral imagery. *2010 International Conference on Computer Application and System Modeling (ICCASM 2010).* (Taiyuan, China, 22-24 Oct.), p V7-459-V457-464.

Chindapan N, Sablani SS, Chiewchan N, Devahastin S (2013) Modeling and optimization of electrodialytic desalination of fish sauce using artificial neural networks and genetic algorithm. *Food and Bioprocess Technology* 6(10):2695-2707.

Chuang LC, Luo CH, Huang SW, Wu YC, Huang YC (2011) Photocatalytic degradation mechanism and kinetics of caffeine in aqueous suspension of nano-TiO2. *Advanced Materials Research* 214:97-102.

Comninellis C (1994) Electrocatalysis in the electrochemical conversion/combustion of organic pollutants for waste water treatment. *Electrochimica Acta* 39(11):1857-1862.

Cotillas S, de Vidales MJM, Llanos J, Sáez C, Cañizares P, Rodrigo MA (2016) Electrolytic and electro-irradiated processes with diamond anodes for the oxidation of persistent pollutants and disinfection of urban treated wastewater. *Journal of Hazardous materials* 319:93-101.

Curteanu S, Cazacu M (2007) Neural Networks and Genetic Algorithms used for modeling and optimization of the siloxane-siloxane copolymers synthesis. *Journal of Macromolecular Science Part A-Pure and Applied Chemistry* 45(1):23-36.

Curteanu S, Godini K, Piuleac CG, Azarian G, Rahmani AR, Butnariu C (2014) Electro-Oxidation method applied for activated sludge treatment: Experiment and simulation based on supervised machine learning methods. *Industrial & Engineering Chemistry Research* 53(12):4902-4912.

Curteanu S, Piuleac CG, Godini K, Azaryan G (2011) Modeling of electrolysis process in wastewater treatment using different types of neural networks. *Chemical Engineering Journal* 172(1):267-276.

Da Silva IN, Hernane Spatti D, Andrade Flauzino R, Liboni LHB, Dos Reis Alves SF (2017) *Artificial Neural Networks : A practical course.* Springer, Cham, Switzerland

da Silva Ribeiro T, Grossi CD, Merma AG, dos Santos BF, Torem ML (2019) Removal of boron from mining wastewaters by electrocoagulation method: Modelling experimental data using artificial neural networks. *Minerals Engineering* 131:8-13.

Daghrir R, Drogui P, François Blais J, Mercier G (2012) Hybrid process combining electrocoagulation and electro-oxidation processes for the treatment of restaurant wastewaters. *Journal of Environmental Engineering* 138(11):1146-1156.

Daghrir R, Drogui P, Tshibangu J (2014) Efficient treatment of domestic wastewater by electrochemical oxidation process using bored doped diamond anode. *Separation and Purification Technology* 131:79-83.

Daghrir R, Drogui P, Zaviska F (2013) Effectiveness of a hybrid process combining electro-coagulation and electro-oxidation for the treatment of domestic wastewaters using response surface methodology. *Journal of Environment Science and Health, Part A Environmental Science* 48(3):308-318.

Dalmázio I, Santos LS, Lopes RP, Eberlin MN, Augusti R (2005) Advanced oxidation of caffeine in water: On-line and real-time monitoring by electrospray ionization mass spectrometry. *Environmental Science & Technology* 39(16):5982-5988.

Daneshvar N, Khataee AR, Djafarzadeh N (2006) The use of artificial neural networks (ANN) for modeling of decolorization of textile dye solution containing C. I. Basic Yellow 28 by electrocoagulation process. *Journal of Hazardous materials* 137(3):1788-1795.

David C, A K, Thangavelu A (2020) Degradation of distillery effluent by twisted-type Iron electrodes: experimental with ANN approach. *International Journal of Environmental Analytical Chemistry* 10.1080/03067319.2020.1807532:1-13.

de Barros Machado M, Santiago VMJ (2014) Electrodialysis treatment of refinery wastewater. *Electrodialysis and Water Reuse: Novel Approaches,* Springer, (eds). Berlin, Heidelberg. p 77-90.

Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6(2):182-197.

Dia O, Drogui P, Buelna G, Dube R, Ihsen BS (2017) Electrocoagulation of bio-filtrated landfill leachate: Fractionation of organic matter and influence of anode materials. *Chemosphere* 168:1136-1141.

Ding S-F, Qi B-J, Tan H-Y (2011a) An overview on theory and algorithm of support vector machines. *Journal of University of Electronic Science and Technology of China* 40(1):2-10.

Ding S, Su C, Yu J (2011b) An optimizing BP neural network algorithm based on genetic algorithm. *Artificial Intelligence Review* 36(2):153-162.

Drogui P, Blais J-F, Mercier G (2007) Review of electrochemical technologies for environmental applications. *Recent Patents on Engineering* 1:257-272.

Durán FE, de Araújo DM, do Nascimento Brito C, Santos EV, Ganiyu SO, Martínez-Huitle CA (2018) Electrochemical technology for the treatment of real washing machine effluent at pre-pilot plant scale by using active and non-active anodes. *Journal of Electroanalytical Chemistry* 818:216-222.

Eberhart RC, Shi Y (1998) Comparison between genetic algorithms and particle swarm optimization. *Evolutionary Programming VII.* (Berlin, Heidelberg, Springer Berlin Heidelberg, p 611-616.

El-Hanafi N, Mehibel L, Li HZ, Poncin S, Bensadok K (2014) Mineralization of the pharmaceutical β-blocker Atenolol by means of indirect electrochemical advanced oxidation process: parametric and kinetic study. *Separation Science and Technology* 49(18):2942-2950.

Elazzouzi M, Haboubi K, Elyoubi MS (2017) Electrocoagulation flocculation as a low-cost process for pollutants removal from urban wastewater. *Chemical Engineering Research and Design* 117:614-626.

Elhalil A, Elmoubarki R, Farnane M, Machrouhi A, Sadiq M, Mahjoubi FZ, Qourzal S, Barka N (2018) Photocatalytic degradation of caffeine as a model pharmaceutical pollutant on Mg doped ZnO-Al2O3 heterostructure. *Environmental Nanotechnology, Monitoring & Management* 10:63-72.

Elmas S, Ambroz F, Chugh D, Nann T (2016) Microfluidic chip for the photocatalytic production of active chlorine. *Langmuir* 32(19):4952-4958.

Elsheikh AH, Sharshir SW, Abd Elaziz M, Kabeel AE, Guilan W, Haiou Z (2019) Modeling of solar energy systems using artificial neural network: A comprehensive review. *Solar Energy* 180:622-639.

Entezari MH, Kruus P (1994) Effect of frequency on sonochemical reactions. I: Oxidation of iodide. *Ultrasonics Sonochemistry* 1(2):S75-S79.

Farzin S, Nabizadeh Chianeh F, Valikhan Anaraki M, Mahmoudian F (2020) Introducing a framework for modeling of drug electrochemical removal from wastewater based on data mining algorithms, scatter interpolation method, and multi criteria decision analysis (DID). *Journal of Cleaner Production* 266:122075.

Feng L, Hong W (2008) Classification error of multilayer perceptron neural networks. *Neural Computing and Applications* 18(4):377-380.

Feng Y, Yang L, Liu J, Logan BE (2016) Electrochemical technologies for wastewater treatment and resource reclamation. *Environmental Science: Water Research & Technology* 2(5):800-831.

Ferre-Aracil J, Valcárcel Y, Negreira N, de Alda ML, Barceló D, Cardona SC, Navarro-Laboulais J (2016) Ozonation of hospital raw wastewaters for cytostatic compounds removal. Kinetic modelling and economic assessment of the process. *Science of the Total Environment* 556:70-79.

Gallant SI (1993) *Neural Network Learning and Expert Systems.* MIT Press, Cambridge, MA, USA

Ganzenko O, Huguenot D, van Hullebusch ED, Esposito G, Oturan MA (2014) Electrochemical advanced oxidation and biological processes for wastewater treatment: a review of the combined approaches. *Environmental Science and Pollution Research* 21(14):8493-8524.

Ganzenko O, Oturan N, Huguenot D, van Hullebusch ED, Esposito G, Oturan MA (2015) Removal of psychoactive pharmaceutical caffeine from water by electro-Fenton process using BDD anode: Effects of operating parameters on removal efficiency. *Separation and Purification Technology* 156:987-995.

Ganzenko O, Trellu C, Oturan N, Huguenot D, Péchaud Y, van Hullebusch ED, Oturan MA (2020) Electro-Fenton treatment of a complex pharmaceutical mixture: Mineralization efficiency and biodegradability enhancement. *Chemosphere* 253:126659.

Garcia-Segura S, Ocon JD, Chong MN (2018) Electrochemical oxidation remediation of real wastewater effluents — A review. *Process Safety and Environmental Protection* 113:48-67.

Garson GD (1991) Interpreting Neural-Network Connection Weights. *AI Expert* 6:46-51.

Gevrey M, Dimopoulos I, Lek S (2003) Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling* 160(3):249-264.

Ghaedi AM, Vafaei A (2017) Applications of artificial neural networks for adsorption removal of dyes from aqueous solution: A review. *Advances in Colloid and Interface Science* 245:20-39.
Ghaedi M, Ghaedi AM, Abdi F, Roosta M, Sahraei R, Daneshfar A (2014) Principal component analysis-artificial neural network and genetic algorithm optimization for removal of reactive orange 12 by copper sulfide nanoparticles-activated carbon. *Journal of Industrial and Engineering Chemistry* 20(3):787-795.
Ghanbari F, Moradi M (2015) A comparative study of electrocoagulation, electrochemical Fenton, electro-Fenton and peroxi-coagulation for decolorization of real textile wastewater: electrical energy consumption and biodegradability improvement. *Journal of Environmental Chemical Engineering* 3(1):499-506.
Ghiasi MM, Bahadori A, Zendehboudi S (2014) Estimation of the water content of natural gas dried by solid calcium chloride dehydrator units. *Fuel* 117:33-42.
Ghiasi MM, Bahadori A, Zendehboudi S, Chatzis I (2015) Rigorous models to optimise stripping gas rate in natural gas dehydration units. *Fuel* 140:421-428.
Gholami Shirkoohi M, D. Tyagi R, Vanrolleghem PA, Drogui P (2021) Artificial neural networks and genetic algorithms: An efficient modelling and optimization methodology for active chlorine production using the electrolysis process. *The Canadian Journal of Chemical Engineering* 99(S1):S389-S403.
Gimeno O, García-Araya JF, Beltrán FJ, Rivas FJ, Espejo A (2016) Removal of emerging contaminants from a primary effluent of municipal wastewater by means of sequential biological degradation-solar photocatalytic oxidation processes. *Chemical Engineering Journal* 290:12-20.
Gireli GAS, Maniero MG, Guimarães JR (2019) Influence of pH value on sulfonamide ozonation using caffeine as a contamination indicator. *Water Supply* 20(2):508-515.
Giwa A, Daer S, Ahmed I, Marpu PR, Hasan SW (2016) Experimental investigation and artificial neural networks ANNs modeling of electrically-enhanced membrane bioreactor for wastewater treatment. *Journal of Water Process Engineering* 11:88-97.
Gizaw A, Zewge F, Kumar A, Mekonnen A, Tesfaye M (2021) A comprehensive review on nitrate and phosphate removal and recovery from aqueous solutions by adsorption. *Journal of Water Supply: Research and Technology-Aqua* 70(7):921-947.
Goh ATC (1995) Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering* 9(3):143-151.
Gonce N, Voudrias EA (1994) Removal of chlorite and chlorate ions from water using granular activated carbon. *Water Research* 28(5):1059-1069.
Grimm J, Bessarabov D, Sanderson R (1998) Review of electro-assisted methods for water purification. *Desalination* 115(3):285-294.
Gu J, Zhu M, Jiang L (2011) Housing price forecasting based on genetic algorithm and support vector machine. *Expert Systems with Applications* 38(4):3383-3386.
Guitaya L, Azaïs A, Zaviska F, Drogui P, Blais J-F, Gourich B (2017) Electrochemical oxidation as treatment for contaminated wastewaters by carbamazepine: Process optimization through response surface methodology. *Water, Air, & Soil Pollution* 228(10):384-397.
Hafezi R, Shahrabi J, Hadavandi E (2015) A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price. *Applied Soft Computing* 29:196-210.
Hakizimana JN, Gourich B, Chafi M, Stiriba Y, Vial C, Drogui P, Naja J (2017) Electrocoagulation process in water treatment: A review of electrocoagulation modeling approaches. *Desalination* 404:1-21.
Hamed MM, Khalafallah MG, Hassanien EA (2004) Prediction of wastewater treatment plant performance using artificial neural networks. *Environmental Modelling & Software* 19(10):919-928.
Hammerstrom D (1993) Working with neural networks. *IEEE Spectrum* 30(7):46-53.

Hasani G, Daraei H, Shahmoradi B, Gharibi F, Maleki A, Yetilmezsoy K, McKay G (2018) A novel ANN approach for modeling of alternating pulse current electrocoagulation-flotation (APC-ECF) process: Humic acid removal from aqueous media. *Process Safety and Environmental Protection* 117:111-124.

Haykin S (1998) *Neural Networks. A Comprehensive Foundation.* Prentice Hall PTR, USA

Hillebrand O, Nödler K, Licha T, Sauter M, Geyer T (2012) Caffeine as an indicator for the quantification of untreated wastewater in karst systems. *Water Research* 46(2):395-402.

Holland JH (1992) *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.* MIT Press, United States

Hsu G-SW, Hsia C-W, Hsu S-Y (2015a) Effects of electrode settings on chlorine generation efficiency of electrolyzing seawater. *Journal of Food and Drug Analysis* 23(4):729-734.

Hsu G-SW, Hsia C-W, Hsu S-Y (2015b) Effects of process conditions on chlorine generation and storage stability of electrolyzed deep ocean water. *Journal of Food and Drug Analysis* 23(4):735-741.

Hsu G-SW, Hsu S-Y (2016) Effects of electrode gap and electric potential on chlorine generation of electrolyzed deep ocean water. *Journal of Food and Drug Analysis* 24(3):575-578.

Hu X, Wang M, Hu X, Leeson MS (2014) Calculating the complete pareto front for a special class of continuous multi-objective optimization problems. *IEEE Congress on Evolutionary Computation (CEC).* (Beijing, 6-11 Jul.), p 290-297.

Huang W, Liu H, Zhang Y, Mi R, Tong C, Xiao W, Shuai B (2021) Railway dangerous goods transportation system risk identification: Comparisons among SVM, PSO-SVM, GA-SVM and GS-SVM. *Applied Soft Computing* 109:107541.

Huang YUR, Hung Y-C, Hsu S-Y, Huang Y-w, Hwang D-F (2008) Application of electrolyzed water in the food industry. *Food Control* 19:329-345.

Hussein AM (2016) Adaptive Neuro-Fuzzy Inference System of friction factor and heat transfer nanofluid turbulent flow in a heated tube. *Case Studies in Thermal Engineering* 8:94-104.

Indermuhle C, Martin de Vidales MJ, Saez C, Robles J, Canizares P, Garcia-Reyes JF, Molina-Diaz A, Comninellis C, Rodrigo MA (2013) Degradation of caffeine by conductive diamond electrochemical oxidation. *Chemosphere* 93(9):1720-1725.

Isarain-Chávez E, Arias C, Cabot PL, Centellas F, Rodríguez RM, Garrido JA, Brillas E (2010) Mineralization of the drug β-blocker atenolol by electro-Fenton and photoelectro-Fenton using an air-diffusion cathode for H2O2 electrogeneration combined with a carbon-felt cathode for Fe2+ regeneration. *Applied Catalysis B: Environmental* 96(3):361-369.

Jain LC, Medsker LR (1999) *Recurrent Neural Networks: Design and Applications.* CRC Press, Inc. 416 p

Jang JR (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics* 23(3):665-685.

Jang JR, Sun C (1993) Predicting chaotic time series with fuzzy if-then rules. *Second IEEE International Conference on Fuzzy Systems.*28 March-1 April), p 1079-1084 vol.1072.

Jardak K, Dirany A, Drogui P, El Khakani MA (2016) Electrochemical degradation of ethylene glycol in antifreeze liquids using boron doped diamond anode. *Separation and Purification Technology* 168:215-222.

Jardak K, Dirany A, Drogui P, El Khakani MA (2017) Statistical optimization of electrochemical oxidation of ethylene glycol using response surface methodology. *Process Safety and Environmental Protection* 105:12-20.

Jelic A, Gros M, Ginebreda A, Cespedes-Sánchez R, Ventura F, Petrovic M, Barcelo D (2011) Occurrence, partition and removal of pharmaceuticals in sewage water and sludge during wastewater treatment. *Water Research* 45(3):1165-1176.

Jiang BT, Zhou J, Huang XB, Wang PF (2020) Prediction of critical heat flux using Gaussian process regression and ant colony optimization. *Annals of Nuclear Energy* 149:107765.

Jing G, Du W, Guo Y (2012) Studies on prediction of separation percent in electrodialysis process via BP neural networks and improved BP algorithms. *Desalination* 291:78-93.
Jing G, Ren S, Pooley S, Sun W, Kowalczuk PB, Gao Z (2021) Electrocoagulation for industrial wastewater treatment: an updated review. *Environmental Science: Water Research & Technology* 7(7):1177-1196.
Juneja M, Nagar SK (2016) Particle swarm optimization algorithm and its parameters: A review. *International Conference on Control, Computing, Communication and Materials (ICCCCM).* (Allahbad, India, 21-22 Oct.), p 1-5.
Kahl S, Nivala J, van Afferden M, Müller RA, Reemtsma T (2017) Effect of design and operational conditions on the performance of subsurface flow treatment wetlands: Emerging organic contaminants as indicators. *Water Research* 125:490-500.
Kamari A, Mohammadi AH, Bahadori A, Zendehboudi S (2014) A reliable model for estimating the wax deposition rate during crude oil production and processing. *Petroleum Science and Technology* 32(23):2837-2844.
Kamosi M, Hashemi AB, Meybodi MR (2010) A new Particle Swarm Optimization Algorithm for dynamic environments. *Swarm, Evolutionary, and Memetic Computing.* (Berlin, Heidelberg, 2010) Springer, p 129-138.
Karimi H, Ghaedi M (2014) Application of artificial neural network and genetic algorithm to modeling and optimization of removal of methylene blue using activated carbon. *Journal of Industrial and Engineering Chemistry* 20(4):2471-2476.
Karimifard S, Alavi Moghaddam MR (2018) Application of response surface methodology in physicochemical removal of dyes from wastewater: A critical review. *Science of the Total Environment* 640-641:772-797.
Kaur P, Sangal VK, Kushwaha JP (2019) Parametric study of electro-Fenton treatment for real textile wastewater, disposal study and its cost analysis. *International Journal of Environmental Science and Technology* 16(2):801-810.
Kennedy J, Eberhart R (1995) Particle Swarm optimization. *ICNN'95 - International Conference on Neural Networks.* (Perth, WA, Australia, 27 Nov.-1 Dec.), p 1942-1948 vol.1944.
Keskin C, Kirbac E, Özdemir A, Şengil İ (2011) Decolorization Of binary dye mixture by electrocoagulation and modeling of experimental results using Artificial Neural Network. *Fresenius Environmental Bulletin* 20:3206-3216.
Khataee AR, Kasiri MB (2010) Artificial neural networks modeling of contaminated water treatment processes by homogeneous and heterogeneous nanocatalysis. *Journal of Molecular Catalysis A: Chemical* 331(1-2):86-100.
Khouzam KY (2008) Electrolysis of salt water for chlorine production by photovoltaic power. *IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century.* (Pittsburgh, 20-24 Jul.), p 1-8.
Khude P, Majumdar A, Butola BS (2020) Modelling and prediction of antibacterial activity of knitted fabrics made from silver nanocomposite fibres using soft computing approaches. *Neural Computing and Applications* 32(13):9509-9519.
Kobya M, Gengec E, Sensoy MT, Demirbas E (2014) Treatment of textile dyeing wastewater by electrocoagulation using Fe and Al electrodes: optimisation of operating parameters using central composite design. *Coloration Technology* 130(3):226-235.
Kobya M, Omwene PI, Sarabi SM, Yildirim S, Ukundimana Z (2021) Phosphorous removal from anaerobically digested municipal sludge centrate by an electrocoagulation reactor using metal (Al, Fe and Al-Fe) scrap anodes. *Process Safety and Environmental Protection* 152:188-200.
Kolonko KJ, Shapiro RH, Barkley RM, Sievers RE (1979) Ozonization of caffeine in aqueous solution. *The Journal of Organic Chemistry* 44(22):3769-3778.

Komtchou S, Dirany A, Drogui P, Robert D, Lafrance P (2017) Removal of atrazine and its by-products from water using electrochemical advanced oxidation processes. *Water Research* 125:91-103.
Kothari MS, Shah KA (2020) Electrochemical oxidation for decolorization of Rhodamine-B dye using mixed metal oxide electrode: modeling and optimization. *Water Science and Technology* 81(4):720-731.
Kulkarni MNK, Patekar MS, Bhoskar MT, Kulkarni MO, Kakandikar GM, Nandedkar VM (2015) Particle Swarm Optimization applications to mechanical engineering- A review. *Materials Today: Proceedings* 2(4):2631-2639.
Kumari M, Gupta SK (2019) Response surface methodological (RSM) approach for optimizing the removal of trihalomethanes (THMs) and its precursor's by surfactant modified magnetic nanoadsorbents (sMNP) - An endeavor to diminish probable cancer risk. *Scientific Reports* 9(1):18339.
Lavanya A, Sri Krishnaperumal Thanga R (2021) Effective removal of phosphorous from dairy wastewater by struvite precipitation: process optimization using response surface methodology and chemical equilibrium modeling. *Separation Science and Technology* 56(2):395-410.
Lenin Sundar M, Kalyani G, Gokulan R, Ragunath S, Joga Rao H (2021) Comparative adsorptive removal of Reactive Red 120 using RSM and ANFIS models in batch and packed bed column. *Biomass Conversion and Biorefinery* https://doi.org/10.1007/s13399-021-01444-7.
Li M, Mei Q, Han D, Wei B, An Z, Cao H, Xie J, He M (2021) The roles of HO•, ClO• and BrO• radicals in caffeine degradation: A theoretical study. *Science of the Total Environment* 768:144733.
Li S, He B, Wang J, Liu J, Hu X (2020a) Risks of caffeine residues in the environment: Necessity for a targeted ecopharmacovigilance program. *Chemosphere* 243:125343.
Li S, Wen J, He B, Wang J, Hu X, Liu J (2020b) Occurrence of caffeine in the freshwater environment: Implications for ecopharmacovigilance. *Environmental Pollution* 263:114371.
Li Y, Lu D, Liu X, Li Z, Zhu H, Cui J, Zhang H, Mao X (2022) Coupling of cathodic aluminum dissolution and anodic oxidation process for simultaneous removal of phosphate and ammonia in wastewaters. *Chemical Engineering Journal* 427:130944.
Li Z, Guo J, Jiao W, Xu P, Liu B, Zhao X (2020c) Random linear interpolation data augmentation for person re-identification. *Multimedia Tools and Applications* 79(7):4931-4947.
Liu G, Zhou Y, Yan Q, Doekhi-Bennani Y (2021) Graphdiyne as a novel nonactive anode for wastewater treatment: A theoretical study. *Chinese Chemical Letters* 32(9):2819-2822.
Loaiza-Ambuludi S, Panizza M, Oturan N, Özcan A, Oturan MA (2013) Electro-Fenton degradation of anti-inflammatory drug ibuprofen in hydroorganic medium. *Journal of Electroanalytical Chemistry* 702:31-36.
Lobo FG, Lima CF, Michalewicz Z (2007) *Parameter Setting in Evolutionary Algorithms.* Springer Publishing Company, Heidelberg, Germany
Lu H, Zou W, Chai P, Wang J, Bazinet L (2016) Feasibility of antibiotic and sulfate ions separation from wastewater using electrodialysis with ultrafiltration membrane. *Journal of Cleaner Production* 112:3097-3105.
Madhiarasan M, Deepa SN (2017) Comparative analysis on hidden neurons estimation in multi layer perceptron neural networks for wind speed forecasting. *Artificial Intelligence Review* 48(4):449-471.
Maleki A, Daraei H, Shahmoradi B, Razee S, Ghobadi N (2014) Electrocoagulation efficiency and energy consumption probing by artificial intelligent approaches. *Desalination and Water Treatment* 52(13-15):2400-2411.
Maleki A, Mahvi A, Daraei H, Rezaei R, Meihami N, Mohammadi K, Zand S (2015) Influence of selected anions on fluoride removal in electrocoagulation/electroflotation. *Fluoride* 48:37-47.

Manokaran P, Saravanathamizhan R, Basha CA, Kannadasan T (2014) Feed-forward back-propagation neural network for the electro-oxidation of distillery effluent. *Chemical Engineering Communications* 201(10):1404-1416.
Mansour D, Fourcade F, Soutrel I, Hauchard D, Bellakhal N, Amrane A (2015) Mineralization of synthetic and industrial pharmaceutical effluent containing trimethoprim by combining electro-Fenton and activated sludge treatment. *Journal of the Taiwan Institute of Chemical Engineers* 53:58-67.
Manu DS, Thalla AK (2017) Artificial intelligence models for predicting the performance of biological wastewater treatment plant in the removal of Kjeldahl Nitrogen from wastewater. *Applied Water Science* 7(7):3783-3791.
Martin de Vidales MJ, Millan M, Saez C, Perez JF, Rodrigo MA, Canizares P (2015) Conductive diamond electrochemical oxidation of caffeine-intensified biologically treated urban wastewater. *Chemosphere* 136:281-288.
Martín de Vidales MJ, Millán M, Sáez C, Pérez JF, Rodrigo MA, Cañizares P (2015) Conductive diamond electrochemical oxidation of caffeine-intensified biologically treated urban wastewater. *Chemosphere* 136:281-288.
Martín de Vidales MJ, Robles-Molina J, Domínguez-Romero JC, Cañizares P, Sáez C, Molina-Díaz A, Rodrigo MA (2012a) Removal of sulfamethoxazole from waters and wastewaters by conductive-diamond electrochemical oxidation. *Journal of Chemical Technology and Biotechnology* 87(10):1441-1449.
Martín de Vidales MJ, Sáez C, Cañizares P, Rodrigo MA (2012b) Metoprolol abatement from wastewaters by electrochemical oxidation with boron doped diamond anodes. *Journal of Chemical Technology and Biotechnology* 87(2):225-231.
Martínez-Huitle CA, Panizza M (2018) Electrochemical oxidation of organic pollutants for wastewater treatment. *Current Opinion in Electrochemistry* 11:62-71.
Mei Y, Yang J, Lu Y, Hao F, Xu D, Pan H, Wang J (2019) BP–ANN model coupled with Particle Swarm Optimization for the efficient prediction of 2-Chlorophenol removal in an electro-oxidation system. *International Journal of Environmental Research and Public Health* 16(14):2454.
Melin P, Castillo O (2014) A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition. *Applied Soft Computing* 21:568-577.
Menapace A, Zanfei A, Righetti M (2021) Tuning ANN hyperparameters for forecasting drinking water demand. *Applied Sciences* 11(9):4290.
Min K, Choi S, Jang D, Lee J (2019a) Separation of metals from electroplating wastewater using electrodialysis. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 41:1-10.
Min KJ, Choi SY, Jang D, Lee J, Park KY (2019b) Separation of metals from electroplating wastewater using electrodialysis. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 41:1-10.
Mirsoleimani-Azizi SM, Amooey AA, Ghasemi S, Salkhordeh-Panbechouleh S (2015) Modeling the removal of Endosulfan from aqueous solution by electrocoagulation process using Artificial Neural Network (ANN). *Industrial and Engineering Chemistry Research* 54(40):9844-9849.
Moghaddam AA, Seifi A, Niknam T (2012) Multi-operation management of a typical micro-grids using Particle Swarm Optimization: A comparative study. *Renewable and Sustainable Energy Reviews* 16(2):1268-1281.
Mohammadi T, Razmi A, Sadrzadeh M (2004) Effect of operating parameters on Pb2+ separation from wastewater using electrodialysis. *Desalination* 167:379-385.
Mólgora CC, Domínguez AM, Avila EM, Drogui P, Buelna G (2013) Removal of arsenic from drinking water: A comparative study between electrocoagulation-microfiltration and chemical coagulation-microfiltration processes. *Separation and Purification Technology* 118:645-651.

Monteil H, Péchaud Y, Oturan N, Oturan MA (2018) A review on efficiency and cost effectiveness of electro- and bio-electro-Fenton processes: Application to the treatment of pharmaceutical pollutants in water. *Chemical Engineering Journal* 376:119577-119609.

Moreira FC, Boaventura RAR, Brillas E, Vilar VJP (2017) Electrochemical advanced oxidation processes: A review on their application to synthetic and real wastewaters. *Applied Catalysis B: Environmental* 202:217-261.

Moreno-Casillas HA, Cocke DL, Gomes JAG, Morkovsky P, Parga JR, Peterson E (2007) Electrocoagulation mechanism for COD removal. *Separation and Purification Technology* 56(2):204-211.

Murugan AA, Ramamurthy T, Subramanian B, Kannan CS, Ganesan M (2009) Electrocoagulation of Textile Effluent: RSM and ANN Modeling. *International Journal of Chemical Reactor Engineering* 7(1).

Naghibi SA, Salehi E, Khajavian M, Vatanpour V, Sillanpää M (2021) Multivariate data-based optimization of membrane adsorption process for wastewater treatment: Multi-layer perceptron adaptive neural network versus adaptive neural fuzzy inference system. *Chemosphere* 267:129268.

Nandi S, Badhe Y, Lonari J, Sridevi U, Rao BS, Tambe SS, Kulkarni BD (2004) Hybrid process modeling and optimization strategies integrating neural networks/support vector regression and genetic algorithms: study of benzene isopropylation on Hbeta catalyst. *Chemical Engineering Journal* 97(2):115-129.

Nardi AE, Lopes FL, Freire RC, Veras AB, Nascimento I, Valença AM, de-Melo-Neto VL, Soares-Filho GL, King AL, Araújo DM, Mezzasalma MA, Rassi A, Zin WA (2009) Panic disorder and social anxiety disorder subtypes in a caffeine challenge test. *Psychiatry Research* 169(2):149-153.

Nasr M, Ateia M, Hassan K (2016) Artificial intelligence for greywater treatment using electrocoagulation process. *Separation Science and Technology* 51(1):96-105.

Nejatian I, Kanani M, Arabloo M, Bahadori A, Zendehboudi S (2014) Prediction of natural gas flow through chokes using support vector machine algorithm. *Journal of Natural Gas Science and Engineering* 18:155-163.

Niculescu SP (2003) Artificial neural networks and genetic algorithms in QSAR. *Journal of Molecular Structure-THEOCHEM* 622(1):71-83.

Noorani Khomeyrani SF, Ahmadi Azqhandi MH, Ghalami-Choobar B (2021) Rapid and efficient ultrasonic assisted adsorption of PNP onto LDH-GO-CNTs: ANFIS, GRNN and RSM modeling, optimization, isotherm, kinetic, and thermodynamic study. *Journal of Molecular Liquids* 333:115917.

Nourouzi MM, Chuah TG, Choong TS (2011) Optimisation of reactive dye removal by sequential electrocoagulation-flocculation method: comparing ANN and RSM prediction. *Water Science & Technology* 63(5):984-994.

Oh C, Han S, Jeong J (2020) Time-Series Data Augmentation based on Interpolation. *Procedia Computer Science* 175:64-71.

Oliveira FH, Osugi ME, Paschoal FMM, Profeti D, Olivi P, Zanoni MVB (2007) Electrochemical oxidation of an acid dye by active chlorine generated using Ti/Sn(1−x)IrxO2 electrodes. *Journal of Applied Electrochemistry* 37(5):583-592.

Olvera-Vargas H, Oturan N, Oturan MA, Brillas E (2015) Electro-Fenton and solar photoelectro-Fenton treatments of the pharmaceutical ranitidine in pre-pilot flow plant scale. *Separation and Purification Technology* 146:127-135.

Ouarda Y, Tiwari B, Azaïs A, Vaudreuil M-A, Ndiaye SD, Drogui P, Tyagi RD, Sauvé S, Desrosiers M, Buelna G, Dubé R (2018) Synthetic hospital wastewater treatment by coupling submerged membrane bioreactor and electrochemical advanced oxidation process: Kinetic study and toxicity assessment. *Chemosphere* 193:160-169.

P. Telo J, J. S. C. Vieira A (1997) Mechanism of free radical oxidation of caffeine in aqueous solution. *Journal of the Chemical Society, Perkin Transactions 2* 10.1039/A700944E(9):1755-1758.

Padovese LR (2002) Comparison Between Probabilistic and Multilayer Perceptron Neural Networks for Rolling Bearing Fault Classification. *International Journal of Modelling and Simulation* 22(2):97-103.
Panizza M (2010) Importance of Electrode Material in the Electrochemical Treatment of Wastewater Containing Organic Pollutants. *Electrochemistry for the Environment,* Springer New York, NY. p 25-54.
Panizza M, Barbucci A, Delucchi M, Carpanese MP, Giuliano A, Cataldo-Hernández M, Cerisola G (2013) Electro-Fenton degradation of anionic surfactants. *Separation and Purification Technology* 118:394-398.
Panizza M, Dirany A, Sirés I, Haidar M, Oturan N, Oturan MA (2014) Complete mineralization of the antibiotic amoxicillin by electro-Fenton with a BDD anode. *Journal of Applied Electrochemistry* 44(12):1327-1335.
Pendashteh AR, Fakhru'l-Razi A, Chaibakhsh N, Abdullah LC, Madaeni SS, Abidin ZZ (2011) Modeling of membrane bioreactor treating hypersaline oily wastewater by artificial neural network. *Journal of Hazardous materials* 192(2):568-575.
Periyasamy S, Muthuchamy M (2018) Electrochemical degradation of psychoactive drug caffeine in aqueous solution using graphite electrode. *Environmental Technology* 39(18):2373-2381.
Picard RR, Cook RD (1984) Cross-validation of regression models. *Journal of the American Statistical Association* 79(387):575-583.
Picos-Benitez AR, Martinez-Vargas BL, Duron-Torres SM, Brillas E, Peralta-Hernandez JM (2020) The use of artificial intelligence models in the prediction of optimum operational conditions for the treatment of dye wastewaters with similar structural characteristics. *Process Safety and Environmental Protection* 143:36-44.
Picos A, Peralta-Hernández JM (2018) Genetic algorithm and artificial neural network model for prediction of discoloration dye from an electro-oxidation process in a press-type reactor. *Water Science and Technology* 78(4):925-935.
Pillai KC, Kwon TO, Park BB, Moon IS (2009) Studies on process parameters for chlorine dioxide production using IrO2 anode in an un-divided electrochemical cell. *Journal of Hazardous materials* 164(2):812-819.
Pinto AH, Eyng E, Baraldi IJ, Fleck L, Orssatto F, Sabbi L (2018) Artificial neural network feedforward controller applied to a hybrid system for textile effluent treatment. *International Journal of Innovative Computing, Information and Control* 14:657-669.
Piotrowski AP, Napiorkowski JJ, Piotrowska AE (2020) Population size in Particle Swarm Optimization. *Swarm and Evolutionary Computation* 58:100718.
Piuleac C, Saez C, Cañizares P, Curteanu S, Rodrigo M (2012) Hybrid Model of a Wastewater-Treatment Electrolytic Process. *International Journal of Electrochemical Science* 7:771-775.
Piuleac CG, Curteanu S, Rodrigo MA, Sáez C, Fernández FJ (2013) Optimization methodology based on neural networks and genetic algorithms applied to electro-coagulation processes. *Central European Journal of Chemistry* 11(7):1213-1224.
Piuleac CG, Rodrigo MA, Cañizares P, Curteanu S, Sáez C (2010) Ten steps modeling of electrolysis processes by using neural networks. *Environmental Modelling & Software* 25(1):74-81.
Polcaro AM, Mascia M, Palmas S, Vacca A (2004) Electrochemical degradation of diuron and dichloroaniline at BDD electrode. *Electrochimica Acta* 49(4):649-656.
Poli R, Kennedy J, Blackwell T (2007) Particle swarm optimization. *Swarm Intelligence* 1(1):33-57.
Qian Y, Hu H, Tan T (2019) Data augmentation using generative adversarial networks for robust speech recognition. *Speech Communication* 114:1-9.

Radwan M, Gar Alalm M, Eletriby H (2018) Optimization and modeling of electro-Fenton process for treatment of phenolic wastewater using nickel and sacrificial stainless steel anodes. *Journal of Water Process Engineering* 22:155-162.

Rahman MF, Yanful EK, Jasim SY (2009) Occurrences of endocrine disrupting compounds and pharmaceuticals in the aquatic environment and their removal from drinking water: Challenges in the context of the developing world. *Desalination* 248(1):578-585.

Rajeshwar K, Ibanez JG (1997) *Environmental electrochemistry: Fundamentals and applications in pollution abatement.* Academic Press, San Diego, USA

Rao Y, Long H, Hao J (2021) The oxidative degradation of Caffeine in UV/Fe(II)/persulfate system—Reaction kinetics and decay pathways. *Water Environment Research* 93(4):559-569.

Ratnaweera A, Halgamuge SK, Watson HC (2004) Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. *IEEE Transactions on Evolutionary Computation* 8(3):240-255.

Ridha H, Njah M, Chtourou M (2008) Multilayer perceptron training using an evolutionary algorithm. *International Journal of Modelling, Identification and Control* 5:305-312.

Rosal R, Rodríguez A, Perdigón-Melón JA, Petre A, García-Calvo E, Gómez MJ, Agüera A, Fernández-Alba AR (2009) Degradation of caffeine and identification of the transformation products generated by ozonation. *Chemosphere* 74(6):825-831.

Rosales E, Buftia G, Pazos M, Lazar G, Sanromán MA (2018) Highly active based iron-carbonaceous cathodes for heterogeneous electro-Fenton process: Application to degradation of parabens. *Process Safety and Environmental Protection* 117:363-371.

Rui J, Zhang H, Zhang D, Han F, Guo Q (2019) Total organic carbon content prediction based on support-vector-regression machine with particle swarm optimization. *Journal of Petroleum Science and Engineering* 180:699-706.

Ryan C (2003) Evolutionary Algorithms and Metaheuristics. *Encyclopedia of Physical Science and Technology (Third Edition)*, Meyers RA (Édit.) Academic Press, New York. p 673-685.

Sadrzadeh M, Mohammadi T, Ivakpour J, Kasiri N (2007a) Modeling of electrodialysis using neural network. in *Proceedings of the 1st WSEAS International Conference on Computational Chemistry* (World Scientific and Engineering Academy and Society (WSEAS), Cairo, Egypt), p 83-91.

Sadrzadeh M, Mohammadi T, Ivakpour J, Kasiri N (2008) Separation of lead ions from wastewater using electrodialysis: Comparing mathematical and neural network modeling. *Chemical Engineering Journal* 144(3):431-441.

Sadrzadeh M, Mohammadi T, Ivakpour J, Kasiri N (2009) Neural network modeling of Pb2+ removal from wastewater using electrodialysis. *Chemical Engineering and Processing: Process Intensification* 48(8):1371-1381.

Sadrzadeh M, Razmi A, Mohammadi T (2007b) Separation of different ions from wastewater at various operating conditions using electrodialysis. *Separation and Purification Technology* 54(2):147-156.

Sahoo PK, Pattnaik S, Sutar MK (2020) Parametric optimization of permeability of green sand mould using ANN and ANFIS methods. *Advances in Materials and Manufacturing Engineering.* Springer Singapore, p 495-501.

Sangal V, Kaur P, Kushwaha J (2015) Modeling and evaluation of electro-oxidation of dye wastewater using artificial neural networks. *RSC Advances* 5:34663-34771.

Saradhi VV, Karnick H (2007) Classifier complexity reduction by support vector pruning in kernel matrix learning. *Computational and Ambient Intelligence.* (Berlin, Heidelberg, Springer, p 268-275.

Särkkä H, Bhatnagar A, Sillanpää M (2015) Recent developments of electro-oxidation in water treatment — A review. *Journal of Electroanalytical Chemistry* 754:46-56.

Schaffer JD, Whitley D, Eshelman LJ (1992) Combinations of genetic algorithms and neural networks: a survey of the state of the art. *COGANN-92: International Workshop on Combinations of Genetic Algorithms and Neural Networks.* (Baltimore, MD, USA, 6 Jun.), p 1-37.

Senghor F, Drogui P, Seyhi B (2015) A combined electrocoagulation-electroperoxidation process for the tertiary treatment of domestic wastewaters. *Water, Air, & Soil Pollution* 226(11):373-383.

Severeyn E, Velásquez J, Herrera H, Wong S (2019) Random sub-sampling cross validation for empirical correlation between heart rate variability, biochemical and anthropometrics parameters. *Information and Communication Technologies of Ecuador (TIC.EC).* (Cuenca, Ecuador, November 27-29) Springer International Publishing, p 347-357.

Shahi VK, Thampy SK, Siddhanta AK, Rangarajan R (2002) Separation of sodium sulfate and p-toluene sulfonic acid by electrodialysis. *Separation Science and Technology* 37(14):3273-3289.

Shi Y, Eberhart RC (1999) Empirical study of particle swarm optimization. *1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406).* (Washington, DC, USA, 6-9 July), p 1945-1950 Vol. 1943.

Shirkoohi MG, Doghri M, Duchesne S (2021) Short-term water demand predictions coupling an artificial neural network model and a genetic algorithm. *Water Supply* 21(5):2374-2386.

Singh KP, Basant A, Malik A, Jain G (2009) Artificial neural network modeling of the river water quality—A case study. *Ecological Modelling* 220(6):888-895.

Singh KP, Basant N, Malik A, Jain G (2010) Modeling the performance of "up-flow anaerobic sludge blanket" reactor based wastewater treatment plant using linear and nonlinear approaches—A case study. *Analytica Chimica Acta* 658(1):1-11.

Singh V, Pencina M, Einstein AJ, Liang JX, Berman DS, Slomka P (2021) Impact of train/test sample regimen on performance estimate stability of machine learning in cardiovascular imaging. *Scientific Reports* 11(1):14490.

Sinha T, Sikka K, Lall R (2021) Artificial neural networks and bayesian techniques for flip-chip package thermo-mechanical analysis. *IEEE 71st Electronic Components and Technology Conference (ECTC).*1 June-4 July), p 1442-1449.

Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Statistics and Computing* 14(3):199-222.

Soloman P, Basha C, Manickam V, Subramanian B (2010) Electro oxidation of Malachite Green and Modeling Using ANN. *Chemical and Biochemical Engineering Quarterly* 24.

Sorlini S, Collivignarelli C (2005) Chlorite removal with granular activated carbon. *Desalination* 176(1):255-265.

Souza PR, Dotto GL, Salau NPG (2018) Artificial neural network (ANN) and adaptive neuro-fuzzy interference system (ANFIS) modelling for nickel adsorption onto agro-wastes and commercial activated carbon. *Journal of Environmental Chemical Engineering* 6(6):7152-7160.

Sriamornsak P, Kennedy RA (2007) Effect of drug solubility on release behavior of calcium polysaccharide gel-coated pellets. *European Journal of Pharmaceutical Sciences* 32(3):231-239.

Stadler RH, Richoz J, Turesky RJ, Welti DH, Fay LB (1996) Oxidation of caffeine and related methylxanthines in ascorbate and polyphenol-driven Fenton-type oxidations. *Free Radical Research* 24(3):225-240.

Starling MCVM, Souza PP, Le Person A, Amorim CC, Criquet J (2019) Intensification of UV-C treatment to remove emerging contaminants by UV-C/H2O2 and UV-C/S2O82−: Susceptibility to photolysis and investigation of acute toxicity. *Chemical Engineering Journal* 376:120856.

Suidan MT, Snoeyink VL, Schmitz RA (1977) Reduction of aqueous free chlorine with granular activated carbon - pH and temperature effects. *Environmental Science & Technology* 11(8):785-789.

Taheri M, Alavi Moghaddam MR, Arami M (2013) Techno-economical optimization of Reactive Blue 19 removal by combined electrocoagulation/coagulation process through MOPSO using RSM and ANFIS models. *Journal of Environmental Management* 128:798-806.

Taheri M, Moghaddam MRA, Arami M (2015) Improvement of the /Taguchi/ design optimization using artificial intelligence in three acid azo dyes removal by electrocoagulation. *Environmental Progress & Sustainable Energy* 34(6):1568-1575.

Talebi A, Molaei MA, Sheikh MJ (2010) Performance investigation and comparison of two evolutionary algorithms in portfolio optimization: Genetic and particle swarm optimization. *IEEE International Conference on Information and Financial Engineering.*17-19 Sept.) Chongqing, China, p 430-437.

Tian Y, Shi Y, Liu X (2012) Recent advances on support vector machines research. *Technological and Economic Development of Economy* 18(1):5-33.

Tiwari B, Sellamuthu B, Ouarda Y, Drogui P, Tyagi RD, Buelna G (2017) Review on fate and mechanism of removal of pharmaceutical pollutants from wastewater using biological approach. *Bioresource Technology* 224:1-12.

Torres-Sospedra J, Hernández-Espinosa C, Fernández-Redondo M (2006) Combining MF Networks: A Comparison Among Statistical Methods and Stacked Generalization. *Artificial Neural Networks in Pattern Recognition.* (Berlin, Heidelberg, 31 Aug.- 2 Sept.) Springer, p 210-220.

Tran L-H, Drogui P, Mercier G, Blais J-F (2010) Comparison between Fenton oxidation process and electrochemical oxidation for PAH removal from an amphoteric surfactant solution. *Journal of Applied Electrochemistry* 40(8):1493-1510.

Tran N, Drogui P (2013) Electrochemical removal of microcystin-LR from aqueous solution in the presence of natural organic pollutants. *J Environ Manage* 114:253-260.

Tran N, Drogui P, Blais J-F, Mercier G (2012) Phosphorus removal from spiked municipal wastewater using either electrochemical coagulation or chemical coagulation as tertiary treatment. *Separation and Purification Technology* 95:16-25.

Tran N, Drogui P, Brar SK (2015a) Sonoelectrochemical oxidation of carbamazepine in waters: optimization using response surface methodology. *Journal of Chemical Technology and Biotechnology* 90(5):921-929.

Tran N, Drogui P, Nguyen L, Brar SK (2015b) Optimization of sono-electrochemical oxidation of ibuprofen in wastewater. *Journal of Environmental Chemical Engineering* 3(4):2637-2646.

Tran NH, Li J, Hu J, Ong SL (2014) Occurrence and suitability of pharmaceuticals and personal care products as molecular markers for raw wastewater contamination in surface water and groundwater. *Environmental Science and Pollution Research* 21(6):4727-4740.

Trovó AG, Silva TFS, Gomes O, Machado AEH, Neto WB, Muller PS, Daniel D (2013) Degradation of caffeine by photo-Fenton process: Optimization of treatment conditions using experimental design. *Chemosphere* 90(2):170-175.

USEPA (2002) Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms. (EPA-821-R-02-014, Washington DC, USA).

Valencia F, Arcos H, Quilumba F (2021) Prediction of stress in power transformer winding conductors using artificial neural networks: Hyperparameter analysis. *Energies* 14(14):4242.

Valente GFS, Mendonça RCS, Pereira JAM, Felix LB (2014) Artificial neural network prediction of chemical oxygen demand in dairy industry effluent treated by electrocoagulation. *Separation and Purification Technology* 132:627-633.

Van Messem A, Christmann A (2010) A review on consistency and robustness properties of support vector machines for heavy-tailed distributions. *Advances in Data Analysis and Classification* 4(2):199-220.

Vapnik V (1963) Pattern recognition using generalized portrait method. *Automation and Remote Control* 24:774-780.

Viana DF, Salazar-Banda GR, Leite MS (2018) Electrochemical degradation of Reactive Black 5 with surface response and artificial neural networks optimization models. *Separation Science and Technology* 53(16):2647-2661.
Wang D, Tan D, Liu L (2018) Particle swarm optimization algorithm: an overview. *Soft Computing* 22(2):387-408.
Wang Z, Wiebe S, Shang H (2016) Fault detection and diagnosis of an industrial copper electrowinning process. *The Canadian Journal of Chemical Engineering* 94(3):415-423.
Whitley D (1994) A genetic algorithm tutorial. *Statistics and Computing* 4(2):65-85.
Wikoff D, Welsh BT, Henderson R, Brorby GP, Britt J, Myers E, Goldberger J, Lieberman HR, O'Brien C, Peck J, Tenenbein M, Weaver C, Harvey S, Urban J, Doepker C (2017) Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children. *Food and Chemical Toxicology* 109:585-648.
Wolpert DH (1992) Stacked generalization. *Neural Networks* 5(2):241-259.
Wu D, Chen GQ, Hu B, Deng H (2019) Feasibility and energy consumption analysis of phenol removal from salty wastewater by electro-electrodialysis. *Separation and Purification Technology* 215:44-50.
Yang Y (2020) Recent advances in the electrochemical oxidation water treatment: Spotlight on byproduct control. *Frontiers of Environmental Science & Engineering* 14(5):85.
Yap CL, Gan S, Ng HK (2011) Fenton based remediation of polycyclic aromatic hydrocarbons-contaminated soils. *Chemosphere* 83(11):1414-1430.
Yu R-F, Chen H-W, Cheng W-P, Hsieh P-H (2009) Dosage Control of the Fenton Process for Color Removal of Textile Wastewater Applying ORP Monitoring and Artificial Neural Networks. *Journal of Environmental Engineering* 135(5):325-332.
Yu R-F, Chen H-W, Liu K-Y, Cheng W-P, Hsieh P-H (2010) Control of the Fenton process for textile wastewater treatment using artificial neural networks. *Journal of Chemical Technology and Biotechnology* 85(2):267-278.
Yu R-F, Lin C-H, Chen H-W, Cheng W-P, Kao M-C (2013) Possible control approaches of the Electro-Fenton process for textile wastewater treatment using on-line monitoring of DO and ORP. *Chemical Engineering Journal* 218:341-349.
Yuan S, Xiao M, Zheng G, Tian M, Lu X (2006) Quantitative structure-property relationship studies on electrochemical degradation of substituted phenols using a support vector machine. *SAR and QSAR in Environmental Research* 17(5):473-481.
Zarei M, Khataee AR, Ordikhani-Seyedlar R, Fathinia M (2010) Photoelectro-Fenton combined with photocatalytic process for degradation of an azo dye using supported TiO2 nanoparticles and carbon nanotube cathode: Neural network modeling. *Electrochimica Acta* 55(24):7259-7265.
Zarra T, Galang MG, Ballesteros F, Belgiorno V, Naddeo V (2019) Environmental odour management by artificial neural network – A review. *Environment International* 133:105189.
Zaviska F, Drogui P, Blais J-F, Mercier G (2009) In situ active chlorine generation for the treatment of dye-containing effluents. *Journal of Applied Electrochemistry* 39(12):2397-2408.
Zaviska F, Drogui P, Blais J-F, Mercier G (2012a) Electrochemical treatment of bisphenol-A using response surface methodology. *Journal of Applied Electrochemistry* 42(2):95-109.
Zaviska F, Drogui P, Blais J-F, Mercier G, d'Auzay SDLR (2013) Electrochemical Oxidation of Chlortetracycline Using Ti/IrO2 and Ti/PbO2 Anode Electrodes: Application of Experimental Design Methodology. *Journal of Environmental Engineering* 139(6):810-821.
Zaviska F, Drogui P, Pablo G (2012b) Statistical optimization of active chlorine production from a synthetic saline effluent by electrolysis. *Desalination* 296:16-23.
Zendehboudi S, Rezaei N, Lohi A (2018) Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review. *Applied Energy* 228:2539-2566.

Zeng Y, Fu Xe, Ren Z (2012) The effects of residual chlorine on the behavioural responses of Daphnia magna in the early warning of drinking water accidental events. *Procedia Environmental Sciences* 13:71-79.
Zhang C, Guisasola A, Baeza JA (2022) A review on the integration of mainstream P-recovery strategies with enhanced biological phosphorus removal. *Water Research* 212:118102.
Zhang J, Teng Y-F, Chen W (2019) Support vector regression with modified firefly algorithm for stock price forecasting. *Applied Intelligence* 49(5):1658-1674.
Zhang LP, Yu HJ, Hu SX (2005) Optimal choice of parameters for particle swarm optimization. *Journal of Zhejiang University: Science* 6 A(6):528-534.
Zhang Y, Ghyselbrecht K, Meesschaert B, Pinoy L, Van der Bruggen B (2011) Electrodialysis on RO concentrate to improve water recovery in wastewater reclamation. *Journal of Membrane Science* 378(1):101-110.
Zhang Z, Friedrich K (2003) Artificial neural networks applied to polymer composites: a review. *Composites Science and Technology* 63(14):2029-2044.
Zheng T, Wang J, Wang Q, Meng H, Wang L (2017) Research trends in electrochemical technology for water and wastewater treatment. *Applied Water Science* 7(1):13-30.