**Big Data Mining and Classification of Intelligent Material**

**Science Data Using Machine Learning**

**Advantage:**

* Created a big data storage for material science data.
* Developed data mining techniques to retrieve the required data from a database to perform data analytics.
* Developed machine learning prediction model to determine the strength of metals.

**Disadvantage:**

* Query document to retrieve tensile properties of metals resulting in JSON file format.
* Convert JSON to CSV file format and prepare the dataset to feed.
* The machine learning model we developed the first ever big data storage for the Mg-alloy data, which is highly essential to develop an intelligent database, automate .
* The machine learning models for determination of strong Mg-alloy materials, and further perform material science analytics.

**Predicting drug properties with parameter-free machine learning:**

**pareto- optimal embedded modeling (POEM)**

**Advantage:**

* The algorithms publicly disclosed in this investigation were developed and commercialized by Cyclic a Inc.
* All data are publicly available, with information about data location reported in the supplementary information.
* The above notwithstanding, there are important trade-offs associated with the POEM approach. Mainly, POEM has high algorithmic complexity associated with the generation of an **M** *×* **M** dominance matrix.

**Disadvantage:**

* The prediction of absorption, distribution, metabolism, excretion, and toxicity (ADMET) of small molecules from their molecular structure is a central problem in medicinal chemistry with great practical importance in drug discovery.
* Here, we describe pareto-optimal embedded modeling (POEM), a similarity-based method for predicting molecular properties.

**Characterization and identification of lysine glutarylation based on intrinsic interdependence between positions in the substrate sites**

**Advantage:**

* We proposed a bioinformatics method for characterization and identification of glutarylation sites using substrate site specificity.
* The investigation using two-sample logo revealed that the most conspicuous feature

of glutarylation sites is an enrichment of positively charged As stated previously.

* The main purpose of this study was to explore the substrate motifs of glutarylation sites based on amino acid sequences.

**Disadvantage:**

* To accrued a variety of physiological and biological processes.
* As the number of experimentally identified glutarylated peptides increases, it becomes imperative
* To investigate substrate motifs to enhance the study of protein glutarylation.

**Competitive Deep-Belief Networks for Underwater**

**Acoustic Target Recognition**

**Advantage:**

* The proposed CDBN method integrates a new competitive learning mechanism into deep-belief networks to learn more robust and discriminative features for underwater acoustic target recognition.
* The proposed CDBN method can make use of unlabeled samples to solve the small-sample-size problem of underwater acoustic target recognition.
* Compared with traditional hand-engineered feature-extraction methods, the proposed method can learn features from datasets automatically, and does not require prior knowledge.

**Disadvantage:**

* RBM was pre trained with 20,000 unlabeled data points in an unsupervised manner.
* The hidden units of RBM were grouped using 2800 labeled training data points.
* Competitive learning was conducted to construct a CRBM.
* A 2048-500-500-50-50 CDBN was constructed by greedy layer-wise training and supervised fine-tuning to obtain CDBN features.

**Ensembles of Regularized Linear Models**

**Advantage:**

* A novel method for forming ensembles of linear regression models.
* Examples using real and synthetic data-sets show that the approach systematically improves the prediction accuracy of the base estimators being ensembled.
* The experimental results demonstrated that the proposed CDBN method is effective for underwater acoustic target recognition.
* It can significantly reduce the random noise and enhance the line-spectrum characteristics of ship noises, and the CDBN features have better classification performance than other hand-engineered features.

**Disadvantage:**

* We prove the consistency of our method in possibly high-dimensional linear models, where the number of predictors can increase with the sample size.
* SVM was used to evaluate the classification performance of CDBN features.
* The classification performance of CDBN features was compared with four widely used traditional hand-engineered feature sets.

**House Price Prediction Using Machine Learning Algorithm**

**Advantage:**

* It is simple to understand, interpret and visualize.
* Little effort required for data preparation.
* It can handle both numerical and categorical data.
* After training and testing of datasets with all models, the random forest classifier and ridge classifier models performs better than the simple linear regression model.

**Disadvantage:**

* We tested a regression models such as Simple Linear Regression, Ridge Regression, Lasso Regression, Support Vector Regression, Random Forest Regression, Decision Tree Algorithm and selected the best fit among the algorithm.
* Most of the trees can provide correct prediction of class for most part of the data.
* The tree are making mistakes at different places.

**Predicting drug side effects by multi-label learning and ensemble learning**

**Advantage:**

* This paper transforms the side effect prediction as a multi-label learning task.
* We propose a novel multilabel learning method for side effect prediction, named ‘feature selection-based multi-label k nearest neighbor
* In order to combine various features effectively, we construct individual feature-based FS-MLKNN models and use them as base predictors.

**Disadvantage:**

* We propose a novel method ‘feature selection-based multi-label k-nearest neighbor method’ (FS-MLKNN), which can simultaneously determine critical feature dimensions and construct high-accuracy multi-label prediction models.
* Since researchers collected drug data and compile them in the public databases, computational methods were developed for the side effect prediction.

**Underwater target recognition methods based on the framework**

**of deep learning: A survey**

**Advantage:**

* With the development of the cluster system, multiple AUVs are used for collaborative work to collect target information from different angles and reduce the limitation of collecting information from a single perspective.
* For the development of diversified shapes of underwater dangerous targets, as well as the shapes of unknown enemy dangerous targets, the accuracy of target recognition cannot be guaranteed only by training dataset

**Disadvantage:**

* This article systematically describes the application of deep learning in underwater image analysis in the past few years and briefly expounds the basic principles of various underwater target recognition methods.
* The technical problems of AUV underwater dangerous target recognition methods are analyzed, and corresponding solutions are given. At the same time, we prospect the future development trend of AUV underwater target recognition.