Deconstructing Triboinformatics: A Conceptual Framework for the Intersection of Machine Learning and Tribology

[1*]Amir Kordijazi ✉

[1]Department of Engineering, University of Southern Maine, Gorham, ME 04038, USA

Abstract

Tribology has traditionally functioned as an empirical discipline, heavily reliant on physical experimentation rather than unified first principles. To address the resulting epistemic bottlenecks, the paradigm of "Triboinformatics" has emerged, leveraging machine learning (ML) to extract structured knowledge from high-entropy, multiscale datasets. Moving beyond conventional systematic reviews that primarily catalog bibliographic trends, this paper presents a conceptual framework that deconstructs triboinformatics into its fundamental building blocks. We identify three major pillars: the Machine Learning Component (the data-driven computational engine encompassing data sources, algorithms, and task modes), the Physics Component (the domain context defined by system scale and physical laws), and the dynamic Interface between them. This paper critically analyzes this bidirectional Interface, conceptualizing two primary pathways of innovation. Direction A explores how ML algorithms drive epistemic gain through surrogate modeling, inverse materials design, and explainable AI. Conversely, Direction B examines how the strict laws of tribology compel methodological innovation, resulting in Physics-Informed Machine Learning (PIML) architectures that constrain algorithms within thermodynamically and mechanically plausible boundaries. Ultimately, this framework demonstrates that the future of triboinformatics lies not in replacing physical experiments with black-box algorithms, but in their symbiotic integration, highlighting the transition from statistical correlation to causal discovery as the next critical frontier.

Keywords: Triboinformatics, Machine Learning, Friction, Wear, Lubrication

1. Introduction

Since the term "tribology" was formally introduced in 1966, the field has continuously expanded, branching into specialized domains such as nanotribology, biotribology, and ecotribology. Despite this evolution, a fundamental challenge in surface engineering persists: while there exists a wealth of experimental data concerning the surface characteristics of diverse materials, the discipline remains heavily reliant on empirical methods. Unlike fields governed by unified governing equations, tribology often lacks derivation from physical or chemical first principles, rendering it a largely data-driven inductive science. Traditional attempts to formulate universal tribological laws have struggled to encompass the non-linear complexity of wear and friction across different scales (*1*).

To address this epistemic bottleneck, the field has recently witnessed the emergence of a new sub-discipline known as "Triboinformatics." Formalized around 2020, this paradigm shifts the focus from purely physical experimentation to the integration of information science (*1*). Triboinformatics leverages digital technology to transform high-entropy, scattered data into

structured knowledge. By utilizing advancements in artificial intelligence and machine learning, this approach employs inductive statistics to deduce laws, non-linear relationships, and behavioral predictions from extensive datasets that traditional physics-based models cannot easily resolve (*1*, *2*). In this framework, informatics does not merely store data but actively interrogates it. Intelligent algorithms are now employed to extract insights from diverse sources—including tabular experimental records (*3*), time-series sensor data (*4*), and microscopic images (*5*)—to predict tribological properties and design optimized materials.

Distinguishing itself from traditional systematic reviews that largely catalog bibliographic trends, this paper presents a conceptual framework designed to deconstruct the field of triboinformatics into its fundamental building blocks. Rather than merely summarizing literature, we identify and analyze the three major components that constitute this discipline: the Machine Learning Component, the Physics Component, and the Interface where they interact.

The structure of the paper follows this conceptual breakdown, as displayed in **Figure 1**. Section 2 dissects the ML Component, breaking it down into its constituent elements: the Data Source, the Algorithm, and the Task Mode. Section 3 establishes the Physics Component by categorizing studies according to System Scale (Nano, Micro, and Macro) and Domain (Solid Mechanics, Chemistry, Fluid Dynamics). Section 4 explores the Interface, analyzing the bidirectional flow of information between computation and mechanics. This includes Direction A, where ML approximates physical principles (Surrogate Modeling, xAI) and discovers materials (Inverse Design), and, finally, Direction B, where physical constraints actively reshape algorithmic architectures (Physics-Informed Machine Learning).

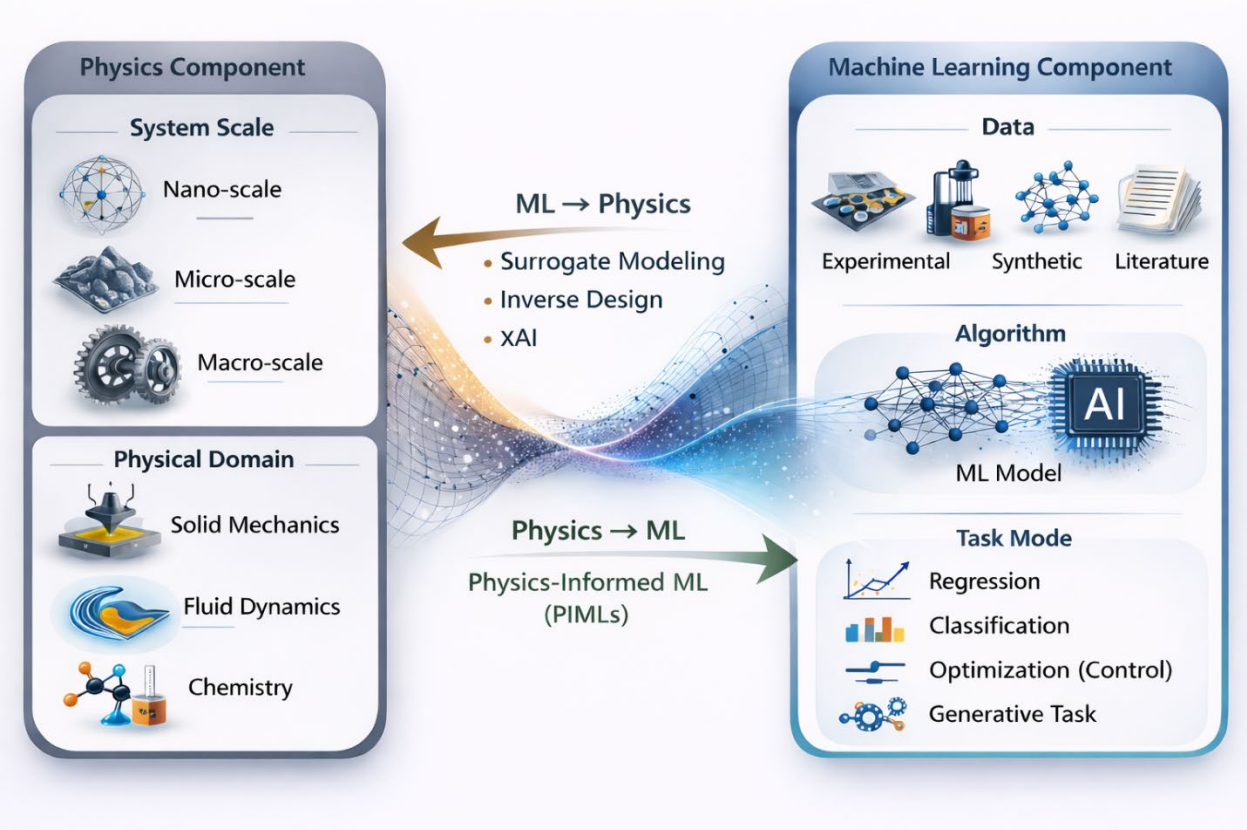

Figure 1. Integrative framework of machine learning in tribology

2. The Machine Learning Component

The Machine Learning (ML) Component constitutes the data-driven and computational framework responsible for extracting patterns from complex tribological datasets. Unlike traditional empirical equations, which rely on fixed physical constants to describe friction and wear, this component operates by dynamically mapping input features to target variables through mathematical optimization. In this section, we deconstruct this framework into its three functional pillars. First, we examine the Data Source, characterizing the provenance and granularity of the information—ranging from atomic simulations to experimental sensor readings—that serves as the foundation for training. Second, we analyze the Algorithm, detailing the specific mathematical architectures, such as decision trees and neural networks, used to process this information. Finally, we define the Task Mode, categorizing the operational objectives of these models, whether they are deployed to classify failure mechanisms, regress quantitative wear rates, or discover underlying regimes.

2.1. Data Source

The efficacy of any machine learning model in tribology is fundamentally constrained by the quality, volume, and provenance of its training data. Unlike computer vision or natural language processing, which benefit from massive, standardized datasets, tribology operates in a data-scarce environment characterized by extreme heterogeneity. Information can be sourced from three distinct reservoirs. This section categorizes these streams: Experimental Data, representing the ground-truth reality of physical interactions; Synthetic Data, generated by physics-based and data-driven simulations to populate sparse design spaces; and Literature/Text Data, a largely untapped resource of unstructured knowledge embedded in decades of academic publications.

2.1.1. Experimental Data

Experimental data serves as the foundational "ground truth" for triboinformatics, distinguishing itself from synthetic data by its inherent physical fidelity despite being constrained by high acquisition costs and time-consuming generation. Unlike the "big data" regimes found in commercial AI applications, experimental tribology typically operates within a "small data" paradigm, where datasets are manually curated from standardized mechanical tests (e.g., pin-on-disk, reciprocating rigs) and surface characterization. These datasets are highly heterogeneous, requiring the mapping of complex input vectors—spanning material composition, mechanical properties, and contact conditions—to tribological performance metrics. Consequently, the utility of this data is defined not just by its volume, but by its structure and dimensionality. Research in this domain can be broadly classified into three data modalities: (1) Static Tabular Datasets, which correlate aggregated test parameters with average performance metrics; (2) Image-Based Data, which utilizes computer vision to extract features from worn surfaces and debris; and (3) Acoustic and Time-Series Data, which captures the dynamic temporal evolution of the friction process.

2.1.1.1. Static Tabular Datasets

A predominant share of triboinformatics research relies on small, structured datasets generated from standard tribological testers, where the primary objective is to map operational parameters and material compositions to performance metrics. This approach is most extensively applied in the development of Metal Matrix Composites (MMCs) and light alloys, where the non-linear influence of reinforcement phases often defies simple analytical modeling. For instance, in the domain of aluminum alloys, Nagarajan & Arumugam(*6*) employed pin-on-disc tribometry to generate datasets varying in load and velocity, successfully training an ensemble model comprising ANN, LSTM, and Random Forest to predict the specific wear rate of additively manufactured Al5356. Similarly, Aydin(*7*) utilized reciprocating wear test data to isolate the effects of particle size and load on the volume loss of AA7075/Al_2O_3 composites. Moving to the nanoscale, Prasanth et al.(*8*) correlated graphene content with tribological outputs in AA7075 nanocomposites. This data-driven methodology extends to magnesium systems as well; Pasha et al.(*9*) integrated worn surface characteristics into their prediction of friction and wear for Mg/Si_3N_4 composites, while Haldar et al.(*10*) focused on the frictional behavior of magnesium hybrid composites. Furthermore, the robust wear behavior of Zinc-Aluminum alloys has been rigorously modeled by Sheikh and Khan (*11*, *12*), who utilized tabular data to predict high-stress abrasive wear in TiC-reinforced ZA27 and ZA37 alloys (**Figure 2(a)**). Their work culminated in a comparative study (Sheikh et al. (*13*)) that mapped oil-additive interactions to the wear performance of ZA37 composites versus grey cast iron under lubricated conditions.

Beyond bulk composites, machine learning has been widely adopted to optimize coatings and surface engineering processes, where the dataset features often shift from material composition to deposition parameters. In the field of thermal spraying, Jagadeeshanayaka et al.(*14*) generated wear rate datasets from High-Velocity Oxygen Fuel (HVOF) sprayed hydroxyapatite coatings, using process parameters like flow and feed rates to predict performance via Gradient Boosting Regression Trees. Similarly, Wang et al. (*15*) created a multi-parameter dataset for WS_2 coatings, accounting for environmental temperatures alongside load and velocity. The complexity of multi-principal element alloys was addressed by Kang et al.(*16*), who linked alloy composition (Al, Ti, Cu) to the wear resistance of AlCoCrFeNi-based high-entropy alloy coatings. In hardfacing applications, Ulas et al.(*17*) utilized wear loss data from Plasma Transfer Arc Welding (PTAW) coatings to compare the predictive efficacy of Extreme Learning Machines (ELM) against traditional ANNs. Furthermore, Shabana et al.(*18*) applied Gaussian Process Regression to model the wear depth of WC-Co cermet coatings.

Finally, the application of tabular data extends to non-metallic systems and lubricant formulation, where chemical and environmental interactions play a pivotal role. In polymer tribology, Wang et al. (*19*) and Yan et al.(*20*) analyzed the performance of PTFE composites, using experimental inputs to predict wear rates under diverse operational conditions. A unique study by Qiao et al.(*21*) on glass tribology incorporated chemical factors, specifically relative humidity, alongside mechanical loads to predict the mechanochemical wear of borosilicate and phosphate glasses. In the domain of lubrication, Gupta et al.(*22*) employed ML to predict the friction-induced characteristics of Steel 316L sliding against 100Cr6 under various lubricating conditions. Additionally, Xia et al.(*23*) demonstrated the utility of ensemble learning in formulation chemistry, optimizing the ratios of extreme-pressure anti-wear additives in amide-based greases based on reciprocating friction data (**Figure 2(b)**). Collectively, these studies

illustrate the versatility of static tabular data in capturing the "composition-process-structure-property-performance" linkage across a vast spectrum of tribological systems.

2.1.1.2. Image-Based Data

While tabular data remains the standard for performance modeling, image-based datasets represent a paradigm shift towards intelligent, automated diagnostics. In these studies, raw visual information—captured via optical microscopy or ferrography—replaces manual measurements as the primary input, enabling algorithms to "see" and interpret wear features directly. This modality typically requires Deep Learning architectures, such as Convolutional Neural Networks (CNNs), which can automatically extract complex spatial features (e.g., texture, shape, color) that are difficult to quantify numerically.

Research in this domain generally targets two applications: automated fault diagnosis and direct wear estimation. In the context of condition monitoring, Shah et al. (*5*) curated a dataset of ferrographic images containing distinct wear particle morphologies (cutting, spherical, fatigue, and sliding). By training an ensemble of deep learning models (ResNet, VGG), they achieved automated classification of machine faults, eliminating the subjectivity inherent in human visual inspection. Conversely, for direct quantification, Truong et al. (*24*) developed a computer vision framework for monitoring tool wear during the milling of Inconel 718. Instead of measuring flank wear width manually, they fed high-resolution images of worn cutting edges directly into a ResNet50 model, allowing the system to estimate the tool's condition based solely on visual cues (**Figure 2(c)**). These studies demonstrate that image data can serve as a rich, high-fidelity source for tribological insight, provided that sufficient labeled examples are available to train data-hungry vision models.

2.1.1.3. Acoustic & Time-Series Sensor Data

A novel and rapidly evolving frontier in triboinformatics is the utilization of high-frequency sensor data and time-series trajectories, moving beyond the limitations of static, post-mortem averages. This approach treats friction and wear not as single scalar values, but as dynamic processes with distinct temporal signatures. In the realm of non-intrusive monitoring, Zhao et al. (*4*) demonstrated that friction noise contains latent information directly correlated with material loss. By extracting features from acoustic signals, they trained machine learning algorithms to predict wear rates without physical inspection, offering a pathway for real-time condition monitoring (**Figure 2(d)**). Similarly, shifting the focus from final-state prediction to process evolution, Granja and Higgs III (*25*) developed a data-driven model for lubricated four-ball tests that predicts the time-dependent trajectory of both the Coefficient of Friction (CoF) and Wear Scar Diameter (WSD). This capability to forecast the progression of failure, rather than just the endpoint, marks a critical step toward establishing "digital twins" for tribological systems.

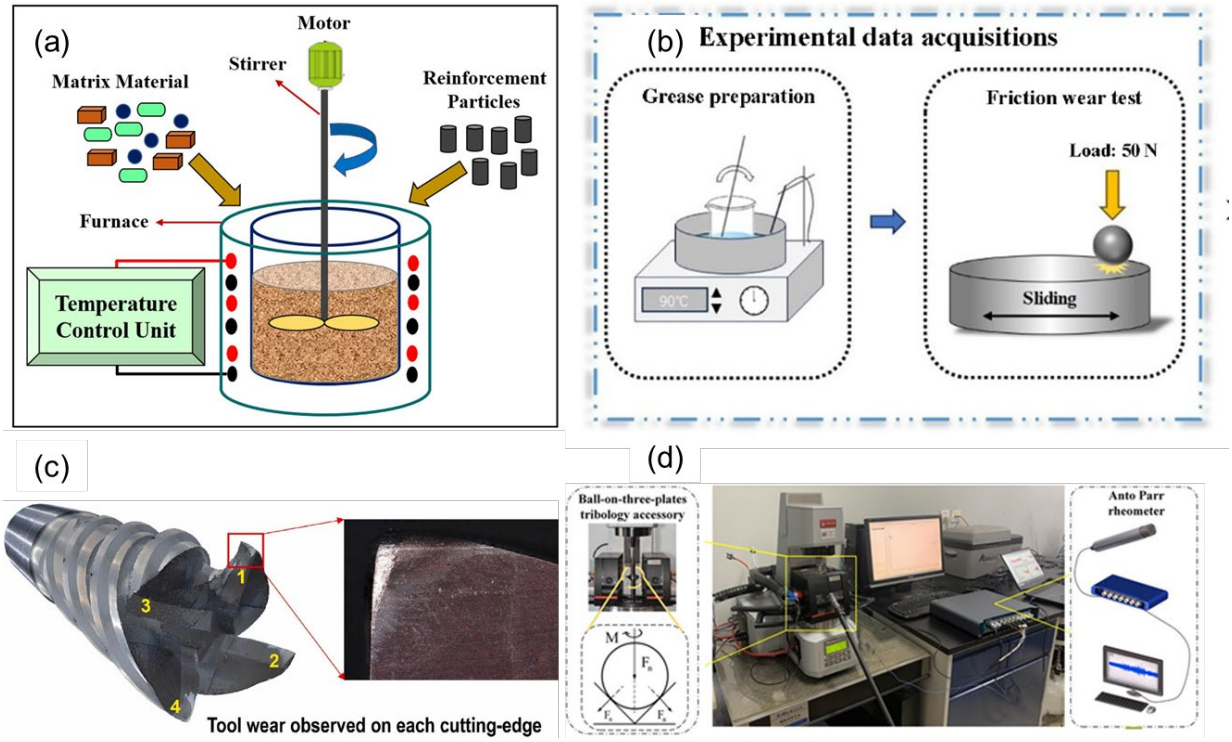

Figure 2. Classification of experimental data sources in triboinformatics. (a) Structured tabular data for Metal Matrix Composites (MMCs) correlating composition with wear rates (*13*). (b) Chemical formulation datasets for lubricant additive optimization (*23*). (c) Unstructured image-based data for wear mechanism classification (*24*). (d) Time-series acoustic emission signals to predict wear rates (*4*).

2.1.2. Synthetic Data

To overcome the "small data" bottleneck inherent in experimental tribology, a growing body of research has turned to synthetic data generated via physics-based simulations. Unlike experimental datasets, which are constrained by physical resources and acquisition time, synthetic data offers a high-volume, low-cost alternative where the underlying laws of mechanics and chemistry serve as the generative engine. This approach allows for the massive exploration of parameter spaces—ranging from macroscopic contact mechanics to atomic-level lubrication mechanisms—that would be infeasible to test empirically. Furthermore, simulations provide access to "hidden" variables, such as interfacial binding energies or internal stress distributions, which are often inaccessible to physical sensors but critical for training robust machine learning models. Research in this domain is stratified by physical scale, progressing from Continuum Mechanics (FEA) for component-level analysis, to Molecular Dynamics (MD) for nanoscale interactions, and finally to First-Principles (DFT) for electronic structure predictions.

2.1.2.1. Continuum Mechanics Data

At the macroscopic scale, Finite Element Analysis (FEA) serves as the primary engine for generating synthetic data related to component-level wear and contact mechanics. These

simulations allow researchers to populate the design space of stress distributions and deformation profiles that are experimentally elusive. For instance, Zhu et al. (*26*) utilized explicit dynamic FEA to construct a comprehensive database of line-contact friction behaviors. By systematically varying surface roughness parameters (R_a, R_q) and normal loads, they trained a Random Forest model to predict the coefficient of friction for rough surfaces. Moving beyond static prediction to dynamic monitoring, Dai et al. (*27*) leveraged transient FEA simulations to build a "Digital Twin" for sliding bearings (**Figure 3(a)**). Their model generated time-series data of wear depth profiles and contact pressures, which were then used to train a parallel hybrid model of finite element and deep neural network (PFENN), demonstrating how continuum simulations can fill the data gaps required for real-time prognostic systems.

2.1.2.2. Molecular Dynamics (MD) Data

Descending to the atomic scale, Molecular Dynamics (MD) simulations provide a unique source of synthetic data that captures the fundamental origins of friction and wear—phenomena often lost in macroscopic continuum models. Unlike experimental studies limited by sensor resolution, MD generates time-step trajectories of atoms, allowing researchers to extract "hidden" features such as interfacial binding energies and molecular conformational changes to train machine learning models (*28*).

This capability is particularly powerful for high-throughput materials screening. Quach et al. (*29*) demonstrated the scalability of this approach by performing a massive combinatorial screening of 9,747 unique monolayer films, totaling over 116,000 simulations using the MoSDeF framework. By feeding variables like terminal group chemistry and backbone stiffness into a Random Forest model, they successfully predicted tribological properties across a vast design space that would be impossible to explore experimentally.

Beyond screening, MD data is essential for elucidating complex interfacial mechanisms. Qiu et al. (*30*) simulated a graphene/base oil lubrication system to generate synthetic data on the interplay between oil chain length, contact pressure, and graphene layers (**Figure 3(b)**). Their ML analysis revealed that longer oil chains increase interlayer sliding resistance, providing a mechanistic explanation for empirical observations. Furthermore, MD allows for the engineering of novel input features; Yuan et al. (*31*) introduced the concept of "energy fingerprints," extracting Van der Waals energies and bond length changes from simulations of functionalized graphene/NBR composites. These physics-based features served as robust inputs for a Random Forest regressor to predict wear rates, effectively bridging the gap between atomic-scale energetics and macroscopic wear performance.

2.1.2.3. First-Principles / DFT Data

At the most fundamental level, First-Principles-Density Functional Theory (DFT) calculations generate data based on quantum mechanical interactions, providing the highest fidelity for understanding electronic and atomic origins of friction. These datasets are computationally expensive to generate but offer "pure" descriptors—such as charge transfer, orbital overlaps, and potential energy surfaces (PES)—that are devoid of experimental noise.

This data modality is pivotal for the design of 2D solid lubricants. Niu et al. (*32*) calculated sliding energy barriers and charge transfer rates for a wide array of elemental 2D materials. By training machine learning models on this electronic-level data, they successfully predicted macro-scale friction coefficients, effectively bridging the gap between angstrom-scale quantum mechanics and engineering-scale performance (**Figure 3(c)**). Similarly, Liu et al. (*33*) investigated the friction of corrugated graphene, utilizing DFT to compute interfacial binding energies and PES corrugation. Their work linked these quantum parameters to frictional shear stress, providing a predictive framework for tuning the lubricity of textured nanostructures. Furthermore, Barik and Woods (*34*) expanded this scope by compiling a dataset of crystallographic and electronic properties for diverse two-dimensional materials. Their data-driven model established explicit links between lattice parameters and frictional energy dissipation, demonstrating that tribological behavior can be forecasted directly from the fundamental electronic fingerprint of a material.

2.1.2.4. Thermodynamic Simulation Data

While dynamic simulations capture the process of wear, thermodynamic simulations focus on predicting the material state—specifically, phase stability and microstructural evolution. Methods such as CALPHAD (Calculation of Phase Diagrams) enable the generation of synthetic datasets that map chemical composition to equilibrium phases, providing critical features that are often too costly to characterize for every potential alloy variant.

In this context, Tong et al. (*35*) leveraged Thermo-Calc software to generate a comprehensive synthetic dataset for C-V-Cr-Mo wear-resistant steels. By calculating phase diagrams, solid solution contents, and precipitation amounts (specifically MC and M_2C carbides), they augmented their limited experimental data with physics-based microstructural descriptors. This synergy allowed them to build a robust "composition-microstructure-performance" model, effectively using thermodynamic simulation to bridge the gap between alloy chemistry and macroscopic wear resistance (**Figure 3(d)**).

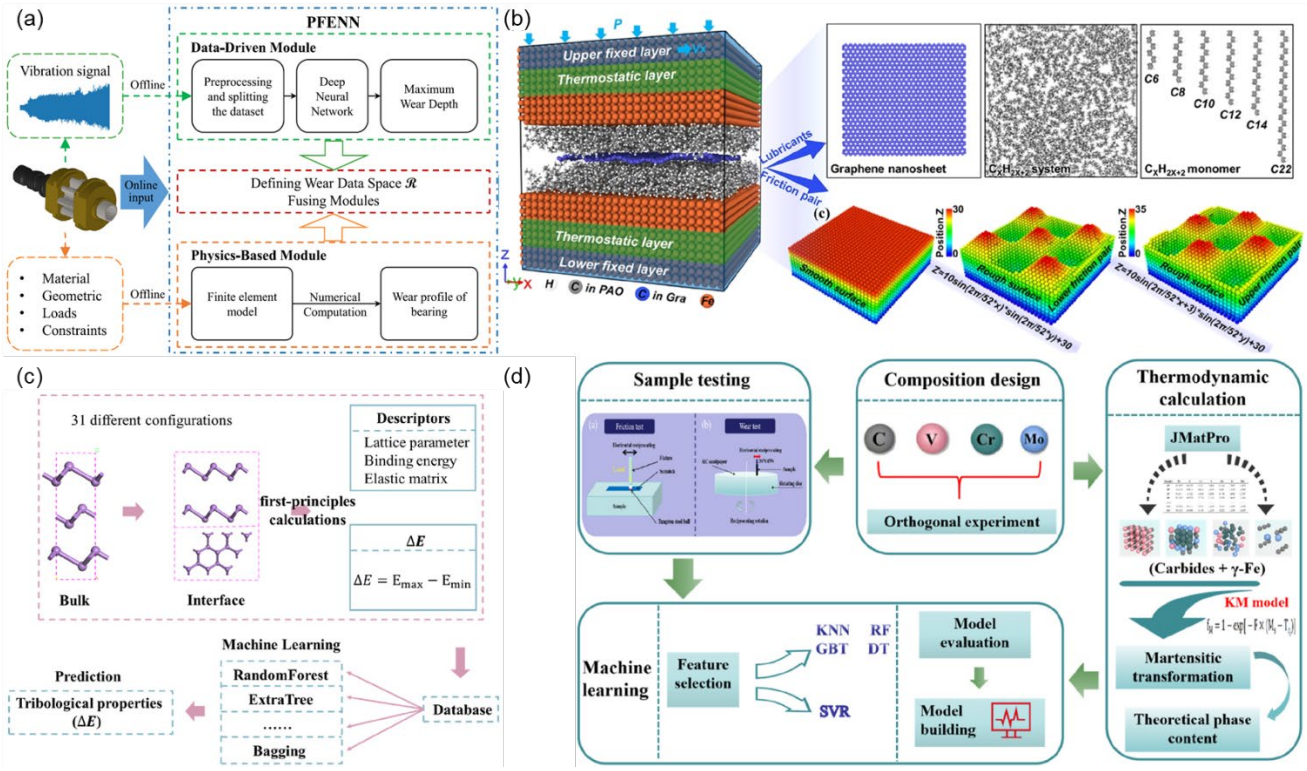

Figure 3. Applying synthetic tribological data for training machine learning models. (a) PFENN for wear degradation of sliding bearing (*27*), (b) MD simulation models of lubrication system and lubricant (*30*), (c) Tribological properties prediction by machine learning models trained DFT data (*32*), and (d) building composition-microstructure-performance model based on experimental testing and thermodynamic calculations (*35*).

2.1.2.4. Data-Driven Synthetic Data

Finally, it is critical to distinguish between physics-based simulations (FEA, MD, DFT) and data-driven synthetic data. While the former relies on explicit physical laws to generate data, a novel class of "Generative AI" methods—specifically Generative Adversarial Networks (GANs) and Diffusion models—has emerged to address data scarcity. These algorithms learn the underlying statistical probability of a dataset to generate realistic new samples, offering a powerful solution for augmenting small datasets or correcting class imbalances (e.g., the scarcity of failure data). As this represents a distinct operational capability of the machine learning pipeline, the specific mechanisms and applications of these generative models are discussed in **Section 1.3.4 (Generative Tasks)**.

2.1.3. Literature/Text Data

A significant bottleneck in traditional tribology is the fragmentation of knowledge across thousands of isolated publications, where valuable experimental results remain trapped in unstructured text and tables. Literature-mined data addresses this limitation by converting vast archives of historical research into structured "Tribo-Data," enabling the discovery of universal

trends that no single experimental study could reveal. Unlike primary data generation, which is constrained by cost and equipment, this approach leverages Natural Language Processing (NLP) and manual curation to aggregate diverse datasets. This allows researchers to train machine learning models on "big data" scales—often exceeding thousands of data points—to identify global "process-structure-property-performance" relationships, standardize non-uniform testing conditions, and accelerate material selection by learning from the collective intelligence of the past.

Research in this domain has been particularly fruitful in the optimization of Metal Matrix Composites (MMCs), where data mining has successfully established design rules for self-lubricating systems. Hasan et al. (*36–38*) pioneered the application of triboinformatics to aluminum-graphite composites, constructing extensive datasets from the literature to compare machine learning predictions against traditional empirical equations. Their work demonstrated that data-driven models (e.g., KNN, SVM, ANN) could accurately map complex non-linear behaviors that classical mechanics failed to capture. Furthermore, by applying Principal Component Analysis (PCA) to mined data, they identified distinct clusters representing the transition between solid and liquid lubrication regimes, effectively mapping the "genes" of tribological failure (*37*). Feature importance analysis on these datasets quantitatively confirmed that graphite content is the dominant variable for predicting wear in composites (*36*, *38*), whereas operational parameters (load/speed) are more critical for unreinforced alloys (*3*). Building on this foundation, Ning et al. (*39*) expanded the scope to include copper-matrix composites, aggregating over 275 data points to rigorously compare the tribological efficiency of Cu/Gr versus Al/Gr systems using Random Forest and Gradient Boosting algorithms.

Beyond bulk composites, literature mining has proven essential for surface engineering and advanced alloys, where the high cost of fabrication limits experimental volume. In the field of Diamond-Like Carbon (DLC) coatings, Cherguy et al. (*40*) addressed the critical issue of data heterogeneity by compiling a massive dataset of approximately 4,100 points from diverse studies. Despite the lack of standardization in reported test conditions, their Deep Learning models successfully normalized these variations to predict dry friction coefficients, proving that ML can find robust signals amidst "noisy" literature data. Similarly, Sivaraman and Radhika (*41*) reviewed and aggregated wear performance data for High Entropy Alloy (HEA) coatings, enabling the modeling of wear rates for complex multi-principal element systems that are too expensive to characterize exhaustively. In the domain of steel design, Liang et al. (*42*) established a "composition-process-performance" database for Duplex Stainless Steels (DSS). By training Random Forest models on this historical data, they significantly accelerated the alloy design process, accurately predicting microhardness, wear, and corrosion potential without the need for extensive trial-and-error experimentation.

The **Figure 4** illustrates the provenance of data used in triboinformatics by categorizing data sources into three main groups: experimental, synthetic, and literature/text data. Each category represents a distinct pathway through which tribological knowledge is generated or compiled. Experimental data include tabular measurements, imaging data, and acoustic or time-series sensor signals, while synthetic data originate from physics-based simulations and computational models. Literature and text data represent knowledge extracted from published studies, reports, and databases.

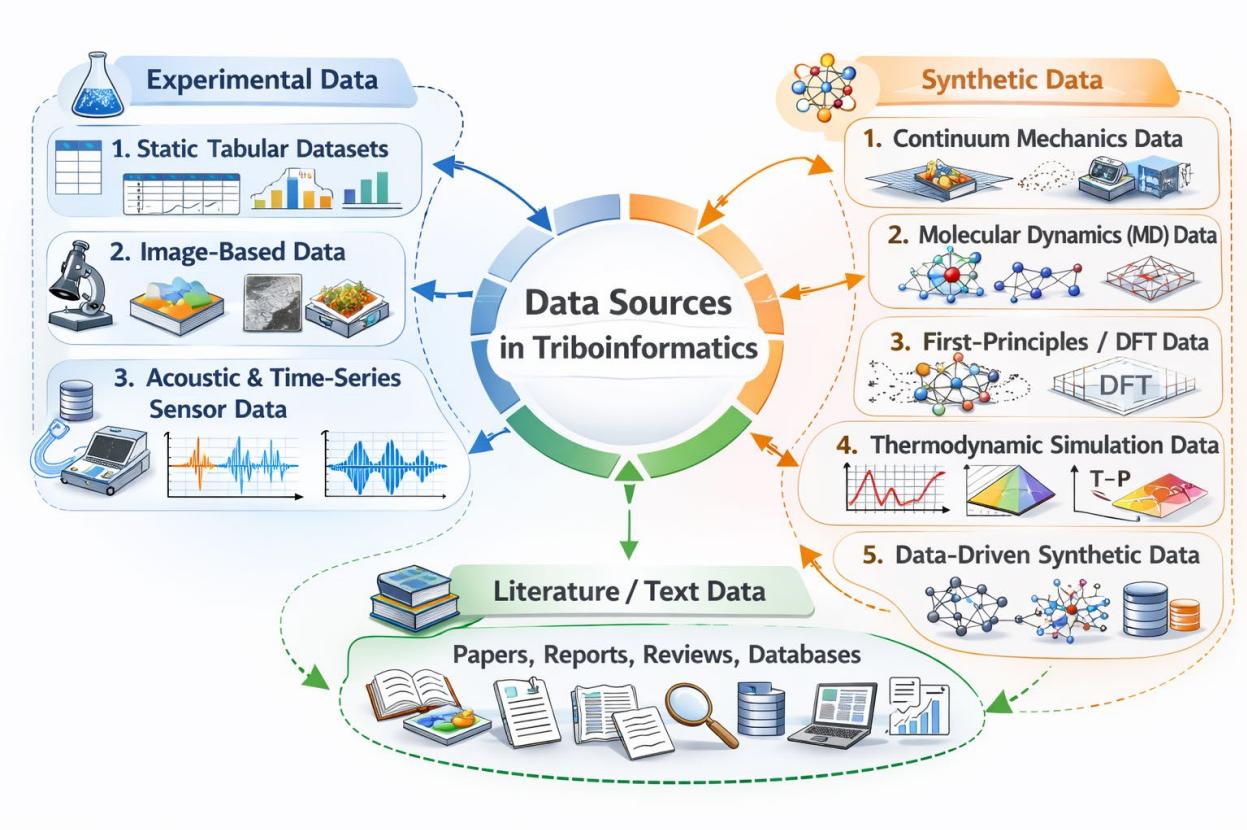

Figure 4. Provenance of data in triboinformatics, showing the three primary data sources—experimental data, synthetic data, and literature/text data—and the main data types within each category.

2.2. The Algorithm

The algorithmic component acts as the engine of the triboinformatics pipeline, transforming raw data into predictive insights. While comprehensive taxonomies of these algorithms are available in recent systematic reviews (*43–45*), the field is operationally divided into distinct modeling strategies based on data complexity. For the majority of experimental tribology, which operates in "small data" regimes (typically fewer than 1,000 data points), researchers prioritize "shallow" learning algorithms that offer robustness and interpretability over computational intensity. Basic instance-based learning methods, such as K-Nearest Neighbors (KNN), and linear models, such as Ridge or Linear Regression (LR), are frequently employed as baselines to establish simple correlations. For instance, Liang et al. (*42*) utilized Ridge Regression and KNN alongside decision trees to model the properties of duplex stainless steels, leveraging these algorithms for their simplicity and low computational cost when dealing with limited composition datasets. When the decision boundaries become more complex, Support Vector Machines (SVM) are the standard choice due to their effectiveness in high-dimensional spaces (*46*); this was demonstrated by Ji et al. (*47*) to link heat treatment parameters to wear resistance, and by Shabana et al. (*18*) to model friction in cermet coatings.

As the non-linearity of the data increases, the field shifts toward ensemble methods and simple neural architectures. Random Forest (RF) (*48*) and Gradient Boosting (*49*) are currently the dominant algorithms for tabular tribological data because they handle noise well and provide intrinsic measures of feature importance, a critical capability for material design. This utility is evident in the work of Ning et al. (*39*) and Quach et al. (*29*), who deployed these ensembles to screen composites and molecular films, respectively. Alongside these ensembles, shallow Artificial Neural Networks (ANN)—specifically simple Multi-Layer Perceptrons (MLP)—remain a popular choice for their universal approximation capabilities. While prone to overfitting on small datasets, they are effective when sufficient historical data is available. Hasan et al. (*3*) and Nagarajan and Arumugam (*6*) successfully employed simple ANNs to map non-linear wear behaviors in aluminum alloys, often benchmarking them against classical regression to highlight the superior flexibility of the neural approach. Cherguy et al. (*40*) further illustrated this by comparing ANNs with tree-based models for DLC coatings, finding that while ANNs are powerful, ensemble trees often yield more robust generalizations on heterogeneous literature data.

Conversely, Deep Learning is reserved for scenarios where manual feature extraction is infeasible, specifically for unstructured perceptual data like images or time-series signals. In these "high-fidelity" applications, the architecture must automatically learn spatial or temporal features. Convolutional Neural Networks (CNNs) are the standard for visual tasks, as seen in Truong et al. (*24*) and Shah et al. (*5*), where deep architectures bypassed manual measurements to estimate tool condition and classify wear particles directly from raw pixels. Similarly, for dynamic processes, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are utilized to capture time-dependent evolutions. Granja and Higgs III (*25*) and Dai et al. (*27*) leveraged these sequence-based models to predict the future trajectory of friction and wear, effectively enabling the creation of digital twins that can forecast failure before it occurs.

2.3. Task Mode (The Output)

The final stage of the machine learning pipeline is defined by the specific task the model is designed to solve. As displayed in **Figure 4**, these tasks generally fall into four categories: quantifying continuous variables (Regression), diagnosing discrete states (Classification), discovering optimal parameters (Optimization), or generating synthetic data (Creation).

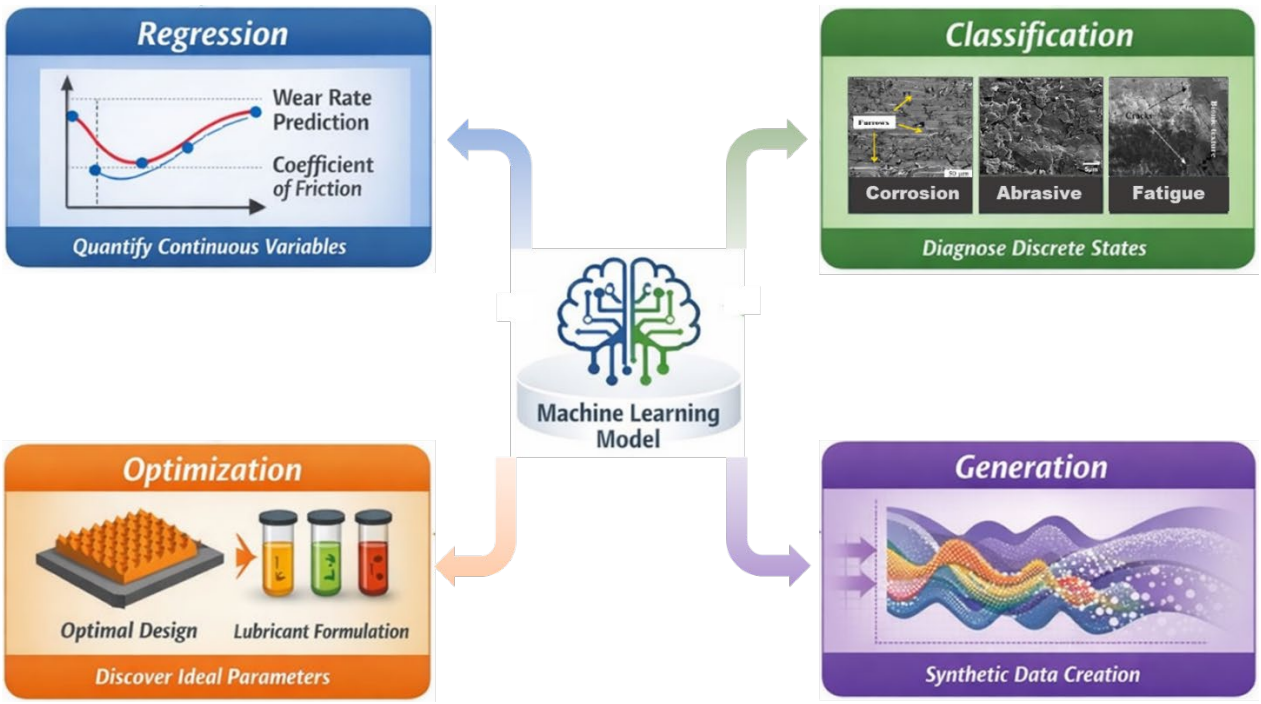

Figure 4. Overview of the triboinformatics pipeline illustrating the four primary machine learning outputs in tribological research: regression for predicting wear and coefficient of friction, classification for identifying wear mechanisms or system states, optimal control for prescribing parameters that minimize friction and wear, and generation for creating synthetic tribological data (*44*).

2.3.1. Regression (Quantification)

Regression is the most prevalent task in triboinformatics, serving as the digital successor to traditional empirical equations like Archard's Law. In this mode, the machine learning model functions as a universal approximator, mapping a complex vector of inputs—spanning material composition, surface topology, and operating conditions—to a continuous numerical output. Unlike analytical models which assume fixed power-law relationships, ML regression captures highly non-linear interactions to predict performance metrics with high fidelity, ranging from macroscopic wear rates to atomic energy barriers.

The majority of research utilizes regression to predict scalar averages of tribological performance, specifically wear rates and coefficients of friction (CoF). In the domain of metal matrix composites, Nagarajan and Arumugam (*6*) and Aydin (*7*) successfully applied regression models to predict specific wear rates and volume losses for aluminum alloys, while Sheikh and Khan (*11*, *12*) extended this approach to model high-stress abrasive wear in Zinc-Aluminum alloys. This capability is equally critical for self-lubricating systems; Ning et al. (*39*) and Hasan et al. (*36*, *37*) utilized regression to quantify the wear-reducing efficiency of graphite in Aluminum and Copper matrices. Similarly, Wang et al. (*19*) and Haldar et al. (*10*) targeted wear rates in PTFE and Magnesium hybrids, respectively. Beyond bulk materials, regression is extensively used to forecast the performance of surface coatings. Jagadeeshanayaka et al. (*14*)

predicted the wear rate of HVOF-sprayed hydroxyapatite, while Kang et al. (*16*) and Sivaraman and Radhika (*41*) modeled the wear resistance of high-entropy alloys. Furthermore, Cherguy et al. (*40*) trained models to predict the dry friction of Diamond-Like Carbon (DLC) coatings, and Zhu et al. (*26*) regress the CoF of rough surfaces based on contact mechanics parameters.

Beyond direct performance metrics, regression is increasingly employed to predict intermediate material properties that govern tribological behavior, effectively acting as a virtual characterization tool. Liang et al. (*42*) and Qiao, Zhu, and Inoue (*50*) used regression to predict microhardness and corrosion potential in Duplex Stainless Steels and Fe-Ni-Cr-Al alloys, respectively, utilizing these predictions to screen for wear-resistant compositions before physical testing. This predictive power extends to the atomic scale, where regression models link quantum-mechanical descriptors to macroscopic friction. Niu et al. (*32*) and Liu et al. (*33*) employed regression to predict fundamental sliding energy barriers and interfacial binding energies from DFT data.

Finally, advanced regression models are moving beyond static averages to predict dynamic, time-dependent trajectories, functioning as "virtual sensors" for real-time monitoring. Instead of predicting a single final value, Dai et al. (*27*) utilized a hybrid FEA-Neural Network approach to regress the entire time-series profile of wear depth in sliding bearings. Similarly, Granja and Higgs III (*25*) developed models to forecast the future trajectory of both CoF and Wear Scar Diameter (WSD) in lubricated contacts, allowing for the prediction of failure progression. In a cross-domain application, Zhao et al. (*4*) demonstrated that regression could map acoustic friction noise signals directly to continuous wear rates, establishing a non-intrusive method for quantifying material loss.

2.3.2. Classification (Diagnosis)

While regression focuses on quantifying *how much* wear has occurred, classification aims to identify *what kind* of tribological phenomenon is taking place. In this mode, the machine learning model acts as a diagnostic expert, categorizing input data into discrete states such as specific wear mechanisms, lubrication regimes, or machine health conditions (e.g., "Healthy" vs. "Critical"). This approach is particularly valuable for automating failure analysis, where the goal is to replace subjective human interpretation with objective, data-driven labeling.

One of the most prominent applications of classification is the automated identification of wear debris and surface morphologies. Shah et al. (*5*) demonstrated this by applying deep learning to ferrography, a domain traditionally reliant on the trained eye of a microscopist. By training deep learning architectures on images of wear particles, they successfully classified distinct debris types—such as cutting, spherical, fatigue, and sliding particles—thereby automating the diagnosis of the underlying machine fault. Similarly, as highlighted by Wang et al. (*44*), deep learning is increasingly deployed to classify wear mechanisms directly from images of worn surfaces, distinguishing between adhesive, abrasive, and oxidative modes which is critical for root-cause analysis.

Beyond visual inspection, classification is used to map the physical "state" or regime in which a tribosystem operates. Hasan et al. (*37*) utilized unsupervised classification techniques to analyze

the transition between solid and liquid lubrication in aluminum-graphite composites. By clustering the experimental data, they were able to identify distinct operational zones, effectively classifying the system's state based on the dominance of the graphite film versus the liquid lubricant. This type of "regime classification" is essential for defining the safe operating envelopes of materials and lubricants, ensuring that systems do not inadvertently drift into severe wear states.

2.3.3. Optimal Control (Action)

The most advanced capability of the triboinformatics pipeline is "Action," where the model transcends passive prediction to actively prescribe the optimal parameters required to achieve a desired performance target. While true real-time control of machinery (e.g., adjusting magnetic bearings on the fly) remains the frontier, the current literature is heavily focused on "Inverse Design"—a form of offline control where algorithms solve the backward problem of finding the ideal material or geometric configuration to minimize friction and wear. This is particularly evident in the domain of surface texturing, where the design space is too vast for empirical testing. Li et al. (*51*) demonstrated this by coupling Artificial Neural Networks with Gradient Boosting Decision Trees (GBDT) to create a "reverse model." Instead of predicting friction from texture, their system inputs a desired friction coefficient and outputs the specific texture dimensions (depth, size, coverage) required to achieve it. Similarly, Ge et al. (*52*) conducted a comparative analysis of predictive models to optimize micro-textured surfaces, effectively using ML to “tune” the surface topography for maximum tribological efficiency before a single part is manufactured.

This operational mode extends beyond geometry to the optimization of lubricants and materials. In the chemical domain, Wen et al. (*53*) integrated eXtreme Gradient Boosting (XGBoost) with Particle Swarm Optimization (PSO), a metaheuristic control algorithm, to pinpoint the exact concentration of lubricant additives that minimizes wear volume. This automates the formulation process, replacing trial-and-error mixing with algorithmic prescription. Wan et al. (*54*) applied a similar logic to the molecular design of ester lubricants, using Quantitative Structure-Property Relationship (QSPR) models to identify optimal molecular structures with low friction coefficients. On the materials front, Zahoor et al. (*55*) proposed a layered machine learning framework that not only predicts wear rates for Cu/Al-graphite composites but also acts as a decision support system, recommending specific alloying or coating strategies when baseline materials fail to meet performance thresholds. Collectively, these studies represent the shift from descriptive tribology to prescriptive engineering, where the "Action" is the algorithmic generation of an optimal design.

2.3.4. Generative Tasks (Creation)

The most novel and rapidly emerging operational mode in triboinformatics is "Creation," where the objective is not to analyze existing data but to generate new, synthetic data that mimics the statistical properties of real-world samples. This task relies on Generative Adversarial Networks (GANs) and Diffusion models to address the critical "small data" and class imbalance problems inherent in tribology, particularly the scarcity of failure data in industrial settings. Unlike physics-based simulations (FEA or MD) which generate data based on explicit physical laws,

generative models learn the underlying probability distribution of the data itself, effectively generating realistic new samples to augment training sets.

Recent literature demonstrates the power of this approach in generating synthetic sensor signals to train robust predictive models. Jiang et al. (*56*) tackled the challenge of predicting CNC machine tool wear, where collecting run-to-failure data is expensive and time-consuming. They employed a GAN-based data augmentation technique to generate synthetic multi-sensor signals—comprising current, vibration, and acoustics—that statistically matched real wear states. This allowed them to train a deep learning model on a balanced dataset, significantly improving prediction accuracy in real-world scenarios. Similarly, Zeng et al. (*57*) utilized a GAN to expand the dataset of acoustic emission signals for circular saw blades. By generating synthetic wear signatures, they overcame the severe imbalance between "healthy" and "worn" samples, enabling a downstream CNN-LSTM model to accurately predict the wear status of the blades. These studies highlight the role of generative AI not just as an analytical tool, but as a creative engine that manufactures the "fuel" required for data-driven tribology.

3. The Physics Component (The Context)

While the machine learning algorithms described previously serve as the computational engine, the Physics Component acts as the navigational system, imposing domain constraints that prevent the model from drifting into physically impossible states. As the theoretical foundations of these constraints have been exhaustively categorized in recent systematic reviews (*2*, *58–60*), we will not reiterate the fundamental governing equations here. Instead, we classify the specific literature analyzed in this study according to the physical context in which the data was generated, organizing the field by the scale of the system and the dominant physical domain.

3.1. System Scale

Tribological data is inherently multiscale, necessitating different modeling strategies depending on the geometric magnitude of the interaction. At the Nano-Scale, research focuses on the intrinsic potential energy surfaces that govern friction before mechanical wear begins. In this domain, Sattari Baboukani et al. (*61*) and Niu et al. (*32*) utilized Maximum Energy Barriers (MEB) and charge transfer rates derived from MD and DFT to predict the lubricity of 2D materials. Similarly, Wan et al. (*54*) operated at the molecular level to link the chemical structure of ester molecules directly to their friction-reducing capabilities. Scaling up to the Micro-Scale, the focus shifts to contact mechanics and surface topography, where the physical features are defined by the geometry of asperities and textures. Li et al. (*51*) and Ge et al. (*52*) focused on optimizing micro-dimple arrays, utilizing parameters such as depth, diameter, and area coverage, while Zhu et al. (*26*) bridged this gap by using FEA to simulate stress concentrations at individual asperity contacts. Finally, at the Macro-Scale, the data represents the cumulative effect of millions of micro-events. This includes the bulk wear of composites modeled by Hasan et al. (*36*) and Ning et al. (*39*), as well as the system-level dynamics of rotating machinery. For instance, Dai et al. (*27*) and Jiang et al. (*56*) modeled entire bearing assemblies and CNC machine tools, respectively, where the physical context involves vibration, acoustic emission, and bulk temperature changes rather than atomic forces.

3.2. Physical Science Domain

The "Context" is further defined by the dominant branch of physical science—Solid Mechanics, Chemistry, or Fluid Dynamics—that governs the failure mode. The majority of the reviewed literature operates in the domain of Solid Mechanics, where performance is dictated by stress, strain, and hardness. Hasan et al. (*38*) and Liang et al. (*42*) demonstrated that for metal matrix composites and steels, the physics can be effectively compressed into mechanical descriptors such as microhardness, yield strength, and plasticity indices. However, for systems relying on boundary lubrication or coatings, the physical context shifts to Tribochemistry. Wen et al. (*53*) highlighted this by optimizing lubricant additives, where the governing factors are concentration and chemical reactivity rather than load-bearing capacity. Sivaraman and Radhika (*41*) also operated in this domain, linking the corrosion potential of High Entropy Alloys to their wear performance, acknowledging that material degradation is often a synergistic mechano-chemical process. Finally, in lubricated contacts, the physics shifts to Fluid Dynamics and rheology. Hasan et al. (*37*) explicitly modeled the transition from solid-to-liquid lubrication, identifying the specific regimes where fluid viscosity supersedes surface roughness as the dominant variable. Similarly, the work by Dai et al. (*27*) on sliding bearings integrated the Reynolds equation into their predictive maintenance models, ensuring that the machine learning predictions respected the laws of fluid film formation.

4. The Interface

The Interface represents the critical junction where the computational power of the "Engine" (**Section 2**) meets the domain constraints of the "Context" (**Section 3**). It is not merely a boundary but a dynamic, bidirectional channel where data-driven insights and physical laws exchange information. As highlighted in recent systematic reviews (*45*, *60*), the true potential of triboinformatics lies not in replacing physical experiments with black-box algorithms, but in this symbiotic integration. This section explores that duality: Direction A examines how machine learning acts as a tool for epistemic gain, revealing new tribological phenomena and accelerating discovery, while Direction B explores how the strict laws of tribology are forcing the invention of new, physics-constrained machine learning architectures.

4.1. Direction A: ML to Tribology (Epistemic Gain)

This direction characterizes the flow of "Epistemic Gain"—the acquisition of new tribological knowledge generated by algorithms that was previously inaccessible through traditional experimentation or simulation alone. Here, machine learning transcends its role as a passive calculator to become an active instrument of discovery. By processing high-dimensional datasets that overwhelm human intuition, ML models are now identifying complex process-structure-property relationships in materials and lubricants. The literature demonstrates that this gain manifests primarily through three avenues: acceleration, inversion, and explanation.

4.1.1. Surrogate Modeling (Acceleration)

The primary mechanism for epistemic gain in triboinformatics is the creation of "surrogate models"—lightweight machine learning approximators that replace computationally or

experimentally expensive procedures. In the computational domain, exploring the vast design space of surface textures or molecular compositions is limited by the execution time of FEA or MD, which can take hours per iteration. Surrogate models bridge this gap by learning the mapping between input parameters and simulation outputs, effectively "memorizing" the physics to provide instantaneous predictions.

This acceleration capability allows researchers to conduct multi-scale modeling that was previously infeasible. Recent studies have successfully trained neural networks on FEA datasets to predict the contact mechanics of rough surfaces (*26*) and the wear depth of sliding bearings (*27*) in real-time, bypassing iterative numerical solvers. Crucially, this logic extends to experimental research, where the "expensive function" is the physical test itself. By training models on historical experimental data, researchers create "virtual tribometers" that predict the wear performance of metal matrix composites (*36*) or lubricant additives (*53*) without requiring new physical samples. This transforms the research workflow from a linear series of tests into a high-throughput screening process, allowing the "Context" (physics) to inform the "Engine" (ML) at speeds compatible with iterative design.

4.1.2. Inverse Design (Materials Discovery)

While surrogate modeling accelerates the evaluation of existing designs, Inverse Design represents a more profound epistemic gain: it inverts the traditional scientific method to discover entirely new materials. Rather than simulating the performance of a known structure ("Forward Problem"), machine learning algorithms are tasked with the "Inverse Problem"—taking a desired tribological target (e.g., a specific friction coefficient or minimal wear rate) as the input and generating the required chemical composition or geometric configuration as the output. This capability transforms the role of the researcher from a "tester" to a "specifier," effectively automating the formulation process.

In the domain of lubricants, this approach has been successfully applied to reverse-engineer optimal lubricants. Recent investigations identified the specific molecular chain lengths and functional groups of ester lubricants required to minimize boundary friction (*54*). Similarly, the exact concentration ratios of lubricant additives have been mathematically derived that yield the minimum wear volume, replacing heuristic trial-and-error mixing (**Figure 5(a)**) (*53*).

The approach scales seamlessly to surface engineering and metallurgy. Research on micro-textured surfaces has deployed "inverse models" that input a target friction coefficient and output the precise dimple diameter, depth, and area density needed to generate the necessary hydrodynamic lift (*51*). Furthermore, in the development of wear-resistant composites, layered machine learning frameworks now function as decision support systems; instead of merely predicting failure, they recommend specific alloying strategies or coating thicknesses to salvage materials that fall short of performance thresholds (*55*). Through inverse design, the algorithm provides the "recipe" for materials that mathematically satisfy the physical constraints of the application.

4.1.3. Explainable AI (XAI) for Mechanism Identification

One of the most significant barriers to the adoption of machine learning in tribology is the "Black Box" problem—the opacity of complex algorithms like Neural Networks and Gradient Boosting, which offer high predictive accuracy but zero transparency regarding their decision-making process. This opacity is unacceptable in safety-critical applications where engineers must understand *why* a failure is predicted. Consequently, the field has moved toward Explainable AI (XAI), a set of techniques designed to open the black box and reveal which physical parameters actually drive the model's output. This represents a critical epistemic gain: rather than relying on assumed theoretical equations, researchers use XAI to extract "feature importance" rankings directly from experimental data, effectively allowing the physics to emerge from the statistics (*62*, *63*).

In the domain of chemical tribology, XAI has proven instrumental in identifying the specific atomic structures that govern anti-wear performance. Wang et al. (*64*) applied SHAP (SHapley Additive exPlanations) to quantitative structure-property relationship for oligoether ester lubricants. The analysis revealed that specific molecular descriptors—such as the number of ether linkages and the count of functional groups—were the primary drivers of wear volume, providing chemical insights that guided the synthesis of optimized base oils. Similarly, in the optimization of lubricant additives, Wen et al. (*53*) utilized the same technique to rank the influence of various additive concentrations, mathematically confirming that specific interactions between dispersants and anti-wear agents were more critical than their individual contributions.

For macroscopic systems like brakes and clutches, XAI serves to disentangle the complex, non-linear interactions between operating conditions. Wu et al. (*65*) employed interpretable machine learning on wet friction components, demonstrating via feature contribution plots how the influence of sliding velocity and interface temperature varies non-linearly across different engagement phases. This provided a data-driven validation of the Stribeck curve transitions without explicitly programming them. Sellami et al. (*66*) extended this to brake pad materials, using Random Forest feature importance to quantify exactly how much formulation parameters (e.g., phenolic resin content) outweigh operating parameters (e.g., braking pressure) in determining the final friction coefficient. Similarly, Hulipalled et al. (*67*) used ensemble learning to isolate the specific contribution of Manganese content in ZA-27 alloys, offering a clear metallurgical roadmap for minimizing wear rates.

Perhaps the most profound application of XAI lies in visual diagnostics, where it bridges the "trust gap" in automated failure analysis. In deep learning applications for gear wear, Herwig et al. (*68*) applied Layer-wise Relevance Propagation (LRP) to convolutional neural networks. This technique generated "heatmaps" indicating exactly which pixels the AI focused on when classifying a fault. The study confirmed that the model was correctly identifying physical features like micropitting and spalling—rather than learning background noise or lighting artifacts—thereby validating the model's alignment with human expert reasoning (**Figure 5(b)**). Through these methods, XAI transforms machine learning from a passive prediction tool into an active generator of physical theory, distinguishing between meaningful causal drivers and mere statistical correlations.

4.2. Direction B: Physics to ML (Methodological Innovation)

While the previous section focused on how algorithms accelerate tribological discovery, this direction explores the reciprocal influence: how the unique constraints of tribology—specifically the laws of conservation and the high cost of data generation—are forcing the invention of new machine learning architectures. Standard "off-the-shelf" algorithms (like standard Neural Networks) are often insufficient for tribology because they are "physics-blind"; they can statistically minimize error while predicting physically impossible states, such as negative friction coefficients. Consequently, the field is moving toward Methodological Innovation, where physical laws are not just inputs but are embedded into the mathematical structure of the algorithm itself.

4.2.1. Physics-Informed Machine Learning (PIMLs)

The most prominent innovation in this space is Physics-Informed Machine Learning. PIML fundamentally alters the learning process by embedding physical laws as an Inductive Bias—modifying the architecture or loss function to ensure that the model respects conservation laws and empirical relationships (e.g., Archard's Law) (*69*).

The most common implementation is the modification of the Loss Function. In standard ML, the model minimizes the error between predictions and data (L_{data}). In PIML, a residual term ($L_{physics}$) is added to penalize physically impossible predictions. This approach offers two critical values. First, it acts as a regularizer, ensuring that predictions remain physically consistent even in regimes where no training data exists (extrapolation). Second, it drastically reduces the data requirement. Because the model is "guided" by differential equations, it does not need to relearn basic physics from scratch, making it feasible to train robust models on the sparse datasets typical of tribological experiments (*43*). A striking example of this is the "Physics-Informed Random Forest" developed by Li et al. (*70*) for carbon-ceramic composites. They modified the decision tree splitting criterion to include a penalty for violating Archard's wear equation and energy conservation principles. This ensured that even with small datasets, the model's predictions for friction and wear remained thermodynamically consistent, avoiding the "overfitting" common in pure data-driven models.

Beyond loss functions, the physical topology of the system is increasingly dictating the network architecture itself. This is evident in the shift toward Graph Neural Networks (GNNs) for granular and suspension tribology. Aminimajd et al. (*71*) (2025) argued that standard neural networks fail to capture the complex, multi-body interactions in dense suspensions. By modeling the system as a graph—where particles are nodes and frictional contacts are edges—the ML architecture was forced to mirror the actual physical connectivity of the granular media. This structure-preserving approach allowed the model to accurately predict frictional contact networks and extrapolate to flow conditions far beyond the training set (**Figure 5(c)**). This also extends to hyperparameter optimization. Yin et al. (*72*) demonstrated that the architecture of the neural network itself can be constrained by physics. They introduced a "Physics-Informed Sparrow Search Algorithm (PISSA)" to tune a CNN-LSTM network for seawater-lubricated bearings. Instead of allowing the network to select random configurations, the search algorithm was mathematically constrained by the hydrodynamic relationships between salinity, load, and

friction. This ensures that the model's complexity scales consistently with the governing fluid dynamics, preventing the "black box" from becoming overly complex and overfitting the noise.

An alternative methodological innovation involves restructuring the entire modeling workflow into a Sequential Framework. In this approach, analytical physical models are not discarded but are used as "pre-processors" or "generators" that feed high-quality, domain-specific features into the machine learning algorithm. This effectively solves the "Cold Start" problem, where pure ML models fail due to a lack of initial physical intuition. A prime example of this is the "Sequential PIML" approach developed by Pashmforoush (*73*) for predicting tool wear. Recognizing that pure data-driven models struggle to learn complex thermo-mechanical interactions from scratch, the authors first utilized an analytical physics model to calculate cutting forces and stress distributions. These theoretically derived values were then fed as "physics-enriched features" into the machine learning model. This hybrid architecture outperformed pure ML models because the algorithm did not need to "re-discover" basic cutting mechanics; it could focus entirely on learning the non-linear wear residuals that the analytical model missed.

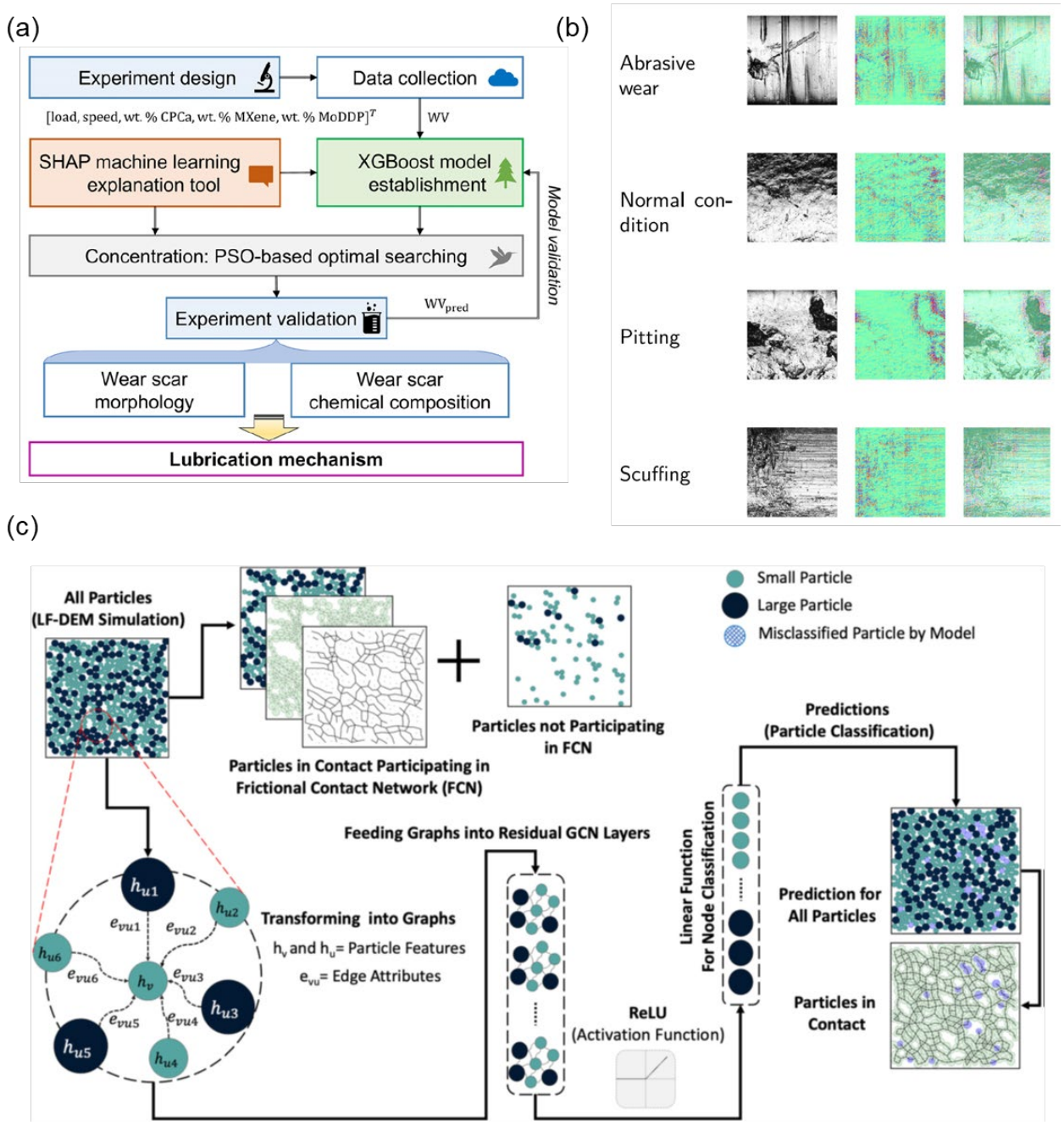

Figure 5. (a) ML-based concentration optimization of the lubricant additive (*53*), (b) CNN classification of mould wear mechanism and respective heat map (*68*), (c) Development of graph neural network (GNN) for predicting frictional contact networks (FCNs) (*71*).

Conclusion:

The transition of tribology from a purely empirical science to a data-driven discipline—Triboinformatics—represents a profound paradigm shift in surface engineering. Rather than

providing a standard bibliographic summary, this paper has established a conceptual framework to deconstruct this rapidly evolving field into three distinct but interacting elements. First, the Machine Learning Component serves as the computational engine, driven by heterogeneous Data Sources (experimental, synthetic, and literature), processed through sophisticated Algorithms, and deployed via specific Task Modes (regression, classification, control, and generation). Second, the Physics Component establishes the essential domain boundaries, contextualizing the data based on the System Scale (nano to macro) and the governing Physical Science Domain (solid mechanics, tribochemistry, or fluid dynamics). Finally, the Interface represents the dynamic, bidirectional junction where computation and physical reality meet.

Our analysis reveals that the true potential of triboinformatics is realized only when the computational power of the ML engine is firmly grounded in its physical context. Through Direction A (ML to Tribology), we observe machine learning acting as an epistemic catalyst. It accelerates research via surrogate modeling, facilitates the inverse design of novel composites and lubricants, and unmasks complex wear mechanisms through Explainable AI (XAI). Conversely, Direction B (Physics to ML) highlights an equally critical trend: the unique, multiscale constraints of tribology are forcing the evolution of AI itself. Through the development of Physics-Informed Machine Learning (PIML), sequential frameworks, and topology-aware architectures like Graph Neural Networks, the discipline is actively moving away from physics-blind "black boxes" toward models that are structurally forced to respect conservation laws and contact mechanics.

Looking forward, the maturation of triboinformatics hinges on overcoming the inherent limitations of inductive statistics. While current state-of-the-art tools excel at identifying complex correlations and feature importances, they fundamentally lack the ability to independently determine mechanism directionality. The next major frontier at this computation-mechanics interface is the integration of Causal Discovery—algorithms capable of mathematically deriving the direction of cause-and-effect from observational data without relying on prior human assumptions. By bridging the gap between statistical prediction and true causal understanding, triboinformatics can ultimately transcend advanced data fitting to uncover the fundamental, universal laws governing friction and wear.

Acknowledgment: The Maine Space Grant Consortium (award number SG-24-29 and EP-26-06) supported this research. The author gratefully acknowledge their supports.

Conflict of Interest: The author declare that they have no conflict of interest.

References:

1. A. Kordijazi, H. M. Roshan, A. Dhingra, M. Povolo, P. K. Rohatgi, M. Nosonovsky, Machine-learning methods to predict the wetting properties of iron-based composites. *Surface Innovations*, 1–9 (2020).

2. Z. Zhang, N. Yin, S. Chen, C. Liu, Tribo-informatics: Concept, architecture, and case study. *Friction* **9**, 642–655 (2021).

3. M. S. Hasan, A. Kordijazi, P. K. Rohatgi, M. Nosonovsky, Triboinformatic modeling of dry friction and wear of aluminum base alloys using machine learning algorithms. *Tribology International* **161**, 107065 (2021).

4. H. Zhao, Z. Yang, B. Zhang, C. Xiang, F. Guo, Machine Learning Algorithms for Predicting Wear Rates on the Basis of Friction Noise. *Tribology Transactions* **67**, 730–743 (2024).

5. R. Shah, N. V. Sridharan, T. K. Mahanta, A. Muniyappa, S. Vaithiyanathan, S. M. Ramteke, M. Marian, R. Shah, N. V. Sridharan, T. K. Mahanta, A. Muniyappa, S. Vaithiyanathan, S. M. Ramteke, M. Marian, Ensemble Deep Learning for Wear Particle Image Analysis. *Lubricants* **11** (2023).

6. M. Nagarajan, M. Arumugam, Advanced ensemble machine learning and response surface methodology for optimizing and predicting tribological performance of CMT-WAAM fabricated Al5356 alloy. *Int J Interact Des Manuf* **19**, 4535–4563 (2025).

7. F. Aydin, The investigation of the effect of particle size on wear performance of AA7075/Al2O3 composites using statistical analysis and different machine learning methods. *Advanced Powder Technology* **32**, 445–463 (2021).

8. I. S. N. V. R. Prasanth, P. Jeevanandam, P. Selvaraju, K. Sathish, S. K. Hasane Ahammad, P. Sujatha, M. Kaarthik, S. Mayakannan, B. Sasikumar, Study of Friction and Wear Behavior of Graphene-Reinforced AA7075 Nanocomposites by Machine Learning. *Journal of Nanomaterials* **2023**, 5723730 (2023).

9. M. B. Pasha, R. N. Rao, S. Ismail, M. Gupta, P. S. Prasad, Tribo-informatics approach to predict wear and friction coefficient of Mg/Si3N4 composites using machine learning techniques. *Tribology International* **196**, 109696 (2024).

10. B. Haldar, H. Joardar, A. K. Mondal, N. H. Alrasheedi, R. Khan, M. P. Papathi, B. Haldar, H. Joardar, A. K. Mondal, N. H. Alrasheedi, R. Khan, M. P. Papathi, Machine-Learning-Driven Analysis of Wear Loss and Frictional Behavior in Magnesium Hybrid Composites. *Crystals* **15** (2025).

11. K. A. Sheikh, M. M. Khan, Tribo-informatics analysis of in-situ TiC reinforced ZA27 alloy: Microstructural insights and wear performance modeling by machine learning. *Tribology International* **199**, 110021 (2024).

12. K. A. Sheikh, M. M. Khan, Predictive modeling of abrasive wear in in-situ TiC reinforced ZA37 alloy: A machine learning approach. *Tribology International* **202**, 110291 (2025).

13. K. A. Sheikh, M. M. Khan, M. N. Bhat, Comparative study of wear behaviour of ZA37 alloy, ZA37/SiC composite, and grey cast iron under lubricated conditions: Predictive modeling by machine learning. *Tribology International* **207**, 110623 (2025).

14. N. Jagadeeshanayaka, M. S. Negi, S. C. Jambagi, Tribological and microstructural evaluation of HVOF-sprayed hydroxyapatite coatings with machine learning-based wear rate prediction. *Tribology International* **213**, 111038 (2026).

15. G. Wang, J. Song, G. Zhao, Q. Ding, T. Yin, H. Wang, Tribological performance prediction of WS2 coating under different conditions by machine learning. *Wear* **532–533**, 205092 (2023).

16. J. Kang, Y. Niu, Y. Zhou, Y. Fan, G. Ma, J. Kang, Y. Niu, Y. Zhou, Y. Fan, G. Ma, Wear Resistance Prediction of AlCoCrFeNi-X (Ti, Cu) High-Entropy Alloy Coatings Based on Machine Learning. *Metals* **13** (2023).

17. M. Ulas, O. Altay, T. Gurgenc, C. Özel, A new approach for prediction of the wear loss of PTA surface coatings using artificial neural network and basic, kernel-based, and weighted extreme learning machine. *Friction* **8**, 1102–1116 (2020).

18. S. Shabana, J. K. Jatavallabhula, R. N. Kumar, R. Chinthalapudi, B. Pappula, Towards Intelligent Tribology: Predicting Wear and Friction in WC–Co Coatings with Machine Learning. *J Bio Tribo Corros* **11**, 88 (2025).

19. Q. Wang, X. Wang, X. Zhang, S. Li, T. Wang, Tribological properties study and prediction of PTFE composites based on experiments and machine learning. *Tribology International* **188**, 108815 (2023).

20. Y. Yan, J. Du, S. Ren, M. Shao, Y. Yan, J. Du, S. Ren, M. Shao, Prediction of the Tribological Properties of Polytetrafluoroethylene Composites Based on Experiments and Machine Learning. *Polymers* **16** (2024).

21. Q. Qiao, H. He, J. Yu, Y. Zhang, H. Qi, Applicability of machine learning on predicting the mechanochemical wear of the borosilicate and phosphate glass. *Wear* **476**, 203721 (2021).

22. M. K. Gupta, M. E. Korkmaz, A. Karolczuk, N. S. Ross, G. Vashishtha, J. B. Królczyk, R. Demirsöz, G. M. Królczyk, A study on friction induced tribological characteristics of steel 316 L against 100 cr6 alloy under different lubricating conditions with machine learning model. *Tribology International* **195**, 109599 (2024).

23. Y. Xia, Z. He, X. Feng, Stacking ensemble learning framework for predicting tribological properties and optimal additive ratios of amide-based greases. *Friction* **13**, 9440982 (2025).

24. T. T. Truong, J. Airao, S. Fattahi, B. Azarhoushang, P. Karras, R. Aghababaei, Image-based machine learning model for tool wear estimation in milling Inconel 718. *Wear* **571**, 205865 (2025).

25. V. Granja, C. F. H. Iii, V. Granja, C. F. H. Iii, Data-Driven AI Model for Time-Based Prediction of Friction and Wear in Lubricated Tribosystems. *Lubricants* **14** (2026).

26. S. Zhu, Y. Liu, B. Zhao, S. Zhang, Friction Behaviors of Rough Line Contact Using Machine Learning-Assisted Finite Element Analysis. *Acta Mech. Solida Sin.*, doi: 10.1007/s10338-025-00613-9 (2025).

27. J. Dai, L. Tian, T. Han, H. Chang, Digital Twin for wear degradation of sliding bearing based on PFENN. *Advanced Engineering Informatics* **61**, 102512 (2024).

28. T. Yin, G. Wang, Z. Guo, Y. Pan, J. Song, Q. Ding, G. Zhao, T. Yin, G. Wang, Z. Guo, Y. Pan, J. Song, Q. Ding, G. Zhao, Molecular Dynamics Simulation on Polymer Tribology: A Review. *Lubricants* **12** (2024).

29. C. D. Quach, J. B. Gilmer, D. Pert, A. Mason-Hogans, C. R. Iacovella, P. T. Cummings, C. McCabe, High-throughput screening of tribological properties of monolayer films using molecular dynamics and machine learning. *J. Chem. Phys.* **156**, 154902 (2022).

30. F. Qiu, H. Song, Z. Yang, Z. Lu, M. Jiang, X. Hu, Molecular dynamics simulation and machine learning prediction of tribological properties of graphene solid-liquid two-phase lubrication system. *Tribology International* **202**, 110347 (2025).

31. Q. Yuan, Y. Li, E. He, B. Yang, S. Wang, Coupling molecular dynamics–derived energy fingerprints with random forest to reveal wear mechanisms in functionalized nanocomposites. *Journal of Reinforced Plastics and Composites*, 07316844251390999 (2025).

32. Y. Niu, Y. Wang, M. He, W. He, Z. Zhao, Z. Lu, Calculation and prediction of sliding energy barriers by first-principles combined with machine learning. *Ceramics International* **49**, 24752–24761 (2023).

33. Z. Liu, X. Zhao, H. Wang, Y. Ma, L. Gao, H. Huang, Y. Yan, Y. Su, L. Qiao, A first-principles and machine learning combined method to investigate the interfacial friction between corrugated graphene. *Modelling Simul. Mater. Sci. Eng.* **29**, 035011 (2021).

34. R. K. Barik, L. M. Woods, Frictional Properties of Two-Dimensional Materials: Data-Driven Machine Learning Predictive Modeling. *ACS Appl. Mater. Interfaces* **16**, 40149–40159 (2024).

35. S. Tong, S. Zhang, C. Chen, T. Jiang, P. Li, S. Wei, Building a composition-microstructure-performance model for C–V–Cr–Mo wear-resistant steel via the thermodynamic calculations and machine learning synergy. *Journal of Materials Research and Technology* **36**, 4308–4320 (2025).

36. M. S. Hasan, A. Kordijazi, P. K. Rohatgi, M. Nosonovsky, Triboinformatics Approach for Friction and Wear Prediction of Al-Graphite Composites Using Machine Learning Methods. *Journal of Tribology* **144** (2021).

37. M. S. Hasan, A. Kordijazi, P. K. Rohatgi, M. Nosonovsky, Machine learning models of the transition from solid to liquid lubricated friction and wear in aluminum-graphite composites. *Tribology International* **165**, 107326 (2022).

38. M. S. Hasan, A. Kordijazi, P. K. Rohatgi, M. Nosonovsky, "Application of Triboinformatics Approach in Tribological Studies of Aluminum Alloys and Aluminum-Graphite Metal Matrix Composites" in *Metal-Matrix Composites*, T. S. Srivatsan, P. K. Rohatgi, S. Hunyadi Murph, Eds. (Springer International Publishing, Cham, 2022)*The Minerals, Metals & Materials Series*, pp. 41–51.

39. H. Ning, F. Chen, Y. Su, H. Li, H. Fan, J. Song, Y. Zhang, L. Hu, Modeling and prediction of tribological properties of copper/aluminum-graphite self-lubricating composites using machine learning algorithms. *Friction* **12**, 1322–1340 (2024).

40. O. Cherguy, R. Chmielowski, E. Hachem, I. Lahouij, Deep Learning Prediction of Dry Friction in DLC Coatings Using Literature-Derived Data. *Tribol Lett* **73**, 125 (2025).

41. S. Sivaraman, N. Radhika, Predictive analytics of wear performance in high entropy alloy coatings through machine learning. *Phys. Scr.* **99**, 076014 (2024).

42. J. Liang, N. Lv, Z. Xie, X. Yin, S. Chen, C. Liu, Study on the Composition Design, Microstructure, Wear and Corrosion Resistant of Duplex Stainless Steels Based on Machine Learning. *Met. Mater. Int.* **30**, 3402–3417 (2024).

43. U. M. R. Paturi, S. T. Palakurthy, N. S. Reddy, The Role of Machine Learning in Tribology: A Systematic Review. *Arch Computat Methods Eng* **30**, 1345–1397 (2023).

44. X. Wang, B. Xiao, B. Zhang, R. Sun, L. Cui, P. K. Liaw, Deep Learning Applications in the Analysis of Wear Mechanisms in Metallic Materials: A Review. *ACS Materials Lett.* **7**, 2936–2954 (2025).

45. N. Yin, Z. Xing, K. He, Z. Zhang, Tribo-informatics approaches in tribology research: A review. *Friction* **11**, 1–22 (2023).

46. M. Kolev, svm_friction: A Python based software for calculating, data analysis and modeling the coefficient of friction of aluminum metal matrix composites using support vector regression. *Software Impacts* **17**, 100561 (2023).

47. Y. Ji, S. Ma, W. Peng, Z. Cai, Research on alloy composition-process-wear properties of medium manganese steel based on machine learning. *Tribology International* **200**, 110164 (2024).

48. M. Kolev, COF-RF-Tool: A Python software for predicting the coefficient of friction of open-cell AlSi10Mg-SiC composites using Random Forest model. *Software Impacts* **17**, 100520 (2023).

49. M. Kolev, XGB-COF: A machine learning software in Python for predicting the friction coefficient of porous Al-based composites with Extreme Gradient Boosting. *Software Impacts* **17**, 100531 (2023).

50. L. Qiao, J. Zhu, J. Inoue, Machine learning assisted design of Fe-Ni-Cr-Al based multi-principal elements alloys with ultra-high microhardness and unexpected wear resistance. *Journal of Materials Research and Technology* **33**, 8222–8231 (2024).

51. Z. Li, J. Li, B. An, R. Li, The design method for surface texture of sliding friction pairs based on machine learning under mixed lubrication. *Tribology International* **194**, 109563 (2024).

52. Z. Ge, Q. Hu, R. Wang, H. Fei, Y. Zhu, Z. Wang, Z. Ge, Q. Hu, R. Wang, H. Fei, Y. Zhu, Z. Wang, Machine Learning-Driven Optimization of Micro-Textured Surfaces for Enhanced Tribological Performance: A Comparative Analysis of Predictive Models. *Coatings* **14** (2024).

53. G. Wen, W. Liu, X. Wen, P. Wei, H. Cao, P. Bai, Y. Tian, Effective tribological performance-oriented concentration optimization of lubricant additives based on a machine learning approach. *Tribology International* **197**, 109770 (2024).

54. Z. Wan, Q.-D. Wang, D. Liu, J. Liang, Discovery of ester lubricants with low coefficient of friction on material surface via machine learning. *Chemical Physics Letters* **773**, 138589 (2021).

55. A. Zahoor, A. Rahman, M. Mursaleen, Accelerating tribological material discovery: A machine learning framework with predictive insights. *Tribology International* **211**, 110862 (2025).

56. Y. Jiang, B. Drescher, G. Yuan, A GAN-Based Multi-Sensor Data Augmentation Technique for CNC Machine Tool Wear Prediction. *IEEE Access* **11**, 95782–95795 (2023).

57. C. Zeng, C. Wang, X. Xiong, X. Wang, S. Xiao, Circular saw blade wear status prediction based on generative adversarial network and CNN-LSTM model. *PLOS ONE* **20**, e0326044 (2025).

58. N. Yin, P. Yang, S. Liu, S. Pan, Z. Zhang, AI for tribology: Present and future. *Friction* **12**, 1060–1097 (2024).

59. A. T. Sose, S. Y. Joshi, L. Kumar Kunche, F. Wang, S. A. Deshmukh, A review of recent advances and applications of machine learning in tribology. *Physical Chemistry Chemical Physics* **25**, 4408–4443 (2023).

60. M. Marian, S. Tremmel, M. Marian, S. Tremmel, Current Trends and Applications of Machine Learning in Tribology—A Review. *Lubricants* **9** (2021).

61. B. Sattari Baboukani, Z. Ye, K. G. Reyes, P. C. Nalam, Prediction of Nanoscale Friction for Two-Dimensional Materials Using a Machine Learning Approach. *Tribol Lett* **68**, 57 (2020).

62. F. Oviedo, J. L. Ferres, T. Buonassisi, K. T. Butler, Interpretable and Explainable Machine Learning for Materials Science and Chemistry. *Acc. Mater. Res.* **3**, 597–607 (2022).

63. R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, R. Ranjan, Explainable AI (XAI): Core Ideas, Techniques, and Solutions. *ACM Comput. Surv.* **55**, 194:1-194:33 (2023).

64. H. Wang, C. Zhang, X. Yu, Y. Li, Evaluating Wear Volume of Oligoether Esters with an Interpretable Machine Learning Approach. *Tribol Lett* **71**, 43 (2023).

65. J. Wu, P. Zhao, J. Cui, L. Wang, C. Yang, J. Ouyang, Data-Driven Prediction of Coefficient of Friction in Wet Friction Components: A Model Development and Interpretability Analysis. *J. Tribol* **147** (2024).

66. A. Sellami, M. Rekik, C. B. Njima, R. Elleuch, Towards an Explainable Multi-Target Regression, for Wear and Friction Prediction for Brake Pad Materials. *Acta Polytechnica Hungarica* **21**, 155–174 (2024).

67. P. Hulipalled, V. Algur, V. Lokesha, S. Saumya, Satyanarayan, Interpretable ensemble machine learning framework to predict wear rate of modified ZA-27 alloy. *Tribology International* **188**, 108783 (2023).

68. N. Herwig, Z. Peng, P. Borghesani, Bridging the trust gap: Evaluating feature relevance in neural network-based gear wear mechanism analysis with explainable AI. *Tribology International* **187**, 108670 (2023).

69. M. Marian, S. Tremmel, M. Marian, S. Tremmel, Physics-Informed Machine Learning—An Emerging Trend in Tribology. *Lubricants* **11** (2023).

70. Y. Li, B. Liu, Z. Zhang, W. Ouyang, W. Shi, Z. Chang, Z. Hu, K. Zhang, B. Wei, Physics-informed machine learning for explainable prediction of friction and wear in carbon-ceramic composites. *Ceramics International* **51**, 56400–56416 (2025).

71. A. Aminimajd, J. Maia, A. Singh, Scalability of a graph neural network in accurate prediction of frictional contact networks in suspensions. *Soft Matter* **21**, 2826–2835 (2025).

72. F. Yin, H. Luo, S. Nie, H. Ji, Z. Ma, Physics-Informed machine learning for tribological properties prediction of S32750/CFRPEEK tribopair under seawater lubrication via PISSA-CNN-LSTM. *Tribology International* **199**, 109965 (2024).

73. F. Pashmforoush, A. Ebrahimi Araghizad, E. Budak, Physics-informed tool wear prediction in turning process: A thermo-mechanical wear-included force model integrated with machine learning. *Journal of Manufacturing Systems* **77**, 266–283 (2024).