UQAC
Université du Québec
à Chicoutimi

Approches d'apprentissage automatique pour prédire le temps d'endurance des fluides antigivrage pour aéronefs à partir de données de neige naturelle, artificielle et synthétiques

par
Andres Cardona Paniagua

Sous la direction de la professeure Gelareh Momen
Sous la codirection de la professeure Haïfa Nakouri et du professeur Jean-Denis Brassard

Mémoire présenté à l'Université du Québec à Chicoutimi en vue de l'obtention du grade de Maître ès sciences appliquées (M. Sc. A.) en ingénierie

Jury d'évaluation :
Professeur Hamdi Ben Abdessalem, Département d'informatique et de mathématique de l'UQAC, membre interne
Professeur Elyes Manai, Département d'informatique et de mathématique de l'UQAC, membre interne
Professeure Gelareh Momen, Département des sciences appliquées de l'UQAC, membre interne
Professeur Jean-Denis Brassard, Département des sciences appliquées de l'UQAC, membre interne
Professeure Haïfa Nakouri, Département d'informatique et de mathématique de l'UQAC, membre interne

Québec, Canada
© Andres Cardona Paniagua, 2026

Résumé

Les opérations aériennes hivernales représentent un défi important pour l'aviation civile en raison de l'accumulation de neige et de glace sur les surfaces des aéronefs. Afin d'assurer la sécurité des opérations, des fluides antigivrage sont appliqués sur les ailes et les surfaces critiques pour prévenir la formation de glace et maintenir les performances aérodynamiques de l'appareil. Toutefois, l'efficacité de ces fluides est limitée dans le temps et dépend fortement des conditions météorologiques, notamment du type et de l'intensité des précipitations hivernales. L'évaluation expérimentale du temps d'endurance de ces fluides demeure coûteuse et dépend largement des conditions environnementales naturelles, ce qui rend les essais difficiles à reproduire et limite la disponibilité des données expérimentales.

Dans ce contexte, cette recherche vise à développer une approche basée sur l'apprentissage automatique afin de prédire la performance des fluides antigivrage appliqués sur les surfaces d'aéronefs. L'étude utilise des ensembles de données expérimentales obtenus à partir d'essais réalisés sous neige naturelle et neige artificielle, incluant plusieurs variables environnementales et opérationnelles influençant le comportement des fluides. Différents modèles d'apprentissage automatique ont été étudiés afin d'identifier les approches les plus performantes pour la prédiction du temps d'endurance des fluides. Parmi ces modèles, XGBoost s'est distingué comme l'approche la plus performante, offrant les meilleurs résultats en termes de précision et de capacité de généralisation. Les modèles ont été optimisés à l'aide de techniques d'optimisation avancées et soumis à une stratégie de validation croisée afin d'en assurer la stabilité des performances, d'attester de leur généralisation et de prévenir le surapprentissage.

Par ailleurs, afin de pallier la limitation du volume de données expérimentales disponibles, des techniques de génération de données synthétiques ont été explorées à l'aide de réseaux antagonistes génératifs (GAN). Plusieurs architectures ont été étudiées afin d'améliorer la stabilité de l'entraînement et la qualité statistique des données générées. Les ensembles de données synthétiques ont ensuite été évalués à l'aide de différentes métriques statistiques et de leur utilité prédictive dans des modèles d'apprentissage automatique.

Dans les ensembles de données étudiés, les résultats montrent que les modèles d'apprentissage automatique peuvent prédire le temps d'endurance des fluides antigivrage à partir de variables météorologiques et opérationnelles. Les données synthétiques générées par GAN permettent d'examiner la fidélité statistique, le transfert prédictif et l'augmentation des données; leur bénéfice dépend toutefois de la configuration et du protocole d'évaluation.

Cette étude établit la faisabilité de l'approche pour les données et les conditions examinées. Sa généralisation à d'autres fluides, campagnes, sites et conditions opérationnelles demeure à confirmer par une validation externe avant toute utilisation comme outil d'aide à la décision.

Mots clés : Aviation hivernale, fluides antigivrage, apprentissage automatique, XGBoost, génération de données synthétiques, réseaux antagonistes génératifs (GAN), neige naturelle et artificielle, sécurité aérienne.

TABLE DES MATIÈRES

Résumé i
Liste des figures vi
Liste des tableaux viii
Liste des symboles et abréviations ix
Remerciements x

1 Introduction **1**
1.1 Problématique 1
1.2 Objectifs 3
1.3 Méthodologie 4
1.4 Originalité de la recherche 5
1.5 Structure du mémoire 7

2 Revue de littérature **9**
2.1 Introduction 9
2.2 Givrage des aéronefs et fluides antigivrage 10
2.3 Différences entre essais en laboratoire et conditions réelles 12
2.4 Apprentissage automatique pour la prédiction des phénomènes de givrage 16
2.5 Génération de données synthétiques pour les données environnementales 20
2.6 Évaluation des données synthétiques 23
2.6.1 Similarité des distributions statistiques 24
2.6.2 Préservation des dépendances entre variables 24
2.6.3 Utilité prédictive des données synthétiques 25
2.7 Lacunes dans la littérature 26
Note de l'auteur concernant le chapitre publié 26

3 Prédiction du temps d'endurance des fluides antigivrage pour aéronefs par apprentissage automatique : étude comparative entre données de neige naturelle et artificielle **28**
3.1 Résumé 29
3.2 Abstract 29
3.3 Introduction 30
3.4 Methodology 33
3.4.1 Data Collection 33
3.4.2 Computational Methods 38
3.5 Results 44
3.5.1 Data Analysis 44

3.5.2 Prediction Performance 49
3.5.3 Model Explainability using SHAP Analysis 55
3.6 Discussion 58
3.7 Conclusion 60
3.8 Remerciements 61

4 Génération de données synthétiques pour améliorer la modélisation prédictive du temps d'endurance des fluides antigivrage en conditions hivernales 62
4.1 Résumé 62
4.2 Abstract 63
4.3 Introduction 64
4.4 Literature Review 68
4.4.1 GANs for generating meteorological and climate data 68
4.4.2 Data augmentation using generative adversarial networks (GANs) for meteorology-related engineering tasks 69
4.4.3 Protocols for the generation and evaluation of synthetic tabular data 70
4.4.4 Gaps in research and the relevance of data on anti-icing endurance 71
4.5 Methodology 71
4.5.1 Description of the dataset 72
4.5.2 Data preparation and preprocessing 73
4.5.3 Architecture of Generative Models 75
4.5.4 Training process 79
4.5.5 Generation of synthetic data 82
4.6 Implementation and evaluation of the experiment 84
4.6.1 Experimental design 84
4.6.2 Quantitative statistical analysis 85
4.6.3 Multivariate Dependency Evaluation 86
4.6.4 Predictive Utility Evaluation (TSTR) 86
4.7 Results 87
4.7.1 Training Dynamics 87
4.7.2 Initial GAN Baseline Performance 89
4.7.3 Distribution Fidelity 90
4.7.4 Multivariate Dependency Preservation 92
4.7.5 Predictive Utility 93
4.7.6 Model Comparison 96
4.8 Discussion 99
4.8.1 Impact of Batch Normalization on Adversarial Stability 99
4.8.2 Limitations of Dropout in Small Tabular GANs 100
4.8.3 Extended Training Duration: Objective-Dependent Results . . . 100
4.8.4 Comparison Between GAN and VAE Generative Models 101
4.8.5 Fidelity vs. Predictive Utility 103
4.8.6 Implications for Aviation Engineering Datasets 104
4.9 Conclusion 104

5 Conclusion générale et recommandations **107**
5.1 Conclusion générale 107
5.2 Contributions scientifiques 109
5.3 Limites 110
5.4 Perspectives et recommandations 110

LISTE DES FIGURES

2.1 Formation de glace sur les surfaces critiques d'un aéronef et impact sur l'écoulement aérodynamique. Source : Organisation de l'Aviation Civile Internationale (OACI), *Manual of Aircraft Ground De-icing/Anti-icing Operations*, Doc 9640, 3e édition, 2018. 10

2.2 Exemples d'échec du temps d'endurance des fluides antigivrage observés lors d'essais sous précipitations solides. Les images thermiques (à gauche) illustrent l'évolution de la température de surface au cours du temps, tandis que les images visibles (à droite) montrent l'accumulation progressive de neige sur la surface testée. L'échec du fluide est identifié lorsque l'accumulation de neige dépasse un seuil critique sur la surface. Source : Laboratoire International des Matériaux Antigivre (LIMA), Université du Québec à Chicoutimi. 14

2.3 Cadre conceptuel pour la sélection des modèles d'apprentissage automatique dans la prédiction du givrage en aviation. Figure élaborée par l'auteur à partir de [2, 16, 20, 52, 63]. 20

2.4 Schéma simplifié de l'architecture d'un Generative Adversarial Network (GAN). Le générateur produit des données synthétiques à partir d'un vecteur de bruit aléatoire, tandis que le discriminateur apprend à distinguer les données réelles des données générées. L'entraînement adversarial entre ces deux réseaux permet d'améliorer progressivement le réalisme des données synthétiques. 21

3.1 Data Processing and Optimization Pipeline for Snow Classification and Fluid Endurance Prediction 34

3.2 Support Structure for Test Plate and Snow Deposit 34

3.3 Setup for Natural Snow Test 35

3.4 Setup for Artificial Snow Test 37

3.5 Visual Fluid Failure, According to SAE ARP5485B Criteria 38

3.6 Distribution of the Variables 45

3.7 Boxplot Endurance time by fluids and snow type 46

3.8 Endurance time vs. snow deposition rate by fluids from temperatures 0 to -20°C 47

3.9 Matrix of Correlations: Temperature, Snow Rate, Endurance 48

3.10 SHAP summary plot for the tuned XGBoost model trained on combined natural and artificial snow data. 56

3.11 SHAP Waterfall Plots Illustrating Feature Contributions for Representative Low, Median, and High Endurance Time Predictions. 57

4.1 Overview of the synthetic data generation and evaluation framework used in this study. The workflow begins with the experimental dataset, followed by generative model training (GAN/VAE) and synthetic data generation 72
4.2 Training loss evolution for discriminator (top row) and generator (bottom row) across the three evaluated GAN configurations: Dropout, Batch Normalization, and extended 500-epoch training. 88
4.3 Correlation matrices of synthetic datasets generated using different stabilization strategies 94

LISTE DES TABLEAUX

2.1 Synthèse critique des principales approches d'apprentissage automatique pertinentes pour le givrage aéronautique. Les mentions « non précisée ici » signalent qu'une valeur comparable n'est pas disponible dans la synthèse consultée et évitent d'inférer une taille ou une performance non documentée. 18

3.1 Dataset Characteristics and Testing Conditions 39
3.2 Descriptive statistics (minimum, maximum, mean, and standard deviation) for the input and engineered features used in model development. 43
3.3 Hyperparameter configurations and training details for each ML model. XGBoost parameters were optimized using Bayesian and genetic algorithms. 50
3.4 Model Comparison — Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) for Decision Tree (DT), Random Forest (RF), and XGBoost (XGBoost) 50
3.5 Model Comparison with Feature Engineering — for Decision Tree (DT), Random Forest (RF), and XGBoost (XGBoost) 52
3.6 Performance Metrics for Optimization Techniques: R^2, RMSE, RMSE (%), MSE, MAE, MAE (%) 53

4.2 Initial predictive performance obtained using the baseline GAN synthetic data generation pipeline. 90
4.3 Corrected statistical-fidelity metrics for the evaluated GAN stabilization strategies. Dropout and BatchNorm were trained for 200 epochs, while the extended configuration was trained for 500 epochs. Mean WD denotes the mean Wasserstein distance; Corr. F and MI F denote the Frobenius norms of the differences between the real and synthetic correlation and mutual-information matrices, respectively. Lower values indicate stronger alignment with the real dataset. 90
4.4 Predictive-utility evaluation using the Train-on-Synthetic, Test-on-Real (TSTR) protocol. All models were evaluated on the same reserved natural-snow test set ($n = 34$). Bold values in the synthetic-data rows identify the strongest TSTR result. The Real-to-Real model is included as a reference baseline. Higher R^2 and lower MAE, RMSE, MSE, and MAPE indicate stronger predictive transfer. 95

4.5 Predictive performance of XGBoost models trained on baseline, VAE-augmented, and GAN-augmented datasets. Every model was evaluated on the same reserved natural-snow test set ($n = 34$). RMSE and MAE are reported in minutes, and their normalized percentages use the common test-set mean $\overline{y}_{\text{test}} = 64.20$ minutes. 97

Liste des symboles et abréviations

Symbol	Definition
T	Ambient air temperature (°C)
R	Snowfall deposition rate (g/dm^2/h)
y	Observed endurance time (minutes)
$\hat{y}$	Predicted endurance time (minutes)
T^2	Quadratic transformation of temperature
r^2	Quadratic transformation of snow rate
$T \times R$	Interaction term between temperature and snow rate
$\|T\|$	Absolute temperature (°C)
MAPE	Mean Absolute Percentage Error (%)
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
R^2	Coefficient of Determination
SHAP	SHapley Additive exPlanations (model interpretability method)
VAE	Variational Autoencoder (generative model)
GAN	Generative Adversarial Network (generative model)
CGAN	Conditional GAN
CV	Cross-Validation
HOT	Holdover Time
SAE ARP5485B	Standard procedure for endurance time evaluation
AMS1424	Standard for Type I deicing fluids
AMS1428	Standard for Type II, III, and IV anti-icing fluids
$E[f(x)]$	Expected model output for a given sample
$f(x)$	Specific model prediction value for a given instance
ELBO	Evidence Lower Bound (VAE loss function objective)

Remerciements

Ce travail a été réalisé dans le cadre de mes études de maîtrise en ingénierie à l'Université du Québec à Chicoutimi, au sein du Laboratoire International des Matériaux Antigivre (LIMA), en collaboration avec Transport Canada.

Tout d'abord, je tiens à exprimer mes plus sincères remerciements à ma directrice de recherche, la professeure Gelareh Momen, ainsi qu'à mes codirecteurs, la professeure Haïfa Nakouri et le professeur Jean-Denis Brassard, pour leur encadrement, leur disponibilité et leurs précieux conseils tout au long de ce projet. Leur expertise scientifique, leur rigueur et leur engagement ont été des sources constantes d'inspiration et ont grandement contribué à la qualité de ce travail. Je leur suis profondément reconnaissant pour la confiance qu'ils m'ont accordée ainsi que pour l'environnement de recherche stimulant qu'ils ont su créer.

Je souhaite également remercier chaleureusement Transport Canada pour son soutien à ce projet ainsi que pour sa contribution au développement de solutions innovantes dans le domaine de la sécurité aéronautique hivernale. Mes remerciements s'adressent également à toute l'équipe du Laboratoire International des Matériaux Antigivre (AMIL), pour leur aide technique, leur expertise et leur collaboration lors des campagnes expérimentales. Leur professionnalisme et leur esprit d'équipe ont été essentiels à la réalisation de cette recherche.

Je tiens à exprimer ma profonde gratitude envers ma famille pour leur soutien inconditionnel tout au long de mon parcours académique. À mes parents, pour leurs sacrifices, leurs encouragements constants et leur confiance en moi. À ma sœur et à ma grand-mère, pour leur présence, leur affection et leur soutien moral, qui m'ont permis de persévérer dans les moments les plus exigeants.

Enfin, je souhaite adresser un remerciement tout particulier à mon partenaire pour son soutien, sa patience et sa compréhension tout au long de cette étape importante de ma vie. Sa présence, ses encouragements et sa confiance ont été une source de motivation et de force au quotidien.

Pour conclure, je remercie toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce projet, et qui ont rendu cette expérience aussi enrichissante sur le plan académique que personnel.

CHAPITRE 1

INTRODUCTION

1.1 Problématique

Les opérations aériennes en conditions hivernales représentent un défi majeur pour l'industrie aéronautique en raison de l'accumulation potentielle de neige, de givre ou de pluie verglaçante sur les surfaces critiques des aéronefs. La présence de contaminants sur les ailes ou les surfaces de contrôle peut modifier l'écoulement de l'air, augmenter la traînée et réduire la portance. Même de faibles quantités de glace ou de neige peuvent entraîner une dégradation significative des performances aérodynamiques et compromettre la sécurité des vols [14, 28].

Afin de réduire ces risques, des procédures de dégivrage et d'antigivrage sont appliquées avant le décollage. Ces opérations consistent à éliminer les contaminants présents sur les surfaces de l'aéronef et à appliquer des fluides spécialisés capables de retarder l'accumulation de nouvelles précipitations. Ces fluides sont largement utilisés dans l'aviation civile et sont essentiels pour garantir la sécurité des opérations dans les régions où les conditions hivernales sont fréquentes [3].

La performance expérimentale de ces fluides est évaluée à l'aide du temps d'endurance (*Endurance Time*), soit le résultat mesuré jusqu'au critère de défaillance dans des conditions d'essai définies. Le *Holdover Time* (HOT) désigne plutôt une durée opérationnelle indicative établie à partir d'un ensemble de données et de règles applicables; il ne doit donc pas être confondu avec une mesure individuelle du temps

d'endurance. Les essais sont réalisés selon des protocoles normalisés, notamment ceux définis par la norme SAE ARP5485, afin d'évaluer la dégradation progressive du fluide sous différentes intensités de précipitations [1].

Malgré l'existence de ces protocoles standardisés, l'évaluation expérimentale du temps d'endurance présente plusieurs limitations. Les essais réalisés sous neige naturelle dépendent fortement des conditions météorologiques et sont difficiles à reproduire. La variabilité naturelle des précipitations, notamment la taille des particules, la densité de la neige et les conditions thermiques, peut entraîner une dispersion importante des résultats expérimentaux [59].

Les essais réalisés sous neige artificielle permettent de mieux contrôler certains paramètres expérimentaux et d'obtenir des conditions plus reproductibles. Toutefois, ces essais ne reproduisent pas toujours fidèlement la complexité des précipitations naturelles et peuvent présenter des différences dans la structure ou les propriétés physiques de la neige générée en laboratoire [61].

Un autre défi important concerne la disponibilité limitée des ensembles de données expérimentales. Les essais d'endurance nécessitent des installations spécialisées et des conditions météorologiques spécifiques, ce qui limite souvent le volume de données disponibles pour l'analyse scientifique. Cette contrainte peut rendre difficile l'application de méthodes avancées d'analyse de données et de modélisation prédictive.

Dans ce contexte, les techniques d'apprentissage automatique offrent des perspectives intéressantes pour analyser ces ensembles de données expérimentales et identifier les relations complexes entre les variables environnementales et les performances des fluides antigivrage. Les modèles d'apprentissage automatique sont capables de capturer des relations non linéaires entre plusieurs variables et ont été appliqués avec succès dans différents domaines de l'aéronautique et des sciences atmosphériques [32, 52].

Par ailleurs, des approches récentes explorent l'utilisation de techniques de génération de données synthétiques afin d'augmenter les ensembles de données disponibles pour

l'entraînement des modèles prédictifs. Les réseaux antagonistes génératifs (*Generative Adversarial Networks*, GAN) constituent une approche prometteuse pour générer de nouvelles observations réalistes tout en conservant les propriétés statistiques des données originales [33]. Ces techniques peuvent contribuer à améliorer la robustesse des modèles prédictifs et à réduire la dépendance aux essais expérimentaux coûteux réalisés sous conditions hivernales réelles.

1.2 Objectifs

L'objectif principal de cette recherche est de développer et d'évaluer des modèles d'apprentissage automatique capables de prédire le temps d'endurance des fluides anti-givrage appliqués sur les surfaces d'aéronefs dans différentes conditions de précipitations hivernales.

Cet objectif général est décliné en trois objectifs principaux. Les six objectifs initiaux sont conservés comme sous-objectifs afin d'en faciliter le suivi :

1. **Constituer, caractériser et interpréter les données expérimentales.**
    - Analyser et caractériser les données issues d'essais d'endurance sous neige naturelle et artificielle.
    - Identifier les variables environnementales déterminantes et analyser leur influence sur la dégradation des fluides.
2. **Développer, comparer et interpréter les modèles prédictifs.**
    - Développer et comparer plusieurs modèles d'apprentissage automatique pour prédire le temps d'endurance.
    - Évaluer leurs performances à l'aide de métriques statistiques appropriées et interpréter la contribution des variables.

3. **Évaluer la génération et l'augmentation de données synthétiques.**

   - Explorer des GAN pour générer des observations synthétiques à partir des données expérimentales disponibles.
   - Distinguer la fidélité statistique, le transfert synthétique-vers-réel et l'effet de l'augmentation sur les modèles prédictifs.

1.3 Méthodologie

La méthodologie adoptée dans cette recherche repose sur l'intégration de données expérimentales issues d'essais d'endurance de fluides antigivrage et l'utilisation de techniques d'apprentissage automatique afin de développer des modèles prédictifs capables d'estimer la performance de ces fluides dans différentes conditions hivernales.

Les données utilisées dans cette étude proviennent d'essais réalisés conformément au protocole SAE ARP5485, largement utilisé dans l'industrie aéronautique pour évaluer la performance des fluides de dégivrage et d'antigivrage appliqués sur les surfaces des aéronefs [1]. Ces essais permettent de mesurer le temps d'endurance des fluides en exposant une surface recouverte de fluide à différentes intensités de précipitations hivernales. Les données expérimentales incluent des observations obtenues sous neige naturelle ainsi que sous neige artificielle, permettant ainsi de capturer différentes conditions environnementales représentatives des opérations aériennes hivernales.

Dans un premier temps, une phase de prétraitement et d'analyse exploratoire des données est réalisée afin d'identifier les caractéristiques principales des ensembles de données et de mieux comprendre les relations entre les variables environnementales et la performance des fluides antigivrage. Cette étape inclut l'analyse statistique des variables, la gestion des données manquantes ainsi que la normalisation des variables utilisées pour l'entraînement des modèles.

Par la suite, plusieurs modèles d'apprentissage automatique sont développés afin de

prédire le temps d'endurance des fluides antigivrage à partir des variables environnementales disponibles. Les techniques d'apprentissage automatique sont particulièrement adaptées à ce type de problématique puisqu'elles permettent de capturer des relations complexes et potentiellement non linéaires entre plusieurs variables d'entrée et une variable cible [32, 52].

En raison du volume limité de données expérimentales disponibles, cette recherche explore également l'utilisation de techniques de génération de données synthétiques afin d'augmenter les ensembles de données utilisés pour l'entraînement des modèles prédictifs. Plus précisément, des réseaux antagonistes génératifs (*Generative Adversarial Networks*, GAN) sont utilisés pour générer de nouvelles observations synthétiques présentant des caractéristiques statistiques similaires aux données expérimentales originales [33].

Enfin, l'impact de ces données synthétiques sur la performance des modèles d'apprentissage automatique est évalué en comparant les résultats obtenus avec différentes configurations de données d'entraînement. Cette approche permet d'analyser le potentiel des techniques de génération de données pour améliorer la robustesse des modèles prédictifs dans le contexte de l'évaluation des fluides antigivrage en conditions hivernales.

1.4 Originalité de la recherche

L'originalité de cette recherche réside dans l'intégration de techniques d'apprentissage automatique et de génération de données synthétiques pour améliorer la prédiction du temps d'endurance des fluides antigivrage utilisés dans les opérations aériennes hivernales.

Elle répond directement à trois lacunes relevées dans la littérature : le nombre limité d'études comparant, dans un même cadre, la neige naturelle et la neige artificielle; le manque de données naturelles disponibles pour une modélisation robuste; et l'évaluation

souvent fragmentée des données synthétiques. Les contributions correspondantes sont la comparaison contrôlée des sources de données, l'interprétation des facteurs de prédiction et un cadre séparant explicitement fidélité statistique, transfert TSTR et utilité de l'augmentation.

Traditionnellement, l'évaluation des performances des fluides de dégivrage et d'antigivrage repose principalement sur des essais expérimentaux réalisés selon des protocoles normalisés. Bien que ces essais soient essentiels pour la certification et l'évaluation des fluides, ils présentent certaines limitations liées au coût élevé des expérimentations, à la dépendance aux conditions météorologiques et au volume limité de données expérimentales disponibles.

La présente recherche propose une approche complémentaire basée sur l'analyse de données et l'apprentissage automatique afin de mieux exploiter les informations contenues dans les ensembles de données expérimentales existants. Contrairement aux approches traditionnelles reposant uniquement sur l'analyse expérimentale, cette étude explore l'utilisation de modèles prédictifs capables d'identifier les relations complexes entre les variables environnementales et la dégradation des fluides antigivrage.

Une autre contribution importante de ce travail réside dans l'utilisation conjointe de données obtenues sous neige naturelle et sous neige artificielle. Cette combinaison permet d'analyser la complémentarité de ces deux types d'essais et d'évaluer leur contribution dans le développement de modèles prédictifs pour les opérations aériennes hivernales.

Enfin, cette recherche explore l'utilisation de techniques de génération de données synthétiques basées sur des réseaux antagonistes génératifs (*Generative Adversarial Networks*, GAN). L'objectif est d'augmenter les ensembles de données expérimentales disponibles tout en préservant leurs caractéristiques statistiques principales. Cette approche permet d'étudier l'impact de l'augmentation des données sur la performance des modèles d'apprentissage automatique et d'évaluer le potentiel de ces techniques pour améliorer la robustesse des modèles prédictifs dans un contexte où les données

expérimentales sont limitées.

Ainsi, cette recherche contribue à l'avancement des méthodes d'analyse appliquées aux opérations aériennes en conditions hivernales en combinant des approches expérimentales et des techniques modernes d'intelligence artificielle. Elle ouvre également de nouvelles perspectives pour l'utilisation de méthodes basées sur les données dans l'évaluation et la prédiction des performances des fluides antigivrage.

1.5 Structure du mémoire

Ce projet de recherche est structuré en cinq chapitres, chacun organisé de manière à présenter progressivement le contexte, les méthodes et les contributions de cette étude. Cette section présente brièvement le contenu des chapitres qui composent ce mémoire.

Chapitre 2

Le deuxième chapitre présente la revue de la littérature qui constitue le fondement théorique de cette étude. Il aborde d'abord les défis liés aux opérations aériennes en conditions hivernales ainsi que l'impact de la contamination par la neige et la glace sur les performances des aéronefs. Ce chapitre présente également les différentes catégories de fluides de dégivrage et d'antigivrage utilisés dans l'industrie aéronautique ainsi que les protocoles expérimentaux utilisés pour évaluer leur performance.

Par ailleurs, ce chapitre introduit les concepts fondamentaux liés aux techniques d'apprentissage automatique et leur application dans l'analyse de données environnementales et aéronautiques. Il permet ainsi d'établir un cadre scientifique pour l'utilisation de ces méthodes dans la prédiction du temps d'endurance des fluides antigivrage.

Chapitre 3

Le troisième chapitre présente le premier article scientifique issu de ce projet de recherche. Cet article porte sur l'utilisation de modèles d'apprentissage automatique pour prédire le temps d'endurance des fluides antigivrage à partir de données expérimentales

obtenues sous neige naturelle et neige artificielle. Il décrit les ensembles de données utilisés, les techniques de prétraitement des données ainsi que les modèles de prédiction développés et évalués dans cette étude.

Chapitre 4

Le quatrième chapitre présente le deuxième article scientifique de ce mémoire. Il explore l'utilisation de techniques de génération de données synthétiques basées sur des réseaux antagonistes génératifs (GAN) afin d'augmenter les ensembles de données expérimentales disponibles. Ce chapitre analyse l'impact de ces données synthétiques sur la performance des modèles d'apprentissage automatique et discute des résultats obtenus.

Chapitre 5

Enfin, le cinquième chapitre présente les conclusions générales de cette recherche ainsi que les principales contributions scientifiques du projet. Il propose également des recommandations pour les travaux futurs visant à améliorer l'utilisation des techniques d'apprentissage automatique dans l'évaluation et la prédiction des performances des fluides antigivrage en conditions hivernales.

CHAPITRE 2

REVUE DE LITTÉRATURE

2.1 Introduction

Les opérations aériennes en conditions hivernales nécessitent l'utilisation de fluides de dégivrage et d'antigivrage afin de prévenir l'accumulation de neige et de glace sur les surfaces critiques des aéronefs. L'efficacité de ces fluides est généralement évaluée à l'aide du temps d'endurance, qui représente la durée pendant laquelle le fluide protège la surface contre l'accumulation de précipitations hivernales.

Traditionnellement, l'évaluation de ces performances repose sur des essais expérimentaux réalisés selon des protocoles normalisés. Toutefois, ces essais présentent plusieurs limitations, notamment la dépendance aux conditions météorologiques et la disponibilité limitée de données expérimentales. Dans ce contexte, les techniques d'apprentissage automatique et les méthodes de génération de données synthétiques offrent de nouvelles perspectives pour améliorer la modélisation et la prédiction de la performance des fluides antigivrage.

Ce chapitre présente d'abord les bases scientifiques liées au givrage des aéronefs et aux méthodes d'évaluation des fluides antigivrage. Il examine ensuite les approches d'apprentissage automatique utilisées pour la prédiction des phénomènes de givrage en aviation, avant de présenter les avancées récentes dans la génération de données synthétiques appliquées aux données environnementales et aux ensembles de données tabulaires.

2.2 Givrage des aéronefs et fluides antigivrage

Le givrage des aéronefs représente l'un des risques les plus critiques pour la sécurité aérienne lors des opérations hivernales. La présence de neige, de glace ou de givre sur les surfaces aérodynamiques d'un avion peut altérer significativement les performances aérodynamiques de l'appareil. Même de faibles quantités de contamination sur les ailes ou les surfaces de contrôle peuvent perturber l'écoulement de l'air, entraîner une diminution importante de la portance et provoquer une augmentation de la traînée aérodynamique [4, 14]. Des études ont montré que de fines couches de givre peuvent réduire la portance jusqu'à 30 % et augmenter la traînée de manière significative, ce qui peut compromettre la capacité de l'aéronef à décoller en toute sécurité [28].



(a) Exemple 1 - Dégivrage (b) Exemple 2 - Dégivrage

Figure 2.1: Formation de glace sur les surfaces critiques d'un aéronef et impact sur l'écoulement aérodynamique. Source : Organisation de l'Aviation Civile Internationale (OACI), *Manual of Aircraft Ground De-icing/Anti-icing Operations*, Doc 9640, 3e édition, 2018.

Afin de réduire ces risques, les procédures opérationnelles de dégivrage et d'antigivrage sont systématiquement appliquées dans les aéroports situés dans des régions soumises à des conditions hivernales. Le dégivrage consiste à éliminer la contamination existante sur les surfaces de l'aéronef, tandis que l'antigivrage vise à prévenir la formation ou l'accumulation de nouvelles précipitations solides avant le décollage [4, 17]. Ces procédures impliquent l'application de fluides spécialisés formulés

pour maintenir les surfaces critiques exemptes de glace pendant une période limitée.

Les fluides antigivrage utilisés dans l'aviation sont classés selon plusieurs spécifications industrielles, notamment les normes AMS1424 et AMS1428 développées par la Society of Automotive Engineers (SAE). Les fluides conformes à la norme AMS1424 correspondent généralement aux fluides de type I, principalement utilisés pour le dégivrage et caractérisés par un comportement rhéologique newtonien. En revanche, les fluides de type II, III et IV, spécifiés dans la norme AMS1428, sont conçus pour des applications d'antigivrage et possèdent des propriétés pseudoplastiques qui leur permettent d'adhérer aux surfaces de l'aéronef tout en étant éliminés lors du décollage sous l'effet du cisaillement aérodynamique [37, 79].

La performance de ces fluides est généralement évaluée à l'aide d'essais d'endurance, qui mesurent la durée pendant laquelle un fluide demeure capable de protéger une surface contre l'accumulation de neige ou de glace. Ces essais sont réalisés conformément au protocole standardisé SAE ARP5485, qui définit les conditions expérimentales permettant de simuler des précipitations hivernales contrôlées et de déterminer le moment où la défaillance du fluide se produit [78]. Les résultats obtenus servent à établir les tables de temps de protection, communément appelées *Holdover Time* (HOT), qui fournissent aux opérateurs aéronautiques des indications essentielles pour la prise de décision avant le décollage dans des conditions météorologiques hivernales [17].

Malgré l'existence de protocoles expérimentaux standardisés, la performance réelle des fluides antigivrage peut varier considérablement en fonction de nombreux facteurs environnementaux, tels que la température ambiante, le type de neige, l'intensité des précipitations, ainsi que les propriétés physiques de la neige, notamment sa densité et son taux de dépôt [24, 92]. Cette variabilité rend la prédiction du temps d'endurance particulièrement complexe et souligne la nécessité de développer des approches analytiques avancées capables de modéliser les interactions entre les conditions environnementales et le comportement des fluides antigivrage.

2.3 Différences entre essais en laboratoire et conditions réelles

Malgré l'existence de protocoles expérimentaux standardisés pour l'évaluation des fluides antigivrage, plusieurs études ont mis en évidence des différences significatives entre les résultats obtenus lors d'essais en laboratoire et les performances observées dans des conditions opérationnelles réelles. Ces écarts s'expliquent principalement par la complexité des phénomènes physiques associés aux précipitations hivernales et aux interactions thermiques entre la neige, le fluide antigivrage et la surface de l'aéronef.

Dans les essais en laboratoire, les conditions environnementales sont généralement contrôlées afin d'assurer la reproductibilité des expériences. Les paramètres tels que la température de l'air, le taux de précipitation, la vitesse de dépôt de la neige et la géométrie de la surface d'essai sont maintenus dans des plages définies par les protocoles expérimentaux, notamment la norme SAE ARP5485 [78]. Ces conditions permettent de comparer la performance relative de différents fluides antigivrage dans un environnement contrôlé.

Cependant, les précipitations naturelles présentent une variabilité beaucoup plus importante que celle reproduite en laboratoire. La neige naturelle peut varier considérablement en termes de structure cristalline, de densité, de teneur en eau liquide et de taille des flocons, ce qui influence directement les interactions thermiques et mécaniques avec les fluides antigivrage [24, 92]. Par exemple, la densité de la neige peut varier de manière significative selon les conditions atmosphériques, ce qui modifie la quantité de chaleur absorbée lors du contact avec le fluide ainsi que le taux d'accumulation sur la surface.

Les interactions thermiques jouent également un rôle déterminant dans la dégradation des fluides antigivrage. Lorsqu'un flocon de neige entre en contact avec le fluide, un transfert de chaleur se produit entre la neige froide et le fluide relativement plus

chaud. Ce phénomène entraîne une diminution rapide de la température du fluide et peut accélérer sa dilution ou sa dégradation. Des études expérimentales ont montré que la saturation de la neige et la quantité d'eau liquide contenue dans les précipitations peuvent fortement influencer ce processus thermique [92].

Les essais d'endurance permettent d'observer l'évolution du comportement des fluides antigivrage au cours du temps sous précipitations solides. La Figure 2.2 illustre deux exemples typiques d'échec du temps d'endurance observés lors d'essais expérimentaux. Les images thermiques permettent de visualiser l'évolution de la température de surface, tandis que les images optiques montrent l'accumulation progressive de neige sur la surface testée.

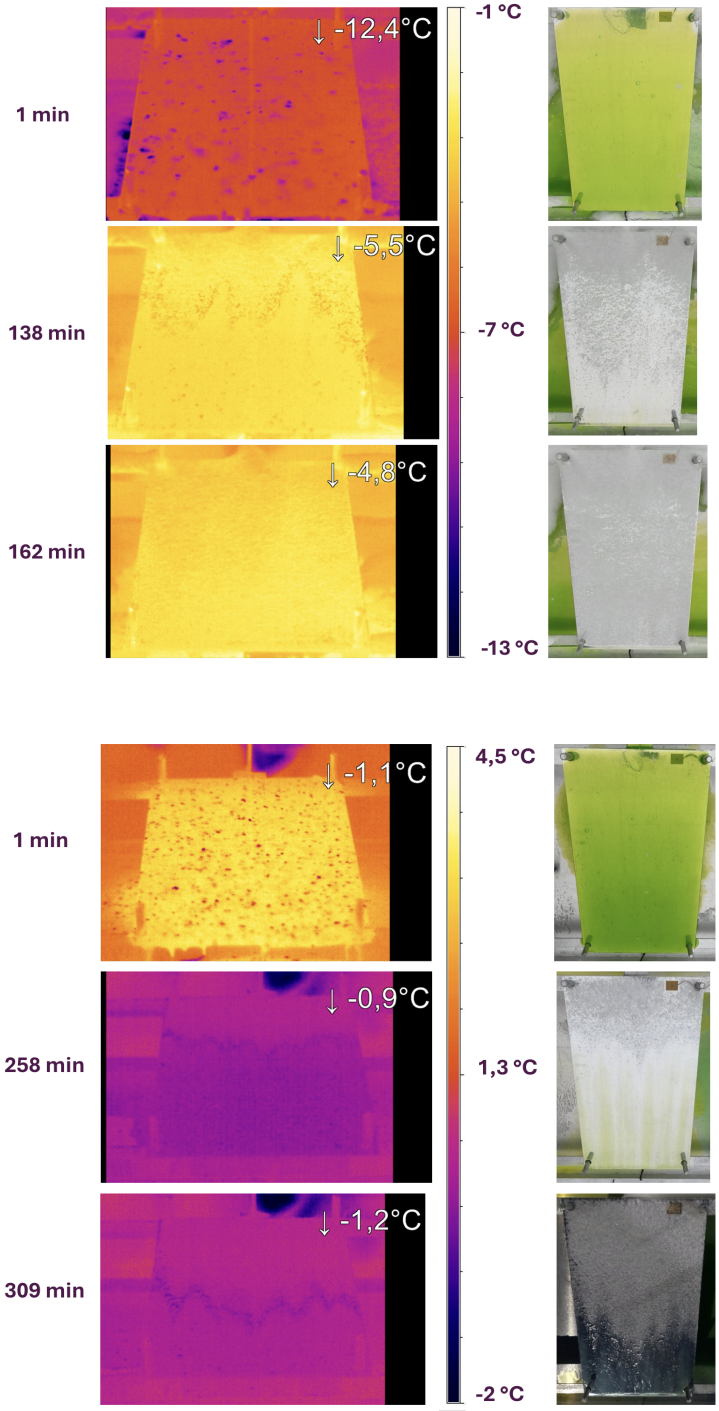


Figure 2.2: Exemples d'échec du temps d'endurance des fluides antigivrage observés lors d'essais sous précipitations solides. Les images thermiques (à gauche) illustrent l'évolution de la température de surface au cours du temps, tandis que les images visibles (à droite) montrent l'accumulation progressive de neige sur la surface testée. L'échec du fluide est identifié lorsque l'accumulation de neige dépasse un seuil critique sur la surface. Source : Laboratoire International des Matériaux Antigivre (LIMA), Université du Québec à Chicoutimi.

Dans ces exemples, l'échec du fluide est caractérisé par l'accumulation progressive de neige sur la surface malgré la présence du fluide antigivrage. Ce phénomène est généralement associé à la dilution du fluide par les précipitations, à une perte de propriétés rhéologiques ou à une diminution du transfert thermique entre le fluide et les particules de neige.

Les systèmes de génération de neige artificielle utilisés en laboratoire permettent de contrôler certains paramètres physiques, tels que le taux de précipitation ou la taille moyenne des particules de neige. Ces dispositifs offrent un environnement expérimental reproductible, ce qui est essentiel pour comparer les performances des fluides dans des conditions similaires. Toutefois, ces systèmes ne reproduisent pas entièrement la variabilité des précipitations naturelles. En particulier, les propriétés microstructurales de la neige artificielle, telles que la forme des cristaux et la distribution des tailles de particules, diffèrent souvent de celles observées dans la neige naturelle [18].

En conséquence, les résultats obtenus lors d'essais sous neige artificielle peuvent différer de ceux observés dans des conditions naturelles, même lorsque les taux de précipitation et les températures sont similaires. Cette divergence souligne l'importance d'intégrer des données issues à la fois d'essais en laboratoire et d'observations en conditions naturelles afin d'obtenir une compréhension plus complète du comportement des fluides antigivrage.

Face à ces limitations expérimentales, plusieurs travaux récents ont proposé l'utilisation de techniques d'analyse de données et d'apprentissage automatique afin d'exploiter les ensembles de données expérimentales disponibles et d'identifier les relations complexes entre les variables environnementales et la performance des fluides antigivrage. Ces approches permettent de compléter les méthodes expérimentales traditionnelles en fournissant des outils prédictifs capables d'intégrer la variabilité des conditions météorologiques hivernales.

2.4 Apprentissage automatique pour la prédiction des phénomènes de givrage

L'apprentissage automatique est de plus en plus utilisé pour modéliser des phénomènes complexes dans les domaines de l'aéronautique et des sciences atmosphériques. Les phénomènes de givrage sont influencés par de multiples variables environnementales, telles que la température de l'air, la structure des précipitations, l'humidité relative et les propriétés physiques des surfaces exposées. En raison de la nature fortement non linéaire et multivariée de ces interactions, les méthodes traditionnelles basées sur des modèles physiques simplifiés peuvent présenter des limitations dans certaines conditions opérationnelles. Les approches d'apprentissage automatique permettent d'identifier des relations complexes entre ces variables et la formation de glace, offrant ainsi de nouvelles perspectives pour la prédiction et la gestion des risques liés au givrage [13, 25].

Plusieurs approches basées sur les réseaux de neurones profonds ont été proposées pour détecter ou prédire la formation de glace sur les surfaces aéronautiques. Les Deep Belief Networks (DBN) et les Stacked Autoencoders (SAE) ont notamment été utilisés pour modéliser des relations non linéaires complexes dans des systèmes de détection du givrage. Ces architectures profondes permettent d'extraire automatiquement des caractéristiques pertinentes à partir de données météorologiques ou de capteurs embarqués, ce qui peut améliorer la précision des systèmes de détection [55]. Toutefois, ces méthodes présentent également certaines limitations, notamment une complexité computationnelle élevée et une interprétabilité limitée, ce qui peut compliquer leur utilisation dans des systèmes opérationnels critiques.

D'autres approches combinant l'apprentissage automatique et des techniques d'optimisation ont également été proposées. Par exemple, les réseaux de neurones optimisés par des algorithmes d'optimisation par essaim de particules (PSO-BP Neural Networks) ont été appliqués à la modélisation des performances de systèmes antigivrage.

Ces méthodes permettent d'améliorer la convergence de l'apprentissage et d'optimiser les paramètres du modèle, mais elles peuvent également introduire une complexité supplémentaire dans le processus de modélisation [102].

Dans les applications environnementales et industrielles, l'intégration de réseaux de neurones artificiels avec des systèmes de capteurs connectés (IoT) a également été explorée afin de permettre la surveillance en temps réel des conditions favorables au givrage. Ces approches reposent généralement sur des données collectées en continu par des capteurs météorologiques et permettent de fournir des alertes précoces en cas de conditions propices à la formation de glace [93].

Parallèlement aux approches basées sur les réseaux de neurones, les méthodes d'apprentissage automatique basées sur les arbres de décision ont gagné en popularité dans les applications environnementales et aéronautiques. Les modèles d'ensemble tels que Random Forest et Gradient Boosting sont particulièrement adaptés aux ensembles de données tabulaires contenant des variables hétérogènes et des relations non linéaires. Ces méthodes offrent généralement une bonne robustesse face au bruit des données et permettent également d'évaluer l'importance relative des variables dans les prédictions [16, 20].

Plusieurs modèles d'apprentissage automatique ont été appliqués à la prédiction des phénomènes de givrage et aux systèmes antigivrage en aéronautique. Le Tableau 2.1 présente une comparaison des principales approches rapportées dans la littérature, ainsi que l'approche XGBoost retenue dans cette recherche pour la prédiction du temps d'endurance des fluides antigivrage.

Référence et modèle	**Application**	**Données et taille**	**Variables**	**Validation et performance**	**Limites**
Yi et al.; Li et al. – DBN/SAE [55, 101]	Détection du givrage	Données de capteurs; taille non précisée dans la présente synthèse	Variables météorologiques et signaux de capteurs	Validation propre aux études; résultats non directement comparables au temps d'endurance	Coût computationnel élevé; interprétabilité limitée
Mao et al.; Zhang et al. – PSO-BP [63, 102]	Performance de systèmes antigivrage	Données numériques ou expérimentales; taille non précisée ici	Paramètres thermiques et opérationnels	Validation rapportée par les auteurs; protocoles différents de la présente étude	Interprétabilité limitée; complexité d'optimisation
Abdelghany et al.; Wang et al. – ANN/IoT [2, 93]	Surveillance et alerte de givrage	Flux de capteurs; taille variable ou non précisée ici	Température, humidité et variables de capteurs	Évaluation applicative; métriques hétérogènes entre études	Dépendance aux capteurs; transférabilité limitée
Breiman – Random Forest [16]	Méthode générale de régression et classification	Jeux tabulaires; sans jeu d'endurance aéronautique dans l'article méthodologique	Variables tabulaires hétérogènes	Validation dépendante de l'application	Référence méthodologique, non validation directe sur l'Endurance Time
Chen et Guestrin; Li et al. – XGBoost [20, 54]	Modélisation non linéaire sur données tabulaires	Taille dépendante de l'application; non précisée ici	Variables environnementales et opérationnelles	Validation croisée ou jeu de test selon l'étude	Réglage des hyperparamètres; extrapolation non garantie

Table 2.1: Synthèse critique des principales approches d'apprentissage automatique pertinentes pour le givrage aéronautique. Les mentions « non précisée ici » signalent qu'une valeur comparable n'est pas disponible dans la synthèse consultée et évitent d'inférer une taille ou une performance non documentée.

Dans ce contexte, les méthodes d'ensemble basées sur les arbres de décision, telles que Random Forest et XGBoost, offrent un compromis intéressant entre précision,

robustesse et interprétabilité. Ces modèles sont capables de capturer efficacement des relations non linéaires complexes entre les variables d'entrée tout en conservant une structure relativement simple à interpréter. De plus, les techniques basées sur le *gradient boosting*, telles que XGBoost, permettent d'améliorer progressivement les performances du modèle en corrigeant les erreurs des arbres précédemment construits, ce qui peut conduire à des performances prédictives élevées même sur des ensembles de données de taille modérée [20].

Afin de synthétiser les différents critères influençant le choix d'un modèle d'apprentissage automatique dans le contexte de la prédiction du givrage en aviation, la Figure 2.3 présente un cadre conceptuel de sélection des modèles. Ce diagramme illustre comment le choix d'une approche dépend de plusieurs contraintes opérationnelles, notamment la taille des ensembles de données disponibles, les ressources computationnelles, la nécessité de produire des prédictions en temps réel ainsi que le niveau d'interprétabilité requis. Dans de nombreuses applications aéronautiques, les modèles offrant un bon compromis entre précision, efficacité computationnelle et interprétabilité sont privilégiés.

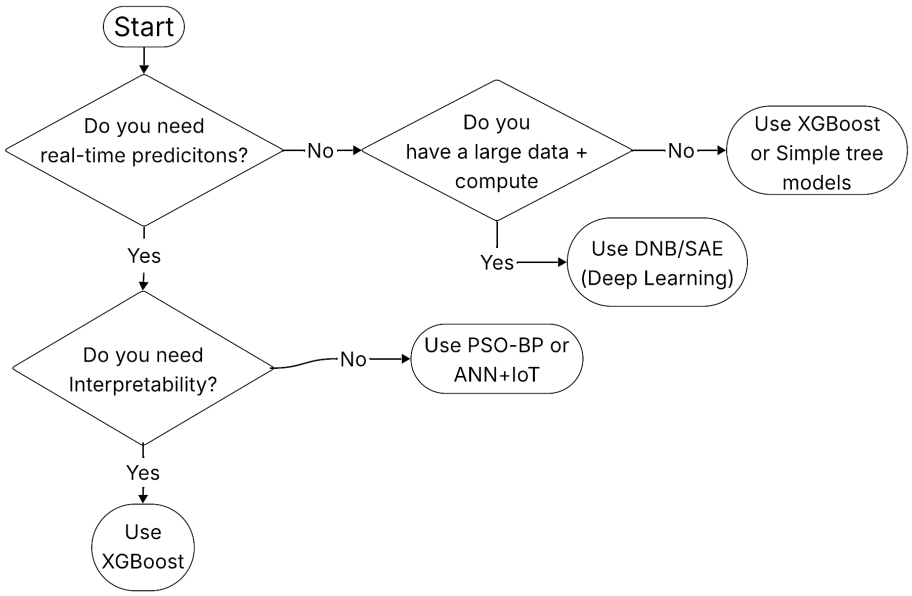


Figure 2.3: Cadre conceptuel pour la sélection des modèles d'apprentissage automatique dans la prédiction du givrage en aviation. Figure élaborée par l'auteur à partir de [2, 16, 20, 52, 63].

2.5 Génération de données synthétiques pour les données environnementales

Les progrès récents en intelligence artificielle et en apprentissage automatique ont favorisé le développement de techniques de génération de données synthétiques capables de reproduire les caractéristiques statistiques de jeux de données réels. Ces approches sont particulièrement utiles dans les domaines où la collecte de données expérimentales est coûteuse, limitée ou dépend fortement de conditions environnementales difficiles à reproduire. Dans ce contexte, les méthodes de génération de données permettent d'augmenter artificiellement les ensembles de données disponibles et d'améliorer la robustesse des modèles prédictifs [26, 33].

Parmi les différentes techniques proposées, les *Generative Adversarial Networks* (GAN) constituent l'une des approches les plus populaires pour la génération de données

synthétiques. Introduits par [33], les GAN reposent sur un processus d'apprentissage adversarial impliquant deux réseaux neuronaux : un générateur et un discriminateur. Le générateur apprend à produire des échantillons synthétiques à partir d'un bruit aléatoire, tandis que le discriminateur tente de distinguer les données réelles des données générées. Au cours de l'entraînement, les deux réseaux s'améliorent simultanément, permettant au générateur de produire des données synthétiques de plus en plus réalistes.

La Figure 2.4 présente un schéma simplifié du fonctionnement des GAN afin d'illustrer l'interaction entre le générateur et le discriminateur lors du processus d'apprentissage adversarial.

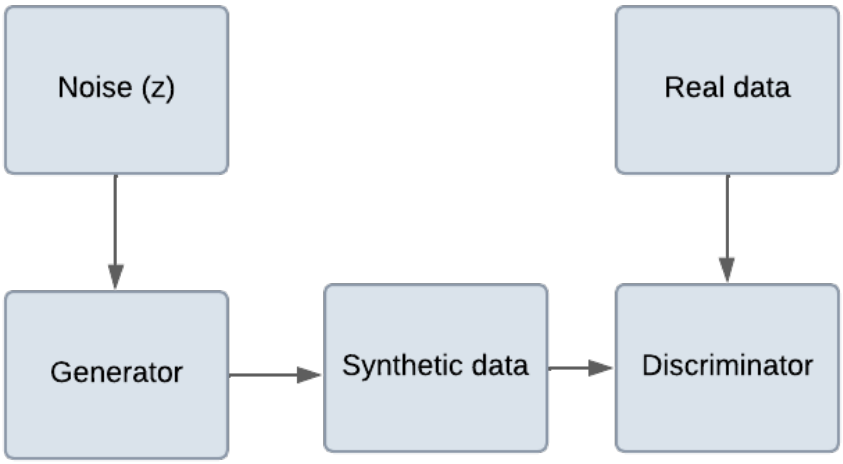


Figure 2.4: Schéma simplifié de l'architecture d'un Generative Adversarial Network (GAN). Le générateur produit des données synthétiques à partir d'un vecteur de bruit aléatoire, tandis que le discriminateur apprend à distinguer les données réelles des données générées. L'entraînement adversarial entre ces deux réseaux permet d'améliorer progressivement le réalisme des données synthétiques.

Dans le domaine des sciences atmosphériques et du climat, les GAN ont été largement utilisés pour améliorer la résolution spatiale des champs météorologiques, un processus connu sous le nom de *downscaling statistique*. Ces modèles permettent de générer des champs météorologiques à haute résolution à partir de données climatiques à basse résolution, tout en préservant les propriétés statistiques des phénomènes at-

mosphériques [30, 35]. Plusieurs études ont démontré que les GAN sont capables de reproduire efficacement la variabilité spatiale et temporelle de variables météorologiques telles que les précipitations, la température ou les champs de vent.

La modélisation des précipitations constitue l'une des applications les plus étudiées des GAN dans les sciences environnementales. En raison de leur caractère fortement non linéaire et de leur variabilité multi-échelle, les précipitations représentent un défi important pour les méthodes statistiques traditionnelles. Des architectures GAN conditionnelles ont été proposées pour générer ou affiner des champs de précipitations en capturant à la fois les distributions d'intensité et les structures spatiales des événements extrêmes [39, 41]. Ces approches ont montré qu'il est possible d'améliorer la représentation statistique des événements rares tout en préservant les corrélations spatiales observées dans les données réelles.

Au-delà des applications basées sur des données spatiales, les méthodes de génération de données synthétiques sont également utilisées pour des ensembles de données tabulaires. Ces ensembles de données contiennent généralement des variables hétérogènes, telles que des variables continues, catégorielles et parfois des indicateurs opérationnels. La génération de données tabulaires réalistes représente un défi particulier, car les modèles doivent préserver non seulement les distributions individuelles des variables, mais également les relations multivariées entre celles-ci [26, 72].

Dans les applications environnementales et d'ingénierie, la génération de données synthétiques est souvent utilisée pour compenser la rareté des observations expérimentales ou pour représenter des conditions extrêmes peu observées dans les données disponibles. Par exemple, dans le domaine de l'énergie éolienne, des approches basées sur les GAN ont été utilisées pour générer des séries temporelles de vent synthétiques afin d'améliorer la précision des modèles de prévision de production d'énergie [57]. Ces travaux montrent que l'augmentation des données à l'aide de modèles génératifs peut améliorer la capacité de généralisation des modèles prédictifs.

Dans le contexte des opérations aériennes hivernales, les ensembles de données expérimentales liés aux essais d'endurance des fluides antigivrage sont souvent de taille limitée. Les essais sous neige naturelle dépendent fortement des conditions météorologiques, tandis que les essais sous neige artificielle nécessitent des infrastructures expérimentales spécialisées. Cette limitation de données rend difficile l'entraînement de modèles d'apprentissage automatique robustes.

Dans ce contexte, la génération de données synthétiques à l'aide de GAN représente une approche prometteuse pour augmenter les ensembles de données expérimentales tout en préservant les caractéristiques statistiques des variables environnementales et opérationnelles. L'utilisation de ces techniques permet non seulement d'augmenter la diversité des observations disponibles, mais également d'étudier l'impact de différentes architectures de modèles génératifs sur la qualité des données synthétiques et leur utilité pour les tâches de modélisation prédictive.

2.6 Évaluation des données synthétiques

L'évaluation de la qualité des données synthétiques constitue un défi central dans les applications de génération de données, car il est nécessaire de s'assurer que les données générées reproduisent fidèlement les caractéristiques statistiques et structurelles des données réelles. Contrairement aux ensembles de données traditionnels, il n'existe pas de critère unique permettant d'évaluer la qualité des données synthétiques. Pour cette raison, la littérature recommande généralement l'utilisation d'une approche d'évaluation multi-critères combinant plusieurs perspectives complémentaires [26, 72].

Ces critères peuvent être regroupés en trois catégories principales : la similarité statistique entre les distributions réelles et synthétiques, la préservation des dépendances entre variables, et l'utilité des données synthétiques pour les tâches de modélisation prédictive.

2.6.1 Similarité des distributions statistiques

La première étape de l'évaluation consiste à comparer les distributions statistiques des variables réelles et synthétiques. L'objectif est de vérifier si les données générées reproduisent correctement les caractéristiques fondamentales des données originales, telles que la moyenne, la variance, l'asymétrie et la distribution globale des observations.

Plusieurs métriques sont utilisées pour mesurer cette similarité. Parmi les plus courantes, on retrouve la distance de Wasserstein, la divergence de Kullback–Leibler (KL) et la divergence de Jensen–Shannon. Ces mesures permettent de quantifier l'écart entre les distributions réelles et synthétiques [6]. La distance de Wasserstein est particulièrement populaire dans les applications basées sur les GAN, car elle fournit une mesure stable de la différence entre deux distributions de probabilité.

En complément de ces métriques, des analyses visuelles telles que les histogrammes, les diagrammes de densité ou les fonctions de distribution cumulative sont souvent utilisées afin d'évaluer qualitativement la similarité entre les données réelles et synthétiques.

2.6.2 Préservation des dépendances entre variables

Au-delà des distributions individuelles, il est essentiel que les données synthétiques préservent les relations multivariées entre les variables. Dans les ensembles de données environnementales ou expérimentales, les variables sont souvent liées par des relations physiques ou statistiques complexes. Si ces dépendances ne sont pas correctement reproduites, les données synthétiques peuvent devenir inutilisables pour l'analyse ou la modélisation.

Pour évaluer ces relations, les chercheurs utilisent généralement des matrices de corrélation, des mesures d'information mutuelle ou des normes matricielles permettant de comparer les structures de dépendance entre les données réelles et synthétiques [26]. La comparaison des matrices de corrélation constitue une méthode simple mais efficace

pour vérifier si les relations linéaires entre variables sont correctement reproduites. Les mesures d'information mutuelle permettent quant à elles de capturer des dépendances non linéaires plus complexes.

Ces approches sont particulièrement importantes dans les applications scientifiques et d'ingénierie, où les relations entre variables reflètent souvent des mécanismes physiques sous-jacents.

2.6.3 Utilité prédictive des données synthétiques

La troisième catégorie d'évaluation consiste à mesurer l'utilité des données synthétiques pour l'entraînement de modèles d'apprentissage automatique. Cette approche vise à déterminer si les données générées contiennent suffisamment d'information pour permettre le développement de modèles prédictifs efficaces.

Une méthode largement utilisée dans ce contexte est l'approche *Train-on-Synthetic, Test-on-Real* (TSTR). Dans cette procédure, un modèle d'apprentissage automatique est entraîné exclusivement sur les données synthétiques, puis évalué sur un ensemble de données réelles. Si les performances obtenues sont comparables à celles d'un modèle entraîné sur les données réelles, cela indique que les données synthétiques capturent correctement les relations importantes présentes dans les données originales [23].

D'autres variantes de cette approche incluent les stratégies *Train-on-Real, Test-on-Synthetic* (TRTS) ou les évaluations combinant données réelles et synthétiques pour mesurer l'impact de l'augmentation des données sur la performance des modèles [23, 26].

L'utilisation conjointe de ces différentes méthodes d'évaluation permet d'obtenir une analyse plus complète de la qualité des données synthétiques.En combinant des mesures statistiques, des analyses de dépendance et des tests d'utilité prédictive, il devient possible d'évaluer non seulement le réalisme des données générées, mais également leur pertinence pour les applications de modélisation et de prédiction.

2.7 Lacunes dans la littérature

Malgré les avancées significatives dans la modélisation des phénomènes de givrage et l'utilisation de techniques d'apprentissage automatique, plusieurs lacunes subsistent dans la littérature scientifique.

Premièrement, peu d'études ont exploré l'intégration de données provenant à la fois d'essais sous neige naturelle et sous neige artificielle pour la prédiction du temps d'endurance des fluides antigivrage.

Deuxièmement, l'utilisation de modèles d'apprentissage automatique interprétables et adaptés à des ensembles de données expérimentales de petite taille reste limitée.

Troisièmement, les applications de techniques de génération de données synthétiques à des ensembles de données tabulaires liés aux opérations hivernales en aviation demeurent encore peu étudiées.

La présente recherche vise à combler ces lacunes en combinant des techniques d'apprentissage automatique et de génération de données synthétiques afin d'améliorer la modélisation du temps d'endurance des fluides antigivrage dans des conditions hivernales.

Note de l'auteur concernant le chapitre publié

Le chapitre 3 reproduit un article déjà publié et validé; son texte scientifique est conservé dans sa version publiée. Pour son interprétation dans le présent mémoire, il convient toutefois de préciser que les fluides de Type I présentent généralement un comportement newtonien, tandis que les fluides de Types II, III et IV sont généralement pseudo-plastiques et non newtoniens. Les références à SAE ARP5485B et aux seuils d'erreur constituent ici des points de comparaison propres à l'étude et non des critères réglementaires d'acceptation d'un modèle d'apprentissage automatique. Enfin, les performances présentées démontrent une capacité prédictive dans le domaine expérimental étudié;

elles ne constituent ni la validation d'un outil opérationnel ni un substitut aux essais normalisés.

Dans le tableau 3.4 et dans la discussion qui l'accompagne, la valeur 5,135 correspond à l'erreur absolue moyenne (*Mean Absolute Error*, MAE), exprimée en minutes, et non à l'erreur absolue moyenne en pourcentage (*Mean Absolute Percentage Error*, MAPE). Cette correction concerne uniquement la désignation de la métrique et ne modifie ni le résultat numérique ni les conclusions de l'article publié.

L'équation de propagation de l'incertitude présentée au chapitre 3 apparaît sans numéro dans la version publiée de l'article. Elle correspond à l'expression de l'incertitude-type composée obtenue par propagation gaussienne des incertitudes. La numérotation originale de l'article a été conservée afin de ne pas modifier la version publiée.

Concernant le protocole de validation du chapitre 3, les partitions de validation croisée ont été constituées au niveau des observations individuelles à l'aide d'une validation croisée à dix plis avec brassage et une graine aléatoire fixée à 42. L'utilisation d'une graine unique assure la reproductibilité du partitionnement et de l'entraînement, mais elle ne permet pas d'évaluer la sensibilité des performances aux variations aléatoires des plis ou de l'initialisation des modèles. Les métriques rapportées correspondent donc à cette configuration reproductible particulière et ne représentent pas une distribution de performance obtenue sur plusieurs graines. Aucun identifiant permettant de regrouper formellement les observations provenant d'un même essai expérimental n'était disponible dans le jeu de données utilisé. Par conséquent, certaines observations issues d'un même essai peuvent avoir été réparties entre différents plis. Les performances rapportées doivent être interprétées en tenant compte de ces limitations. Une validation groupée par essai ou par campagne expérimentale, répétée sur plusieurs graines aléatoires et accompagnée de moyennes, d'écarts-types ou d'intervalles de confiance, sera nécessaire pour évaluer plus rigoureusement la robustesse et la généralisation des modèles.

CHAPITRE 3

PRÉDICTION DU TEMPS D'ENDURANCE DES FLUIDES ANTIGIVRAGE POUR AÉRONEFS PAR APPRENTISSAGE AUTOMATIQUE : ÉTUDE COMPARATIVE ENTRE DONNÉES DE NEIGE NATURELLE ET ARTIFICIELLE

Avant-propos

État de l'acceptation : Publié.

Journal : Aerospace Science and Technology

Lien d'accès : https://www.sciencedirect.com/science/article/pii/S1270963826004839

https://doi.org/10.1016/j.ast.2026.112103

Référence : Paniagua, A., et al. (2026). "Machine Learning Prediction of Aircraft Anti-Icing Fluid Endurance: A Comparative Study Using Natural and Artificial Snow Data." *Aerospace Science and Technology*: 112103.

Titre en français : Prédiction du temps d'endurance des fluides antigivrage pour aéronefs par apprentissage automatique : étude comparative entre données de neige naturelle et artificielle

3.1 Résumé

Ce travail propose une approche basée sur l'apprentissage automatique pour prédire le temps d'endurance des fluides antigivrage appliqués sur les surfaces d'aéronefs dans des conditions hivernales. L'étude s'appuie sur des données expérimentales obtenues sous neige naturelle et artificielle afin de comparer l'impact de ces environnements sur la performance des fluides. Plusieurs modèles d'apprentissage automatique ont été développés et évalués, permettant d'identifier les variables environnementales les plus influentes dans la dégradation des fluides. Les résultats montrent que les modèles proposés sont capables de capturer les relations non linéaires complexes entre les conditions météorologiques et le temps d'endurance, tout en mettant en évidence des différences significatives entre les données issues de neige naturelle et artificielle. Cette approche offre une alternative prometteuse aux méthodes expérimentales traditionnelles en permettant une meilleure compréhension et prédiction du comportement des fluides antigivrage dans des conditions opérationnelles réelles.

3.2 Abstract

Ground ice poses a critical safety risk to winter aviation operations, primarily because the protective endurance of aircraft anti-icing fluids fluctuates unpredictably with volatile weather conditions. To mitigate these risks, this study develops a robust machine learning framework designed to forecast fluid endurance times using high-fidelity data from both artificial and natural snow environments. Research was conducted via endurance tests following the SAE ARP5485B protocol to train and evaluate multiple decision tree-based architectures, specifically the Decision Tree, Random Forest, and Extreme Gradient Boosting (XGBoost) models. Model reliability was ensured through rigorous K-Fold cross-validation and systematic hyperparameter optimization. Results demonstrate

that the optimized XGBoost model achieved superior predictive accuracy, yielding an RMSPE of 15.38% and a MAPE of 8.88%. Furthermore, SHAP analysis identified snow deposition rate and ambient temperature as the most influential determinants of fluid failure. These findings suggest that integrating machine learning into endurance testing can revolutionize current safety protocols and facilitate precise, data-driven decision-making for winter ground operations.

3.3 Introduction

Ground icing constitutes a considerable and frequently overlooked risk to aviation safety, especially during pre-takeoff procedures. A thin layer of snow, frost, or freezing rain on aerodynamic surfaces, such as wings, tail, or fuselage, can significantly disrupt airflow, resulting in increased drag and decreased lift [27]. Research indicates that even minimal contamination can significantly compromise aerodynamic performance, increasing the likelihood of accidents during critical phases of flight [4]. The crash of Air Ontario Flight 1363 in 1989 exemplified significant hazards: insufficient de-icing resulted in ice buildup on the wings, impairing lift and leading to the aircraft's crash, which caused 24 fatalities [69, 71]. This event highlighted the critical significance of comprehensive ground de-icing and anti-icing procedures.

To mitigate such risks, Transport Canada Civil Aviation (TCCA) enforces compliance with SAE ARP5485 procedures, which determine the endurance time of anti-icing fluids under controlled and real-world conditions [79, 80, 89]. These fluids—classified as Type I (AMS1424) and Type II–IV (AMS1428)—exhibit non-Newtonian, pseudoplastic behaviour, with viscosity dependent on shear rate [80]. Endurance testing is adding fluid to a sloping aluminum plate that simulates an aircraft surface, followed by exposure to controlled snow conditions until failure, characterized by the commencement of surface contamination [78]. Field testing conducted under natural precipitation subsequently

corroborate laboratory findings and develop Holdover Time (HOT) guidelines, assisting operators in determining safe time limits prior to departure [70].

Nevertheless these standardized methods, significant disparities between laboratory and natural test outcomes endure, underscoring the divergence between controlled and real-world settings. A recent study suggests that snow particles interacting with surfaces under various conditions undergo thermal and hydrodynamic interactions [44, 45, 67]. Villeneuve et al [92] assert that ethylene-based fluids have unique cooling and rebalancing phases correlated with the type of snow and saturation level. This underscores the significance of considering snow unpredictability when evaluating endurance length. Although widely utilized, standardized laboratory and outdoor endurance tests are limited by practical, budgetary, and environmental constraints. Natural testing is inherently opportunistic, expensive, and geographically limited, leading to sparse datasets that inadequately represent the complete spectrum of operating winter conditions. In contrast, laboratory-based artificial snow experiments provide controlled repeatability but fail to completely replicate the stochastic variability of natural precipitation, encompassing variations in snow density, particle shape, and thermal interactions at the fluid–surface interface [18, 24]. Consequently, endurance time evaluations based exclusively on a single testing environment sometimes lack generalizability in real-world applications.

Beyond conventional endurance testing, recent research has investigated a broad range of anti-icing and surface protection strategies designed to limit ice and snow accumulation under challenging winter conditions. These efforts include the development of active and passive anti-icing systems, advances in surface engineering, and hybrid thermal–material solutions, reflecting the inherently multidisciplinary nature of aircraft icing mitigation [45, 63, 73]. Collectively, these studies indicate that anti-icing effectiveness is governed not only by fluid formulation, but also by the coupled influence of environmental forcing, surface characteristics, and operational conditions [67, 74]. As a result, there is an increasing need for predictive frameworks capable of integrat-

ing heterogeneous variables to support the assessment and optimization of anti-icing strategies under realistic and variable winter scenarios.

Machine learning provides an additional framework that may integrate various data from both controlled laboratory and natural testing to overcome these limitations. Data-driven models can interpolate in untested scenarios and improve forecast reliability by identifying non-linear relationships among environmental variables, fluid properties, and endurance performance, therefore exceeding the constraints of experimental validation alone. In this regard, machine learning attempts to improve current SAE and Transport Canada testing methodologies by enabling scalable, adaptive, and interpretable forecasts of anti-icing fluid endurance length under diverse winter conditions.

A variety of machine learning algorithms have been tested to simulate ice formation and predict anti-icing effectiveness. Deep learning frameworks, such as Deep Belief Networks (DBN) and Stacked Auto-Encoders (SAE), have achieved notable accuracy in simulating icing phenomena; however, they often require considerable training and computational resources [2, 21, 63, 101]. Tree-based models, especially XGBoost, combine interpretability and computational efficiency, providing strong classification and regression results with moderate data demands [54]. This study introduces a cohesive and interpretable machine learning framework for aircraft icing, trained and validated on both artificial and natural snow datasets, in contrast to previous research that utilized only laboratory or field data. It explicitly assesses generalization, optimization strategies, and feature-level physical interpretability via SHAP analysis. The methodology seeks to uphold a Mean Absolute Percentage Error (MAPE) under 10–15% and a Root Mean Square Error (RMSE) within 5% of the Holdover Time (HOT), in accordance with engineering accuracy standards and validation criteria established in Transport Canada and SAE documents [76, 77, 88, 90, 100].

3.4 Methodology

The study is structured into two phases. Endurance time data is initially collected through standardized research in both natural and artificial snow settings. A machine learning model is subsequently developed to predict endurance time using environmental and operational data.

Three algorithms based on decision trees were assessed: Decision Tree, Random Forest, and XGBoost, utilizing the same input qualities. The model's reliability was evaluated using 10-fold cross-validation. Hyperparameter optimization was conducted using Bayesian Optimization and Genetic Algorithms, as depicted in Figure 3.1. To verify applicability in operational contexts, model outputs were tested against FAA and Transport Canada Holdover Time (HOT) rules.

SHAP analysis indicates that endurance is influenced by a limited set of interpretable, physics-based variables, particularly when accounting for non-linear interactions between temperature and snow.

3.4.1 Data Collection

Data preparation is the first stage in developing the prediction model. To recollect the data, two separate tests must be performed, both with the same parameters, for the Society of Automotive Engineering (SAE) [81]. To collect the snow in both tests, an aluminium support is installed, as depicted in Figure 3.2, which features a right-angle support structure to replicate the inclination of aircraft wings at 10°.

Natural snow endurance tests were conducted at two sites to maximize variability: APS Aviation Inc. (under contract with Transport Canada's Transportation Development Centre) and the Anti-Icing Materials International Laboratory (AMIL). Both facilities followed the SAE ARP5485B protocol of the Aircraft Ground Anti-Icing Fluid Holdover and Endurance Time Testing Program. The setup presented in Figure 3.3 employed an

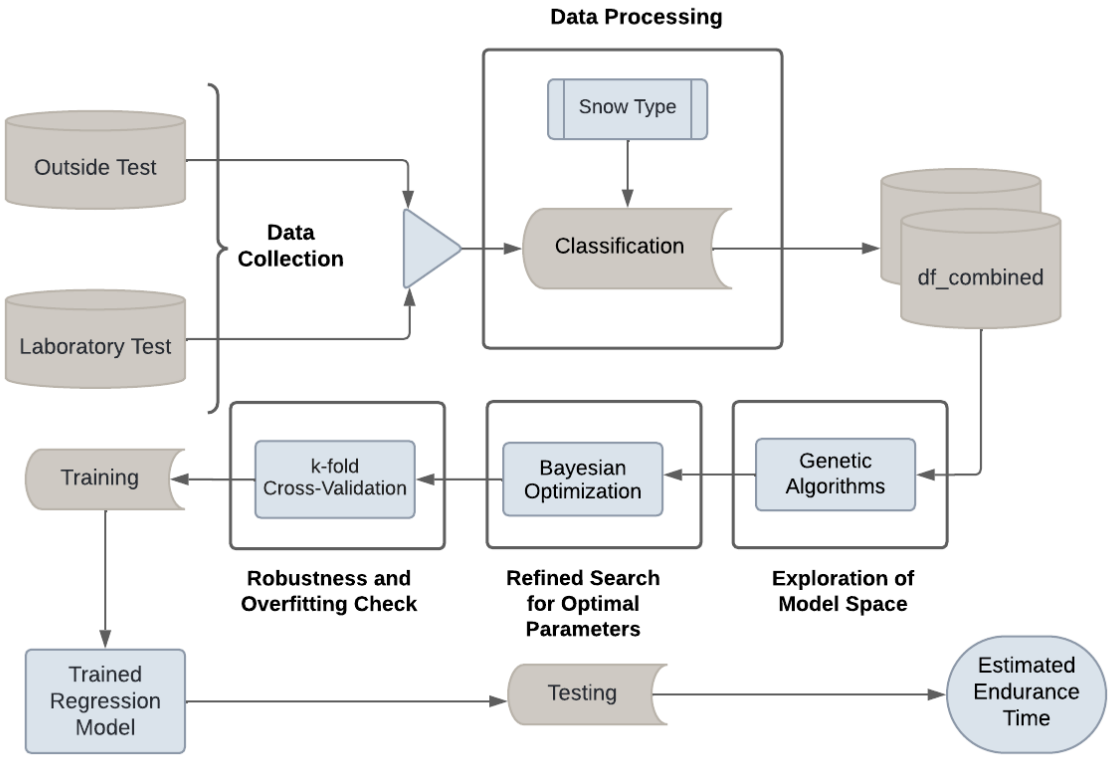


Figure 3.1: Data Processing and Optimization Pipeline for Snow Classification and Fluid Endurance Prediction

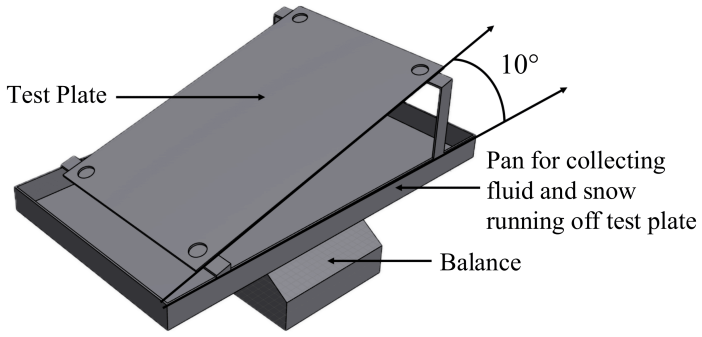


Figure 3.2: Support Structure for Test Plate and Snow Deposit

external aluminum structure to simulate an aircraft surface. Test fluid was released during snowfall, with environmental factors documented at one-minute intervals. Standardized protocols enabled uniformity across locations, facilitating the accurate assessment of



Figure 3.3: Setup for Natural Snow Test

fluid endurance duration under natural snow circumstances.

Artificial snow testing was conducted at the AMIL Laboratory under controlled conditions, in accordance with the SAE ARP5485B standard. The experimental setup consists of a cold room capable of maintaining ambient conditions with a thermal stability of ±0.5°C. Test plates were fitted with Pt100 temperature detectors made of platinum, located at mid-span and linked to an automated data collection system for ongoing monitoring and traceability. All temperature sensors and weighing instruments were calibrated semiannually in accordance with the ISO 10012 requirements.

For reference, the combined standard uncertainty for derived quantities (e.g., snow load) can be expressed in accordance with Gaussian error propagation as:

$$u_y = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial y}{\partial x_i} u_{x_i} \right)^2} \tag{3.1}$$

where u_{x_i} represents the standard uncertainty of each measured variable. However, because the objective of this study is predictive modeling using experimentally observed endurance times rather than deterministic analytical calculation, uncertainty propagation was not explicitly computed for each endurance prediction. Instead, measurement

variability is inherently embedded within the experimental dataset through repeated trials and varying environmental conditions. The use of K-Fold cross-validation further ensures that the reported performance metrics (RMSE, MAE, and MAPE) reflect averaged predictive deviations incorporating both modeling and experimental uncertainties across the full dataset.

The AMIL snow machine features an automated feeding system operating on a 10° sloped test plate (30cm × 50cm). Snow is produced by a top-mounted supply box and released onto a rotating cylinder equipped with small cavities that control the deposition rate and uniformity of artificial snowfall [7]. To ensure the quality and conformity of the artificial snow with the characteristics of natural snowfall, its density is measured twice daily. Volumetric sampling is performed using a calibrated snow dropper, and the deposited mass is recorded using a VWR-203P2 digital precision balance with 0.001g readability and ±0.004g accuracy. For each volume setting, six successive measurements are averaged to ensure repeatability. The SLF SNOWPRO-4040 sensor (FPGA GmbH, Switzerland) is utilized to determine the bulk density of snow. This measurement method adheres to the protocols established in recent research on indoor snow testing [8].

As illustrated in Figure 3.4, the snow machine was initiated until approximately 30% of the test surface was covered by unmelted white snow, and the cold chamber was preconditioned several days in advance to reach the target ambient temperature. During testing, the snow machine performs a back-and-forth motion to uniformly cover the test plate. Snowfall is directed directly beneath the machine, where a scale positioned under the test plate measures the accumulated mass. The data is relayed to a computer that processes the measurements and provides them to the snowfall distribution control system. The testing occurs in a temperature-controlled cold room, with data gathering and control computers positioned externally to prevent heat interference.

Fluid failure, according to the SAE ARP5485B criteria, is considered to occur when

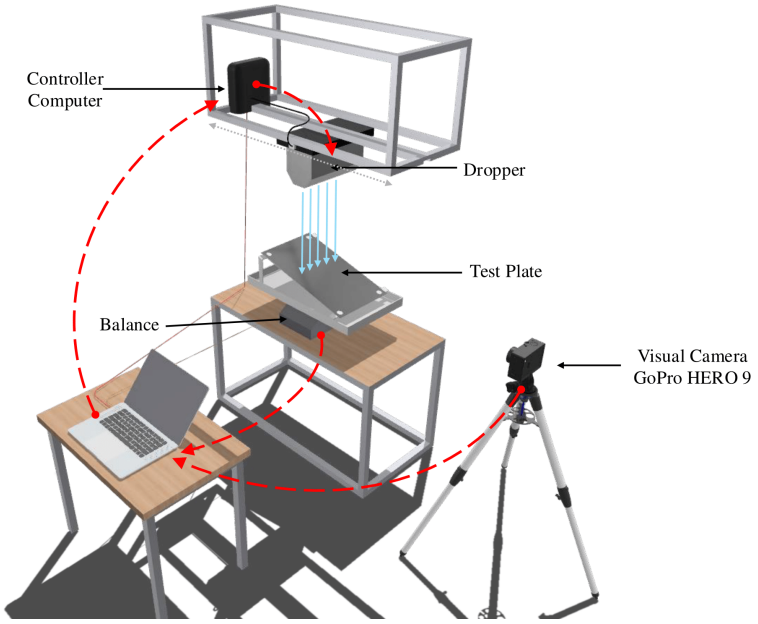


Figure 3.4: Setup for Artificial Snow Test

approximately 30% of the test surface is visually obscured by white snow, indicating that the fluid has reached melt saturation and has lost its anti-icing effectiveness [8]. To reduce subjectivity, the evaluation is carried out by two trained observers following the defined protocol. The visual assessment is reinforced by infrared thermography, which confirms the absence of further temperature decrease upon snow deposition, as well as by the visible accumulation of unmelted snow on the plate surface. In practice, the grid lines drawn on the plate surface facilitate estimation of the 30% coverage threshold, as shown in the images. The implementation of a defined SAE ARP5485B visual technique, alongside independent assessment by two qualified observers and validation by infrared thermography, guaranteed consistency and repeatability in failure detection across all testing.



(a) Beginning of the test (b) After 138 minutes

Figure 3.5: Visual Fluid Failure, According to SAE ARP5485B Criteria

Figure 3.5 illustrates a representative test used to build the ML training dataset. The left image shows the start of the test, with the surface completely covered by the fluid and snow melting upon contact. After 138 minutes, as depicted in the right image, around 30% of white snow, particularly in the central and lower areas, although some fluid persists at the top. This evident alteration in surface coverage illustrates the systematic identification and documentation of failures for dataset construction.

The dynamics of snowflake deposition and their interaction with the test surface are known to affect the fluid's stability and failure threshold. Such interactions, including restitution and flake deformation on impact, have been recently modeled using ML and fluid-mechanical frameworks [44, 46].

3.4.2 Computational Methods

To provide a clearer overview of the experimental setup and dataset structure, Table 3.1 summarizes the key characteristics of the samples collected under both artificial and natural snowfall conditions, as well as the modeling strategies applied in this study.

The combined dataset demonstrates an evident mismatch between artificial snow samples (478 cases) and real snow samples (295 cases), highlighting the practical limitations of field testing under genuine snowfall settings. This disparity reflects

Characteristic	Details
Total Samples	773
Artificial Snow Samples	478 (collected in controlled laboratory environment)
Natural Snow Samples	295 (collected under real snowfall conditions)
Samples per Test (Lab)	Approximately 3 samples per fluid
Samples per Test (Natural)	Approximately 9 samples per fluid
Anti-Icing Fluids Tested	4 fluids (labeled A, B, C, D for confidentiality)
Snow Intensity Range	1.24 to 35.8 g/dm^2/h
Temperature Range	-26.2°C to -1.97°C
Modeling Techniques	Decision Tree, Random Forest, XGBoost
Prediction Target	Endurance time (y) of anti-icing fluids

Table 3.1: Dataset Characteristics and Testing Conditions

practical realities, since controlled laboratory testing facilitates greater data throughput compared to opportunistic outdoor experiments. To alleviate potential bias resulting from disparate sample sizes, various solutions were implemented. Initially, all models conducted evaluation by K-Fold cross-validation to guarantee consistent performance across various data segments. Secondly, model generalization was methodically evaluated by training on artificial, natural, and hybrid datasets, while ensuring assessment on natural snow data to represent authentic operational settings. Ultimately, performance indicators were analyzed across datasets to ensure that predicted accuracy did not unduly favor the predominant artificial snow category. These techniques jointly mitigate sampling-induced bias and enhance the reliability of the proposed machine learning system for predicting endurance duration under varying snowfall circumstances.

In this study, three supervised ML algorithms were implemented to model the endurance time of anti-icing fluids: Decision Tree, Random Forest, and Extreme Gradient Boosting (XGBoost). These algorithms were selected due to their capacity to handle nonlinear interactions between environmental variables such as temperature, snow concentration, and fluid type. Ensemble models like Random Forest and XGBoost further enhance predictive performance by aggregating multiple weak learners and reducing overfitting [15, 20]. Among them, XGBoost demonstrated the most stable and accurate results during cross-validation and was therefore chosen as the final

model. Detailed theoretical descriptions of these models are available in prior literature [34, 51, 96].

Validation strategies for Enhanced Model Training

K-Fold cross-validation is one of the most popular and effective validation strategies employed. It is a technique leveraged to assess a model's generalization capability by partitioning the dataset into K subsets, commonly referred to as folds. During training, the model undergoes training for K iterations, utilizing $K-1$ folds for the training process while reserving the remaining fold for validation purposes [10]. The final performance is determined by calculating the average across all folds, providing a more dependable assessment of the model's accuracy. The formal representation of the cross-validation error calculation is as follows:

$$CV_{error} = \frac{1}{K}\sum_{i=1}^{K} E_i \tag{3.2}$$

In this context, E_i represents the error associated with the i-th validation fold. This methodology facilitates the identification of overfitting and guarantees the robustness of the model, particularly in the context of limited datasets.

Hyperparameter Tuning

Hyperparameter tuning is the process of selecting the optimal values for a ML model's hyperparameters. Bayesian Optimization and Genetic Algorithms are two of the most powerful techniques for hyperparameter tuning.

Bayesian Optimization represents a probabilistic approach to hyperparameter tuning, wherein a surrogate function, commonly modeled as a Gaussian Process (GP), is constructed to facilitate an efficient search process. The process achieves a balance between exploration and exploitation by determining the subsequent hyperparameter values through the application of an acquisition function, exemplified by Expected

Improvement (EI):

$$EI(x) = (\mu(x) - f(x^+))\Phi(Z) + \sigma(x)\phi(Z) \tag{3.3}$$

where $\mu(x)$ and $\sigma(x)$ are the mean and standard deviation of the GP, $f(x^+)$ is the best observed value, and Z is the standard normal variable. Bayesian Optimization presents significant advantages for the tuning of models such as XGBoost, Support Vector Machines, and deep learning networks, primarily by minimizing the number of required function evaluations [95].

Genetic Algorithms (GA) are derived from the principles of biological evolution and are employed to address complex optimization challenges. A population of candidate solutions is developed through successive generations utilizing selection, crossover, and mutation processes. The fitness of each individual is assessed through the application of a function denoted as $F(x)$, with subsequent generations being established in accordance with selection probabilities. The process of evolution can be represented as follows:

$$x_{new} = \alpha x_{parent1} + (1 - \alpha)x_{parent2} + \text{mutation} \tag{3.4}$$

where α serves as a parameter that regulates the crossover rate. Genetic algorithms demonstrate efficacy in contexts characterized by high-dimensional or non-continuous search spaces, including applications such as deep learning architecture search and feature selection. Although it incurs significant computational costs, this approach offers global optimization in contexts where gradient-based methods encounter difficulties [58, 97].

The amalgamation of these methodologies has the potential to improve the efficacy of predictive models, as they fulfill various roles, encompassing global search, hyperparameter optimization, and model assessment. The subsequent approach is characterized by a logical framework:

- Evaluate the robustness of the model and address overfitting by employing **K-Fold Cross-Validation**.
- Conduct an investigation into a diverse array of hyperparameter configurations and potential model architectures through the application of the **Genetic Algorithm**.
- Employ **Bayesian Optimization** to enhance the search process, concentrating on significant hyperparameter regions previously identified by the Genetic Algorithm.

In order to prevent overfitting while employing effective optimization methods, model building followed a methodical and sequential approach. K-Fold cross-validation (k = 10) was employed as the external validation mechanism to assess model generalization across different data partitions. This validation system utilized Genetic Algorithms to thoroughly explore the hyperparameter space, identifying significant regions without dependency on gradient-based assumptions. Bayesian Optimization was subsequently employed to refine hyperparameters inside these domains, focusing on local exploitation and convergence efficiency. Optimization was exclusively conducted on training folds, while validation folds remained unexposed during the tuning process. The final model's efficacy was evaluated using natural snow data to confirm its reliability in real-world operational contexts. This stratified method integrates exploration, exploitation, and validation, thus reducing the danger of overfitting while enhancing prediction stability.

The optimization of hyperparameters, coupled with the establishment of robust validation protocols, is complemented by the implementation of advanced feature engineering techniques, which are crucial for enhancing the predictive capability of the model. The integration of polynomial and nonlinear features, feature interactions, and feature transformations enables the model to effectively capture the complex relationships that are inherent in the physical behaviour of snow-fluid interactions. For example, polynomial expansions, including the squaring of temperature values, facilitate the model's ability to incorporate nonlinear effects, wherein variations in

temperature may not exert a uniform influence on endurance time [40]. In a similar vein, interaction terms (for instance, the product of temperature and snow rate) are essential for accurately depicting synergistic effects that arise solely when variables function simultaneously at extreme values [50]. Furthermore, log transformations significantly reduce skewness in distributions, hence improving numerical stability and promoting better model convergence during the optimization process [12].

The incorporation of Rule-based Features and Domain-Informed Features serves to augment the model's ability to accurately represent real-world phenomena. Binary indicators, which are informed by domain knowledge, such as a threshold-based cold condition flag, contribute to the identification of critical regimes in which fluid performance is adversely affected [68]. Furthermore, the calculated features, including the snow load obtained from density and mass, incorporate physics-based reasoning directly into the ML pipeline [10]. The integration of domain-specific factors improves the model's interpretability and realism, essential for evaluating endurance time predictions in safety-critical contexts, such as the efficacy of aircraft anti-icing fluids. The fusion of these feature engineering techniques with the optimization framework yields a model distinguished by accuracy and grounded in both physical and statistical validity.

Table 3.2 presents the descriptive statistics (minimum, maximum, mean, and standard deviation) of the main input and engineered features used in model development.

Feature	Min	Max	Mean	Std Dev
Temperature	−26.08	−1.98	−14.20	7.13
Rate	1.40	35.78	18.56	10.12
Temp_Squared	3.91	680.03	252.33	206.20
Rate_Squared	1.96	1280.19	446.83	383.62
Temp_Rate_Interaction	−930.35	−5.03	−266.63	209.79
Abs_Temperature	1.98	26.08	14.20	7.13
Is_Very_Cold	0.00	1.00	0.86	0.34

Table 3.2: Descriptive statistics (minimum, maximum, mean, and standard deviation) for the input and engineered features used in model development.

These statistical ranges confirm the dataset's heterogeneity across environmental

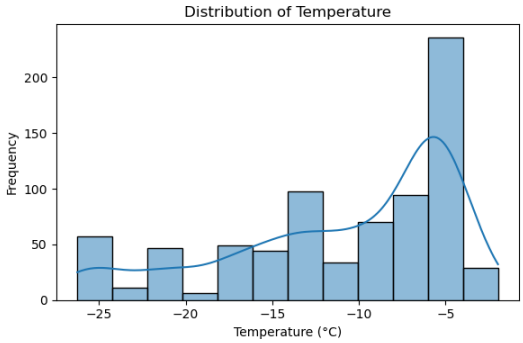
and operational conditions, supporting the model's ability to generalize under varied snow scenarios.

3.5 Results

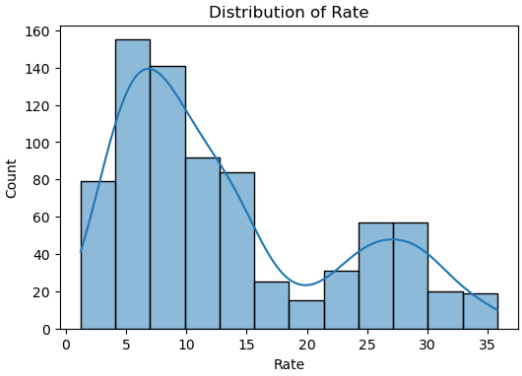
3.5.1 Data Analysis

Comprehending the distribution of critical variables, including temperature, snow application rate, and endurance time, is imperative prior to the interpretation of model outcomes. The configuration of these distributions elucidates fundamental trends within the dataset; for example, it indicates whether specific conditions were disproportionately represented or if outlier values might affect the model's performance. A skewed distribution in endurance time, for instance, suggests that a limited number of test scenarios result in considerably longer durations, which may signify optimal fluid performance under particular environmental conditions. Identifying these patterns serves to affirm the reliability of the data and establishes the framework for understanding the distribution of critical variables, including temperature, snow application rate, and endurance time, is crucial before interpreting the subsequent graphical results.

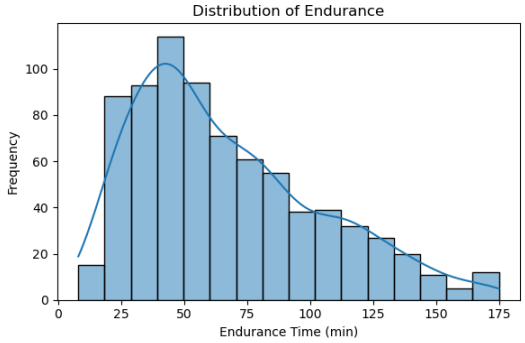
The distribution of key experimental variables, shown in Figure 3.6, offers valuable insight into the operating conditions under which the performance of anti-icing fluid was assessed. These distributions were derived from the combined natural and artificial datasets, thereby capturing a comprehensive range of testing scenarios. The temperature variable exhibits a right-skewed distribution, characterized by a significant concentration of observations near $-5\,^{\circ}\mathrm{C}$. This suggests that a significant percentage of the tests were conducted in proximity to the upper threshold of standard freezing conditions, wherein the efficacy of the fluid is anticipated to be of greater importance. The rate of snow deposition demonstrates a bimodal and positively skewed distribution, characterized by primary and secondary peaks located approximately at 5 to 10 units and 25 to 30



(a) Distribution of Temperature



(b) Distribution of Rate



(c) Distribution of Endurance

Figure 3.6: Distribution of the Variables

units, respectively. This indicates that the dataset encompasses distinct clusters that correspond to scenarios of low and high snowfall intensity. The distribution of fluid endurance time exhibits a right-skewed pattern, with the majority of values concentrated within the range of 30 to 70 minutes, accompanied by a gradual decrease in frequency as endurance time extends. This pattern underscores that, although the majority of fluids failed within a discernible time frame, a specific subset demonstrated prolonged endurance under particular environmental conditions. The distributions presented collectively illustrate the variability and heterogeneity inherent in real-world anti-icing conditions, thereby providing a foundational basis for the training of the predictive models.

Figure 3.7 illustrates the differences in endurance time distributions among various fluid types and snow conditions, specifically emphasizing produced snow. The median

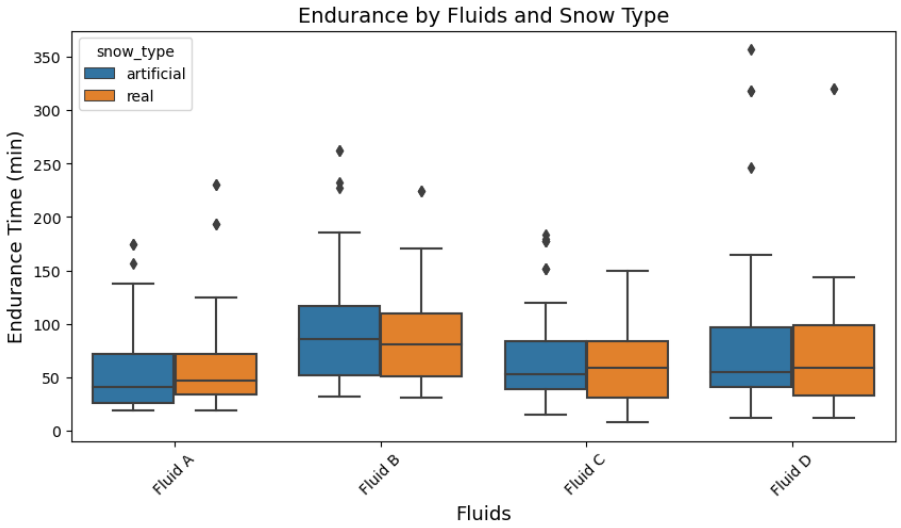


Figure 3.7: Boxplot Endurance time by fluids and snow type

endurance times exhibit significant stability across various snow types; however, controlled environments reveal more pronounced outliers and wider interquartile ranges, particularly for Fluids B and D. Fluid B exhibits endurance durations exceeding 200 minutes on artificial snow, suggesting notable variability rather than a uniform trend of enhancement. The visual evidence suggests that dispersion increases under controlled snow conditions, likely due to laboratory-specific factors such as flake uniformity or deposition rate. The observed variation may indicate that the laboratory setting successfully replicated a broader spectrum of natural variables. This observation does not imply that artificial snow inherently improves performance. Rather, the variability may reflect a wider range of environmental scenarios simulated in the laboratory setting. It is important to clarify that the identification of subsets showing prolonged endurance cannot be made solely from Figure 3.7; further statistical or model-based analysis would be required. Therefore, interpretations based solely on interquartile range should be made cautiously.

The correlation between endurance time and the rate of snow deposition demonstrates a distinct inverse relationship across all fluid categories. The scatter plot, Figure 3.8, together with logarithmic regression fits, demonstrates a steady decrease in en-

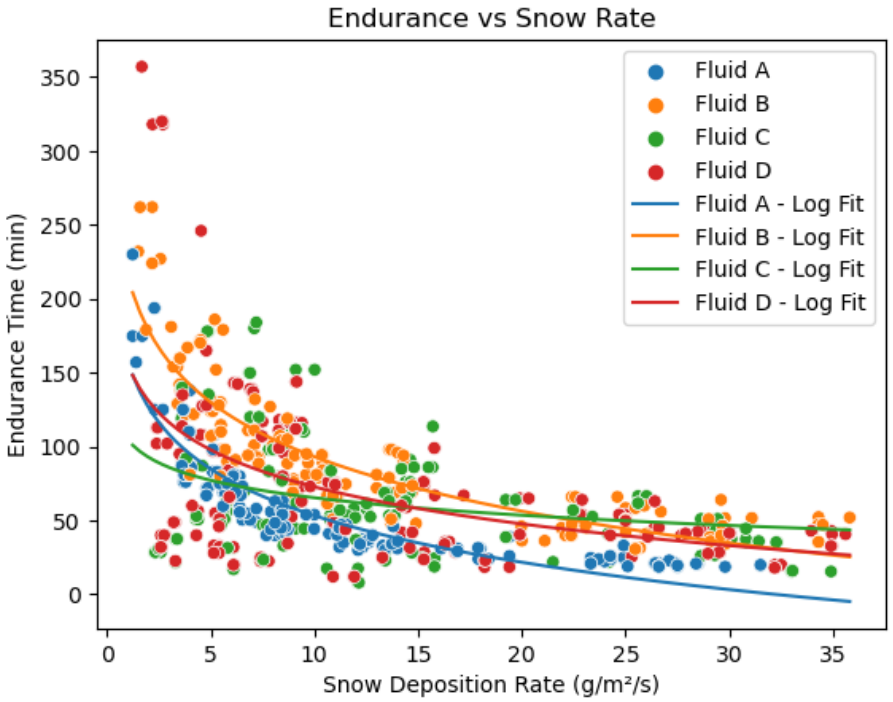


Figure 3.8: Endurance time vs. snow deposition rate by fluids from temperatures 0 to -20°C

durance as snow rate increases throughout the temperature range of 0 to -20°C. This observation suggests that increased snowfall contributes to an expedited occurrence of fluid failure. Although all fluids adhere to this overarching trend, the rate of decline exhibits variability, with Type A and Type D fluids demonstrating a more pronounced decrease in endurance as the rates increase. Conversely, Type B demonstrates a tendency to sustain comparatively elevated endurance values in low to moderate snowfall conditions, which is consistent with prior observations regarding its wider performance spectrum. The existence of high-endurance outliers at low snow rates, particularly for Type B and Type D, underscores the potential for prolonged protection during light snowfall conditions. The results of this study substantiate the significance of the snow rate as an essential predictive variable for the modelling of fluid endurance, This helps differentiate of fluid formulations in relation to environmental load conditions. This variability aligns with findings from recent snow impact studies, which highlight the

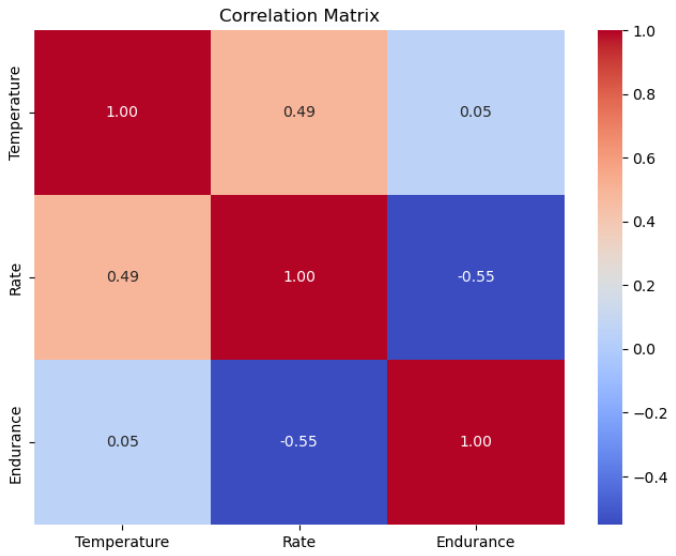


Figure 3.9: Matrix of Correlations: Temperature, Snow Rate, Endurance

non-linear influence of snow particle trajectory and surface restitution on retention and fluid endurance [45, 46].

The correlation matrix (Figure 3.9) depicts the relationships among key environmental and response variables. A moderate negative correlation (r = -0.55) was observed between snow deposition rate and endurance duration, indicating that higher snowfall rates reduce fluid efficacy. The temperature and snow rate exhibit a moderate positive correlation (r = 0.49), while the temperature and endurance time show a weak association (r = 0.05). The robustness of this link will be further evaluated using the SHAP analysis in Section 3.5.3.

The exploratory analysis of variable distributions and their interrelationships reveals significant patterns that inform the modeling strategy. The considerable influence of snow deposition rate on endurance time, coupled with the observed variability among various fluid and snow types, highlights the necessity for models capable of accurately depicting the complex, non-linear interactions inherent in the data. The findings articulated in this study not only corroborate the experimental dataset but also offer direction for the selection and assessment of appropriate ML methodologies. Building upon the analytical framework previously established, the subsequent section assesses

the predictive efficacy of three supervised learning models, Decision Tree (DT), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), across different test conditions, with the aim of identifying the most effective approach for estimating fluid endurance time in relation to various environmental loads.

3.5.2 Prediction Performance

In order to evaluate the predictive capability of different algorithms for estimating endurance time, an assessment was conducted utilizing three supervised learning models: Decision Tree (DT), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). The unrefined models, developed without feature engineering or post-processing, were evaluated across natural, artificial, and hybrid datasets. Table 3.4 delineates the performance metrics utilized: the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), offering an exhaustive assessment of the accuracy and reliability of each model. It is important to note that, for each model, separate training was conducted on the natural, artificial, and combined datasets. The performance metrics presented in tables 3.4,3.5,3.6 correspond exclusively to predictions made on real (natural) snow data, allowing for a consistent evaluation of model generalization under authentic environmental conditions.

The configuration of hyperparameters for each model, along with training time and cross-validation details, is summarized in Table 3.3. These values were obtained either from default settings (for DT and RF) or through optimization procedures such as Bayesian and genetic algorithms in the case of XGBoost.

Evaluation Metrics

The coefficient of determination (R^2) is used to measure how well a prediction model fits the actual data. A value closer to 1 indicates that the model captures most of the variability in the target variable, while values closer to 0 suggest poor predictive

Hyperparameter	Decision Tree [51]	Random Forest [15]	XGBoost [20]
Max Depth	5	10	5
n Estimators	–	100	100
Learning Rate	–	–	0.1
Subsample	–	–	1.0
Colsample Bytree	–	–	1.0
Alpha (α)	–	–	0.0
Lambda (λ)	–	–	1.0
Random State	42	42	42
Cross-Validation Folds (k)	10	10	10
Training Time (s)	0.41	2.23	3.98

Table 3.3: Hyperparameter configurations and training details for each ML model. XGBoost parameters were optimized using Bayesian and genetic algorithms.

power. RMSE, which gives higher weight to larger errors, indicates how concentrated the prediction errors are around the true values; a lower RMSE reflects higher accuracy. Likewise, MAE measures the average absolute deviation between expected and actual values, enhancing its interpretability in real units. Lower MAE values are advantageous, signifying that the model exhibits reduced average prediction errors.

Model Selection

	DT			RF			XGBoost		
Dataset	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
Natural	0.899	11.193	4.109	0.887	14.302	6.745	0.899	11.195	4.033
Artificial	0.897	11.795	4.409	0.888	13.826	6.340	0.896	11.953	4.688
Combined	0.898	11.591	4.290	0.886	13.874	6.440	0.896	12.105	5.135

Table 3.4: Model Comparison — Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) for Decision Tree (DT), Random Forest (RF), and XGBoost (XGBoost)

Gradient Boosting, namely XGBoost, was selected for its ability to enhance fluid behavior identification [87] and its expertise in modeling complex, non-linear interactions between environmental variables and endurance time. Cross-validation and comparative assessments of Decision Tree and Random Forest models confirmed their robustness.

The consolidated dataset analysis revealed that XGBoost attained a $R^2 = 0.896$, an RMSE = 12.105, and a MAPE = 5.135. While both Decision Tree and Random Forest occasionally attained slightly improved R^2 or reduced RMSE values (see Table 3.4), XGBoost consistently produced lower MAE, demonstrating greater prediction stability. The model exhibited strong performance after implementing advanced feature engineering and optimization, confirming its selection for further refinement and interpretability assessment, as detailed in Section 3.5.3.

Feature Engineering and Model Refinement

In order to improve the model's representation of physical interactions, several features were added: `Temp_Squared` (to describe the nonlinear temperature–endurance relationship), `Temp` $\times$ `SnowRate` (to account for the combined effects of temperature and snow deposition rate), `Log_SnowAge` (to rectify distributional skewness), `Is_Cold` (a binary indicator for essential temperature thresholds), and `Density_Mass` (a physics-informed feature that indicates snow load).

The inclusion of these engineered variables enhanced the input space of the XGBoost model. As shown in Table 3.5, the updated model achieved an $R^2 = \mathbf{0.897}$, an RMSE = **11.908**, and MAPE = **4.972** on the combined dataset, compared with the previous $R^2 = \mathbf{0.896}$, RMSE = **12.105**, and MAPE = **5.135**. Considering the slight numerical improvements, these findings validate that feature engineering helped improve model robustness and interpretability. The diminished prediction error and more consistent metrics suggest that improved physical representations, specifically snow load and temperature-snow interactions, augmented the model's ability to accurately reflect essential endurance dynamics. Subsequent enhancements may augment precision and applicability.

The current model exhibits considerable predictive power; however, it is crucial to further reduce residual errors to enhance resilience and generalizability, particularly in

	DT			RF			XGBoost		
Dataset	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
Natural	0.899	11.193	4.099	0.887	14.229	6.702	0.899	11.194	4.123
Artificial	0.897	11.795	4.409	0.883	14.548	6.716	0.897	11.892	4.648
Combined	0.898	11.590	4.290	0.888	13.288	6.787	0.897	11.908	4.972

Table 3.5: Model Comparison with Feature Engineering — for Decision Tree (DT), Random Forest (RF), and XGBoost (XGBoost)

safety-critical applications. This project will employ sophisticated optimization methods, such as Bayesian optimization and evolutionary algorithms, along with an efficient K-Fold cross-validation strategy to improve hyperparameter tuning, avoid overfitting, and minimize prediction error.

The R^2 and RMSE values displayed in Tables 3.4 and 3.5 provide complementary insights into model performance and reliability. High R^2 values (about 0.89–0.90 across datasets) indicate that the models accurately capture the primary variability in endurance time affected by environmental and operational factors. The minor variations in R^2 among natural, artificial, and mixed datasets suggest that the essential physical correlations learned by the models remain consistent across many testing environments.

RMSE values, expressed in minutes, denote the absolute deviation between predicted and actual endurance durations. In aviation ground operations, endurance times generally extend over several tens of minutes; hence, RMSE values of roughly 10 to 12 minutes align with forecast errors that remain within the permissible operating parameters established in SAE ARP5485B. The decrease in MAE noted for XGBoost compared to Decision Tree and Random Forest models further signifies enhanced stability and less sensitivity to outliers. Collectively, these measures indicate that the chosen modeling methods attain a balance between goodness-of-fit and operationally significant accuracy.

	10-Fold CV			Bayesian			10-Fold CV + Bayesian			Model Refinement Method		
Training Dataset	Artificial	Natural	Combined	Artificial	Natural	Combined	Artificial	Natural	Combined	Artificial	Natural	Combined
R^2	0.90	0.90	0.90	0.92	0.81	0.79	0.92	0.91	0.92	0.92	0.91	0.92
RMSE	11.36	12.07	11.70	10.10	15.71	8.75	10.12	11.60	10.29	10.10	11.29	10.65
RMSPE(%)	16.99	18.21	17.48	17.51	18.81	18.02	15.13	17.51	15.38	15.11	17.04	15.91
MSE	129.16	145.69	136.90	629.49	661.48	351.75	102.45	134.59	105.91	102.11	127.46	113.33
MAE	7.16	6.84	7.86	14.87	8.04	13.56	5.63	6.41	5.94	5.78	6.52	6.45
MAPE(%)	10.71	10.33	8.88	23.07	8.16	22.50	8.64	9.68	8.88	8.64	9.83	9.64

Table 3.6: Performance Metrics for Optimization Techniques: R^2, RMSE, RMSE (%), MSE, MAE, MAE (%)

Optimized Models Prediction Performance

The model went through additional optimization aimed at improving generalization and forecast accuracy, utilizing the feature collection that was developed earlier. The optimization phase focused on adjusting hyperparameters rather than introducing new variables, maintaining the same physical and statistical characteristics described in the preceding section. This included nonlinear and interaction terms: `Temp_Squared` (to capture the nonlinear temperature–endurance relationship), `Temp` $\times$ `SnowRate` (for the joint influence of temperature and snow deposition rate), `Log_SnowAge` (to correct distributional skewness), `Is_Cold` (a binary indicator for critical temperature thresholds), and `Density_Mass` (a physics-informed feature representing snow load).

The enhanced feature representation provided a solid foundation for the optimization phase, enabling the XGBoost algorithm to discern more complex relationships between environmental variables and endurance time. Four optimization methodologies were utilized to evaluate model performance across artificial, natural, and hybrid datasets. The initial method employed 10-fold cross-validation (10-Fold CV), selecting $k = 10$ for its equilibrium between bias, variance, and computational efficiency [29]. The second employed Bayesian optimization, the third merged 10-Fold cross-validation with Bayesian optimization, and the fourth amalgamated genetic algorithms with both methods—designated as the Model Refinement Method.

The evaluation of model generalization in artificial and natural snow environments was conducted by training and validating models on distinct datasets, as well as on a

unified dataset. The results in Table 3.6 indicate that models trained on the combined dataset exhibit consistently robust performance when assessed on natural snow data, with minimal deterioration compared to artificial snow forecasts. The optimized XGBoost model has similar RMSPE and MAPE values across testing scenarios, suggesting that the established relationships are not confined to controlled laboratory settings. Artificial snow data offer more consistent sampling and reduced variability, but real snow data demonstrate more stochasticity resulting from environmental changes. The model's capacity to uphold expected accuracy under these circumstances validates its robust generalization and endorses the utilization of synthetic snow data as a dependable adjunct to natural snowfall testing within a cohesive machine learning framework.

The study targets a prediction accuracy within 10% of actual endurance time, consistent with industry standards in SAE Aerospace Recommended Practice ARP5485B [78]. The integrated optimization framework, incorporating Bayesian optimization, and K-Fold cross-validation, yields superior overall performance. The consolidated dataset analysis revealed that XGBoost achieved a coefficient of determination (R^2) of 0.92, an RMSPE of 15.38%, and a MAPE of 8.88% after optimization, surpassing alternative methodologies. This hybrid methodology effectively succeeded by Bayesian optimization to concentrate on high-performance areas. The slight disparity in RMSE between artificial (10.12) and natural datasets (11.60) affirms robust generalization facilitated by the K-Fold framework.

Bayesian Optimization alone performs well on artificial data ($R^2 = 0.92$) but exhibits limited generalization, as reflected by its high MAPE (23.07%). The Genetic Algorithm shows larger variability (MAE = 8.52 for artificial and 6.52 for natural data), whereas the integrated approach balances errors across all datasets (MAE = 5.78, 6.52, and 6.45, respectively). The 10-Fold Cross-Validation baseline produces reliable results (RMSE = 11.70, MAPE = 8.88%) but does not fully exploit the parameter space. Overall, the comprehensive optimization framework reduces RMSE by roughly 9% and

increases R^2 from 0.90 to 0.92, demonstrating effective overfitting control and improved predictive stability under diverse snow conditions.

3.5.3 Model Explainability using SHAP Analysis

Comprehending the influence of each factor on model predictions is crucial for understanding outcomes and directing further experiments. Upon determining XGBoost as the optimal model, feature significance was assessed utilizing SHapley Additive exPlanations (SHAP) values [62].

Before conducting SHAP analysis, continuous input variables (e.g., temperature, snow deposition rate, and designed interaction terms) were utilized in their original physical units, as tree-based models like XGBoost are unaffected by monotonic feature scaling. During model training, categorical variables denoting anti-icing fluid type were encoded via binary indicator (one-hot) encoding. SHAP values were derived directly from the trained XGBoost model via the TreeExplainer framework, hence ensuring alignment between the prediction model and the explainability study. This method enables the interpretation of both continuous and categorical characteristics in a physically significant way, without the bias introduced by scaling.

Figure 3.10 illustrates the SHAP summary plot, emphasizing that the interaction between temperature and snow deposition rate (`Temp Rate Interaction`) is the predominant factor influencing endurance duration. Elevated snowfall rates coupled with temperature fluctuations substantially influence fluid dynamics. Additional significant contributors comprise `Rate`, `Fluid B`, and engineered variables such as `Rate Squared` and `Temp Squared`, highlighting the significance of non-linear and interaction effects.

SHAP waterfall plots were created for three representative cases of endurance time: low, median, and high, to effectively demonstrate individual forecasts (Figure 3.11). These charts illustrate the positive or negative contributions of each attribute to certain predictions. In low-endurance settings, significant adverse effects from low temperatures

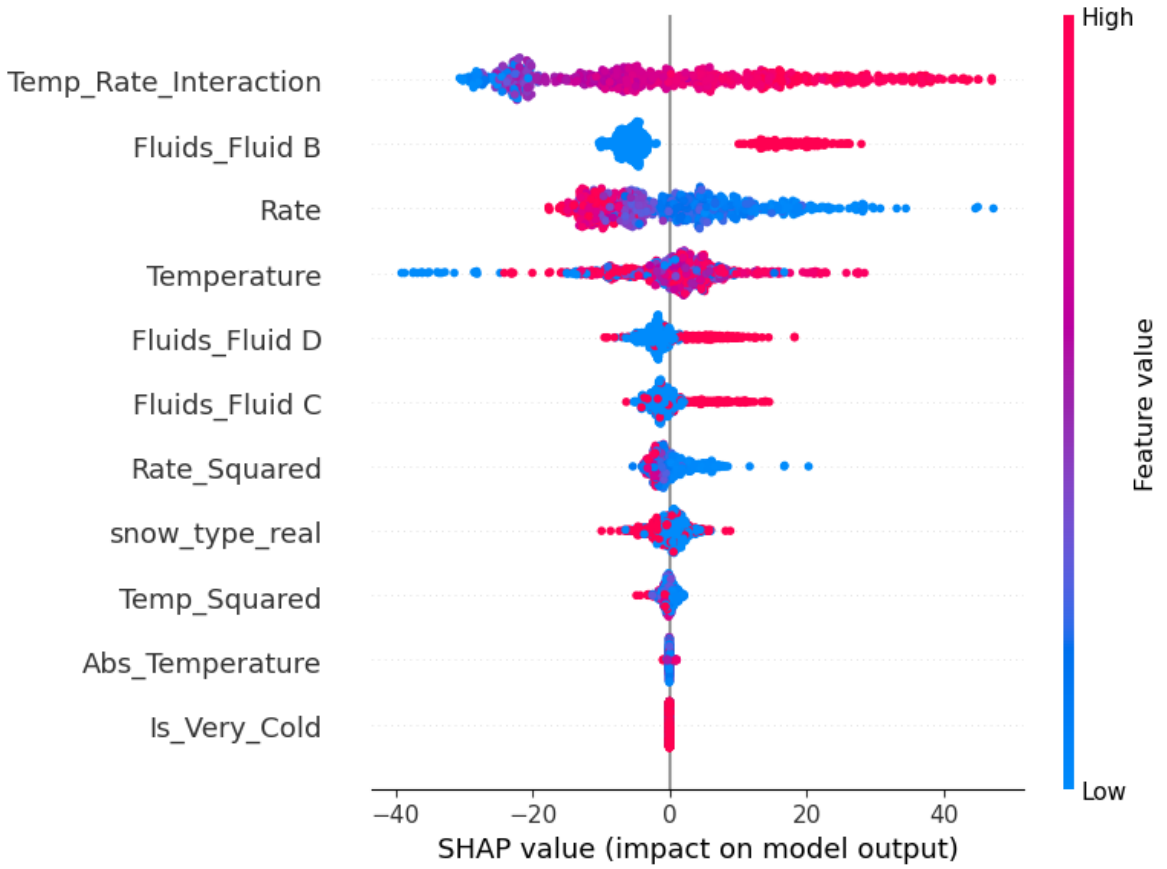
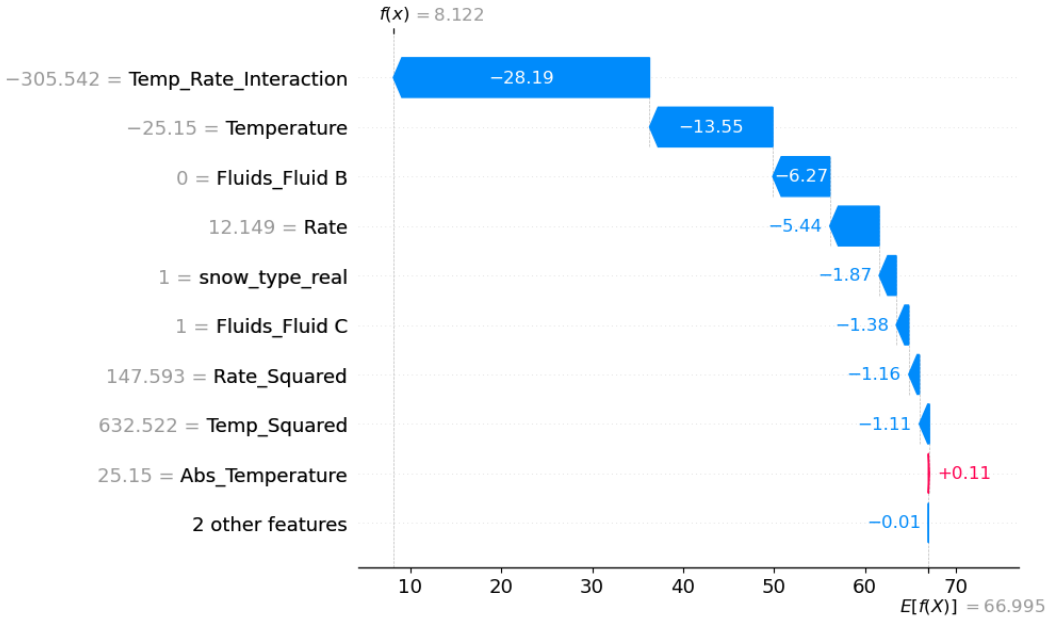


Figure 3.10: SHAP summary plot for the tuned XGBoost model trained on combined natural and artificial snow data.

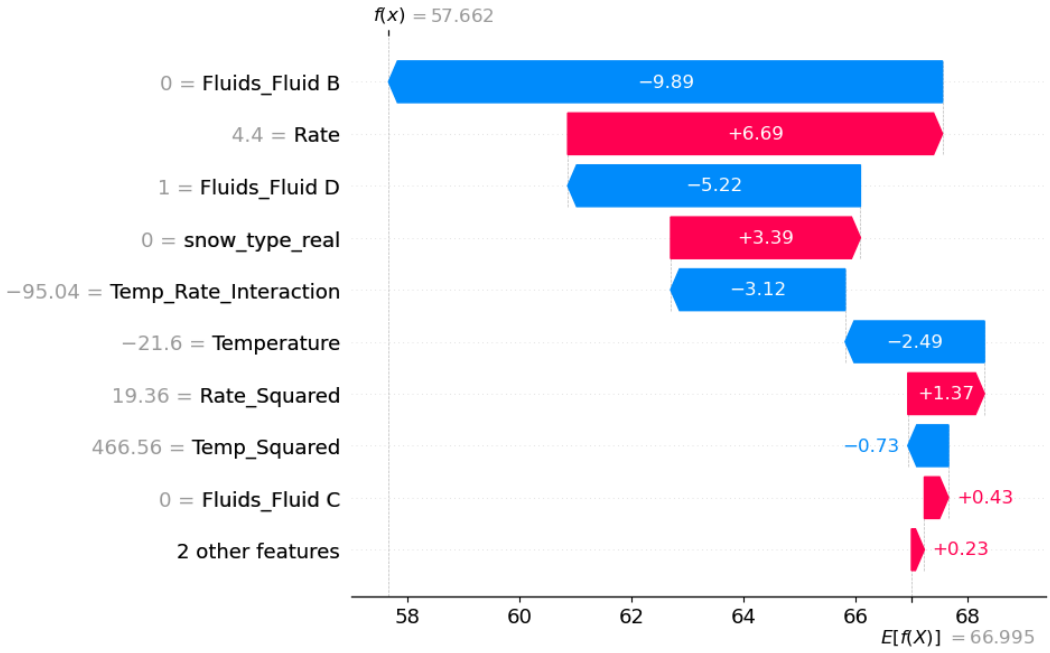
and elevated deposition rates prevail. Conversely, high endurance forecasts are influenced by moderate snowfall rates, conducive temperatures, and beneficial interactions among environmental variables and fluid characteristics.

SHAP analysis indicates that endurance is influenced by a limited set of interpretable, physics-based variables, particularly when accounting for non-linear interactions between temperature and snow.

The SHAP analysis indicates that endurance time is primarily affected by a combination of environmental, operational, and fluid-related factors. In both natural and artificial snow datasets, the primary influencing factors are the snow deposition rate and external temperature, with heightened snowfall intensities and diminished temperatures consistently impacting fluid failure. Operational characteristics are represented by



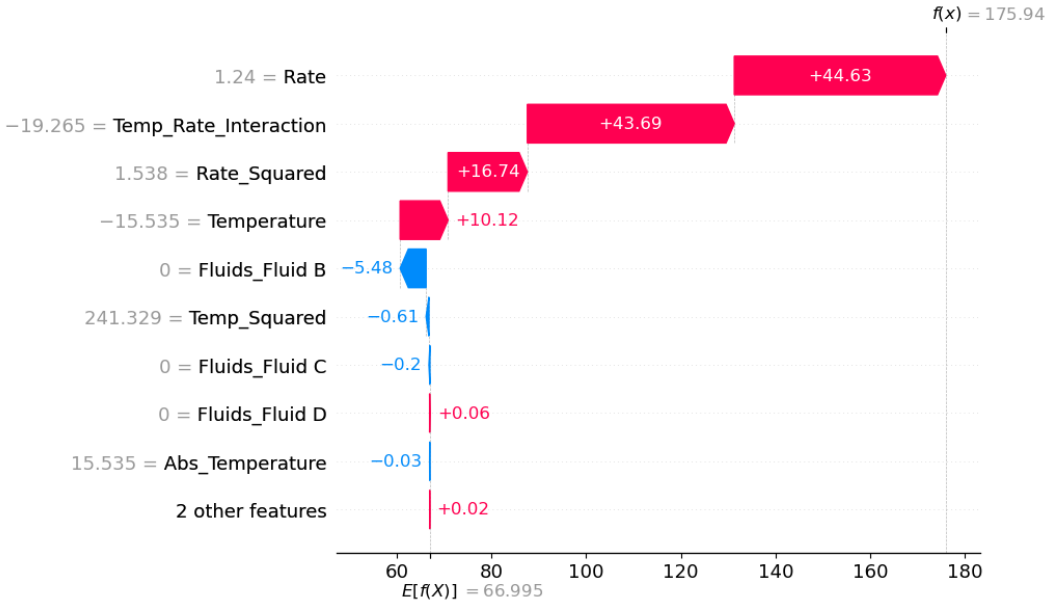
(a) SHAP Waterfall Plot for the Lowest Predicted Endurance Time



(b) SHAP Waterfall Plot for the Median Predicted Endurance Time

Figure 3.11: SHAP Waterfall Plots Illustrating Feature Contributions for Representative Low, Median, and High Endurance Time Predictions.

interaction terms, particularly the temperature-snow rate interaction, which highlights the non-linear correlation between thermal conditions and snow loading. Fluid-specific



(c) SHAP Waterfall Plot for the Highest Predicted Endurance Time

Figure 3.11: SHAP Waterfall Plots (continued).

properties also affect endurance performance, revealing differences in formulation reactions to the same environmental stimuli. The relative importance of these factors remains consistent in both natural and artificial snow environments; however, research on manufactured snow reveals a slightly less variability in SHAP contributions due to controlled deposition rates and uniform snow properties. In contrast, natural snow data demonstrate heightened variability, indicating a broader range of environmental interactions. The findings indicate that machine learning models accurately depict physically significant systems while maintaining robustness across many testing scenarios.

3.6 Discussion

This study illustrates that machine learning can yield precise and comprehensible forecasts of anti-icing fluid longevity under various climatic conditions. By merging natural and artificial snow datasets for training, the model bridges the disparity between controlled laboratory experiments and real-world settings.

The integration of XGBoost, Bayesian optimization, and K-Fold cross-validation improved predictive accuracy and mitigated overfitting, particularly under fluctuating snow and temperature conditions. These findings align with recent work emphasizing the significance of snow-surface interactions and energy dissipation processes in endurance behavior [43, 67].

Moreover, feature engineering significantly influenced the model's performance. The inclusion of polynomial terms, interaction effects, and domain-specific indicators (such as `Temp Rate Interaction` and `Density Mass`) improved the model's physical realism and reduced residual errors. These properties allowed the model to generalize across various contexts while maintaining interpretability.

The SHAP study further shows that endurance time is primarily influenced by the interaction of snow deposition rate and temperature, together with fluid-specific characteristics. The waterfall diagrams illustrated how these interactions might either accelerate or delay fluid collapse. Most operational scenarios often occur within a moderate endurance range, where environmental factors exert a measurable but manageable influence.

Despite its strong performance, the model is limited by the relatively small size of the real snow dataset, which affects its ability to encounter rare environmental edge cases. Future studies should expand natural testing and examine additional physical properties, including surface roughness and phase transitions of precipitation. The application of generative models like VAEs and GANs may assist in data augmentation.

The prevailing influence of snow deposition rate and ambient temperature on endurance duration aligns with recognized physical principles that dictate anti-icing fluid failure. Heightened snowfall intensity expedites fluid dilution and saturation, whereas diminished temperatures impair melting efficiency and augment snow accumulation, resulting in an earlier decline in fluid efficacy. Comparable dependencies have been documented in experimental and numerical investigations of snow-surface interactions

and fluid degradation in winter conditions, wherein the non-linear interplay between thermal and hydrodynamic effects is pivotal to endurance behavior [45, 67, 92].

From a modeling standpoint, the robust efficacy of tree-based machine learning methods demonstrates their capacity to include non-linear relationships and interaction effects that are challenging to depict through empirical correlations alone. Prior research utilizing machine learning for aircraft icing and anti-icing systems has shown that ensemble techniques like XGBoost provide an advantageous equilibrium between predictive precision and interpretability under conditions of limited data availability [54]. The uniformity of model performance across artificial and natural snow datasets corroborates existing research that indicates controlled laboratory data, when suitably integrated, can augment prediction robustness without compromising physical relevance.

3.7 Conclusion

This study presented a machine learning methodology to forecast the endurance duration of airplane anti-icing fluids, integrating artificial and natural snow data to reconcile differences between controlled experiments and actual environmental conditions. Utilizing XGBoost in conjunction with sophisticated optimization methods, including Bayesian optimization, and cross-validation, the model attained robust predictive efficacy ($R^2 = 0.92$, RMSPE = 15.38%, MAPE = 8.88%). Engineered characteristics, such as temperature-snow rate relationships and domain-informed variables, enhanced interpretability and congruence with physical behavior. SHAP studies confirmed that fluid degradation is affected by temperature and the interplay between snowfall and deposition. These results highlight AI's potential to enhance operational safety and refine fluid decision-making in winter aviation settings.

3.8 Remerciements

This work was supported by Transport Canada. Also like to express their sincere gratitude to the Anti-icing Materials International Laboratory (AMIL) for providing the facilities, technical support, and valuable insights that made this research possible.

CHAPITRE 4

GÉNÉRATION DE DONNÉES SYNTHÉTIQUES POUR AMÉLIORER LA MODÉLISATION PRÉDICTIVE DU TEMPS D'ENDURANCE DES FLUIDES ANTIGIVRAGE EN CONDITIONS HIVERNALES

Avant-propos

État de l'acceptation : En préparation pour soumission.

Journal : À déterminer

Titre en anglais : Synthetic data generation to improve predictive modeling of anti-icing fluid endurance time under winter conditions

Titre en français : Génération de données synthétiques pour améliorer la modélisation prédictive du temps d'endurance des fluides antigivrage en conditions hivernales

4.1 Résumé

Ce travail propose l'utilisation de modèles génératifs pour améliorer la modélisation prédictive du temps d'endurance des fluides antigivrage appliqués sur les surfaces aéronautiques. L'étude se concentre sur la génération de données synthétiques à partir d'un ensemble de données tabulaire de petite taille, caractéristique des essais expérimentaux en conditions hivernales. Plusieurs configurations de réseaux antagonistes génératifs (GAN), incluant la régularisation par Dropout, l'intégration de la normalisation par lots

(Batch Normalization) et l'augmentation du nombre d'itérations d'entraînement, sont comparées. L'évaluation des données synthétiques repose sur un cadre multi-critères combinant des métriques de similarité statistique (divergences de Kullback–Leibler, Jensen–Shannon et distance de Wasserstein), la préservation des dépendances multivariées (corrélation et information mutuelle) ainsi que l'utilité prédictive via l'approche Train-on-Synthetic, Test-on-Real (TSTR). Les résultats montrent que l'intégration de la Batch Normalization permet d'améliorer significativement la stabilité de l'entraînement et la qualité des données générées, conduisant à une meilleure préservation des structures statistiques et à des performances prédictives accrues. Cette étude met en évidence le potentiel des techniques de génération de données synthétiques pour pallier les limitations liées à la taille réduite des ensembles de données expérimentales dans les applications aéronautiques hivernales.

4.2 Abstract

Generative modeling of small structured engineering datasets remains challenging due to adversarial instability and limited representation capacity. This study evaluates the impact of architectural stabilization strategies on synthetic data generation for a tabular aviation dataset related to anti-icing fluid endurance performance under winter precipitation conditions. Three Generative Adversarial Network (GAN) configurations—Dropout regularization, Batch Normalization, and extended training duration, were systematically compared using a multi-criteria validation framework incorporating divergence based metrics (KL, Jensen–Shannon, Wasserstein), multivariate dependency preservation (correlation and mutual information Frobenius norms), higher-order moment alignment (skewness and kurtosis error), and predictive utility via Train-on-Synthetic, Test-on-Real (TSTR) evaluation.

Results demonstrate that Batch Normalization yields the most stable adversarial

training dynamics and achieves superior statistical alignment with the real dataset, including the lowest mean Wasserstein distance (7.741) and minimal distortion of correlation structure. Although synthetic datasets did not fully replicate baseline predictive performance, the BatchNorm configuration preserved structural relationships more effectively than Dropout or extended training alone. These findings underscore the importance of architectural stabilization in GAN-based augmentation of small aerospace datasets and establish Batch Normalization as a preferred strategy for distribution-preserving generative modeling in weather-constrained aviation performance studies.

4.3 Introduction

Winter precipitation required the application of anti-icing fluids during aircraft ground operations in order to avoid the buildup of snow and ice on critical aerodynamic surfaces. The protective efficiency of these fluids is measured by endurance time, or holdover time, which denotes the period during which the fluid retains effective anti-ice properties despite intense precipitation. Accurate prediction of endurance duration is essential for ensuring operational safety, reducing departure delays, and optimizing fluid usage during winter operations.

Endurance time testing is generally conducted under natural or artificial snow conditions in accordance with standardized aerospace protocols. Natural snow experiments are subject upon meteorological conditions, expensive, and time-limited. Artificial snow testing improves stability and compliance with environmental regulations; yet, it may not fully replicate the variability and microstructural characteristics of real precipitation. Consequently, experimental datasets are often constrained in size and frequently exhibit imbalance, presenting challenges for data-driven modeling techniques. In small-sample scenarios, machine learning models demonstrate heightened vulnerability to distributional bias and constrained variance representation [31, 98].

The use of synthetic snow data in this study is motivated by the inherent limitations of experimental datasets obtained under natural and controlled laboratory conditions. Natural snow testing is highly dependent on weather availability, leading to limited data collection opportunities and restricted variability. Similarly, artificial snow experiments, while more controllable, still involve significant operational constraints and costs.

As a result, the available dataset remains relatively small and may not fully capture the range of environmental conditions required for robust machine learning model training. To overcome these limitations, generative models such as Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN) are employed to generate synthetic data that preserves the statistical characteristics of real snow conditions [33, 48]. This approach enables dataset augmentation, improves model generalization, and supports the evaluation of anti-icing fluid performance under a broader range of scenarios [85, 99]. This synthetic data generation framework forms the foundation of the methodology developed in this chapter.

Modern developments in machine learning have facilitated the predictive modeling of tricky nonlinear engineering phenomena with limited experimental data. Ensemble methods such as Extreme Gradient Boosting (XGBoost) have demonstrated considerable effectiveness in regression problems involving structured environmental datasets [19]. However, the robustness of predictions and the capacity to generalize are heavily dependent on the size and variability of the dataset. When training data is limited, models may overfit dominant patterns and find it difficult to generalize to atypical climatic conditions.

Synthetic data-generation methods have received increased focus in engineering fields to tackle data shortage challenges. Generative models, particularly Generative Adversarial Networks (GANs), are capable of learning the fundamental probability distribution of a real dataset and generating statistically significant artificial samples [33]. Synthetic data augmentation has been successfully employed in aerospace fault

diagnosis, materials characterization, and reliability analysis, especially in contexts where the gathering of experimental data is expensive or operationally constrained [84, 94]. These methodologies offer a viable avenue to enhance model robustness without necessitating further physical experiments.

The dataset used for synthetic data generation consists of input variables describing environmental and physical conditions affecting anti-icing fluid performance. These include temperature, snow precipitation rate, and additional experimental parameters. The target variable corresponds to the endurance time of the fluid.

Each observation can be represented as a mapping:

X = T, r, ... $\rightarrow$ y

where T denotes temperature, r represents snow intensity, and y is the corresponding endurance time. The objective of the generative models is to learn the joint distribution of these variables in order to generate realistic synthetic samples [33, 99].

Theoretical Background of Generative Models

Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) are deep generative models that operate through a minimax game using two neural networks: a generator G and a discriminator D [33]. The generator establishes a correspondence between a latent variable space $z \sim p_z(z)$ and the data space x, generating synthetic samples $G(z)$ that seek to replicate the actual data distribution $p_{data}(x)$. The discriminator seeks to differentiate between authentic samples and generated samples by estimating the probability that a given input originates from the real data distribution.

The training procedure is formulated as the following minimax optimization problem:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]. \tag{4.1}$$

Under ideal convergence conditions, the generator distribution $p_g(x)$ approximates the true data distribution $p_{data}(x)$, rendering the discriminator unable to distinguish between real and synthetic samples. From a probabilistic perspective, GAN training minimizes the Jensen–Shannon divergence between the two distributions [33].

However, GAN training is known to exhibit instability, including issues such as mode collapse and vanishing gradients, particularly when applied to small or structured datasets [6, 82]. To address these challenges, several stabilization strategies have been proposed, including architectural regularization techniques such as batch normalization, dropout, and extended training procedures.

Autoencoders (AE)

Autoencoders are neural network architectures designed to learn compressed representations of data through an encoder–decoder structure. The encoder maps the input data x into a latent representation z, while the decoder reconstructs the original input from this latent space. The model is trained by minimizing a reconstruction loss, typically measured using mean squared error, encouraging the latent space to capture the most informative features of the data distribution [36].

Although autoencoders are effective for dimensionality reduction and feature extraction, they do not explicitly model the underlying data distribution. As a result, their capacity to generate diverse and realistic synthetic samples is limited compared to probabilistic generative models.

Variational Autoencoders (VAE)

Variational Autoencoders (VAEs) extend the autoencoder framework by introducing a probabilistic interpretation of the latent space. Instead of learning a deterministic mapping, the encoder approximates a posterior distribution $q(z|x)$, while the decoder reconstructs the data by sampling from this latent distribution. The training objective is defined as the maximization of a variational lower bound, combining a reconstruction term and a regularization term based on the Kullback–Leibler divergence [48, 75].

Compared to GANs, VAEs offer more stable training dynamics due to their single-objective optimization and well-defined probabilistic framework. However, they may produce overly smooth samples and may struggle to accurately represent complex multimodal distributions, particularly in structured tabular datasets with limited variability.

4.4 Literature Review

4.4.1 GANs for generating meteorological and climate data

Generative Adversarial Networks (GANs) have become a prominent approach for learning complex, high-dimensional environmental data distributions and producing realistic synthetic samples. Early work in meteorological generative modeling focused on spatial and spatiotemporal refinement (downscaling), leveraging the connection between image super-resolution and geophysical fields. Recent studies demonstrate that GAN-based downscaling can improve realism in precipitation fields by better capturing spatial structure and intensity distributions compared with purely deterministic regressors, particularly for extremes and fine-scale variability [30, 35].

Precipitation has been a major target application due to its intermittency, non-Gaussianity, and strong multi-scale behavior. GAN-based approaches have been proposed

to generate or downscale rainfall fields using conditional architectures that map low-resolution inputs to higher-resolution outputs while preserving intensity statistics and spatial patterns [39, 41, 53]. The recent RainScaleGAN framework, for example, applies conditional adversarial learning to precipitation downscaling and evaluates the ability to reconstruct fine-scale precipitation structure in a perfect-model setting [39]. Similarly, global spatiotemporal precipitation downscaling has been demonstrated using GAN-based models designed to enhance resolution while maintaining realistic rainfall distributions and temporal patterns [30]. These works emphasize that generative modeling can outperform traditional downscaling pipelines in reproducing the distributional properties of precipitation fields, including rare and extreme events.

Beyond precipitation, GANs have been applied to related atmospheric and environmental variables (e.g., wind fields and weather-driven geospatial datasets). Conditional GANs have been used to translate coarser numerical weather prediction outputs into higher-resolution wind patterns with CFD-like detail, offering a data-driven surrogate to bridge resolution gaps [66]. More broadly, recent geospatial downscaling studies highlight the importance of stochastic generation for representing uncertainty, particularly when refining very low-resolution climate inputs [53].

4.4.2 Data augmentation using generative adversarial networks (GANs) for meteorology-related engineering tasks

In applied energy and engineering settings, GAN-based synthetic augmentation has been explored as a method to improve predictive performance under data scarcity or rare-event regimes. For example, GAN augmentation strategies have been proposed for wind-related applications, where abrupt weather changes and limited extreme-event samples can degrade forecasting performance; GAN-generated samples have been reported to improve downstream wind power prediction accuracy compared with baseline training sets [57]. Similar lines of work in weather-dependent forecasting emphasize that the

primary value of generative augmentation is not only marginal distribution matching but also the preservation of informative structure for the target task.

These studies motivate synthetic augmentation in aviation winter operations: experimental endurance testing under natural snowfall is costly, weather-dependent, and often produces limited sample sizes. As in wind and precipitation applications, generative modeling offers a potential path to expand dataset variability while preserving physically relevant dependency structure.

4.4.3 Protocols for the generation and evaluation of synthetic tabular data

While many weather and climate GAN applications use gridded fields (image-like tensors), aviation endurance datasets are typically *tabular* and often include mixed variable types (continuous environmental variables, categorical product indicators, and configuration flags). Synthetic generation for tabular data introduces additional challenges, including mixed distributions, nonlinear dependencies, and semantic/logic constraints [26, 72]. Comparative studies on tabular synthesis methods show that GAN-based approaches can provide privacy and flexibility benefits but may underperform on utility or dependency preservation if the model is not carefully stabilized and evaluated [72].

Because no single metric fully characterizes synthetic realism, recent tabular synthesis literature recommends combining multiple evaluation perspectives: (i) distributional similarity (e.g., divergences, Wasserstein), (ii) dependency preservation (correlation structure and information-theoretic measures), and (iii) task-based utility tests such as Train-on-Synthetic, Test-on-Real (TSTR) [26, 72]. This multi-criteria evaluation philosophy aligns with climate downscaling practices, where realism is judged both by statistical agreement and by the ability to reproduce physically meaningful patterns [30, 35].

4.4.4 Gaps in research and the relevance of data on anti-icing endurance

Despite extensive GAN literature in precipitation downscaling and weather-driven augmentation, the application of GAN stabilization strategies to *small-sample, tabular aviation winter-performance datasets* remains limited. Existing weather GAN studies primarily focus on gridded spatial fields [30, 39, 41], whereas endurance datasets require accurate preservation of joint relationships among environmental conditions and categorical fluid/product indicators. This gap motivates a controlled comparison of stabilization strategies (Dropout, Batch Normalization, and extended training) with rigorous fidelity and utility evaluation, to identify architectures that better preserve both marginal statistics and multivariate dependency structure in aviation anti-icing endurance experiments.

4.5 Methodology

The study presents a generative modeling technique for creating synthetic datasets that are statistically relevant for evaluating anti-icing-fluid endurance time under natural-snow conditions. The aim is to estimate the joint distribution of the measured variables and to produce synthetic samples that reflect the statistical properties of the original dataset.

The process comprises three fundamental components: dataset preparation, generative model architecture design, and model training. These steps enable the production of synthetic samples while seeking to preserve the statistical structure of the experimental data.

The overall workflow of the proposed framework is illustrated in Figure 4.1. The methodology begins with preprocessing of the experimental dataset, followed by GAN-based synthetic data generation, statistical validation, and predictive utility evaluation.

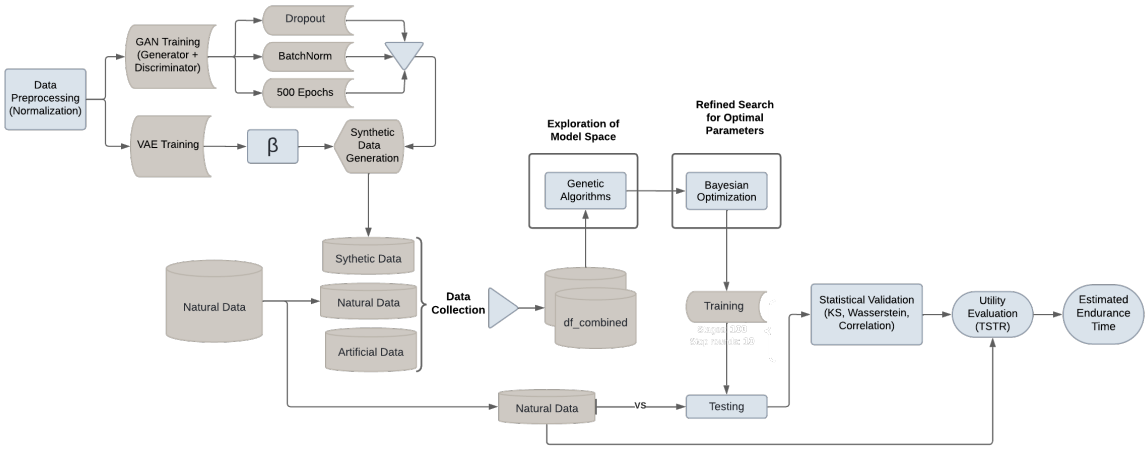


Figure 4.1: Overview of the synthetic data generation and evaluation framework used in this study. The workflow begins with the experimental dataset, followed by generative model training (GAN/VAE) and synthetic data generation

It is important to note that, within the TSTR evaluation framework, model testing is conducted exclusively on natural snow data in order to ensure that performance assessment reflects real operational conditions.

4.5.1 Description of the dataset

The baseline dataset comprises experimentally tested endurance time values acquired under regulated winter precipitation conditions. The trials replicate authentic aircraft ground icing conditions, wherein anti-icing fluids are subjected to descending snow until failure transpires.

Each observation encompasses environmental variables and fluid-related characteristics that affect endurance behavior. Critical variables encompass air temperature and snow precipitation rate, which directly influence the dilution and degradation of anti-icing solutions. The principal response variable is the endurance time, defined as the length during which the fluid effectively prevents ice accumulation.

Replicating natural snow studies consistently is difficult due to their significant cost, logistical intricacies, and dependence on meteorological factors. Therefore, the available

dataset is rather constrained. Small-sample regimes pose challenges for generative modeling, since inadequate data can lead to overfitting or unstable training dynamics. Thus, precise model development and regularization methods are crucial to ensure reliable learning and genuine synthetic data generation.

4.5.2 Data preparation and preprocessing

Prior to generative modeling, all continuous variables were normalized using min–max scaling to ensure numerical stability during model training. The transformation is defined as:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4.2}$$

where x_{min} and x_{max} represent the minimum and maximum values of each feature in the experimental dataset.

Feature scaling is particularly important for deep learning models due to the heterogeneous physical magnitudes of the variables involved. In the present dataset, features include temperature (°C), snow precipitation rate ($g/dm^2/h$), and endurance time (minutes), each operating on different numerical scales. Without normalization, variables with larger magnitudes may dominate gradient updates, leading to unstable convergence and reduced training efficiency [31].

Min-max normalization constrains each feature to a specified range, often $[0, 1]$, ensuring that all variables contribute equitably throughout optimization. This restricted representation also improves comprehension of the fundamental data distribution, particularly in situations with limited samples where the likelihood of overfitting is increased.

The endurance time objective variable was standardized throughout training to maintain alignment with the scaled input features. Following the generation of synthetic data, an inverse transformation was performed to recover the original physical units:

$$x = x_{scaled}(x_{max} - x_{min}) + x_{min} \tag{4.3}$$

Scaling parameters were exclusively obtained from the real experimental dataset to prevent information leakage and uphold statistical integrity during the synthesis of synthetic data.

The raw database contained 478 artificial-snow observations and 336 natural-snow observations. For the controlled natural-snow evaluation, the 336 natural observations were divided with `test_size=0.10` and a fixed random seed of 42, yielding 302 development observations and a final test set of 34 observations.

The difference between the natural-snow sample counts reported in Chapters 3 and 4 reflects distinct stages of data preparation rather than a contradiction. The 336 observations correspond to the complete raw natural-snow dataset before partitioning. The 295 natural-snow observations reported in Chapter 3 correspond to the processed development subset obtained after reserving the 34 test observations and applying the outlier-filtering procedure to the remaining 302 observations. This procedure retained observations satisfying both the interquartile-range criterion and the absolute standardized-score criterion ($|z| < 3$), excluding seven observations. Consequently, the 773 observations reported in Chapter 3 comprise 478 artificial-snow observations and 295 filtered natural-snow development observations.

The 34 test observations were reserved before generative-model fitting and were not used to train the GANs, fit preprocessing transformations, tune XGBoost, or construct augmented datasets. Each scaler was fitted exclusively on the 302-observation development set and then applied without refitting to the test set. Each GAN configuration generated 302 synthetic observations. Any 75/25 partition reported in a separate synthetic-data diagnostic concerns only that internal diagnostic and is not the final real-data evaluation protocol.

4.5.3 Architecture of Generative Models

Synthetic samples were generated using a Generative Adversarial Network (GAN), a deep generative modeling framework based on adversarial learning between two neural networks: a generator G and a discriminator D [33]. GANs learn to approximate the underlying probability distribution of real data by training these two networks in opposition. The generator attempts to produce synthetic samples that resemble real observations, while the discriminator learns to distinguish between real and generated data.

Architecture GAN

The generator $G(z;\theta_g)$ learns a nonlinear mapping from a latent noise distribution $z \sim p_z(z)$ to the data space, producing synthetic samples $G(z)$ intended to approximate the real data distribution. The discriminator $D(x;\theta_d)$ outputs the probability that a given input sample originates from the real experimental dataset rather than from the generator.

The adversarial training process is formulated as a two-player minimax game:

$$\min_G \max_D \mathbb{E}_{x\sim p_{data}}[\log D(x)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z)))] \tag{4.4}$$

During training, the discriminator is refined to enhance its capacity to differentiate between authentic and synthetic samples, whereas the generator is optimized to diminish this objective by generating progressively realistic outputs. This competitive dynamic compels the generator to comprehend the fundamental joint probability distribution $p_{data}(x)$ of the experimental variables.

Under ideal convergence conditions and assuming sufficient model capacity, the generator distribution $p_g(x)$ converges toward the true data distribution by implicitly minimizing the Jensen–Shannon (JS) divergence between $p_g(x)$ and $p_{data}(x)$ [31, 33]. At

Nash equilibrium, the discriminator cannot distinguish real from generated samples and outputs $D(x) = 0.5$ for both.

However, practical GAN training is known to be unstable, particularly in small-sample engineering datasets. Common issues include:

- **Mode collapse**, where the generator produces limited variability and fails to capture the full data manifold;
- **Vanishing gradients**, when a highly accurate discriminator prevents effective generator updates;
- **Training oscillations**, resulting from non-convex optimization dynamics.

These instability phenomena are linked to divergence behavior and gradient flow properties in adversarial optimization [5]. In particular, the original GAN objective based on Jensen–Shannon divergence may yield poor gradient information when real and generated distributions have limited overlap.

The latent input vector z was sampled from a multivariate standard normal distribution $z \sim \mathcal{N}(0, I)$, enabling stochastic sampling of the learned data manifold. The dimensionality of z was selected empirically to balance representation capacity and overfitting risk.

Variational Autoencoder (VAE) Architecture

In addition to GAN-based generative modeling, Variational Autoencoders (VAEs) were implemented as a probabilistic generative framework for synthetic data generation. Unlike GANs, which rely on adversarial training, VAEs are based on an encoder–decoder architecture that learns a latent representation of the data by maximizing a variational lower bound on the data likelihood [48, 75].

The encoder network maps input data x to a latent distribution $q(z|x)$, typically modeled as a multivariate Gaussian characterized by mean $\mu(x)$ and variance $\sigma^2(x)$. The

decoder then reconstructs the input by sampling from this latent space and generating $\hat{x}$:

$$z \sim q(z|x), \quad \hat{x} = f_\theta(z) \tag{4.5}$$

The training objective is defined as the maximization of the Evidence Lower Bound (ELBO), which consists of two components:

$$\mathcal{L}_{VAE} = \mathbb{E}_{q(z|x)}[\log p(x|z)] - D_{KL}(q(z|x)||p(z)) \tag{4.6}$$

The first term represents the reconstruction loss, ensuring that generated samples remain close to the input data, while the second term enforces regularization by aligning the learned latent distribution with a prior distribution, typically $p(z) \sim \mathcal{N}(0, I)$.

In this study, different values of the regularization parameter β were explored to control the trade-off between reconstruction accuracy and latent space regularization. Increasing β enforces stronger latent constraints but may lead to loss of fine-grained data structure, particularly in small tabular datasets.

Compared to GANs, VAEs offer more stable and tractable training due to their single-objective optimization. However, they may produce overly smooth samples and exhibit limited capability in capturing complex multimodal distributions, which can affect their ability to preserve predictive relationships in engineering datasets.

Architectural Stabilization Strategies

The natural snow endurance dataset employed in this study is constrained in size, resulting in a significant risk of overfitting and unstable adversarial training. GANs exhibit heightened sensitivity to data scarcity, since the discriminator may quickly memorize training samples, resulting in vanishing gradients for the generator [5, 22]. To address these concerns, three architectural configurations were assessed.

GAN-Dropout Dropout regularization was implemented in the hidden layers of both the generator and discriminator networks. Dropout randomly disables neurons during training with a predetermined probability, inhibiting co-adaptation of feature representations and diminishing memorization of training samples [86].

In compact tabular datasets, discriminators can rapidly approach near-optimal classification between real and synthetic samples. Dropout mitigates discriminator overconfidence and fosters stochastic representation learning, prompting the generator to explore a wider region of the data manifold [11, 98].

GAN-BatchNorm Batch Normalization (BatchNorm) was implemented to stabilize internal covariate shift and enhance gradient flow during training [38]. In adversarial learning, minor variations in feature distributions among mini-batches can disrupt the minimax optimization process. Batch normalization standardizes intermediate activations, diminishing variance between training iterations and expediting convergence.

From a theoretical standpoint, BatchNorm enhances the Lipschitz continuity of network layers, hence indirectly facilitating smoother optimization landscapes [83]. Consistent gradient propagation is crucial to prevent oscillatory behavior and divergence in GAN optimization [65].

GAN-Extended Epoch Training Prolonged training epochs were examined to enhance convergence to low-density areas of the learned data manifold. In small-sample scenarios, generators may initially capture the predominant modes of the distribution while neglecting less frequent nevertheless physically significant areas.

Augmenting the number of training epochs facilitates iterative enhancement of the learned distribution and expands coverage of the feature space. Excessive training may result in overfitting or the memorization of training samples [31]. Consequently, epoch extension was assessed in conjunction with statistical validation metrics.

Comparative evaluation The three configurations were assessed with respect to their capacity to preserve the statistical characteristics of the real dataset, including marginal feature distributions, multivariate correlation structures, Wasserstein distance, and overall training stability. This comparison supports the selection of a stabilizing approach for generating synthetic data from small aviation datasets [11, 60].

4.5.4 Training process

The GAN model underwent training via an iterative adversarial optimization process, wherein the generator and discriminator networks were updated alternately [33]. In each training iteration, latent vectors z were drawn from a multivariate standard normal distribution:

$$z \sim \mathcal{N}(0, I) \tag{4.7}$$

These latent vectors function as stochastic inputs for the generator, facilitating the exploration of the learned data manifold [33].

Optimization algorithm The Adam optimizer enhanced the model parameters, since it is widely utilized in deep generative modeling due to its flexible learning rate and efficient convergence properties [47]. Adam modifies parameters by employing estimates of the first and second moments of the gradients, hence improving stability in non-convex optimization problems commonly encountered in adversarial learning.

Training periods Training was executed across several epochs, with each epoch representing a complete traversal of the training dataset. Prolonged training periods were investigated to enhance convergence and enable the generator to incrementally revise its estimation of the actual data distribution [33].

Implementation and Reproducibility Parameters

The following table summarizes the principal implementation parameters used for the GAN and VAE experiments. All preprocessing transformations were fitted exclusively on the corresponding development data. The natural-snow test set containing 34 observations was not used to fit the scalers, train the generative models, select their architectures, or generate the synthetic observations.

The random seed of 42 was used for the natural-snow partition and for the downstream XGBoost models. However, no explicit PyTorch seed was fixed for the neural-network initialization, mini-batch shuffling, or latent-space sampling in the reported GAN and VAE experiments. Consequently, exact reproduction of the reported generative results may be affected by stochastic variation. Future implementations should explicitly set the Python, NumPy, and PyTorch seeds, document the software environment, and report results across repeated training runs.

Implementation parameters of the GAN and VAE models

Parameter	**GAN configurations**	**VAE configurations**
Training observations	302 natural-snow development observations	775 combined development observations
Preprocessing	StandardScaler fitted separately to the predictors and Endurance using development data only	StandardScaler fitted to the complete VAE training table using development data only
Modelled dimension	11 variables: 10 predictors and Endurance	12 variables, including Endurance
Latent dimension	32	3

Implementation parameters of the GAN and VAE models continued

Parameter	**GAN configurations**	**VAE configurations**
Generator or encoder	Generator: $32 \to 128 \to 128 \to 11$, with ReLU activations	Encoder: $12 \to 50 \to 12$, with Batch Normalization and ReLU activations; separate $12 \to 3$ layers for μ and $\log \sigma^2$
Decoder	Not applicable	$3 \to 12 \to 50 \to 12$, with ReLU activations and a linear output layer
Discriminator	$11 \to 128 \to 128 \to 1$, with LeakyReLU activations ($\alpha = 0.2$) and a linear logit output	Not applicable
Regularization variants	Dropout probability of 0.3 after both hidden generator layers; BatchNorm after each hidden linear layer; standard generator for extended training	$\beta \in \{1, 5, 10, 20\}$, with a 20-epoch linear warm-up of the KL-divergence weight
Loss function	Binary cross-entropy with logits for the generator and discriminator	Mean-squared reconstruction error plus β-weighted Kullback–Leibler divergence
Optimizer	Adam, learning rate 2×10^{-4}, $\beta_1 = 0.5$, $\beta_2 = 0.999$	Adam, learning rate 1×10^{-3}, $\beta_1 = 0.9$, $\beta_2 = 0.999$
Batch size	32	32
Training epochs	200 for Dropout and BatchNorm; 500 for the extended-training configuration	50 for each value of β

Implementation parameters of the GAN and VAE models continued

Parameter	**GAN configurations**	**VAE configurations**
Synthetic observations	302 observations per GAN configuration	775 observations per VAE configuration
Evaluation	Common reserved natural-snow test set, $n = 34$	Common reserved natural-snow test set, $n = 34$

The random seed of 42 was used for the natural-snow partition and for the downstream XGBoost models. However, no explicit PyTorch seed was fixed for the neural-network initialization, mini-batch shuffling, or latent-space sampling in the reported GAN and VAE experiments. Consequently, exact reproduction of the reported generative results may be affected by stochastic variation. Future implementations should explicitly set the Python, NumPy, and PyTorch seeds, document the software environment, and report results across repeated training runs.

4.5.5 Generation of synthetic data

Following adversarial training convergence, synthetic samples were generated by sampling latent vectors z from the learned latent space and mapping them through the trained generator network [33]:

$$x_{synthetic} = G(z; \theta_g^*) \tag{4.8}$$

where θ_g^* denotes the optimized generator parameters obtained after training.

Latent Space Sampling A Gaussian latent distribution was chosen due to its isotropic properties and theoretical compatibility with continuous generative modeling frameworks [31]. Sampling from $z \sim \mathcal{N}(0, I)$ allows for seamless exploration of the acquired data manifold and promotes interpolation across generated samples. Gaussian latent spaces

have exhibited strong representation learning capabilities in structured tabular datasets [98].

Size of the synthetic sample The number of generated samples was selected to augment the original experimental dataset while allowing robust statistical comparison between real and synthetic distributions. Synthetic datasets were produced at sizes equal to and larger than the original dataset to analyze distributional scaling behavior.

Inverse normalization Because GAN training was performed on normalized variables, generated samples were transformed back into physical units using inverse min–max scaling:

$$x = x_{scaled}(x_{max} - x_{min}) + x_{min} \tag{4.9}$$

Scaling parameters derived exclusively from the real dataset were used in order to preserve statistical consistency and prevent artificial expansion of the generated distributions.

Physical constraints To ensure engineering realism, generated samples were verified against known physical constraints of the system. These encompass non-negative precipitation rates, plausible temperature ranges, and viable endurance duration limits. Samples that contravened these limitations were excluded from the final synthetic dataset to ensure physical plausibility.

Distribution Target The objective of the synthetic generation process is to approximate the joint probability distribution $p_{data}(x)$ of the experimental dataset. Unlike deterministic interpolation methods, GAN-based generation captures nonlinear feature dependencies and multivariate correlations inherent to endurance behavior. The resulting synthetic dataset therefore represents stochastic samples drawn from the learned

approximation $p_g(x)$ of the real data manifold.

4.6 Implementation and evaluation of the experiment

The validity of synthetic data was assessed through a multi-level statistical evaluation framework designed to quantify distributional fidelity between real (p_{data}) and generated (p_g) datasets. Because no single metric fully characterizes generative quality [11], complementary marginal, multivariate, and predictive validation criteria were employed.

4.6.1 Experimental design

The experimental evaluation aims to assess whether synthetic datasets generated by the GAN architectures accurately reproduce the statistical and structural characteristics of the real endurance dataset. Three GAN configurations were evaluated: GAN-Dropout, GAN-BatchNorm, and GAN-Extended Epoch training.

Synthetic datasets produced by each architecture were compared against the real experimental dataset using a multi-level validation framework. The evaluation focuses on three complementary aspects of generative quality: (i) preservation of marginal variable distributions, (ii) conservation of multivariate dependency structures, and (iii) predictive utility when synthetic data are used for downstream machine learning models.

Statistical fidelity tests were therefore conducted using divergence-based metrics, optimal transport measures, and moment-based comparisons. This comparative design allows identification of the most effective architectural stabilization strategy for generating statistically consistent synthetic datasets in small-sample aviation experiments [11].

4.6.2 Quantitative statistical analysis

Kolmogorov–Smirnov (KS) Test For each continuous variable, the two-sample Kolmogorov–Smirnov (KS) test was used to evaluate whether real and synthetic samples originate from the same underlying distribution [64]. The KS statistic is defined as:

$$D_{KS} = \sup_x |F_{real}(x) - F_{synthetic}(x)| \tag{4.10}$$

where $F_{real}(x)$ and $F_{synthetic}(x)$ are the empirical cumulative distribution functions (ECDFs). Smaller D_{KS} values indicate closer alignment between distributions.

Wasserstein Distance To quantify distributional discrepancy in a metric space, the first-order Wasserstein distance (Earth Mover's Distance) was computed:

$$W(p_{data}, p_g) = \inf_{\gamma \in \Pi(p_{data}, p_g)} \mathbb{E}_{(x,y)\sim\gamma}[\|x - y\|] \tag{4.11}$$

where $\Pi(p_{data}, p_g)$ denotes the set of all joint distributions with marginals p_{data} and p_g [5, 91].

Kullback–Leibler (KL) Divergence The Kullback–Leibler divergence was computed to measure the information loss when the synthetic distribution approximates the real data distribution [49]:

$$D_{KL}(p_{data}||p_g) = \sum_x p_{data}(x) \log \frac{p_{data}(x)}{p_g(x)} \tag{4.12}$$

Lower KL divergence values indicate stronger agreement between the two distributions.

Jensen–Shannon (JS) Divergence The Jensen–Shannon divergence provides a symmetric and bounded measure of distribution similarity [56]:

$$D_{JS}(p_{data}||p_g) = \frac{1}{2}D_{KL}(p_{data}||m) + \frac{1}{2}D_{KL}(p_g||m) \tag{4.13}$$

where $m = \frac{1}{2}(p_{data} + p_g)$ represents the mixture distribution.

Preserving the Moment To ensure preservation of central tendency and dispersion characteristics, first- and second-order statistical moments were compared between real and synthetic datasets [42]:

- Mean deviation: $|\mu_{real} - \mu_{synthetic}|$
- Variance ratio: $\frac{\sigma^2_{synthetic}}{\sigma^2_{real}}$

4.6.3 Multivariate Dependency Evaluation

Because endurance behavior depends on nonlinear interactions among environmental variables, preserving multivariate relationships is critical. Pearson correlation matrices were computed for both real and synthetic datasets [9].

Similarity between correlation structures was quantified using the Frobenius norm:

$$\Delta_{corr} = \|R_{real} - R_{synthetic}\|_F \tag{4.14}$$

where R denotes the correlation matrix and $\|\cdot\|_F$ is the Frobenius norm. Smaller values indicate stronger preservation of multivariate dependencies.

This metric evaluates whether the generator successfully captures inter-variable relationships rather than merely reproducing marginal distributions.

4.6.4 Predictive Utility Evaluation (TSTR)

Beyond statistical similarity, the practical usefulness of synthetic data was evaluated using the Train-on-Synthetic, Test-on-Real (TSTR) protocol [23]. In this approach,

machine learning models are trained exclusively on synthetic datasets and subsequently evaluated on real experimental data.

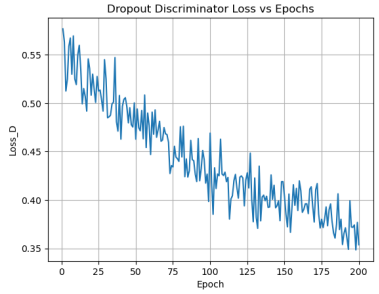
If the synthetic data accurately represent the real data distribution, models trained on synthetic samples should generalize effectively when applied to real observations. Predictive performance was evaluated using regression metrics including RMSE, MAE, and MAPE.

The TSTR evaluation therefore provides an application-oriented validation criterion that assesses whether the synthetic dataset preserves the predictive relationships present in the original endurance dataset.

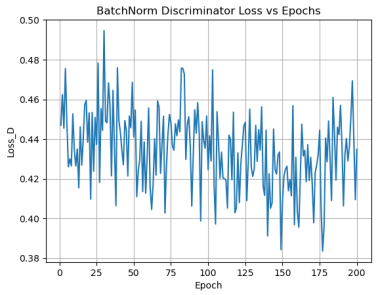
4.7 Results

4.7.1 Training Dynamics

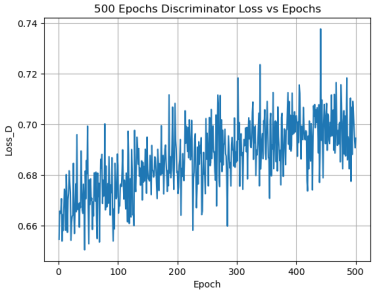
The training dynamics of the evaluated GAN architectures were analyzed through the evolution of generator and discriminator loss functions across training epochs. Figure 4.2 presents the loss trajectories for the Dropout, Batch Normalization (BatchNorm), and extended 500-epoch configurations.



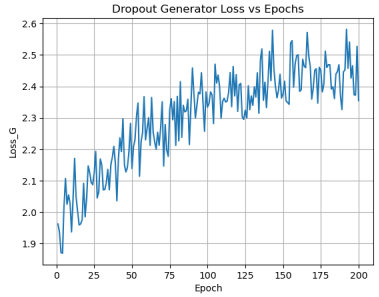
(a) Discriminator loss – Dropout configuration. Progressive decline indicates increasing discriminator dominance over training.



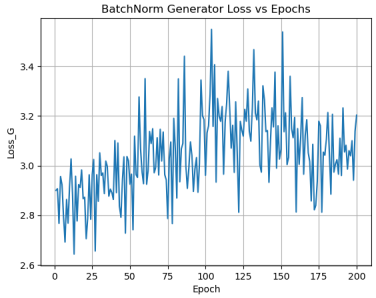
(b) Discriminator loss – BatchNorm configuration. Bounded oscillation suggests balanced adversarial competition.



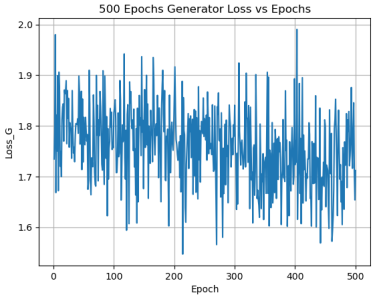
(c) Discriminator loss – Extended 500-epoch training. Mild upward drift indicates persistent discriminator advantage.



(d) Generator loss – Dropout configuration. Increasing trend reflects difficulty in approximating real data distribution.



(e) Generator loss – BatchNorm configuration. Stable oscillatory behavior indicates improved convergence dynamics.



(f) Generator loss – Extended 500-epoch training. Moderate fluctuation without significant convergence improvement.

Figure 4.2: Training loss evolution for discriminator (top row) and generator (bottom row) across the three evaluated GAN configurations: Dropout, Batch Normalization, and extended 500-epoch training.

The Dropout configuration exhibits a gradual decline in discriminator loss accompanied by a steady increase in generator loss. This divergence indicates progressive discriminator dominance, where the discriminator increasingly distinguishes real from generated samples. Such behavior suggests that dropout regularization introduces stochastic perturbations that hinder the generator's ability to effectively approximate the real data distribution.

In contrast, the BatchNorm configuration demonstrates comparatively stable adversarial dynamics. Both generator and discriminator losses oscillate within bounded

intervals without persistent monotonic drift. This oscillatory equilibrium indicates balanced adversarial competition, where neither network consistently overpowers the other. Such behavior is characteristic of stable GAN training and suggests improved gradient propagation during optimization.

The extended 500-epoch configuration exhibits moderate fluctuations in both loss curves without significant convergence improvement relative to shorter training durations. Although prolonged training allows additional refinement of generator parameters, the observed dynamics suggest that increasing training epochs alone does not substantially improve adversarial stability in small-sample tabular datasets.

Overall, the BatchNorm architecture demonstrates the most stable training behavior among the evaluated configurations.

4.7.2 Initial GAN Baseline Performance

Before comparing advanced stabilization strategies such as Batch Normalization, Dropout regularization, and extended training duration, an initial generative pipeline was evaluated using the natural snow dataset. In this baseline framework, a standard GAN was trained to generate synthetic samples from the original natural dataset, and the augmented data were subsequently used to train an XGBoost model optimized through Bayesian Optimization, Genetic Algorithm, and K-Fold cross-validation.

The objective of this preliminary experiment was to determine whether synthetic data generation could provide useful predictive information prior to introducing more sophisticated GAN modifications.

Table 4.2 summarizes the obtained results. The baseline pipeline achieved an R^2 of 0.88 with RMSE = 27.59 and MAE = 21.98. Although these results confirm that the synthetic samples retained part of the predictive structure of the real dataset, the prediction error remained substantially higher than the performance obtained using real data only. This motivated the exploration of improved stabilization techniques aimed at

increasing synthetic data fidelity and downstream predictive performance.

Pipeline	R^2	RMSE	MAE	Synthetic Samples
df_natural → GAN → XGBoost	0.88	27.59	21.98	336

Table 4.2: Initial predictive performance obtained using the baseline GAN synthetic data generation pipeline.

4.7.3 Distribution Fidelity

The GAN-Dropout and GAN-BatchNorm configurations were trained for 200 epochs, whereas the extended configuration was trained for 500 epochs. This comparison makes it possible to assess whether increasing training duration alone improves distributional fidelity to the same extent as architectural stabilization.

To quantify the statistical similarity between the real and generated datasets, several complementary fidelity metrics were calculated. These included Kullback–Leibler (KL) divergence, Jensen–Shannon (JS) divergence, Wasserstein distance, differences between correlation and mutual-information matrices measured using the Frobenius norm, and errors in the third- and fourth-order moments represented by skewness and kurtosis. Table 4.3 summarizes the corrected results. Lower values indicate stronger agreement between the real and synthetic datasets.

Variant	**Mean KL**	**Mean JS**	**Mean WD**	**Corr. F**	**MI F**	**Skew Err.**	**Kurt. Err.**
Dropout	4.268	0.552	27.609	3.891	9.958	1.070	**1.204**
BatchNorm	3.086	**0.472**	**7.741**	**0.549**	**9.225**	**0.443**	2.938
500 Epochs	**2.437**	0.481	9.262	1.209	9.300	0.731	2.261

Table 4.3: Corrected statistical-fidelity metrics for the evaluated GAN stabilization strategies. Dropout and BatchNorm were trained for 200 epochs, while the extended configuration was trained for 500 epochs. Mean WD denotes the mean Wasserstein distance; Corr. F and MI F denote the Frobenius norms of the differences between the real and synthetic correlation and mutual-information matrices, respectively. Lower values indicate stronger alignment with the real dataset.

The corrected results show that BatchNorm achieved the strongest overall distri-

butional fidelity. It produced the lowest mean JS divergence (0.472), mean Wasserstein distance (7.741), correlation-matrix difference (0.549), mutual-information difference (9.225), and mean skewness error (0.443). These results indicate that BatchNorm most effectively preserved both the marginal distributions and the multivariate dependency structure of the real dataset.

The 500-epoch configuration achieved the lowest mean KL divergence (2.437), compared with 3.086 for BatchNorm and 4.268 for Dropout. However, it remained inferior to BatchNorm according to the mean Wasserstein distance, correlation preservation, mutual-information preservation, and skewness error. Its overall performance was therefore intermediate between BatchNorm and Dropout. This finding suggests that extending the training duration can improve selected aspects of distributional similarity, but does not provide the same overall benefit as architectural stabilization through Batch Normalization.

The Dropout configuration exhibited the weakest overall distributional alignment. It produced the largest mean JS divergence (0.552), mean Wasserstein distance (27.609), correlation difference (3.891), mutual-information difference (9.958), and skewness error (1.070). Although Dropout achieved the lowest kurtosis error (1.204), this isolated advantage did not compensate for its substantially weaker performance across the remaining fidelity metrics.

No individual configuration achieved the lowest error for every metric. Nevertheless, BatchNorm provided the most consistent balance across marginal-distribution, dependency-preservation, and higher-order statistical criteria. These findings support Batch Normalization as the preferred configuration when the primary objective is distribution-preserving synthetic-data generation. In the separate exploratory pipeline reported in Table 4.5, the 500-epoch configuration produced the lowest test error. Because that comparison was not generated by the same fitted pipeline as the TSTR baseline, it illustrates that fidelity and apparent downstream utility may rank con-

figurations differently, but it does not by itself quantify a controlled augmentation gain.

4.7.4 Multivariate Dependency Preservation

In addition to reproducing marginal distributions, realistic synthetic-data generation requires the preservation of multivariate relationships among the predictor variables. Figure 4.3 presents the Pearson correlation matrices obtained from the synthetic datasets generated by the Dropout, BatchNorm, and 500-epoch GAN configurations. The matrices include the environmental predictors, engineered variables, and one-hot-encoded fluid indicators. Color intensity represents the magnitude and sign of the Pearson correlation coefficients.

The three configurations reproduce several relationships imposed by the feature-engineering procedure. In particular, the matrices show strong associations between temperature and its derived variables, including squared and absolute temperature, as well as between precipitation rate and squared precipitation rate. These relationships are expected because the engineered variables are deterministic transformations of the original environmental predictors.

Visual comparison across the three synthetic configurations shows that BatchNorm and the 500-epoch GAN produce more coherent and less extreme dependency structures than Dropout. The Dropout matrix contains several strong positive and negative correlations involving the fluid indicators and engineered environmental variables. These pronounced relationships are less evident in the BatchNorm and 500-epoch matrices and suggest that Dropout introduced artificial cross-variable dependencies.

The BatchNorm configuration produced the most balanced correlation structure among the evaluated synthetic datasets. The dominant relationships between the original environmental variables and their engineered transformations were retained, while correlations involving the one-hot-encoded fluid indicators remained comparatively

moderate. The 500-epoch configuration produced a similar overall pattern but exhibited larger residual deviations from the real correlation structure.

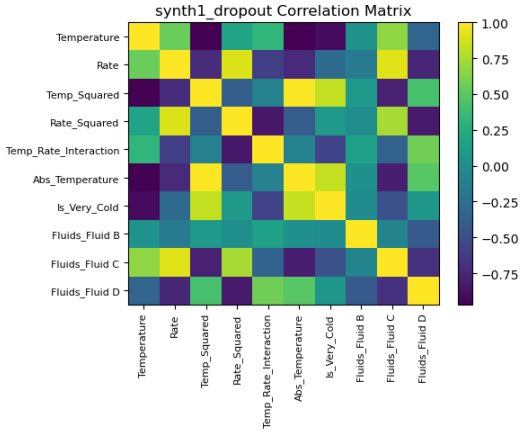
The quantitative comparison confirms these observations. The Frobenius norm of the difference between the real and synthetic correlation matrices was lowest for BatchNorm, with $\Delta_{\text{corr}} = 0.549$. The 500-epoch configuration achieved an intermediate difference of $\Delta_{\text{corr}} = 1.209$, whereas Dropout produced the largest discrepancy, with $\Delta_{\text{corr}} = 3.891$. Because lower values indicate stronger agreement with the real correlation matrix, these results show that BatchNorm most effectively preserved the multivariate linear dependency structure of the predictor space.

It is important to note that the matrices in Figure 4.3 contain predictor variables only and do not include the endurance-time target. Consequently, this analysis evaluates the preservation of dependencies among environmental, engineered, and fluid-related predictors. The preservation of predictive relationships between these variables and endurance time is examined separately in the subsequent predictive-utility subsection using the Train-on-Synthetic, Test-on-Real protocol.

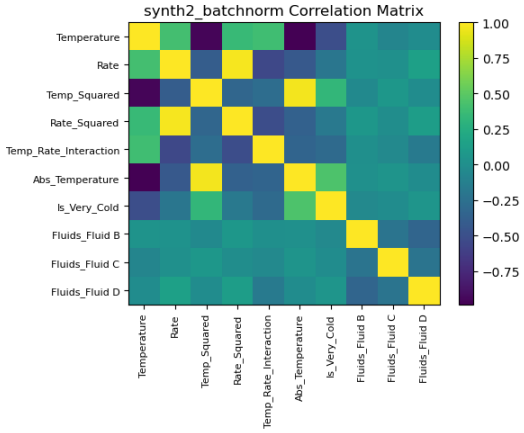
4.7.5 Predictive Utility

Statistical-similarity metrics describe the distributional fidelity of synthetic data but do not directly determine whether the generated observations preserve the predictive relationships present in the real dataset. Therefore, a Train-on-Synthetic, Test-on-Real (TSTR) protocol was implemented to evaluate the predictive transferability of the three GAN configurations.

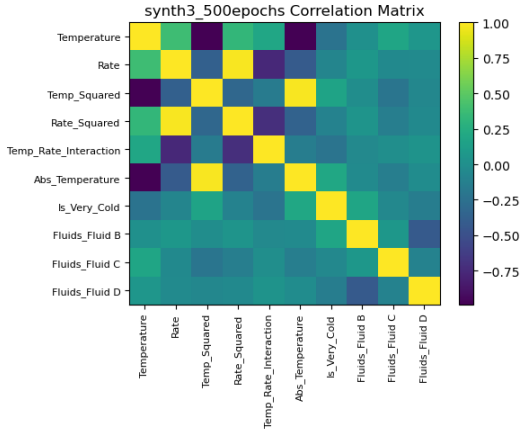
Under the TSTR protocol, an XGBoost regression model was trained exclusively on each synthetic dataset and subsequently evaluated on the same reserved natural-snow test set containing 34 real experimental observations. A Real-to-Real model, trained exclusively on the corresponding real training data and evaluated on the same natural-snow test set, was included as the reference baseline. Predictive performance



(a) Dropout configuration.



(b) BatchNorm configuration.



(c) 500-epoch configuration.

Figure 4.3: Correlation matrices of synthetic datasets generated using different stabilization strategies

was quantified using the coefficient of determination (R^2), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE).

As shown in Table 4.4, the Real-to-Real baseline achieved the strongest predictive performance, with $R^2 = 0.974$, RMSE = 6.596 minutes, MAE = 4.200 minutes, and MAPE = 6.645%. This result establishes the predictive performance attainable when the model is trained and evaluated using real observations from the same experimental domain.

Among the synthetic-data configurations, BatchNorm achieved the strongest TSTR

Variant	R^2	**MAE**	**RMSE**	**MSE**	**MAPE (%)**
Real → Real (Baseline)	**0.974**	**4.200**	**6.596**	**43.506**	**6.645**
Dropout (Synthetic → Real)	0.177	27.353	37.425	1400.610	58.713
BatchNorm (Synthetic → Real)	**0.672**	**19.248**	**23.632**	**558.481**	**40.000**
500 Epochs (Synthetic → Real)	0.306	24.606	34.372	1181.436	49.713

Table 4.4: Predictive-utility evaluation using the Train-on-Synthetic, Test-on-Real (TSTR) protocol. All models were evaluated on the same reserved natural-snow test set ($n = 34$). Bold values in the synthetic-data rows identify the strongest TSTR result. The Real-to-Real model is included as a reference baseline. Higher R^2 and lower MAE, RMSE, MSE, and MAPE indicate stronger predictive transfer.

performance. It obtained $R^2 = 0.672$, RMSE = 23.632 minutes, MAE = 19.248 minutes, and MAPE = 40.000%. The 500-epoch configuration ranked second, with $R^2 = 0.306$ and RMSE = 34.372 minutes, whereas the Dropout configuration produced the weakest transfer result, with $R^2 = 0.177$ and RMSE = 37.425 minutes.

The positive R^2 values obtained by all three synthetic-data configurations indicate that the generated datasets preserved a portion of the predictive relationships present in the real observations. Nevertheless, their larger errors relative to the Real-to-Real baseline demonstrate that the synthetic datasets did not fully reproduce the conditional relationship between the environmental predictors and endurance time.

The TSTR ranking is consistent with the statistical-fidelity analysis. BatchNorm achieved both the strongest overall distributional alignment and the best synthetic-to-real predictive transfer. The 500-epoch configuration demonstrated intermediate fidelity and TSTR performance, while Dropout produced the weakest results under both evaluation perspectives.

The TSTR evaluation should be distinguished from the augmented-data comparison presented in Table 4.5. TSTR evaluates the intrinsic predictive information contained in a synthetic dataset by training exclusively on synthetic observations. In contrast, the augmented-data evaluation combines real and synthetic training observations to determine whether synthetic data improve the final downstream model. Consequently,

BatchNorm achieved the strongest intrinsic synthetic-data quality and transferability, whereas the 500-epoch configuration achieved the strongest downstream augmentation result.

The numerical baselines in Tables 4.4 and 4.5 are not identical because they originate from different notebook pipelines. Table 4.4 uses the corrected TSTR pipeline: the Real-to-Real reference is trained on the 302 reserved natural-snow development observations and evaluated on the 34-observation natural-snow test set. Table 4.5 summarizes a separate exploratory comparison containing artificial-only, natural-only, combined, VAE-augmented, and GAN-augmented training configurations, with its own feature preparation and model-fitting sequence. Thus, the Real-to-Real row in Table 4.4 must not be interpreted as the same fitted baseline as the Natural Data row in Table 4.5. The difference between their scores reflects pipeline differences rather than a scientific contradiction or a measured augmentation effect. A definitive estimate of augmentation gain requires every baseline and augmented configuration to be rerun with identical features, preprocessing, hyperparameter selection, random seeds, and training observations. Until that harmonized rerun is performed, Table 4.5 should be interpreted as an exploratory comparison, whereas Table 4.4 provides the controlled TSTR assessment.

4.7.6 Model Comparison

While the TSTR evaluation measures the intrinsic transferability of the synthetic datasets, it does not represent the predictive performance obtained when synthetic observations are incorporated into the training data. Therefore, a complementary controlled evaluation was conducted using XGBoost regression models trained on baseline, VAE-augmented, and GAN-augmented datasets.

All models reported in Table 4.5 were evaluated on the same reserved natural-snow test set containing 34 observations. No test observation was used to train the

generative models, fit the preprocessing transformations, optimize XGBoost, or construct the augmented training datasets. This common evaluation protocol permits a direct comparison of predictive generalization across the evaluated pipelines.

The normalized percentage errors were calculated relative to the mean endurance value of the common natural-snow test set:

$$\text{RMSE (\%)} = \frac{\text{RMSE}}{\overline{y}_{\text{test}}} \times 100, \tag{4.15}$$

$$\text{MAE (\%)} = \frac{\text{MAE}}{\overline{y}_{\text{test}}} \times 100, \tag{4.16}$$

where $\overline{y}_{\text{test}} = 64.20$ minutes for every row in Table 4.5.

Dataset	R^2	**RMSE**	**RMSE (%)**	**MAE**	**MAE (%)**	**MSE**
Baseline models (no synthetic data)						
Artificial Data	0.5513	27.63	43.03	16.00	24.92	763.32
Natural Data	0.5118	28.82	44.89	9.57	14.91	830.51
Combined Data	0.8044	18.24	28.41	13.29	20.71	332.72
VAE-augmented modelling pipelines						
Combined Data (VAE, $\beta = 1.0$)	0.7920	18.81	29.30	13.83	21.54	353.85
Combined Data (VAE, $\beta = 5.0$)	0.8020	18.35	28.59	13.41	20.89	336.90
Combined Data (VAE, $\beta = 10.0$)	0.7805	19.32	30.10	14.02	21.84	373.33
Combined Data (VAE, $\beta = 20.0$)	0.8096	18.00	28.03	13.04	20.31	323.91
GAN-augmented modelling pipelines						
GAN + Dropout → XGBoost	0.9802	5.80	9.04	4.01	6.24	33.68
GAN + BatchNorm → XGBoost	0.9768	6.28	9.78	4.41	6.87	39.45
GAN + 500 Epochs → XGBoost	**0.9871**	**4.69**	**7.30**	**2.31**	**3.60**	**21.98**

Table 4.5: Predictive performance of XGBoost models trained on baseline, VAE-augmented, and GAN-augmented datasets. Every model was evaluated on the same reserved natural-snow test set ($n = 34$). RMSE and MAE are reported in minutes, and their normalized percentages use the common test-set mean $\overline{y}_{\text{test}} = 64.20$ minutes.

Among the baseline models, training on the combined artificial and natural dataset produced the strongest overall result, with $R^2 = 0.8044$ and RMSE = 18.24 minutes. The models trained exclusively on artificial or natural training observations produced substantially larger errors on the reserved natural-snow test set.

The VAE-augmented configurations produced similar results across the evaluated regularization parameters. The strongest VAE result was obtained with $\beta = 20.0$, which achieved $R^2 = 0.8096$, RMSE = 18.00 minutes, and MAE = 13.04 minutes. Consequently, the corrected results do not indicate a monotonic degradation as β increases. Instead, VAE performance remained within a relatively narrow range, with RMSE values between 18.00 and 19.32 minutes.

Within the exploratory results summarized in Table 4.5, all three GAN-augmented models produced lower errors than the listed baseline and VAE-augmented models. The 500-epoch GAN had $R^2 = 0.9871$, RMSE = 4.69 minutes, and MAE = 2.31 minutes; Dropout and BatchNorm had RMSE values of 5.80 and 6.28 minutes. These numerical differences are descriptive because the underlying pipelines were not fully harmonized.

These results demonstrate that the ranking of GAN configurations depends on the evaluation objective. The 500-epoch configuration achieved the best predictive performance when synthetic data were integrated into the XGBoost training set. In contrast, Batch Normalization achieved the strongest statistical fidelity and synthetic-to-real transferability in the fidelity and TSTR analyses. BatchNorm is therefore retained as the preferred configuration for distribution-preserving synthetic-data generation, whereas the 500-epoch configuration is identified as the strongest configuration for downstream predictive augmentation.

Overall, this exploratory comparison suggests that GAN-generated data may be useful as a complementary augmentation strategy for limited natural-snow datasets. Because the baseline and augmented results were assembled from distinct notebook pipelines, the magnitude of the apparent gain should not be interpreted as a controlled causal effect of augmentation. A harmonized rerun, repeated splits, and external validation are required to confirm the ranking and effect size.

4.8 Discussion

This study evaluated multiple architectural stabilization strategies for generative modeling of a small tabular aviation dataset describing anti-icing fluid endurance under winter precipitation conditions. The comparative analysis demonstrates that architectural design plays a critical role in determining generative model performance in small-sample engineering datasets. Among the evaluated approaches, GAN architectures incorporating Batch Normalization achieved the most consistent results across training stability, statistical fidelity, multivariate dependency preservation, and predictive utility.

4.8.1 Impact of Batch Normalization on Adversarial Stability

Among the evaluated configurations, Batch Normalization consistently achieved superior performance across divergence-based, dependency-based, and utility-based metrics. The improved performance can be attributed to its stabilizing effect on gradient propagation during adversarial training.

GAN optimization is inherently unstable due to its minimax formulation. In small datasets, discriminators can quickly overfit and produce vanishing gradients for the generator. Batch normalization mitigates internal covariate shift and smooths the optimization landscape [38, 83], leading to the bounded adversarial oscillations observed in the training loss trajectories.

The reduced Wasserstein distance and improved preservation of correlation structure suggest that BatchNorm enhances the generator's ability to approximate the joint probability distribution of endurance-related variables rather than merely matching marginal statistics.

These findings align with recent studies emphasizing the importance of architectural stabilization in tabular GANs [11, 98], particularly when training data are limited.

4.8.2 Limitations of Dropout in Small Tabular GANs

Although Dropout is widely used as a regularization strategy in supervised deep learning, its application in adversarial generative modeling appears less effective in this context. The Dropout configuration exhibited progressive discriminator dominance and produced the largest divergence and correlation distortion metrics.

In adversarial training, excessive stochastic regularization may hinder the generator's ability to learn fine-grained multivariate dependencies, particularly when the dataset is already limited. Unlike image-based GANs where high-dimensional representation capacity mitigates such effects, small structured engineering datasets may require stable feature propagation rather than aggressive regularization.

These results suggest that regularization strategies effective in predictive neural networks do not necessarily translate directly to generative adversarial frameworks.

4.8.3 Extended Training Duration: Objective-Dependent Results

Increasing the training duration from 200 to 500 epochs did not improve overall distributional fidelity beyond the BatchNorm configuration. Although the 500-epoch GAN achieved the lowest mean KL divergence (2.437), BatchNorm produced stronger results for mean JS divergence, mean Wasserstein distance, correlation preservation, mutual-information preservation, and skewness error. BatchNorm also achieved stronger synthetic-to-real predictive transfer, with TSTR RMSE values of 23.632 minutes for BatchNorm and 34.372 minutes for the 500-epoch configuration.

These findings indicate that increasing training duration alone does not provide the same overall distributional and transferability benefits as architectural stabilization. Additional epochs may allow the generator to improve selected aspects of the learned distribution, but they do not inherently resolve adversarial instability or guarantee

improved preservation of multivariate dependencies.

However, a different ranking appeared in the exploratory augmented-data evaluation. When synthetic observations were combined with real training data, the 500-epoch configuration produced $R^2 = 0.9871$, RMSE = 4.69 minutes, and MAE = 2.31 minutes on the common natural-snow test set; Dropout and BatchNorm produced RMSE values of 5.80 and 6.28 minutes. This ranking requires confirmation in a harmonized pipeline.

The effect of extended training therefore depends on the evaluation objective. It produced only intermediate statistical fidelity and TSTR transferability, but generated observations that were highly useful when incorporated into downstream XGBoost training. Consequently, training duration cannot be characterized simply as producing diminishing returns; rather, its benefits were more evident for predictive augmentation than for distributional alignment or stand-alone synthetic-to-real transfer.

4.8.4 Comparison Between GAN and VAE Generative Models

In addition to comparing stabilization strategies within the GAN framework, this study evaluated synthetic-data augmentation using Variational Autoencoders (VAE). All reported models were evaluated on the same reserved natural-snow test set; however, differences in upstream feature preparation and fitting mean that the comparison should be considered exploratory rather than fully controlled.

The four VAE configurations produced relatively similar results. Their RMSE values ranged from 18.00 to 19.32 minutes, while their R^2 values ranged from 0.7805 to 0.8096. The strongest VAE result was obtained with $\beta = 20.0$, which achieved $R^2 = 0.8096$, RMSE = 18.00 minutes, and MAE = 13.04 minutes. The corrected results therefore do not demonstrate a monotonic degradation in predictive performance as β increases. Instead, VAE performance remained within a comparatively narrow range across the evaluated regularization parameters.

The GAN-augmented rows had RMSE values between 4.69 and 6.28 minutes,

compared with 18.00 to 19.32 minutes for the VAE rows. Within this exploratory comparison, the 500-epoch GAN had the lowest error, whereas BatchNorm achieved the strongest overall statistical fidelity and synthetic-to-real transferability. A harmonized rerun is needed before attributing the difference to model family or augmentation alone.

These results indicate that the GAN and VAE comparisons also depend on the evaluation objective. The GAN configurations provided greater downstream augmentation utility than the VAE configurations, while the fidelity analysis identified BatchNorm as the most effective GAN architecture for preserving the statistical and dependency structure of the real dataset. The VAE results nevertheless remained comparable to the combined-data baseline, which achieved RMSE = 18.24 minutes, indicating that VAE augmentation neither produced the substantial gains observed with GAN augmentation nor resulted in uniformly poor predictive performance.

Computational efficiency was also considered. Under the reported training configuration, model training required approximately 15 minutes and 53 seconds per run. This computational requirement remained practical for the experimental comparisons conducted in this study. However, training time may vary according to architecture, number of epochs, hardware, optimization procedure, and the number of synthetic observations generated. Consequently, the reported value should be interpreted as an implementation-specific measurement rather than as a general estimate of deployment cost.

Overall, the results suggest that GAN-based augmentation was more effective than VAE-based augmentation for improving endurance prediction in the present dataset. Nevertheless, the small size of the common natural-snow test set requires that this result be confirmed through repeated-split or external validation before it can be generalized to other aviation datasets or operational conditions.

4.8.5 Fidelity vs. Predictive Utility

A central finding of this study is that statistical fidelity, synthetic-to-real transferability, and downstream augmentation utility are related but non-equivalent evaluation objectives. No single GAN configuration achieved the strongest result under every evaluation protocol.

BatchNorm achieved the strongest overall statistical fidelity. It produced the lowest mean JS divergence (0.472), mean Wasserstein distance (7.741), correlation-matrix difference (0.549), mutual-information difference (9.225), and skewness error (0.443). It also achieved the strongest TSTR performance among the synthetic datasets, with $R^2 = 0.672$ and RMSE = 23.632 minutes. These results indicate that BatchNorm most effectively preserved the global statistical structure and transferable predictive relationships of the real data.

In contrast, the 500-epoch configuration achieved weaker fidelity and TSTR results than BatchNorm but had the lowest error in the separate exploratory downstream table. When its synthetic observations were combined with real training data, the reported XGBoost result was $R^2 = 0.9871$ and RMSE = 4.69 minutes on the common natural-snow test set. This observation motivates a harmonized augmentation experiment, but it is not sufficient on its own to establish the size of an augmentation benefit.

This distinction is important because marginal and multivariate distributional similarity do not guarantee predictive equivalence. Likewise, relatively weak stand-alone synthetic-to-real transfer does not necessarily imply that the synthetic observations have no augmentation value when combined with real data. The effect of synthetic data depends on how the generated observations interact with the real training distribution and the downstream learning algorithm.

For engineering applications, generative models should therefore be evaluated according to their intended use. BatchNorm is the preferred configuration when the principal objective is distribution-preserving data generation or stand-alone synthetic-

to-real transferability. The 500-epoch configuration is the preferred option among the evaluated models when the objective is to maximize downstream predictive performance through data augmentation. Both conclusions should be interpreted cautiously because the common natural-snow test set contained only 34 observations.

4.8.6 Implications for Aviation Engineering Datasets

Winter anti-icing endurance experiments are costly, weather-dependent, and logistically constrained. Experimental campaigns require natural snowfall conditions and specialized testing facilities, which limits the availability of large datasets for machine learning development.

The present findings demonstrate that architectural design choices strongly influence the quality of synthetic augmentation for such datasets. Batch Normalization appears particularly well-suited for small structured aviation datasets, offering improved manifold learning without introducing instability.

However, synthetic data should be viewed as complementary augmentation rather than a direct replacement for experimental measurements. Instead, generative models may serve as tools for expanding training datasets, exploring rare environmental conditions, and supporting preliminary model development prior to large-scale experimental validation.

4.9 Conclusion

This study investigated the effect of architectural stabilization strategies on generative modeling for a small tabular aviation dataset describing anti-icing fluid endurance under winter precipitation conditions. Three GAN configurations—Dropout, Batch Normalization, and extended 500-epoch training—were evaluated in terms of training stability, statistical fidelity, dependency preservation, synthetic-to-real transferability,

and downstream predictive augmentation.

The results show that Batch Normalization provided the most stable adversarial training and the highest overall distributional fidelity. This result is consistent with the requirements of small, structured tabular datasets, for which preserving feature distributions and multivariate dependencies is critical. Batch Normalization stabilized gradient propagation and enabled the generator to reproduce the joint relationships between environmental variables and endurance time more effectively. In contrast, Dropout introduced stochastic perturbations that weakened these dependencies, while extending the training duration alone did not provide equivalent improvements in statistical fidelity.

Under the Train-on-Synthetic, Test-on-Real evaluation, none of the synthetic datasets fully matched the performance of the Real-to-Real baseline. Nevertheless, the BatchNorm configuration achieved the strongest synthetic-to-real transfer, with $R^2 = 0.672$ and RMSE = 23.63 minutes, indicating that it preserved the predictive structure of the original data more effectively than the Dropout and 500-epoch configurations.

A different ranking was observed when synthetic observations were incorporated into the downstream XGBoost training process. In this exploratory comparison, all reported models used the same reserved natural-snow test set, but their upstream pipelines were not identical. The 500-epoch GAN produced $R^2 = 0.9871$, RMSE = 4.69 minutes, and MAE = 2.31 minutes; Dropout and BatchNorm produced RMSE values of 5.80 and 6.28 minutes.

These findings demonstrate that the optimal GAN configuration depends on the intended application. BatchNorm was the preferred configuration for distribution-preserving synthetic-data generation and synthetic-to-real transferability, whereas the 500-epoch configuration was the strongest option for downstream predictive augmentation. Statistical fidelity, synthetic-to-real transfer, and augmentation utility should therefore be treated as complementary but non-equivalent evaluation objectives.

Overall, the results support continued investigation of synthetic data as a complementary strategy for limited natural-snow datasets. The exploratory XGBoost results suggest potential utility, but the pipeline differences and the 34-observation test set prevent a firm estimate of improvement. The effect must be confirmed using identical preprocessing and tuning, repeated splits, and external experimental validation.

Future work should investigate conditional generative models, physics-informed constraints, and repeated or nested validation procedures to improve synthetic-data realism, predictive transferability, and the robustness of model comparisons.

CHAPITRE 5

CONCLUSION GÉNÉRALE ET RECOMMANDATIONS

5.1 Conclusion générale

Ce mémoire a exploré l'application des techniques d'apprentissage automatique et de génération de données synthétiques pour la prédiction du temps d'endurance des fluides antigivrage dans des conditions de précipitations hivernales. Ce travail s'inscrit dans un contexte où les méthodes expérimentales traditionnelles présentent des limitations importantes, notamment en raison du coût élevé des essais, de la disponibilité limitée des données sous neige naturelle et des différences observées entre les conditions de laboratoire et les conditions opérationnelles réelles.

Dans un premier temps, des modèles d'apprentissage automatique ont été développés à partir de données expérimentales obtenues sous conditions de neige naturelle et artificielle. Les résultats ont montré que les modèles d'ensemble, en particulier *XGBoost*, offrent une capacité élevée à modéliser les relations non linéaires complexes entre les variables environnementales et la performance des fluides antigivrage. Le modèle *XGBoost* a présenté les meilleures performances globales, avec un coefficient de détermination de $R^2 = 0.92$, une erreur $RMSPE = 15.38$ % et une erreur $MAPE = 8.88$ %. Cette performance supérieure peut être attribuée à sa capacité à capturer les interactions complexes entre les variables, à sa robustesse face aux jeux de données de taille limitée et à ses mécanismes de régularisation, qui contribuent à réduire le risque de surapprentissage.

L'analyse des résultats a également permis d'identifier les variables environnemen-

tales les plus influentes dans la dégradation des fluides antigivrage. En particulier, le taux de précipitation de neige, la température et leurs interactions se sont révélés déterminants dans la prédiction du temps d'endurance. Ces résultats confirment que la performance des fluides ne dépend pas uniquement de l'effet isolé d'une variable, mais plutôt de la combinaison de plusieurs facteurs environnementaux. Les différences observées entre les conditions de neige naturelle et artificielle soulignent ainsi l'importance d'intégrer ces deux types de données afin d'améliorer la fiabilité et la représentativité des modèles prédictifs.

Dans un second temps, ce travail a étudié l'utilisation des réseaux antagonistes génératifs (*Generative Adversarial Networks*, GAN) pour la génération de données synthétiques dans le contexte de jeux de données tabulaires de petite taille. Plusieurs configurations de GAN ont été évaluées, incluant la régularisation par *Dropout*, l'intégration de la normalisation par lots (*Batch Normalization*) et l'augmentation du nombre d'itérations d'entraînement. Les résultats ont démontré que l'architecture du modèle joue un rôle déterminant dans la stabilité de l'entraînement et la qualité statistique des données générées. En particulier, la *Batch Normalization* a conduit aux meilleures performances en termes de similarité statistique, de préservation des dépendances multivariées et d'utilité prédictive. Ce comportement s'explique par sa capacité à stabiliser la propagation des gradients et à réduire les variations internes des distributions durant l'apprentissage, ce qui favorise une meilleure représentation des relations conjointes entre les variables.

Le cadre d'évaluation adopté, combinant des métriques de similarité des distributions, des mesures de dépendance entre variables et l'approche *Train-on-Synthetic, Test-on-Real* (TSTR), a permis de fournir une analyse complète de la qualité et de l'utilité des données synthétiques. Les résultats indiquent que l'intégration de données synthétiques peut contribuer à améliorer la robustesse et la capacité de généralisation des modèles d'apprentissage automatique, notamment en enrichissant la variabilité du jeu de

données et en couvrant des régions peu représentées de l'espace expérimental. Toutefois, les données générées ne peuvent pas remplacer entièrement les données expérimentales réelles, car leur contribution dépend directement de leur fidélité statistique et de leur capacité à préserver les relations physiques sous-jacentes.

Ainsi, les objectifs de cette recherche ont été atteints. Les variables environnementales les plus influentes dans la dégradation des fluides antigivrage ont été identifiées, et leur contribution dans les modèles prédictifs a été analysée. De plus, l'impact de l'intégration de données synthétiques sur la performance et la robustesse des modèles développés a été évalué, démontrant leur potentiel comme outil complémentaire dans un contexte de données expérimentales limitées.

Dans l'ensemble, cette recherche démontre que l'intégration des approches d'apprentissage automatique et de génération de données synthétiques constitue une solution prometteuse pour améliorer la prédiction de la performance des fluides antigivrage, réduire la dépendance aux essais expérimentaux coûteux et contribuer à l'amélioration de la sécurité et de l'efficacité des opérations aéronautiques en conditions hivernales.

5.2 Contributions scientifiques

Ce mémoire apporte quatre contributions principales : (1) une comparaison structurée de modèles prédictifs appliqués à des données de neige naturelle et artificielle; (2) l'identification et l'interprétation des variables environnementales dominantes; (3) un protocole sans fuite de données séparant un ensemble naturel final de 34 observations, avec ajustement des transformations uniquement sur les 302 observations de développement; et (4) une évaluation distincte de la fidélité statistique, du transfert *Train-on-Synthetic, Test-on-Real* (TSTR) et de l'utilité de l'augmentation par GAN.

5.3 Limites

Les conclusions sont limitées par la taille du jeu de test naturel ($n = 34$), l'absence de validation externe indépendante, la représentation restreinte des fluides et des conditions, et l'évaluation sur une seule séparation principale. La graine aléatoire 42 a été utilisée afin d'assurer la reproductibilité du protocole. Toutefois, l'utilisation d'une seule graine ne permet pas d'évaluer la sensibilité des performances aux variations des partitions ou de l'initialisation des modèles. Les résultats devront donc être confirmés au moyen de validations répétées sur plusieurs graines, avec présentation de la moyenne, de l'écart-type ou d'intervalles de confiance.

De plus, la validation du chapitre 3 a été effectuée au niveau des observations individuelles, sans regroupement par essai expérimental, puisqu'un identifiant de groupe approprié n'était pas disponible. Cette configuration peut conduire à une estimation optimiste de la performance lorsque des observations corrélées provenant d'un même essai sont réparties entre l'entraînement et la validation.

Les erreurs de mesure, la répétabilité expérimentale et l'erreur prédictive constituent des sources d'incertitude distinctes; la validation croisée ne remplace pas une estimation de l'incertitude expérimentale.

5.4 Perspectives et recommandations

Les travaux futurs sont proposés dans l'ordre de priorité suivant :

1. **Validation externe.** Évaluer les modèles sur un ensemble entièrement indépendant, acquis hors du processus de développement actuel.

2. **Validation interannuelle.** Entraîner sur certaines saisons hivernales et tester sur une autre afin de mesurer la robustesse aux variations d'une année à l'autre.

3. **Validation *leave-one-fluid-out.*** Exclure successivement chaque fluide de l'entraînement pour évaluer la transférabilité à une formulation non observée.

4. **Validation *leave-one-site-out.*** Tester la transférabilité entre laboratoires, installations ou sites de collecte lorsque de telles données deviennent disponibles.

5. **Quantification des incertitudes.** Distinguer et, lorsque possible, propager l'incertitude de mesure, la répétabilité expérimentale et l'incertitude prédictive du modèle.

6. **Nouvelles campagnes expérimentales.** Accroître la taille et la diversité des données sous neige naturelle, particulièrement aux limites du domaine de température et de taux de précipitation, et utiliser ces campagnes pour confirmer les résultats d'augmentation synthétique dans un pipeline harmonisé. Des modèles conditionnels, des Wasserstein GAN et des contraintes physiques pourront alors être comparés de manière contrôlée.

7. **Prototype opérationnel éventuel.** Envisager seulement après les validations précédentes un prototype intégrant capteurs et prévisions météorologiques. Un tel prototype demeurerait un outil d'étude et ne remplacerait pas les essais normalisés ni les lignes directrices opérationnelles applicables.

Ce travail ouvre la voie à l'intégration de solutions basées sur l'intelligence artificielle dans le domaine de la sécurité aéronautique hivernale, en proposant des outils capables de mieux comprendre et prédire des phénomènes complexes dans des environnements contraints.

BIBLIOGRAPHIE

[1] Sae arp5485: Aircraft ground deicing/anti-icing methods, 2016.

[2] E. S. Abdelghany, M. B. Farghaly, M. M. Almalki, H. H. Sarhan, and M. E.-S. M. Essa. Machine learning and iot trends for intelligent prediction of aircraft wing anti-icing system temperature. *Aerospace*, 10(8):676, 2023.

[3] Federal Aviation Administration. Aircraft ground deicing and anti-icing program. Technical report, FAA, 2018.

[4] Federal Aviation Administration. Aircraft ground deicing and anti-icing program (advisory circular 120-60). Technical report, Federal Aviation Administration, 2020.

[5] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein gan. *International Conference on Machine Learning*, 2017.

[6] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *Proceedings of the International Conference on Machine Learning*, 2017.

[7] A. Beisswenger and J. Perron. Indoor snow testing of aircraft ground anti-icing fluids. Technical report, AMIL/UQAC, AMIL/UQAC: Saguenay, Qc, 2006.

[8] Sanae Benaissa, Derek Harvey, Hassan Khawaja, and Gelareh Momen. Thermographic analysis of ethylene glycol–based aircraft anti-icing fluid: Investigation of fluid failure mechanisms during simulated snow endurance tests. *Cold Regions Science and Technology*, 234:104472, 2025.

[9] Jacob et al. Benesty. Pearson correlation coefficient. *Noise Reduction in Speech Processing*, 2009.

[10] Bernd Bischl, Martin Binder, Michel Lang, Tobias Pielok, Jakob Richter, Stefan Coors, Janek Thomas, Theresa Ullmann, Marc Becker, Anne-Laure Boulesteix, Difan Deng, and Marius Lindauer. Hyperparameter optimization: Foundations, algorithms, best practices and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(2):e1484, 2023.

[11] A. Borji. Pros and cons of gan evaluation measures. *Computer Vision and Image Understanding*, 2022.

[12] George E. P. Box and David R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, 1964.

[13] Michael Bragg, Andy Broeren, and Laura Blumenthal. *Aircraft Icing*. Wiley, 2016.

[14] Michael B. Bragg, Andy P. Broeren, and Larry A. Blumenthal. Aircraft icing. *Progress in Aerospace Sciences*, 41:323–362, 2002.

[15] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[16] Leo Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[17] Transport Canada. Guidelines for aircraft ground icing operations, 2018.

[18] A. Chaput. Artificial snow generation and its application for aircraft anti-icing fluid testing. *Cold Regions Science and Technology*, 2001.

[19] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[20] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the ACM SIGKDD Conference*, 2016.

[21] Z. Chu, J. Geng, Q. Yang, X. Yi, and W. Dong. Prediction of temperature distribution on an aircraft hot-air anti-icing surface by rom and neural networks. *Aerospace*, 11(11):930, 2024.

[22] A. et al. Creswell. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 2018.

[23] Cristobal Esteban, Stephanie Hyland, and Gunnar Rätsch. Real-valued medical time series generation with recurrent conditional gans. *Machine Learning for Healthcare Conference*, 2017.

[24] L. Fay and X. Shi. Performance evaluation of aircraft anti-icing fluids under snowfall conditions. *Cold Regions Science and Technology*, 63:97–108, 2010.

[25] L. Fletcher. Aircraft icing: A review of research and challenges. *Progress in Aerospace Sciences*, 2015.

[26] J. Fonseca et al. Synthetic data generation for tabular datasets: A survey. *ACM Computing Surveys*, 2023.

[27] David Gagnon, Jean-Denis Brassard, Hassan Ezzaidi, and Christophe Volat. Computer-assisted aircraft anti-icing fluids endurance time determination. *Aerospace*, 7(4):39, 2020.

[28] R. W. Gent, N. P. Dart, and J. T. Cansdale. *Aircraft Icing*. 2000.

[29] Hamzeh Ghasemzadeh, Robert E. Hillman, and Daryush D. Mehta. Toward generalizable machine learning models in speech, language, and hearing sciences: Estimating sample size and reducing overfitting. *arXiv preprint arXiv:2308.11197*, 2023.

[30] R. Glawion. Global precipitation downscaling using gans. *Climate Dynamics*, 2025.

[31] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

[32] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. 2016.

[33] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 27, 2014.

[34] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Sebastopol, CA, 3 edition, 2022.

[35] L. Harris. Deep learning for precipitation downscaling. *Journal of Climate*, 2022.

[36] Geoffrey Hinton and Ruslan Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 2006.

[37] SAE International. Ams1428 aircraft anti-icing fluid.

[38] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training. *ICML*, 2015.

[39] A. Iotti. Rainscalegan for precipitation downscaling. *Geoscientific Model Development*, 2025.

[40] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer, 2013.

[41] X. Ji. Gan-based precipitation modeling. *Atmospheric Research*, 2024.

[42] Maurice G. Kendall. *The Advanced Theory of Statistics*. 1943.

[43] E. Khoshbakhtnejad, F. B. Golezani, and H. Sojoudi. Dynamics of snowflakes impacting superhydrophobic surfaces. *Langmuir*, 40(37):11265–11275, 2024.

[44] Ehsan Khoshbakhtnejad, Farshad B. Golezani, Behrouz Mohammadian, Hossein Sojoudi, and A. H. Abou Yassine. Trajectory and impact dynamics of snowflakes: Fundamentals and applications. *Powder Technology*, 2024.

[45] Ehsan Khoshbakhtnejad, Farshad B. Golezani, and Hossein Sojoudi. Dynamics of snowflakes impacting superhydrophobic surfaces. *Langmuir*, 40(37), 2024.

[46] Ehsan Khoshbakhtnejad and Hossein Sojoudi. Predicting coefficient of restitution of ice particles: a machine learning-based approach. *Powder Technology*, 2025.

[47] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*, 2015.

[48] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[49] Solomon Kullback and Richard Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 1951.

[50] Michael H. Kutner, Christopher J. Nachtsheim, John Neter, and William Li. *Applied Linear Statistical Models*. McGraw-Hill Education, 2005.

[51] Brett Lantz. *Machine Learning with R: Expert techniques for predictive modeling*. Packt Publishing, Birmingham, UK, 4 edition, 2021.

[52] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 2015.

[53] G. Li et al. Generative adversarial models for extreme geospatial downscaling. *International Journal of Applied Earth Observation and Geoinformation*, 2025.

[54] S. Li, J. Qin, M. He, and R. Paoli. Fast evaluation of aircraft icing severity using machine learning based on xgboost. *Aerospace*, 7(4):36, 2020.

[55] Y. Li. Deep learning for aircraft icing detection. *Aerospace Science and Technology*, 2018.

[56] Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 1991.

[57] R. Liu et al. Gan-based abrupt weather data augmentation for wind power prediction. *Energies*, 2023.

[58] Zhen Liu, Qiang Zhang, and Yifan Wang. A comprehensive survey on genetic algorithms for hyperparameter optimization. *IEEE Access*, 10:27590–27610, 2022.

[59] Edward P. Lozowski. Aircraft icing: A review of research and operational issues. *Atmospheric Research*, 2000.

[60] M. et al. Lucic. Are gans created equal? a large-scale study. *NeurIPS*, 2018.

[61] Stephen Lukachko. Icing physics and modeling. *Journal of Aircraft*, 2012.

[62] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence*, 2(1):56–67, 2020.

[63] H. Mao, X. Lin, Z. Li, X. Shen, and W. Zhao. Anti-icing system performance prediction using pod and pso-bp neural networks. *Aerospace*, 11(6):430, 2024.

[64] Frank J. Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 1951.

[65] L. et al. Mescheder. Which training methods for gans do actually converge? *ICML*, 2018.

[66] J. Milla-Val et al. An image-to-image adversarial network to generate high-resolution local wind conditions from numerical weather prediction data. *Engineering Applications of Artificial Intelligence*, 2025.

[67] B. Mohammadian, E. Khoshbakhtnejad, and H. Sojoudi. Interfacial phenomena in snow from its formation to accumulation and shedding. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 623:126758, 2021.

[68] Christoph Molnar. *Interpretable Machine Learning*. Christoph Molnar, 2022.

[69] Virgil P. Moshansky. Commission of inquiry into the air ontario crash at dryden, ontario. Technical report, Government of Canada, 1992.

[70] H. H. Oda, P. Adrian, M. Arriaga, L. Davies, J. Hille, T. Sheehan, and E. T. Suter. Safe winter operations. *Aero Magazine*, pages 5–13, 2010.

[71] Transportation Safety Board of Canada. Aviation investigation report a17c0146. Technical report, Transportation Safety Board of Canada, 2017.

[72] S. Pathare. Evaluation metrics for synthetic tabular data generation. *Data Mining and Knowledge Discovery*, 2023.

[73] Hayati Kadir Pazarlioglu et al. Thermohydraulic performance assessment of new alternative methods for anti-icing application against current application in an aircraft. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 2025.

[74] Hayati Kadir Pazarlioglu, Ahmet Umit Tepe, and Kamil Arslan. Optimization of parameters affecting anti-icing performance on wing leading edge of aircraft. *Avrupa Bilim ve Teknoloji Dergisi*, 34:19–27, 2022.

[75] Danilo Jimenez et al. Rezende. Stochastic backpropagation and approximate inference in deep generative models. *ICML*, 2014.

[76] SAE International. Aerospace recommended practice arp5485c: Endurance time tests for aircraft deicing/anti-icing fluids. Technical report, SAE International, 2013.

[77] SAE International. Aerospace recommended practice arp5718: Minimum qualification standards for aircraft deicing/anti-icing fluids. Technical report, SAE International, 2016.

[78] SAE International. ARP5485B: Endurance Time Test Procedures for SAE Type II/III/IV Aircraft Deicing/Anti-Icing Fluids. `https://www.sae.org/standards/content/arp5485b/`, 2017.

[79] SAE International. AMS1424: Deicing/Anti-icing Fluid, Aircraft, SAE Type I. https://www.sae.org/standards/content/ams1424/, 2018. Standard issued by SAE International.

[80] SAE International. AMS1428K: Fluid Aircraft Deicing/Anti-Icing, Non-Newtonian (Pseudoplastic), SAE Types II, III and IV. https://www.sae.org/standards/content/ams1428k/, 2018. SAE International, p. 26.

[81] SAE International. AS5900: Standard Test Method for Aerodynamic Acceptance of SAE AMS1424 and SAE AMS1428 Aircraft Deicing/Anti-icing Fluids. https://www.sae.org/standards/content/as5900/, 2020. Society of Automotive Engineers, p. 33.

[82] Tim et al. Salimans. Improved techniques for training gans. *NeurIPS*, 2016.

[83] S. et al. Santurkar. How does batch normalization help optimization? *NeurIPS*, 2018.

[84] C. Shorten and T. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 2019.

[85] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 2019.

[86] N. et al. Srivastava. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 2014.

[87] Youzhuang Sun, Junhua Zhang, and Yongan Zhang. Fluid classification through well logging is conducted using the extreme gradient boosting model based on the adaptive piecewise flatness-based fast transform feature extraction algorithm. *Physics of Fluids*, 36(1), 2024.

[88] S. Thapa et al. Comparative analysis of snowmelt-driven streamflow forecasting using machine learning techniques. *Water*, 16(15):2095, 2024.

[89] Transport Canada. Advisory Circular: Degree-Specific Holdover Times; Civil Aviation, Standards. https://tc.canada.ca/en/aviation/reference-centre/advisory-circulars, 2021.

[90] Transport Canada. Guidelines for aircraft ground-icing operations (tp 14052e), 2021.

[91] Cédric Villani. *Optimal Transport: Old and New*. Springer, 2008.

[92] E. Villeneuve, C. Charpentier, J. D. Brassard, G. Momen, and A. Lacroix. Aircraft anti-icing fluids endurance under natural and artificial snow: a comparative study. *International Review of Aerospace Engineering (IREASE)*, 15(1):1–11, 2022.

[93] J. Wang. Iot-based monitoring system for aircraft icing conditions. *Sensors*, 2021.

[94] X. et al. Wang. Synthetic data generation for engineering applications: A review. *Engineering Applications of Artificial Intelligence*, 2022.

[95] Xilu Wang, Yaochu Jin, Sebastian Schmitt, and Markus Olhofer. Recent advances in bayesian optimization. *ACM Computing Surveys*, 55(13s):1–36, 2023.

[96] Yan Wang, Huaiqing Wu, and Dan Nettleton. Stability of random forests and coverage of random-forest prediction intervals. In *Advances in Neural Information Processing Systems 2023*, page –, 2023.

[97] Yan Xie, Qian Liu, and Bo Xue. Genetic algorithm for hyperparameter optimization of neural networks. *Applied Soft Computing*, 101:107050, 2021.

[98] L. et al. Xu. Modeling tabular data using conditional gan. *NeurIPS*, 2019.

[99] Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. Modeling tabular data using conditional gan. *NeurIPS*, 2019.

[100] J. Yang et al. Snow depth estimation and historical data reconstruction from satellite and in situ data. *The Cryosphere*, 14(6):1763–1783, 2020.

[101] X. Yi et al. Deep learning-based icing detection using stacked autoencoders and dbn. *Aerospace Science and Technology*, 2021.

[102] H. Zhang. Particle swarm optimization neural network for aircraft icing prediction. *Applied Thermal Engineering*, 2020.