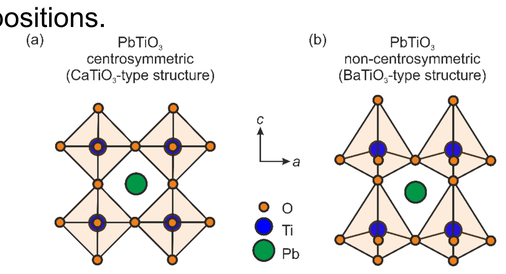
Discovery of Noncentrosymmetric Ternary Compounds from Elemental Composition: A Machine-Learning Approach

Lawrence A. Adutwum, Anton O. Oliynyk, Yuqiao Zhou, and Arthur Mar

BACKGROUND

The absence of an inversion center in crystalline solids is an important prerequisite for many useful electrical and optical properties. Such materials are used in applications such as piezoelectric and nonlinear optical devices. Solids with noncentrosymmetric structures are uncommon, being outnumbered by centrosymmetric structures by a factor of 5. Current strategies aimed at discovering new compounds likely to adopt noncentrosymmetric structures such as the use of asymmetric building blocks, rely on intuition and trial-and-error. Recently, we applied machine-learning (ML) techniques to predict the crystal structures of large classes of solids (AB, ABC, AB_2C) prior to their synthesis. Here we extend these techniques to discover new noncentrosymmetric compounds based solely on their elemental compositions.



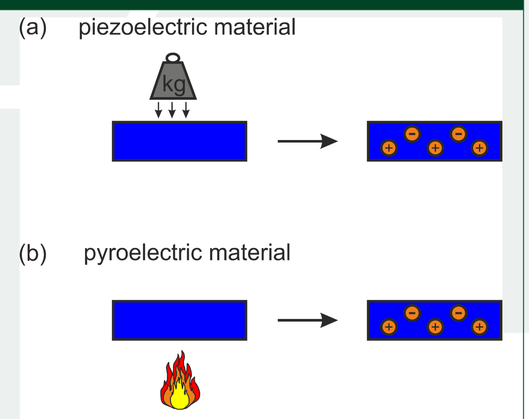
Visualization of the difference between $PbTiO_3$ compound in centrosymmetric (a - $CaTiO_3$-type) and non-centrosymmetric (b - $BaTiO_3$-type) structure.

AIMS AND OBJECTIVES

Short-Term Goal:
To develop a prediction tool for reliable and fast classification of potential structures into centrosymmetric and non-centrosymmetric based on the composition and ratio of constituent elements of a hypothetical compound.

Long-Term Goal:
To develop a qualitative and quantitative prediction model for suggesting potential piezoelectric, pyroelectric, and nonlinear optical materials.

Application:
These materials will be used for green and sustainable energy conversion.

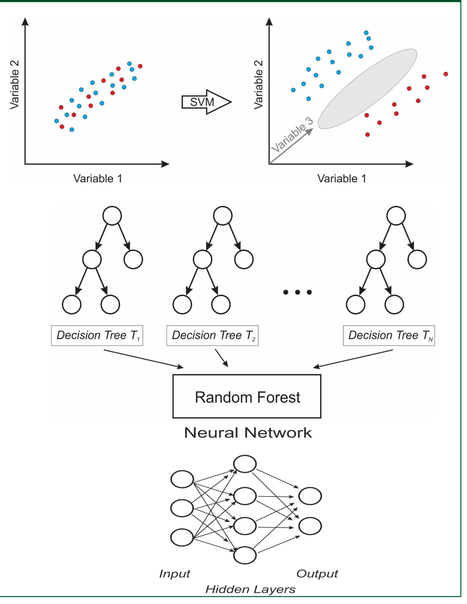


MACHINE LEARNING METHODS AND DATASET

Support Vector Machine (SVM):
SVM algorithm finds a hyperplane in space that separates classes in the training set data. New samples can be projected into the same space as the training set data in order to determine their classes.

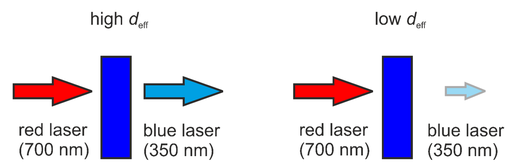
Ensemble Decision Trees (EDT):
This methods generates multiple decision trees using all available features. A selection of trees output the correct class assignments in the training set data are used for future predictions.

Artificial Neural Network(ANN):
ANN relies on a collection of interconnected artificial neurons, which has an input and output. Variations in features going into the network alters the outputs. Known outputs are used to train the ANN.

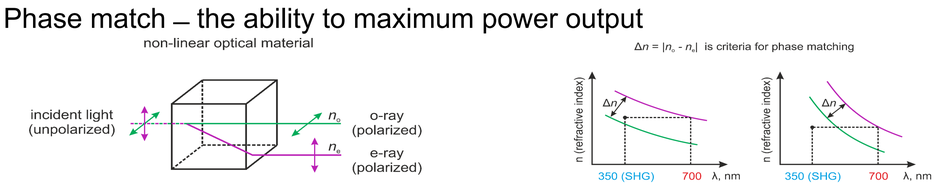


NONLINEAR OPTICS (NLO) APPLICATION FOR NONCENTROSYMMETRIC STRUCTURES

NLO crystals essential characteristics:
(1) NLO coefficient d_{eff} – the efficiency for energy conversion



(2) Birefringence $\Delta n = |n_o - n_e|$ – the capability to maintain phase-matchable
(3) Phase match – the ability to maximum power output



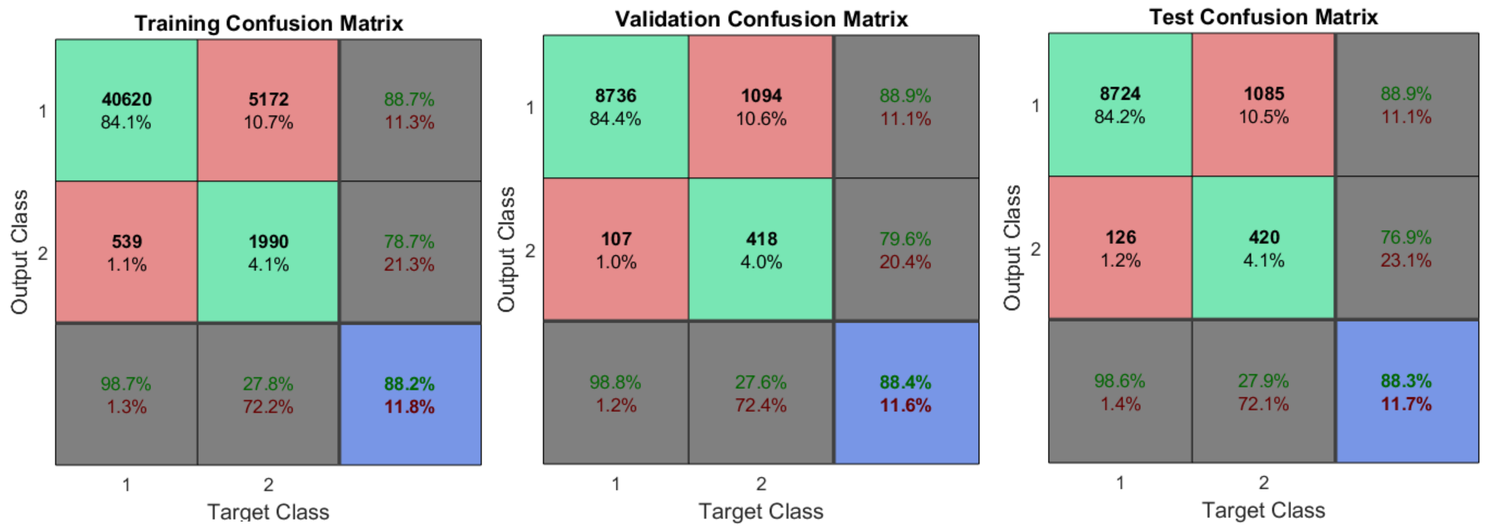
Machine-learning-driven search for new materials with high NLO coefficient d_{eff}, birefringence, and phase match will result in novel NLO materials.

RESULTS

Dataset:
A set of 69031 ternary compounds downloaded from PCD database (2017 Edition). The data was split into 70% for model training and 15% each for model validation and testing. Overall model prediction accuracy for al ML methods tested are show in Table 1.

ML Method	Model Accuracy (%)		
	centrosymmetric	noncentrosymmetric	Mean
SVM	99.2	72.5	85.8
EDT	99.4	65.6	82.5
ANN	98.9	78.1	88.5

Confusion Matrix For Training, Validation And Test Set
(centrosymmetric-Class1 and noncentrosymmetric-Class2)



All three machine learning methods evaluated seem to perform quite well.
Artificial Neural Networks tend to be the optimum Machine-Learning method for this application.

FES PROJECT OVERVIEW

T12-P01 "High-throughput materials discovery through materials genomics"
Discovering better materials is essential for tackling the enormous challenges in developing new renewable energy sources. The experimental variables that must be considered to optimize a given property are too many and their relationships are too complex to allow anything but incremental improvements to be made. So how do we think "outside the box" to find entirely new materials? As part of the larger effort known as the "Materials Genome Initiative," approaches based on data-mining and materials informatics techniques can help screen new compounds with desired properties and features, at greatly accelerated rates, and provide insights into the design principles required to engineer improved materials. These tools will be used, in particular, to find better photovoltaics and catalysts for solar fuels.

Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada

UNIVERSITY OF ALBERTA
FUTURE ENERGY SYSTEMS

CANADA FIRST RESEARCH EXCELLENCE FUND

This research has been undertaken thanks in part to funding from the Canada First Research Excellence Fund.