A Project Report on

Mechanical Property Prediction of High Entropy Alloys using Machine Learning

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology

In

Metallurgical and Materials Engineering

by

Rudraraju Sashank Varma, 185143

Gummapu Nishanth Preetham , 185136

R Sai Swaroop, 185141

Under supervision of:

Dr. V Sreedevi

AssistantProfessor

MME Department, NIT Warangal

Department of Metallurgical and Materials Engineering

National Institute of Technology Warangal

May - 2022

APPROVAL SHEET

This Project Work entitled

"Mechanical Property Prediction of High Entropy Alloys using Machine Learning"

by

Rudraraju Sashank Varma(185143), Nishanth Preetham (185136), R Sai Swaroop(185141) is approved for the degree of Bachelor of Technology in Metallurgical and Materials Engineering

Examiners

Supervisor

Dr. V. Sreedevi

Assistant Professor, MMED, NITW

Chairman

Dr. Mahesh Kumar Talari
HOD, MMED

Date: 10th May 2022

Place: Warangal

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea / data / fact/ source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Rudraraju Sashank Varma, 185143

Gummapu Nishanth Preetham , 185136

R Sai Swaroop, 185141

Date : 09-05-2022

Place : Warangal

Department of Metallurgical and Materials Engineering National Institute of Technology, Warangal - 506004

CERTIFICATE

This is to certify that the project work titled **"Mechanical Property Prediction of High Entropy Alloys using Machine Learning"** is a bonafide record of work carried out by **Rudraraju Sashank Varma(185143), Nishanth Preetham (185136), R Sai Swaroop(185141)** submitted to the faculty of Department of Metallurgical and Materials Engineering in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in Metallurgical and Materials Engineering at National Institute of Technology, Warangal during the academic year 2021-2022.

Project Guide:

Dr. V. Sreedevi

Assistant Professor, MMED, NITW

ACKNOWLEDGEMENT

We would like to extend our gratitude to our project guide **Dr. V Sreedevi**, Assistant Professor, Department of Metallurgical and Materials Engineering, for her constant assistance, support and encouragement throughout the project. We would like to extend our heartfelt thankfulness to the faculty of the department for imparting knowledge that was directly or indirectly involved in the project and for inspiring us all these years. We are grateful to our parents who never failed to motivate us with utmost determination and have inspired us to be what we are today. We are obliged to the institute for shaping our personality and future, for nurturing us in this span of four years and becoming an indispensable part of our lives.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

.

High-entropy alloys (**HEAs**) are alloys that are formed by mixing equal or relatively large proportions of five or more elements. These typical metal alloys consisted of one or two principal components with minor proportions of other elements before these substances are synthesized. HEAs have been receiving immense attention owing to their remarkable mechanical and physical properties. Furthermore, research indicates that some HEAs have relatively better strength-to-weight ratios, with a higher degree of fracture resistance, tensile strength, as well as corrosion and oxidation resistance than conventional alloys. We have a huge amount of data piled up and this data can be analyzed and required patterns can be detected and also new materials can be discovered with desired mechanical properties. Here, we gather all the required data to predict the mechanical properties of high entropy using regression models. These results can be used to analyze the mechanical properties of HEAs that are not yet made through experimentation which is expensive. Hardness and Young's Modulus are predicted using ANN and Random Forest algorithm and accuracies are calculated.

1. INTRODUCTION

Human requirements are growing in our fast-paced environment. In the same way, the world is reshaping itself to make things more comfortable for us. We (humans) desire to have everything at our fingertips. We want tiny things instead of big machinery and equipment, and when it comes to information technology, we need to define specific goals and accomplish them in a smart way rather than doing physical work. Programming languages are now employed in almost every part of our lives, not just in software. Machine learning has been utilized in material science for focused study instead of long processing processes employing machines, among other things[1]. This clever method of constructing a foundation for results-oriented projects will be given more importance.

Machine learning shows exceptional qualities when used in materials science, both in prediction accuracy and time efficiency. As shown in Fig.1, machine learning applications in the field of materials science and engineering can be divided into three main groups: materials discovery, prediction of properties and other purposes. While predicting properties of a material, regression analysis is used. It helps in the prediction of both macroscopic and microscopic properties[1]. Machine learning can also be used to discover new materi. Here a probabilistic model is used to train on combinations of structure and components. A material having desirable properties is then chosen from the dataset using calculations based on the density functional theory (DFT).

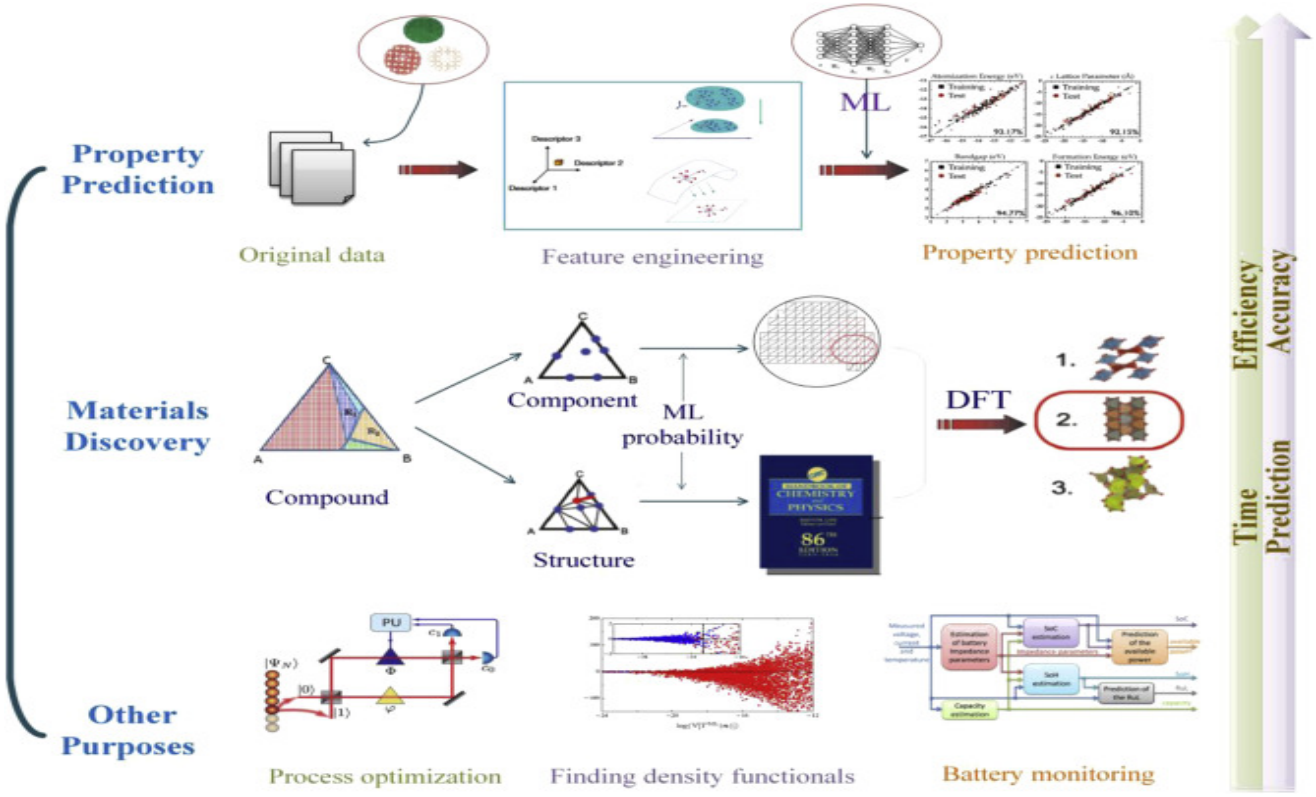

Fig.1.1 Framework for application of machine learning in material Science[1]

Melting point, ionic conductivity, hardness, glass transition temperature, molecular atomization energy, and lattice constant are examples of material qualities that may be represented at a microscopic level. Experimental measurement and computational testing are the two most common methods for determining material qualities. These approaches can't explore intricate links between structure and property since they require a large experimental setup, and they can't simply uncover the components that effect[2]. Furthermore, the findings of the experimental procedures will be gained in the final stages, so if the results obtained are incorrect, the entire experimental budget, cost, and time will be squandered. Furthermore, some material characteristics are difficult to anticipate via experimental measurement and computer testing.

2. Problem Identification

The experiments take huge time and thus we can use machine learning models in property prediction. Application of techniques like AI, Big Data and Machine learning helps in the discovery of new materials and fasten the research. The materials data and ML will give the motivation for this data driven materials invention. Using ML models, it is easy to halve the time and value from the invention to the improvement to employment of recent materials. understanding phases is also important which affects the mechanical properties.

3. OBJECTIVES

It is imperative that we understand the objectives of this project. They are:

- Identify the design parameters in high entropy alloys (HEAs) and apply them as features to build machine learning (ML) models for property prediction in HEAs
- Examine the relationship between the features in the form of a correlation matrix
- Perform data-preprocessing which involves the formation of train, validation and test datasets from the main dataset
- Implement machine learning models to measure property prediction accuracy in HEAs
- Implement feature reduction based on the feature importance as a measure and compare phase prediction accuracies of the above mentioned ML models before and after feature reduction
- Consider other metrics like precision, recall and F-1 score for finalizing the ML model with the highest predictive power

4. LITERATURE REVIEW

4.1 High Entropy Alloys – their phases and properties

Human investigation of metallic materials began with pure metals (e.g., Cu), then progressed to binary (e.g., Fe-C), ternary (e.g., Ni-Co-Al), quaternary (e.g., Ni-Co-Al-Cr), quinary (e.g., Ni-Fe-Cr-Ti-Al), and higher-order alloys to fulfill ever-increasing demands. The dominant technique in alloy creation is, to a significant degree, restricted to a niche, i.e., choose one element as the solvent and add modest amounts of alloying elements as solutes to tune qualities into desired directions. This method has led to the discovery of a large number of alloys over the millennia. Nonetheless, in comparison to the overall number of alloys accessible, the number of alloys currently discovered is still rather tiny.

Following the publishing of five journal publications by Yeh and co-workers and a contemporaneous one by Cantor , the emergence of high-entropy alloys (HEAs) in 2004 sparked new optimism. This alloy family employs a novel method, namely, the utilization of five or more primary elements in equal or nearly equal proportions as alloying elements, with no distinction between solvent and solute atoms. On the one hand, the novel alloy design technique widens alloy composition space to the hitherto unexplored center area of phase diagrams. On the other hand, it creates alloys with substantially simpler structures than multi-component alloys (e.g., disordered face-centered-cubic solid solutions).

4.2 Importance of High Entropy Alloys

Compositional space of alloys vastly expanded: As previously indicated, the number of alloys that can be located using the standard alloy design technique based on one primary element is quite restricted. Only about 30 alloy systems have been identified for common applications during thousands of years of investigation, according to the ASM Handbook. By advancing from the limited region of corners and edges to the huge center area in phase diagrams using the multi-principal-element based design method, the compositional space of alloys may be significantly increased. The total number of possible alloys that can be mapped out of 60 feasible elements, according to Cantor's cautious estimate, is in the order of 10100. HEAs and their variants are unquestionably providing a path to cover a large number of unexploited alloy

systems. In addition it is difficult to predict some material properties by experiments and computational testing.

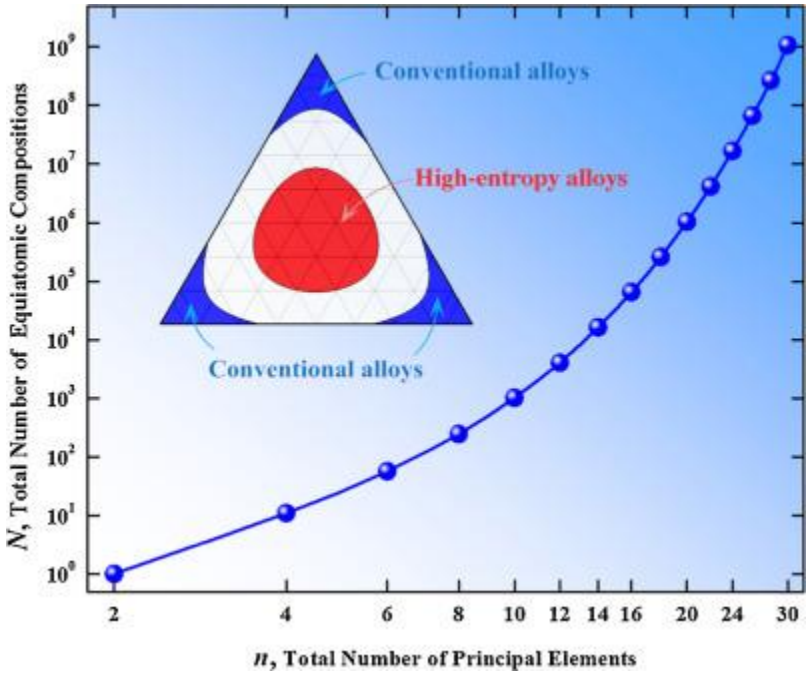

Fig 4.1: Conventional and HEAs[1]

Enormous possibilities in microstructures properties: Expectedly, the expansion of the compositional space alloys will surely bring up a plethora of new prospects in terms of producing affluent, diversified microstructures as well as structural and functional features never seen before. Indeed, numerous intriguing, hitherto unknown features have already been described. The CoCrFeMnNi HEA with a single face-centered-cubic (FCC) phase, for example, has been found to have extraordinarily high fracture toughness, even at cryogenic temperatures.

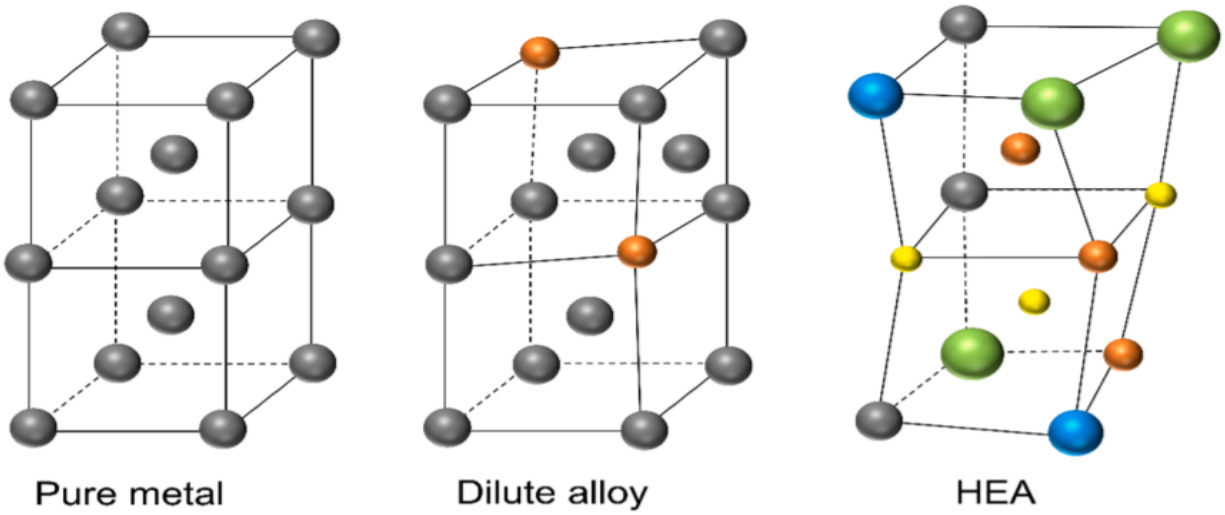

Fig 4.2. Schematics of lattice distortion in body-centered cubic pure metals, conventional dilute alloys, and high-entropy alloys[1]

4.3 Core effects of High Entropy Alloys*:*

High entropy*:* From a thermodynamics approach, the effect of high entropy in favoring the development of simple solid solutions in HEAs may be understood.
Specifically, the Gibbs free energy of mixing that a reaction entails determines whether it can win in a competition with others.

$$\Delta G_{ix} = \Delta H_{ix} - T\Delta S_{ix},$$

where T is the temperature. ΔHmix is the enthalpy of mixing, which is a measure of energy change under constant pressure and temperature. ΔSmix is the entropy of mixing, which is a measure of randomness of a system at the atomic level. In the case of HEAs, consider the competition between a solid solution and intermetallic compounds. In ideal solutions assumed for HEAs, ΔHSS mix,ideal = 0. On the other hand, if assuming that intermetallic compounds are perfectly ordered, ΔSIM mix,ideal = 0. As a result ΔGSS mix,ideal = − TΔSSS mix,ideal for the solid solution, while ΔGIM mix,ideal = ΔHIM mix,ideal for the intermetallic compounds. High entropy makes − TΔSSS mix,ideal < ΔHIM mix,ideal and ΔGSS mix,ideal < ΔGIM mix,ideal, thereby promoting the formation of solid solutions rather than complex intermetallic compounds in HEAs.

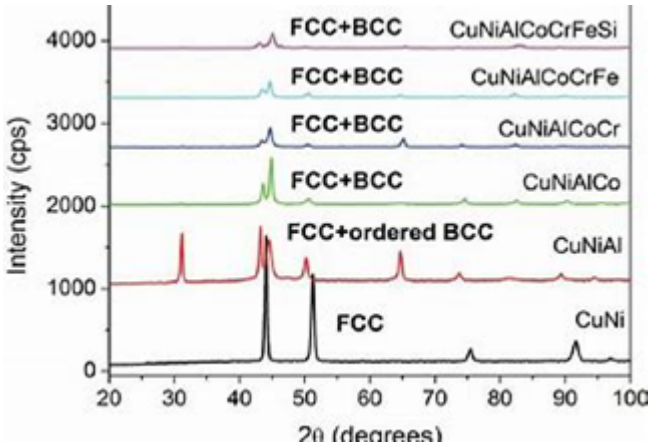

Fig. 4.3: XRD patterns of a series alloy which is designed by the sequential addition of an element to the previous one[3]

Severe lattice distortion: Viewing their relative hard-sphere models of the lattice can reveal the degree of lattice distortion in pure metals, ordinary dilute alloys, and HEAs. A pure metal, as shown schematically in Fig. 2, has the same kind of atoms filling the lattice positions and so creates no lattice distortion. The infiltration of a modest quantity of second atomic species generates mild lattice deformation in ordinary dilute alloys. The enormous concentrations of several different-sized atoms in HEAs cause them to coexist in the lattice at random, resulting in severe deformation.

Sluggish Diffusion: Sluggish diffusion was initially proposed based on the observed tendency of HEAs to produce nano-sized precipitates, as well as nanocrystalline and amorphous phases . By comparing the observed diffusion rates of Ni in HEAs and medium-entropy alloys, later research confirms this fact. Sluggish diffusion in HEAs is thought to be caused by the necessity for cooperative diffusion of distinct types of atoms, and it affects all diffusion-controlled processes, including solidification, grain development, recovery and recrystallization, phase change, and creep.

Cocktail effect: Ranganathan coined the phrase "cocktail effect" to describe amorphous alloys, gum metals, and HEAs. It refers to the fact that combining several distinct element species can result in surprisingly exceptional characteristics. The cocktail effect is rooted in the composite effect — the interactions among all contributing elements – if a HEA is viewed as a composite at

the atomic level. For example, the Vickers hardness of MoNbTaVW is 5250 MPa, significantly more than the 1596 MPa computed using the mixing technique.

Aspects	Conventional dilute alloys	HEAs
Compositional space	Low-dimensional	High- and hyper-dimensional
Composition	Limited	Limitless
Alloy design strategy	Based on one principal element	Based on multiple principal elements
Concentration of secondary elements	Dilute	Concentrated
Solid solution	Distinguishable solvent and solute atoms	Indistinguishable solvent and solute atoms
Phase diagram	Corners and edges	Central region
Configurational entropy of mixing	Low	High
Modulus mismatch	Mild	Severe
Atomic size mismatch	Mild	Severe
Lattice distortion	Mild	Severe
Microstructure	Less diversified	More diversified
Properties	Limited possibilities	Numerous possibilities
Diffusion	Relatively quick	Sluggish
Solid solution strengthening	Weak	Strong

Table 1. Differences between conventional alloys and HEAs[1]

4.4 Prediction of composition and hardness of High Entropy Alloys

Artificial Neural Networks (ANN)-based models are a powerful tool for anticipating and exploiting complex interactions between material attributes and prediction. In this research, ANN is used to predict the composition of a high entropy alloy (HEA) based on non-equimolar AlCoCrFeMnNi in order to achieve hardness values that are higher than the best theoretical value for this alloy system.

Because there is a significant correlation between the input layers, neural networks may determine features. Those that have an impact on the output are combined or reduced to a high-level feature in the network's deeper layers, preventing overfitting. The composition is converted into a list of quantitative characteristics such as solid density, hardness, melting points, cohesive energy, and so on. Because the dataset only contains 91 alloys, capturing the overall trend is more critical than minimizing the discrepancy between experimental and predicted values. As a result, the number of features evaluated for the model's construction was reduced.

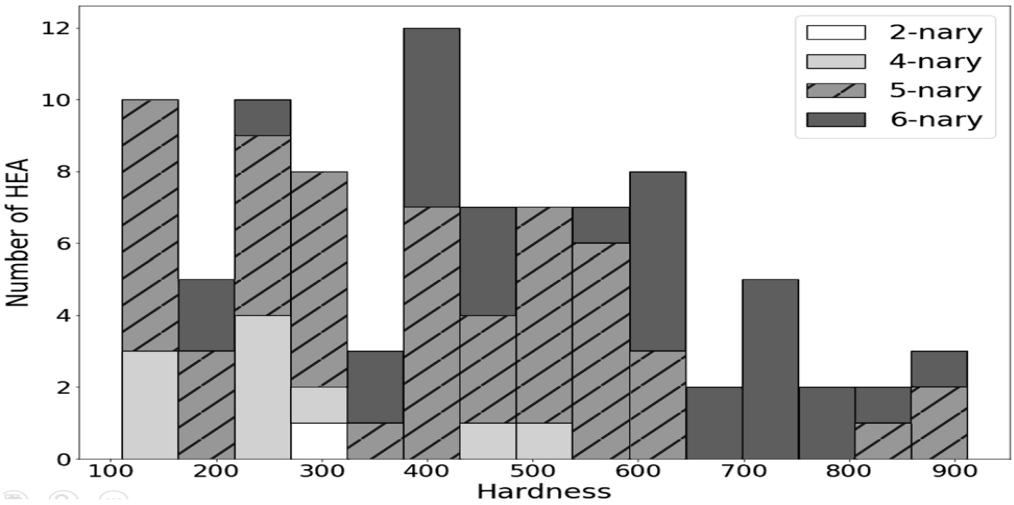

Fig. 4.4 : Data histogram of training set which was used for modeling[2]

4.5 Machine learning and new materials discovery

The goal of materials science is to discover novel materials with unique features. For the time being, some computational and experimental approaches are being used to find novel materials. The compositional and structural search space, on the other hand, appears to be severely confined for both. As a result, both time and resources are used. As a result, machine learning may be utilized in conjunction with computer simulations to develop a method for developing improved materials.

As illustrated in Fig. 3, the complete machine learning process in the discovery of novel materials consists of two primary sections, the first of which is a learning system and the second of which is a prediction system. A learning system is one in which the model cleans the data and selects the features before training and testing. The model is then handed on to the prediction system, which uses a suggestion process to forecast new materials. The suggestion-and-test technique is what it's called.

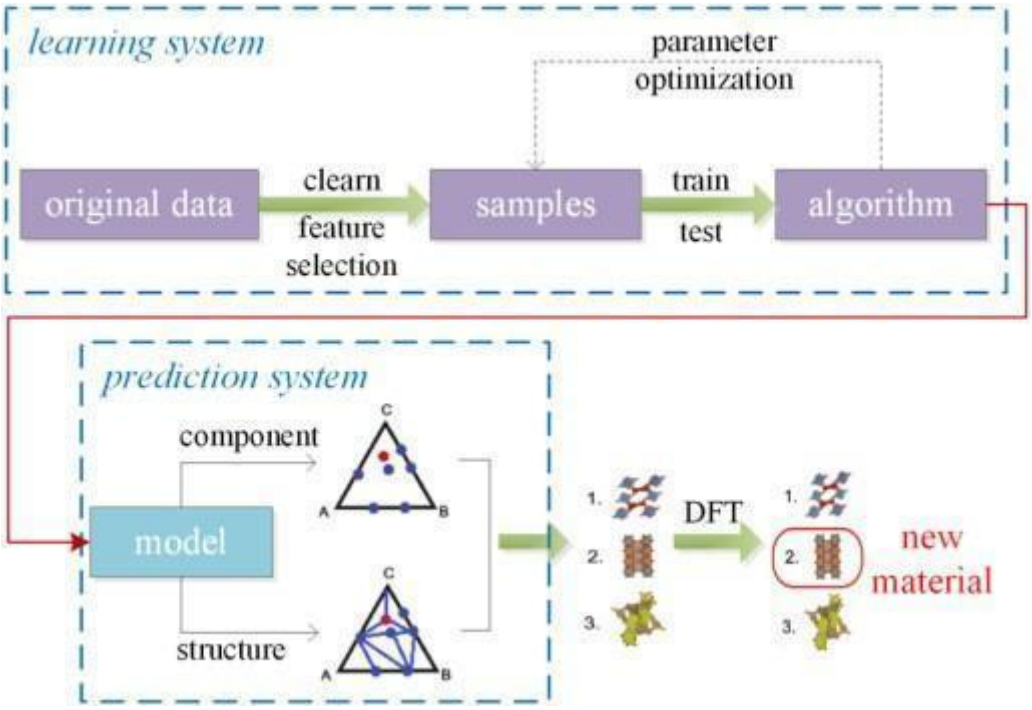

Fig. 4.5: The process of machine learning which results in the new discovery of materials [2]

4.6 Machine Learning Algorithms

There are three major types with the help of which machine learning models learn. They are unsupervised learning, supervised learning and reinforcement learning.

Unsupervised learning is a process in which a machine learning system deduces patterns from a dataset that has no known or labeled outcomes. As a result, this form of learning cannot be utilized for regression or classification algorithms since you have no idea what the values for the output data will be, making training the algorithm very difficult. Unsupervised learning, on the other hand, is still utilized to uncover patterns in the dataset that may be used to further refine feature selection.

Supervised learning is a sort of learning in which the model is given input and output data to train and improve its accuracy for data processing. The purpose of this procedure is to create a mapping function that is accurate enough to be utilized for data prediction in the future.

Reinforcement learning is basically used to train models to be able to make decisions in a sequential manner. The idea is to accomplish the goal in a complex environment. Trial and error are employed by the model to come up with a solution. In order to achieve what the programmer wants the artificial intelligence is given some penalties or rewards for the action it performs. The maximization of rewards is the target to be achieved.

4.7 Machine Learning for High Entropy Alloys

There are two popular machine learning methods that are widely adopted - supervised learning and unsupervised learning. None-the-less, machine learning is now being widely used in the field of Materials Science and Engineering and its usage has brought about a revolution in the way data was visualized in the past.

High entropy alloys have gained attention from scientists due to their remarkable properties. Machine learning models have been built to improve the representativeness of data by taking into consideration the short-range order (SRO) parameter to assess the configuration space. Also, considerable research has been done in order to predict the composition and hardness of HEAs. Not only that, machine learning has also been used to predict the configurational energy and phases that HEAs comprise of. This might lead to the creation of an alternative route to accomplish the design of HEAs. This is also suitable to accelerate the discovery of other metal alloys that could be used for modern engineering applications**.**

Some of the Machine Learning models that are used in our work to predict the mechanical properties:

4.7.1 Random Forest

Random Forest (RF) is a sophisticated learning method. It is commonly used in machine learning for classification and regression approaches. It is based on the ensemble learning approach, which is a method of combining several classifiers to tackle a challenging problem as well as to improve the model.As the name implies, this is a classifier with many calls on diverse variables.

Subsets of the given dataset and uses the best of them to improve the accuracy of prediction of specific data sets rather than relying on a single call, this method uses the prediction.The final output is calculated for each and every set based on the majority.The more pairs in the puzzle, the better the accuracy and the fewer the errors.

Bagging and boosting are the two most common methods of outfit modeling.

Bagging is similar to a group of separate models being trained using a set of rules. A random subset is used to train each of these models. Individually training the models in a sequential manner is beneficial. Both models correct themselves based on the previous model's experience. Random forest can be used for classification as well as regression. This approach is employed as both an outfit mechanism and a decision tree in this situation.

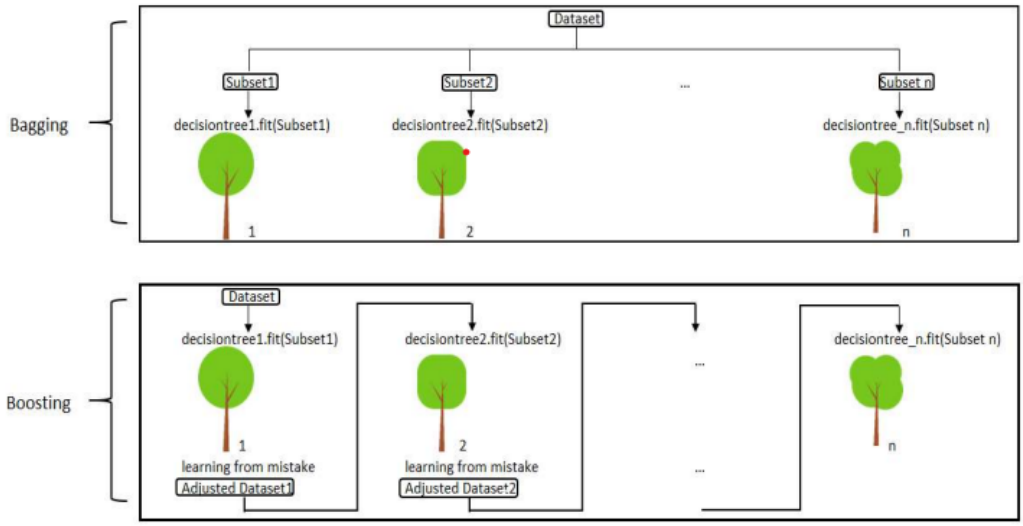

Fig 4.6: Random forest algorithm[9]

4.7.2 Artificial Neural Networks

An input layer, one or more hidden layers, and an output layer make up Artificial Neural Networks, which are brain-inspired models. The artificial networks use numerous neurons as their calculating units in each layer. These artificial neurons are connected by 'synapses,' which are simply weighted values, from one layer to the next.

When creating a neural network model, it is critical to understand the significance of epoch and batch size. When the complete dataset is only processed through the model once, it is called an epoch. It is hard to feed one epoch to the computer at a time since we employ neural networks to cope with a large dataset. As a result, the dataset is usually separated into numerous smaller batches.

The curve swings from underfitting to ideal to overfitting as the number of epochs grows. However, there is no specific answer to the number of epochs that must be used to train the model. After running the model multiple times with different numbers of epochs, you should be able to fix it. Batch size, on the other hand, refers to the amount of data or training examples delivered to each epoch. 32, 64, and 128 samples are common batch sizes.

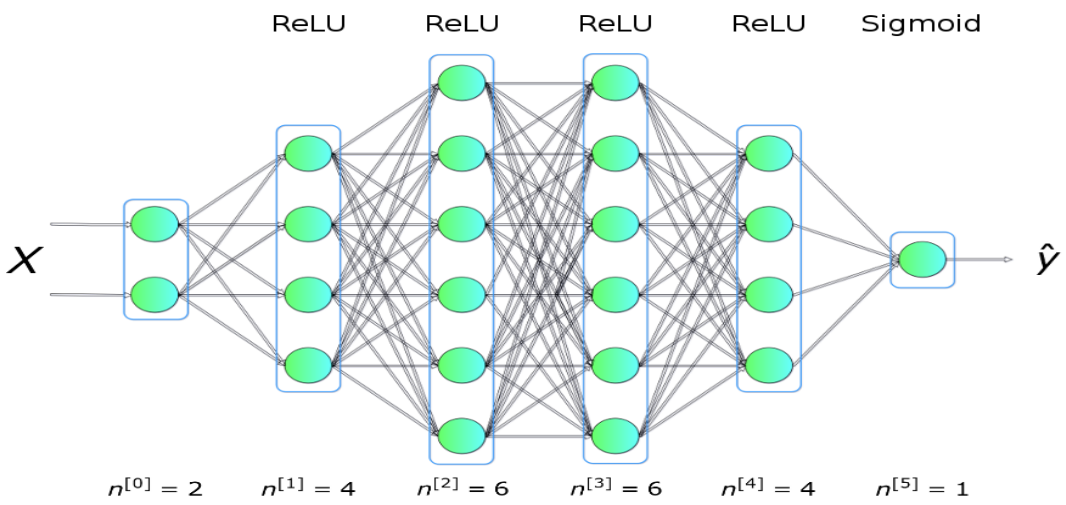

Fig 4.7. A neural network architecture[9]

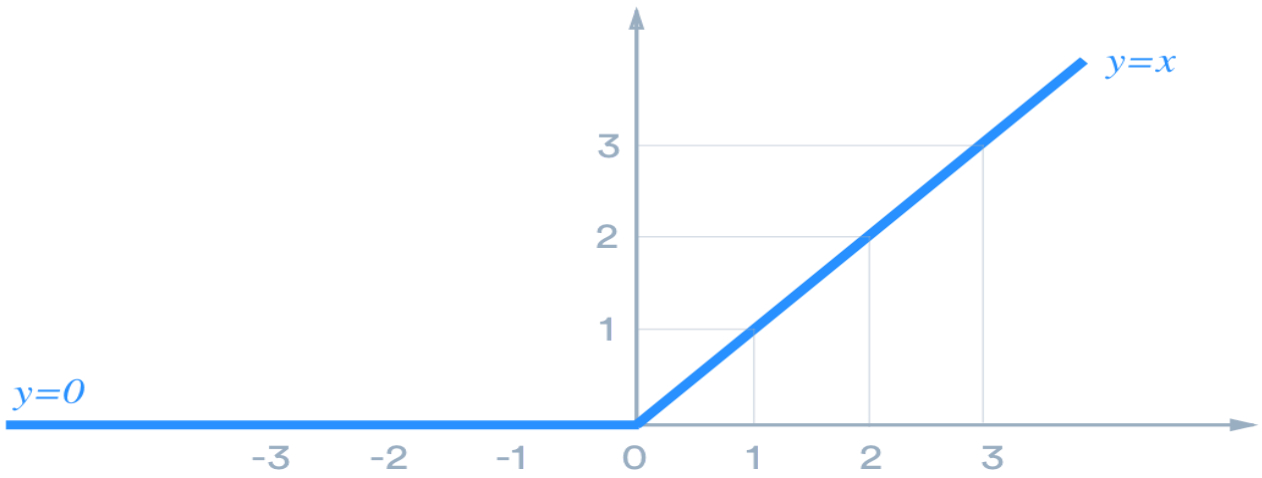

Fig. 4.8. A graph showing the ReLU activation function[9]

We must specify the activation function for each layer. The ReLU function is typically employed as the activation function. The term ReLU refers to a rectified linear unit. This function is expressed as y = max in math (0, x). It looks similar to the graph displayed in the illustration. This function is linear for all positive values and equals zero for all negative values, as can be seen from the graph. Because there is no hard arithmetic involved, this function is simple to compute. The model takes less time to train and generate results. The curve is linear, and the curve's slope does not plateau. The function is only used infrequently.

Sparsity denotes the absence of data and is generally not regarded as a positive indicator for the model. In practice, however, it produces concise models. Models using sparse data are thought to have superior prediction potential and less overfit. In a sparse network, neurons capture crucial elements of the situation. Leaky ReLU and Parametric ReLU are two versions of the ReLU activation function.

All negative values in Leaky ReLU have a preset slope. The slope of a Parametric ReLU is not predetermined, and it is up to the neural network to figure out which parameter is involved.

The loss function, also known as the error function, is used to estimate the model's loss so that it can be lowered in subsequent valuations owing to weight updates. When doing binary classification, the binary cross entropy function is used. Categorical cross entropy is utilised in multi-class categorization. The average difference between the actual and anticipated probability distributions for predicting class 1 (the set of values in a binary classification is 0,1) will be compiled by cross entropy.

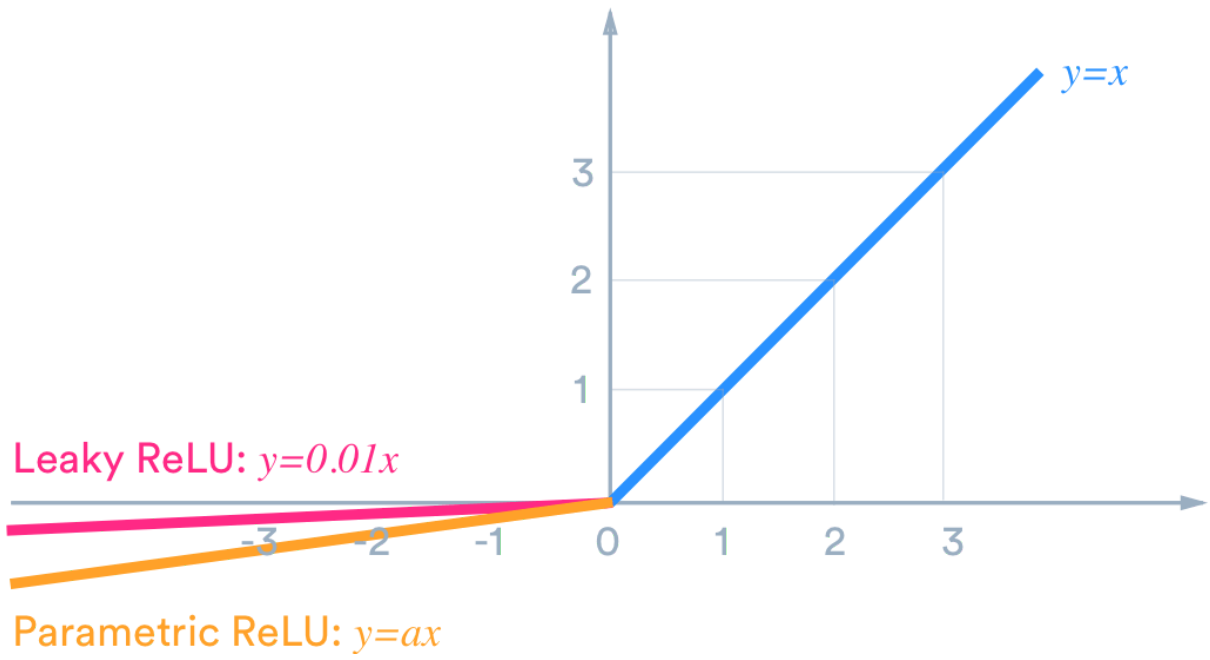

Fig 4.9. A graph showing the Parametric ReLU function and Leaky ReLU function[9]

Artificial neural networks are divided into two categories. The first is feed-forward ANN, and the second is feedback ANN. In the feed-forward mode, data flows from the input nodes to the output nodes via hidden nodes. The neural network contains no loops. The neural network is dynamic in feedback mode, meaning that the output from the output layer is fed back to the input layer to get

the greatest possible model accuracy. Its primary function is to tackle optimization problems.

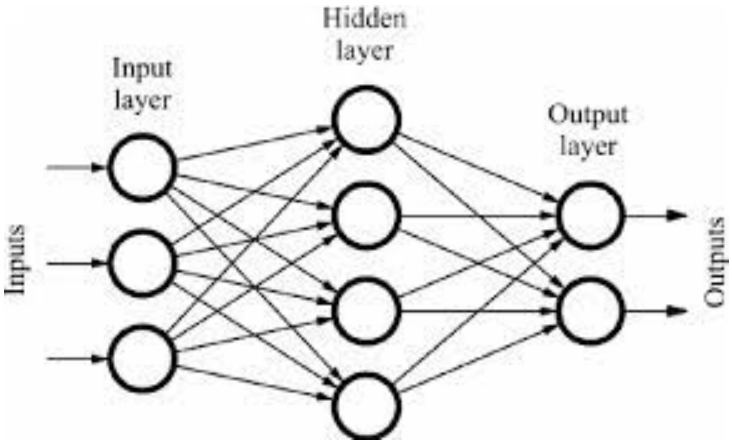

Fig 4.10. : A feed-forward Artificial Neural Network[9]

The backpropagation learning technique is used by neural networks with several layers. This implies that the output values are compared to the real value in order to determine the value of an error-function that has already been created. Various strategies are used to make the fault travel back across the network. After that, the algorithm adjusts the weight of each neuron to lower the error function's value. The neural network converges to a point where the error function's value is minimized to a substantial extent after repeating this process for a large number of training cycles.

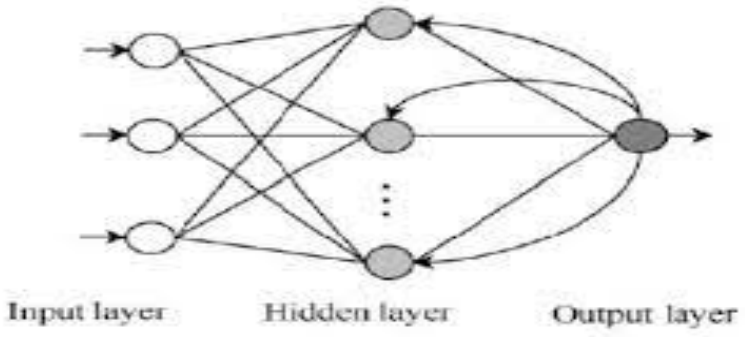

Fig 4.11: A feedback Artificial Neural Network

Often, phase selection in HEAs depends on various parameters. Table **2** gives a list of phase prediction parameters that we have used in our work.

Table 1. The 13 design parameters and the corresponding formula.

Parameters	Formula
Mean atom radius	$a = \sum_{i=1}^{n} c_i r_i$
Atomic size difference	$\delta = \sqrt{\sum_{i=1}^{n} c_i \left(1 - \frac{r_i}{a}\right)^2}$
Average of the melting points of constituent elements	$T_m = \sum_{i=1}^{n} c_i T_{mi}$
Standard deviation of melting temperature	$\sigma_T = \sqrt{\sum_{i=1}^{n} c_i \left(1 - \frac{T_i}{T_m}\right)^2}$
Average mixing enthalpy	$\Delta H_{mix} = 4 \sum_{i \neq j} c_i c_j H_{ij}$
Standard deviation of mixing enthalpy	$\sigma_{\Delta H} = \sqrt{\sum_{i \neq j} c_i c_j \left(H_{ij} - \Delta H_{mix}\right)^2}$
Ideal mixing entropy	$S_{id} = -k_B \sum_{i=1}^{n} c_i \ln c_i$
Electronegativity	$\chi = \sum_{i=1}^{n} c_i \chi_i$
Standard deviation of electronegativity	$\Delta\chi = \sqrt{\sum_{i=1}^{n} c_i (\chi_i - \chi)^2}$
Average VEC	$VEC = \sum_{i=1}^{n} c_i VEC_i$
Standard deviation of VEC	$\sigma_{VEC} = \sqrt{\sum_{i=1}^{n} c_i (VEC_i - VEC)^2}$
Mean bulk modulus	$K = \sum_{i=1}^{n} c_i K_i$
Standard deviation of bulk modulus	$\sigma_K = \sqrt{\sum_{i=1}^{n} c_i (K_i - K)^2}$

Table 2: The parameters used for phase prediction[3]

5. EXPERIMENTAL COMPUTATION

5.1 Data Resource

Database consists of 370 alloys and their mechanical properties. In this study, they combined data from different sources and tabulated the properties of alloys from 2004 to 2016. The database can enable data mining to extract insights and uncover patterns to guide and accelerate the development of HEAs. The database has a tree-like classification which includes four different families: 3d transition metal (3d TM), refractory metal (RHEAs and RCCAs), light metal family, and bronzes and brasses HEAs. Each family is expanded in classes (a class is a unique combination of principal elements), and each class contains members having variations in principal element concentrations. Each member is characterized by a set of attributes which includes: alloy composition, phase content, density, hardness (Vickers), type of mechanical test (tension or compression), yield strength, ultimate strength, elongation, and Young's modulus.

The data set is not directly available, it is made through collection from various research papers and tallying the values of 13 parameters to the already available data of physical properties like hardness, young's modulus.

Composition (atomic)	Ref.	Type of phases	ρ (g/cm^3)	HV	Type of tests	σ^y (MPa)	σ^{max} (MPa)	ε (%)	E (GPa)
3d TM HEAs and CCAs in the Al-Co-Cr-Fe-Mn-Ni system and derivates									
CoFeNi	[4]	FCC	(8.5)	125	C	204			(207)
CoFeNi	[4]	FCC	(8.5)	125	C	209			(207)
CoFeNi	[5]	FCC	(8.5)		T	211	513	31	(207)
CoFeNiSi0.25	[4]	FCC	(7.7)	149	C	196			(194)
CoFeNiSi0.5	[4]	FCC + Im	(7.1)	287	C	476			
CoFeNiSi0.75	[4]	FCC + Im	(6.6)	570	C	1301			
Al0.25CoFeNi	[4]	FCC	(7.9)	138	C	158			(196)
Al0.5CoFeNi	[4]	FCC + BCC	(7.4)	212	C	346			(187)
Al0.75CoFeNi	[4]	FCC + BCC	(7.0)	385	C	794			(179)
CoCrFeNi	[6]	FCC	(8.2)		T	148	413	48	(225)
CoCrFeNi	[7]	FCC	(8.2)	116					(225)
CoCrFeNi	[7]	FCC	(8.2)	113					(225)
CoCrFeMo0.5Ni	[8]	FCC + Im	(8.5)	210					
CoCrFeNb0.103Ni	[6]	FCC + Im	(8.2)		T	318	622	19	
CoCrFeNb0.155Ni	[6]	FCC + Im	(8.2)		T	322	744	23	
CoCrFeNb0.206Ni	[6]	FCC + Im	(8.2)		T	403	807	9	

Table 3 . Database of HEAs[5]

Table 4 . Dataset used for prediction of Young's Modulus

	Type of phases	r# (g/cm3)	a	delta	Tm	D_Tm	Hmix	σHmix	Sid	Elec_ne	D_elec_	VEC	d_VEC	BulkMo	D_Bulk	Phase	Phase_i	E# (G
Al0.25CoFeNi	FCC	7.9	1.2643	0.0613	1,637.89	363.71	-18.8571	11.6220	1.9459	1.8129	0.1158	7.2857	2.8140	143,714,2	37.9839	BCC+FCC	SS	19
Al0.5CoFeNi	FCC + BCC	7.4	1.2643	0.0613	1,637.89	363.71	-18.8571	11.6220	1.9459	1.8129	0.1158	7.2857	2.8140	143,714,2	37.9839	BCC+FCC	SS	18
Al0.75CoFeNi	FCC + BCC	7.0	1.2643	0.0613	1,637.89	363.71	-18.8571	11.6220	1.9459	1.8129	0.1158	7.2857	2.8140	143,714,2	37.9839	BCC+FCC	SS	17
CoCrFeNi	FCC	8.2	1.2468	0.0030	1,871.75	180.37	-3.7500	1.6370	1.3863	1.8200	0.0967	8.2500	1.4790	172,500,0	8.2916	FCC	SS	22
CoCrFeNi	FCC	8.2	1.2468	0.0030	1,871.75	180.37	-3.7500	1.6370	1.3863	1.8200	0.0967	8.2500	1.4790	172,500,0	8.2916	FCC	SS	22
CoCrFeNi	FCC	8.2	1.2468	0.0030	1,871.75	180.37	-3.7500	1.6370	1.3863	1.8200	0.0967	8.2500	1.4790	172,500,0	8.2916	FCC	SS	22
CoCrFeNb0.103Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,0	38.5252	BCC+Lave	IM+SS	
CoCrFeNb0.155Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,0	38.5252	BCC+Lave	IM+SS	
CoCrFeNb0.206Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,0	38.5252	BCC+Lave	IM+SS	
CoCrFeNb0.309Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,0	38.5252	BCC+Lave	IM+SS	
CoCrFeNb0.412Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,0	38.5252	BCC+Lave	IM+SS	
CoCrFeNiTi0.5	FCC	7.2	1.2707	0.0533	1,879.44	171.44	-11.5556	6.8880	1.5811	1.7889	0.1267	7.7778	1.9309	165,555,5	21.1403	FCC+IM	IM+SS	13
Co1.5CrFeNi1.5Ti	FCC + Im	7.4	1.2707	0.0533	1,879.44	171.44	-11.5556	6.8880	1.5811	1.7889	0.1267	7.7778	1.9309	165,555,5	21.1403	FCC+IM	IM+SS	21
Al0.25CoCrFeNi	FCC	7.7	1.2597	0.0376	1,806.29	295.64	-7.2688	3.9105	1.5426	1.8053	0.1075	7.8837	1.9554	165,767,4	25.8518	FCC+L12	IM+SS	21
Al0.25CoCrFeNi	FCC	7.7	1.2597	0.0376	1,806.29	295.64	-7.2688	3.9105	1.5426	1.8053	0.1075	7.8837	1.9554	165,767,4	25.8518	FCC+L12	IM+SS	21

Table 5 . Dataset used for prediction of Hardness values

When dealing with machine learning models, it's critical to scale the features so that the mean is 0 and each feature's variance is centered around zero. If the features aren't scaled, a feature with a high variance may outperform the others in the dataset. This could introduce bias into the model, which is something we don't want. As a result, the dataset was scaled to prevent particular traits from dominating correlated. Here, high correlation means values larger than 0.9 or less than -0.9. We chose all 13 characteristics while training the dataset on models because we didn't have any such values.

It's critical to shuffle the data to avoid bias in the train, validation, and test datasets. This must be completed prior to tweaking the parameters. These are the steps through which data is preprocessed. These methods are used to ensure that the dataset does not introduce any bias into the model during training, which could result in erroneous findings when the model is tested on a completely different dataset. We tweak the models correctly after data preprocessing to produce accurate predictions on the test dataset.

One of the most crucial aspects in any model's construction is feature selection. To locate only the important features, we apply the filtering procedure. The correlation matrix is used to do this. The

Pearson correlation heatmap in Fig. 4.2 shows the correlation between independent variables. Correlation values are used to determine the strength and direction of a linear relationship. The correlation coefficient is typically between -1 and 1. A value near 0 means there is no correlation, a value near 1 means there is a stronger positive correlation, and a value near -1 means there is a stronger negative correlation.

The features with the highest positive or negative correlation are picked in such a way that just one of them is taken into account while developing the model. The features in our heatmap, however, are not substantially connected, as can be observed. Here, high correlation means values larger than 0.9 or less than -0.9. We chose all 13 characteristics while training the dataset on models because we didn't have any such values.

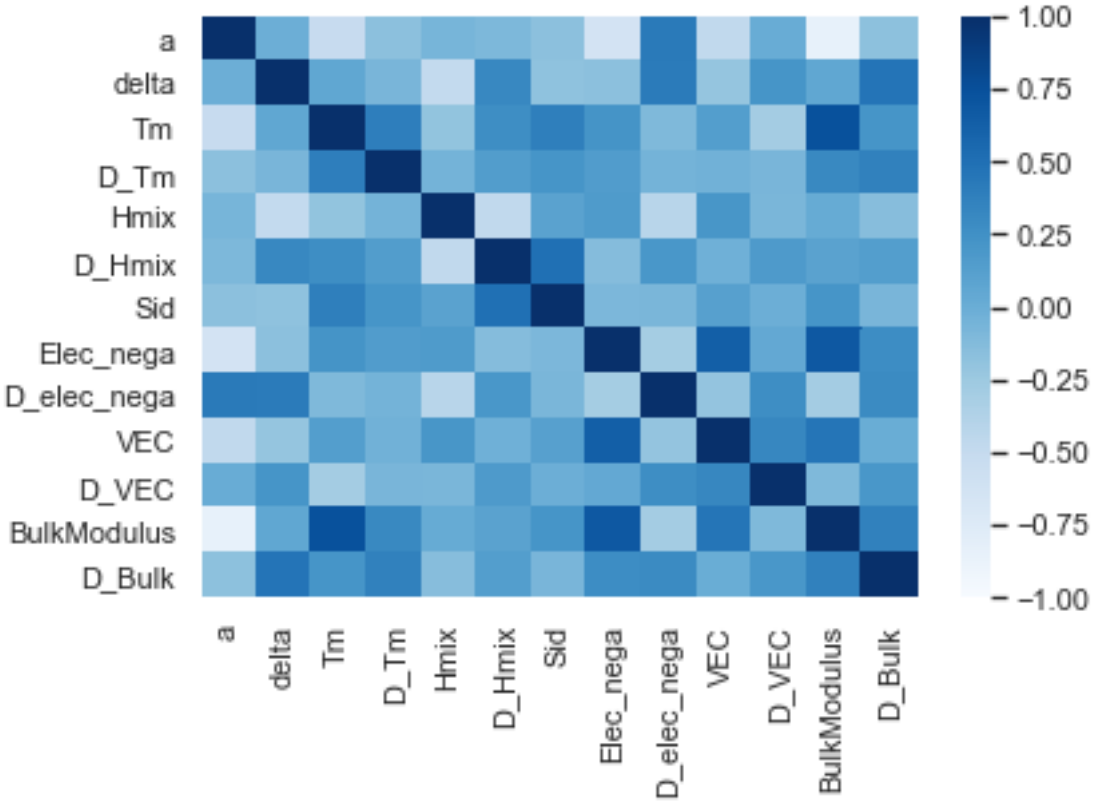

Fig. 5.1. A heatmap that demonstrates the correlation between 13 features[6]

5.2 Implementation of Random Forest

Since hyperparameter tuning is based on experimental data rather than theory, the simplest way to design a set of ideal hyperparameters is to test different combinations of hyperparameters to evaluate the model's performance. However, improving the model only on the training data may result in overfitting. This indicates that the model is exceptionally capable of learning the correlations between features and the attribute to be predicted. However, on the test dataset, it fails. As a result, we've included the use of a validation set in our model. The hyperparameters are recorded and used for the test set if the model performs well on the validation set.

On the training set, random forests are known to perform exceptionally well. We can limit the number of characteristics used for categorization in each decision tree, so that each one uses a different subset of features. To eliminate correlation in the base trees, these attributes are picked at random. The number of trees in our forest has been set to 100 in our model. The tree's depth has not been limited. After a grid search of possible numbers, the optimum number of features for each model was determ2

5.3 Implementation of Artificial Neural Networks

In a tensor flow architecture, the neural network was created by stacking hidden layers, each with a certain number of neurons. A model's number of hidden layers and neurons is different from one another. In our research, we used two hidden layers, each containing 50 neurons. 'ReLU' is the activation function used to convert linear inputs into nonlinear ones at the conclusion of each layer. It is critical that each hidden layer's inputs be converted to non-linear ones.

Because the layers are stacked, if the inputs are linear, the neural network will never be able to understand the dataset's complicated relationships.

Batches and epochs make up the model. Epoch is the number of times the weights are updated to reduce the model's loss, or the number of times the full dataset is processed through the model. The dataset is frequently separated into batches to save time and resources. The best number of batches and epochs is determined by the model and should be determined after a thorough grid search.

The outputs acquired by the neural network are compared to the targets after each batch. The difference between the targets and anticipated values is computed, and the error is then back propagated to update the weights and biases for each hidden layer. This is done to keep the loss to a minimum. The validation set is forward propagated at the conclusion of each epoch in order to compute the validation loss. The categorical cross entropy loss is the loss function in our situation, and the optimization approach is Adaptive Moment Estimation (Adam).

To avoid overfitting, we frequently partition the training dataset into a train set and a validation set. After each period, the validation and training losses are returned. Across epochs, the training loss decreases or remains constant. This could result in overfitting. As a result, the validation loss is continually visible. If the validation loss starts to climb (after initially dropping), the model is overfitting on the training data and should be stopped from being trained further. This is referred to as "early stopping." The patience level is set to 2, which means that if the validation loss increases two times in a row, the model will cease training.

The softmax function is used as the activation function in the output layer in our research. This function returns a vector containing the probability distributions of a list of all potential outcomes.

The following is an example of a function:

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

We divide the exponential values of all the input parameters by the exponential values of the input parameter acquired from the previous layer in the equation above. This procedure is done for each input parameter.

6. RESULTS AND DISCUSSIONS

	Type of phases	r# (g/cm3)	a	delta	Tm	D_Tm	Hmix	σHmix	Sid	Elec_ne	D_elec_	VEC	d_VEC	BulkMo	D_Bulk	Phase	Phase_i	E# (G
Al0.25CoFeNi	FCC	7.9	1.2643	0.0613	1,637.89	363.71	-18.8571	11.6220	1.9459	1.8129	0.1158	7.2857	2.8140	143,714,2	37.9839	BCC+FCC	SS	19
Al0.5CoFeNi	FCC + BCC	7.4	1.2643	0.0613	1,637.89	363.71	-18.8571	11.6220	1.9459	1.8129	0.1158	7.2857	2.8140	143,714,2	37.9839	BCC+FCC	SS	18
Al0.75CoFeNi	FCC + BCC	7.0	1.2643	0.0613	1,637.89	363.71	-18.8571	11.6220	1.9459	1.8129	0.1158	7.2857	2.8140	143,714,2	37.9839	BCC+FCC	SS	17
CoCrFeNi	FCC	8.2	1.2468	0.0030	1,871.75	180.37	-3.7500	1.6370	1.3863	1.8200	0.0967	8.2500	1.4790	172,500,(	8.2916	FCC	SS	22
CoCrFeNi	FCC	8.2	1.2468	0.0030	1,871.75	180.37	-3.7500	1.6370	1.3863	1.8200	0.0967	8.2500	1.4790	172,500,(	8.2916	FCC	SS	22
CoCrFeNi	FCC	8.2	1.2468	0.0030	1,871.75	180.37	-3.7500	1.6370	1.3863	1.8200	0.0967	8.2500	1.4790	172,500,(	8.2916	FCC	SS	22
CoCrFeNb0.103Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,(	38.5252	BCC+Lave	IM+SS	
CoCrFeNb0.155Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,(	38.5252	BCC+Lave	IM+SS	
CoCrFeNb0.206Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,(	38.5252	BCC+Lave	IM+SS	
CoCrFeNb0.309Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,(	38.5252	BCC+Lave	IM+SS	
CoCrFeNb0.412Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,(	38.5252	BCC+Lave	IM+SS	
CoCrFeNiTi0.5	FCC	7.2	1.2707	0.0533	1,879.44	171.44	-11.5556	6.8880	1.5811	1.7889	0.1267	7.7778	1.9309	165,555,5	21.1403	FCC+IM	IM+SS	13
Co1.5CrFeNi1.5Ti	FCC + Im	7.4	1.2707	0.0533	1,879.44	171.44	-11.5556	6.8880	1.5811	1.7889	0.1267	7.7778	1.9309	165,555,5	21.1403	FCC+IM	IM+SS	21
Al0.25CoCrFeNi	FCC	7.7	1.2597	0.0376	1,806.29	295.64	-7.2688	3.9105	1.5426	1.8053	0.1075	7.8837	1.9554	165,767,4	25.8518	FCC+L12	IM+SS	21
Al0.25CoCrFeNi	FCC	7.7	1.2597	0.0376	1,806.29	295.64	-7.2688	3.9105	1.5426	1.8053	0.1075	7.8837	1.9554	165,767,4	25.8518	FCC+L12	IM+SS	21

Table 4 : Prepared dataset for prediction of Hardness

	Type of phases	r# (g/cm3)	a	delta	Tm	D_Tm	Hmix	σHmix	Sid	Elec_ne	D_elec_	VEC	d_VEC	BulkMo	D_Bulk	Phase	Phase_i	E# (G
Al0.25CoFeNi	FCC	7.9	1.2643	0.0613	1,637.89	363.71	-18.8571	11.6220	1.9459	1.8129	0.1158	7.2857	2.8140	143,714,2	37.9839	BCC+FCC	SS	19
Al0.5CoFeNi	FCC + BCC	7.4	1.2643	0.0613	1,637.89	363.71	-18.8571	11.6220	1.9459	1.8129	0.1158	7.2857	2.8140	143,714,2	37.9839	BCC+FCC	SS	18
Al0.75CoFeNi	FCC + BCC	7.0	1.2643	0.0613	1,637.89	363.71	-18.8571	11.6220	1.9459	1.8129	0.1158	7.2857	2.8140	143,714,2	37.9839	BCC+FCC	SS	17
CoCrFeNi	FCC	8.2	1.2468	0.0030	1,871.75	180.37	-3.7500	1.6370	1.3863	1.8200	0.0967	8.2500	1.4790	172,500,(	8.2916	FCC	SS	22
CoCrFeNi	FCC	8.2	1.2468	0.0030	1,871.75	180.37	-3.7500	1.6370	1.3863	1.8200	0.0967	8.2500	1.4790	172,500,(	8.2916	FCC	SS	22
CoCrFeNi	FCC	8.2	1.2468	0.0030	1,871.75	180.37	-3.7500	1.6370	1.3863	1.8200	0.0967	8.2500	1.4790	172,500,(	8.2916	FCC	SS	22
CoCrFeNb0.103Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,(	38.5252	BCC+Lave	IM+SS	
CoCrFeNb0.155Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,(	38.5252	BCC+Lave	IM+SS	
CoCrFeNb0.206Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,(	38.5252	BCC+Lave	IM+SS	
CoCrFeNb0.309Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,(	38.5252	BCC+Lave	IM+SS	
CoCrFeNb0.412Ni	FCC + Im	8.2	1.2907	0.0610	1,734.85	458.77	-14.6576	6.4077	1.7242	1.7695	0.1236	7.0952	2.4670	154,000,(	38.5252	BCC+Lave	IM+SS	
CoCrFeNiTi0.5	FCC	7.2	1.2707	0.0533	1,879.44	171.44	-11.5556	6.8880	1.5811	1.7889	0.1267	7.7778	1.9309	165,555,5	21.1403	FCC+IM	IM+SS	13
Co1.5CrFeNi1.5Ti	FCC + Im	7.4	1.2707	0.0533	1,879.44	171.44	-11.5556	6.8880	1.5811	1.7889	0.1267	7.7778	1.9309	165,555,5	21.1403	FCC+IM	IM+SS	21
Al0.25CoCrFeNi	FCC	7.7	1.2597	0.0376	1,806.29	295.64	-7.2688	3.9105	1.5426	1.8053	0.1075	7.8837	1.9554	165,767,4	25.8518	FCC+L12	IM+SS	21
Al0.25CoCrFeNi	FCC	7.7	1.2597	0.0376	1,806.29	295.64	-7.2688	3.9105	1.5426	1.8053	0.1075	7.8837	1.9554	165,767,4	25.8518	FCC+L12	IM+SS	21

Table 5 : Prepared dataset for prediction of Young's Modulus

6.1 Prediction of Hardness Values using Random Forest Algorithm

In table 6, the alloys and their predicted values, test values and accuracy is tabulated for 10 specific alloys so that we can compare how close the values are to the predicted values. Different accuracies are obtained for different alloys. Random forest values are close when compared to the ANN. The equi-composition alloys have high accuracy compared to the non equi-composition alloys.

Alloys	Predicted values	Test Values	Accuracy(in %)
CoCrCuFeNiTi0.8	449.6	506.0	88.85
CoCrCuFeNiTi	458.1	512.0	89.47
Al0.5CoCrCuFeNi	467.8	517.0	90.48
Al0.5CoCrCuFeNi	472.925	520.0	90.94
Al0.5CoCrCuFeNi	477.8	524.0	91.18
Al0.5CoCrCuFeNi	485.3	530.0	91.56
AlCoCrCuFeNi	490.4	535.0	91.66
AlCoCrCuFeNi	493.1	536.0	92.0
AlCoCrCuFeNi	498.35	537.0	92.8
CoCuFeNi	502.45	539.0	93.21

Table 6: Accuracy values of hardness and comparison using Random Forest

The predicted values are comparatively less than the test values and accuracy can be calculated by ratio of difference between test values and predicted values and test values. The AlCoCrFeNi alloy values have accuracy around 90% and equi composition alloys have more hardness values than the normal alloys. The Vickers hardness doesn't have units.

Using Artificial neural networks, the predicted values are less than the test values and accuracy is calculated and it is around 90%.

The Scatter plot is plotted between the predicted values and test values. The epoch vs loss graph is also plotted. Epoch is nothing but the number of cycles required so that model learns and loss is significantly reduced.

6.2 Prediction of Hardness Values using ANN

In table 7, the alloys and their predicted values, test values and accuracy is tabulated for 10 specific alloys so that we can compare how close the values are to the predicted values. Different accuracies are obtained for different alloys. Using ANN, the values are obtained and they are not that close while compared to Random forest.

Alloy	Predicted values	Test values	Accuracy(in %)
CoCrCuFeNiTi0.8	447.41	504.0	88.77
CoCrCuFeNiTi	451.92	506.0	89.31
Al0.5CoCrCuFeNi	462.8	512.0	90.39
Al0.5CoCrCuFeNi	468.14	517.0	90.54
Al0.5CoCrCuFeNi	472.8	520.0	90.92
Al0.5CoCrCuFeNi	478.3	524.0	91.27
AlCoCrCuFeNi	484.4	530.0	91.39
AlCoCrCuFeNi	491.1	535.0	91.79
AlCoCrCuFeNi	494.35	536.0	92.22
CoCuFeNi	497.45	537.0	92.63

Table 7: Accuracy values of hardness and comparison using ANN

In figure 6.1, the Scatter plot is plotted between the predicted values and test values. The epoch vs loss graph is also plotted. Epoch is nothing but the number of cycles required so that model learns and loss is significantly reduced.

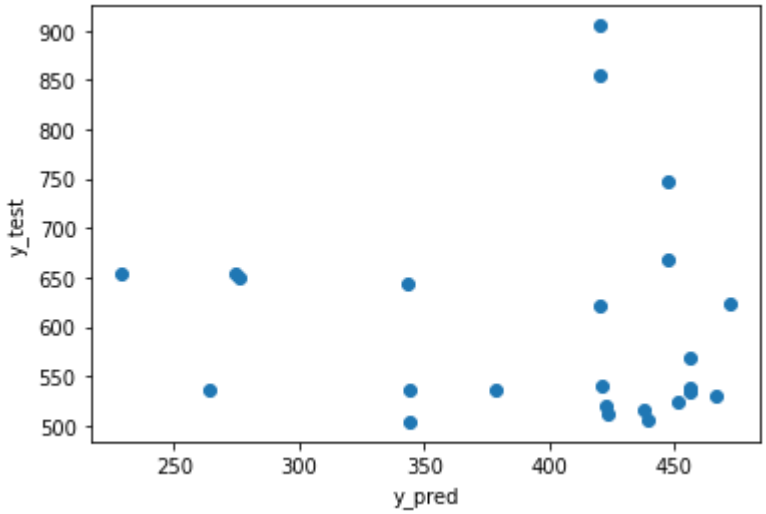

Figure 6.1 . Scatter plot of y_test and y_pred [HV]

In table 8, the R2 value for hardness prediction is 0.931 and RMSLE is 0.029 using random forest algorithm and using ANN the R2 value is 0.921 and RMSLE value is 0.036. By this we can say that for our particular data set random forest algorithm is more feasible than ANN. Random forest is quicker and takes less time.

Table 8: R^2 and RMLSE for prediction of hardness values

SS		Random Forest		ANN
R^2		0.931		0.921
RMSLE		0.029		0.036

In Fig 6.2, the epoch vs loss graph is also plotted. Epoch is nothing but the number of cycles required so that model learns and loss is significantly reduced. So as the number of cycles are increased the model gets to know more about the data and the model is trained so the model is trained and the loss reaches to a saturation point as it can't be trained more and if it over trains, the data gets over fitted and the model would serve no kind.

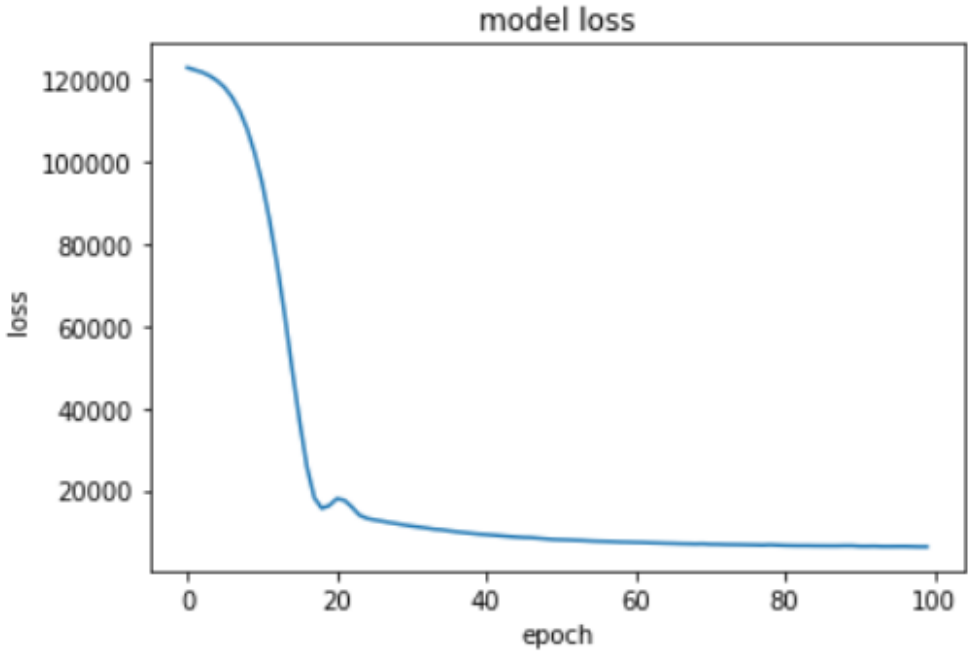

Fig 6.2: Model loss graph of Hardness prediction

6.3 Prediction of Young's Modulus using Random Forest Algorithm

In table 9, the alloys and their predicted values, test values and accuracy is tabulated for 10 specific alloys so that we can compare how close the values are to the predicted values. Different accuracies are obtained for different alloys. Random forest values are close when compared to the ANN. The equi-composition alloys have high accuracy compared to the non equi-composition alloys for Young's Modulus.

Alloys	Predicted data	Test Data	Accuracy
CoCrCuFeNiTi0.8	173.44	187.00	92.8
CoCrCuFeNiTi	169.7	158.00	92.6
Al0.5CoCrCuFeNi	172.1	173	99.5
Al0.5CoCrCuFeNi	159.34	154.00	96.6
Al0.5CoCrCuFeNi	143.52	147.73	97.2
Al0.5CoCrCuFeNi	111.1	111.00	99.1
AlCoCrCuFeNi	185.35	196.00	94.4
AlCoCrCuFeNi	159.39	150.00	94.1
AlCoCrCuFeNi	186.15	190.00	96.2
CoCuFeNi	148.33	139.06	92.3

Table 9: Accuracy values of Young's Modulus and comparison using Random Forest

The predicted values are comparatively less than the test values and accuracy can be calculated by ratio of difference between test values and predicted values and test values. The AlCoCrFeNi alloy values have accuracy around 90% and equi composition alloys have more Young's Modulus values than the normal alloys. The units are GPa.

6.4 Prediction of Young's Modulus using ANN

In table 10, the alloys and their predicted values, test values and accuracy is tabulated for 10 specific alloys so that we can compare how close the values are to the predicted values. Different accuracies are obtained for different alloys. Using ANN, the values are obtained and they are not that close while compared to Random forest. Young's Modulus values are in GPa.

Alloys	Predicted data	Test Data	Accuracy(in %)
CoCrCuFeNiTi0.8	165.73	158.00	95.1
CoCrCuFeNiTi	166.27	173.00	95.9
Al0.5CoCrCuFeNi	160.78	154.00	95.59
Al0.5CoCrCuFeNi	133.90	147.73	90.52
Al0.5CoCrCuFeNi	108.54	111.00	97.78
Al0.5CoCrCuFeNi	195.14	196.00	99.57
AlCoCrCuFeNi	160.60	150.00	93.33
AlCoCrCuFeNi	188.52	135.00	60.75
AlCoCrCuFeNi	134.34	152.20	88.15
CoCuFeNi	148.56	160.00	92.5

Table 10: Accuracy values of Young's Modulus and comparison using ANN

The RMSE and R2 coefficients both indicate how well a regression model fits a dataset. The RMSE measures how well a regression model can predict the response variable's value in absolute terms, whereas the R2 measures how well a model can predict the response variable's value in percentage terms.

In table 11, the R2 value for Young's Modulus prediction is 0.916 and RMSLE is 0.021 using random forest algorithm and using ANN the R2 value is 0.911 and RMSLE value is 0.032. By this we can say that for our particular data set random forest algorithm is more feasible than ANN. Random forest is quicker and takes less time.

Table 11: R^2 and RMLSE for prediction of hardness values

SS	Random Forest	ANN
R^2	0.916	0.911
RMSLE	0.021	0.032

In figure 6.4, the epoch vs loss graph is also plotted. Epoch is nothing but the number of cycles required so that model learns and loss is significantly reduced. So as the number of cycles are increased the model gets to know more about the data and the model is trained so the model is trained and the loss reaches a saturation point as it can't be trained more and if it over trains, the data gets over fitted and the model would serve no kind.

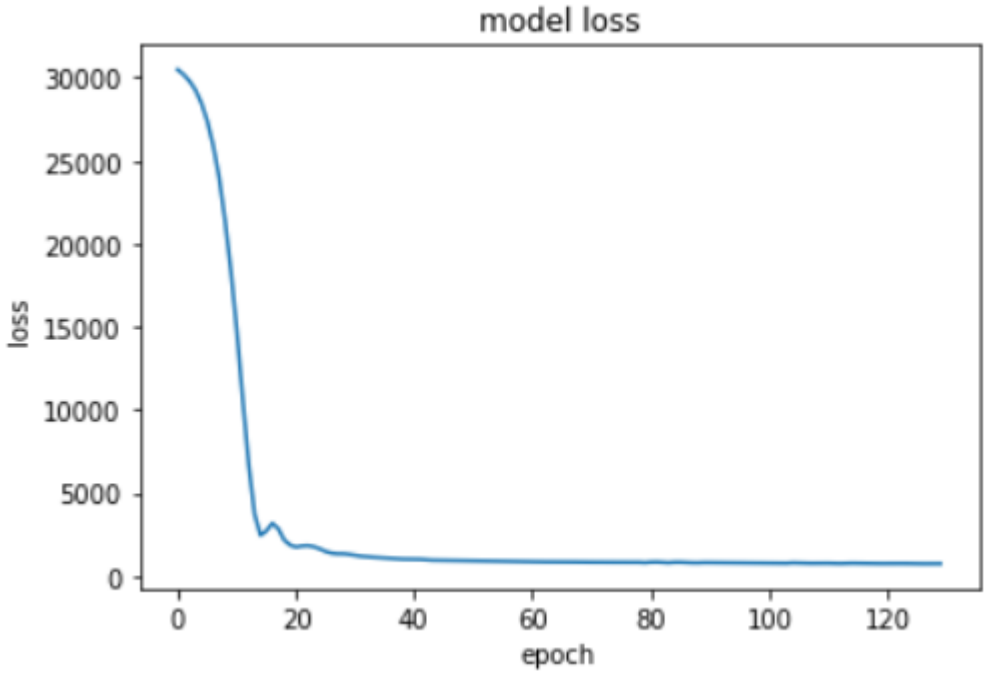

Figure 6.4 . Model loss vs Epoch [Young's Modulus]

Both R square and RMSLE are calculated for both the models in predicting hardness and Young's modulus of high entropy alloys.

7. SUMMARY AND CONCLUSIONS

High entropy alloys have certain exceptional physical and mechanical properties. Data is gathered from various sources of high entropy alloys and merged those data which have common alloy properties. One data set contains mechanical properties and another one contains 13 parameters which are related to phase prediction but have dependence on the mechanical properties. The training accuracy is largely dependent on the hyperparameters and can be improved to a great extent by adjusting the hyperparameters. However, the model's total accuracy is determined by its ability to perform on testing data, which is data that has never been seen before and has not been utilized for training. To quantify the correlation between features, a correlation matrix was created. The models were then trained and tested on carefully constructed datasets that included a balanced proportion of binary, ternary, quaternary, and high order alloys. High accuracies of around 90% are achieved using Random Forest and Ensemble Stacking (in most of the cases). Artificial Neural Networks are capable of learning complex associations in the training dataset with a 90% accuracy rate.

We carried out ANN and random forest models while predicting hardness values and young's modulus. The predicted values are again compared with test values and graphs are plotted between them in order to obtain the accuracy. While predicting hardness values, the R2 value for Random Forest model is 0.931 and for ANN model is 0.921 and RMSLE value for Random Forest model is 0.029 and for ANN model is 0.036. The predicted values are again compared with test values and graphs are plotted between them in order to obtain the accuracy. While predicting Young's Modulus, the R2 value for Random Forest model is 0.916 and for ANN model is 0.911 and RMSLE value for Random Forest model is 0.021 and for ANN model is 0.032. These are the results obtained from various models.

8. FUTURE SCOPE OF WORK

The development of high entropy alloys paved new paths in metallurgy and material science. When compared to typical standard alloys, these alloys feature outstanding qualities and phenomenal properties. In HEAs, there is a lot of room for more research and development, which could lead to new alloys. The following are some important areas where machine learning models should be improved:

- The composition can be taken into consideration when the required amount of data is available.
- Machine learning assisted design of high entropy alloys with desired property such as better elevated temperature strength, better hardness, better strength, toughness, creep resistance, and workability
- More accurate models can be built with more data as data can learn more without overfitting.

9 . References

1. Weidong Li, Di Xie, Dongyue Li, Yong Zhang, Yanfei Gao, Peter K. Liaw, *Mechanical behavior of high-entropy alloys*, Progress in Materials Science, Volume 118, 2021

2. D.B. Miracle, O.N. Senkov, *A critical review of high entropy alloys and related concepts*, Acta Materialia

3. Zhou, Z., Zhou, Y., He, Q. *et al.* Machine learning guided appraisal and exploration of phase design for high entropy alloys. *npj Comput Mater* 5, 128 (2019).

4. D.B. Miracle, O.N. Senkov, *A critical review of high entropy alloys and related concepts*, Acta Materialia

5. Weidong Li, Di Xie, Dongyue Li, Yong Zhang, Yanfei Gao, Peter K. Liaw, *Mechanical behavior of high-entropy alloys*, Progress in Materials Science, Volume 118,

6. S. Gorsse, M.H. Nguyen, O.N. Senkov, D.B. Miracle, *Database on the mechanical properties of high entropy alloys and complex concentrated alloys*, Data in Brief, Volume 21, 2018,

7. Y. Zhang, T.T. Zuo, Z. Tang, M.C. Gao, K.A. Dahmen, P.K. Liaw, *et al. Microstructures and properties of high-entropy alloys* Prog Mater Sci, 61 (2014), pp. 1-93

8. E.P. George, D. Raabe, R.O. Ritchie, *High-entropy alloys*, Nat Rev Mater, 4 (2019), pp. 515-534

9. Chang, YJ., Jui, CY., Lee, WJ. *et al.* Prediction of the Composition and Hardness of High-Entropy Alloys by Machine Learning. *JOM* 71, 3433–3442 (2019). https://doi.org/10.1007/s11837-019-03704-4

10. Liu, Yue & Zhao, Tianlu & Ju, Wangwei & Shi, Siqi. (2017). Materials discovery and design using machine learning. Journal of Materiomics. 3. 10.1016/j.jmat.2017.08.002.

11. Kumar, Nirmal & Goel, Gaurav & Goel, Saurav. (2021). Emergence of machine learning in the development of high entropy alloy and their prospects in advanced engineering applications. Emergent Materials. 10.1007/s42247-021-00249-8.

12. Xiong, Jie & Zhang, Tong-Yi & Shi, San-Qiang. (2020). Analysis of Phase Formations and Mechanical Properties in Complex Concentrated Alloys by Machine Learning Approach.

13. Vaidya, Mayur & Garlapati, Mohan & Murty, Budaraju. (2019). High-entropy alloys by mechanical alloying: A review. Journal of Materials Research. 10.1557/jmr.2019.37.