Supplementary Information: Structure-Aware Graph Neural Network Based Deep Transfer Learning Framework For Enhanced Predictive Analytics On Diverse Materials Datasets

Vishu Gupta[1], Kamal Choudhary[2,3], Brian DeCost[2], Francesca Tavazza[2], Carelyn Campbell[2], Wei-keng Liao[1], Alok Choudhary[1], Ankit Agrawal[1,*]

[1]*Department of Electrical and Computer Engineering, Northwestern University*

[2]*Materials Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD,20899, U.S.A.*

[3]*DeepMaterials LLC, Silver Spring, MD 20906, U.S.A.*

*Correspondence and requests for materials should be addressed to Ankit Agrawal (email: ankitag@eecs.northwestern.edu).

Supplementary Table 1: Data Size for each of the materials property in the Joint Automated Repository for Various Integrated Simulations (JARVIS) 3D Dataset.

Dataset	Property	Abbreviation	Size
JARVIS-3D	Bandgap Optb88vdw (eV)	BgOptb	55711
	Formation Energy ($eV atom^{-1}$)	Deltae	55711
	Kpoints Length Unit (Å)	KLU	55375
	Ehull ($eV atom^{-1}$)	Ehull	55323
	Encut (eV)	Encut	55240
	Magmom Oszicar (μ_B)	MagOszi	52894
	Magmom Outcar (μ_B)	MagOut	52210
	Epsx Refractive Index	Epsx	44490
	Epsy Refractive Index	Epsy	44490
	Epsz Refractive Index	Epsz	44490
	P Powerfact ($\mu W m^{-1} K^{-2}$)	PPf	23210
	N Powerfact ($\mu W m^{-1} K^{-2}$)	NPf	23210
	N Seebeck ($\mu V K^{-1}$)	NSb	23150
	P Seebeck ($\mu V K^{-1}$)	PSb	23125
	N Effective Masses 300K Average (m_0)	NEM300K	21394
	P Effective Masses 300K Average (m_0)	PEM300K	20213
	Elastic Tensor C33 (GPa)	ETC33	20066
	Elastic Tensor C22 (GPa)	ETC22	19941
	Elastic Tensor C11 (GPa)	ETC11	19905
	Bulk Modulus KV (GPa)	Bulk	19542
	Elastic Tensor C13 (GPa)	ETC13	19541
	Elastic Tensor C12 (GPa)	ETC12	19480
	Poisson	Poisson	19455
	Shear Modulus GV (GPa)	Shear	18496
	Bandgap MBJ (eV)	BgMbj	18044
	Elastic Tensor C44 (GPa)	ETC44	17714
	Average Mass Electron (electron mass unit)	AvgME	17642
	Average Mass Hole (electron mass unit)	AvgHE	17642
	Elastic Tensor C55 (GPa)	ETC55	17340
	Elastic Tensor C66 (GPa)	ETC66	17126
	Mepsx Refractive Index	Mepsx	16807
	Mepsy Refractive Index	Mepsy	16807
	Mepsz Refractive Index	Mepsz	16807
	Max (Phonon) Mode (cm^{-1})	MaxM	13876
	Min (Phonon) Mode (cm^{-1})	MinM	13160
	MaxEFG (10^{21} Vm^{-2})	MaxEFG	11865
	Spillage	Spillage	11375
	SLME (%)	SLME	9062
	Max Ir Mode (cm^{-1})	MaxIrM	4803
	Dfpt Piezo Max Eij (cm^{-2})	PMEij	4784
	Dfpt Piezo Max Dielectric	PMDi	4636
	Dfpt Piezo Max Dielectric Electronic (ε_{11})	PMDiEl	4527
	Min Ir Mode (cm^{-1})	MinIrM	4422
	Dfpt Piezo Max Dielectric Ioonic (ε_{11})	PMDiIo	4344
	Dfpt Piezo Max Dij (cm^{-1})	PMDij	3327
	Exfoliation Energy ($meV atom^{-1}$)	Exfoli	812

Supplementary Table 2: Data Size for each of the materials property in the JARVIS-2D Dataset.

Dataset	Property	Abbreviation	Size
JARVIS-2D	Bandgap Optb88vdw (eV)	BgOptb	1074
	Formation Energy ($eV atom^{-1}$)	Deltae	1074
	Kpoints Length Unit (Å)	KLU	1073
	Magmom Outcar (μ_B)	MagOut	1072
	Encut (eV)	Encut	1070
	Magmom Oszicar (μ_B)	MagOszi	1036
	Epsx Refractive Index	Epsx	885
	Epsy Refractive Index	Epsy	885
	Epsz Refractive Index	Epsz	885
	P Powerfact ($\mu W m^{-1} K^{-2}$)	PPf	800
	N Powerfact ($\mu W m^{-1} K^{-2}$)	NPf	800
	Exfoliation Energy ($meV atom^{-1}$)	Exfoli	742
	N Seebeck ($\mu V K^{-1}$)	NSb	733
	P Seebeck ($\mu V K^{-1}$)	PSb	707
	Spillage	Spillage	602
	P Effective Masses 300K Average (m_0)	PEM300K	264
	N Effective Masses 300K Average (m_0)	NEM300K	253
	Mepsz Refractive Index	Mepsz	246
	Mepsx Refractive Index	Mepsx	244
	SLME (%)	SLME	244
	Mepsy Refractive Index	Mepsy	243
	Elastic Tensor C11 (GPa)	ETC11	224
	Elastic Tensor C22 (GPa)	ETC22	223
	Bandgap MBJ (eV)	BgMbj	217
	Elastic Tensor C12 (GPa)	ETC12	212
	Elastic Tensor C44 (GPa)	ETC44	206
	Max (Phonon) Mode (cm^{-1})	MaxM	184
	Min (Phonon) Mode (cm^{-1})	MinM	183
	Elastic Tensor C55 (GPa)	ETC55	180
	Elastic Tensor C66 (GPa)	ETC66	180
	Elastic Tensor C13 (GPa)	ETC13	153
	Elastic Tensor C33 (GPa)	ETC33	135

Supplementary Table 3: Data Size for each of the materials property in the Other density functional theory (DFT) computed and Experimental Datasets.

Dataset	Property	Abbreviation	Size
Flla	Formation Energy (eV)	Deltae	3936
	Formation Energy per atom ($eV atom^{-1}$)	Deltae pa	3936
	Ehull ($eV atom^{-1}$)	Ehull	3927
Dielectric Constant (DC)	Volume (Å^3)	Vol	1054
	Bandgap (eV)	Bg	1054
	N	N	1054
	Poly Electronic	PolyElec	1054
	Poly Total	PolyTotal	1054
Piezoelectric Tensor (PT)	Volume (Å^3)	Vol	940
	Eij Max (cm^{-2})	Eij	936
Experimental Formation Energy (EFE)	Formation Energy ($eV atom^{-1}$)	Deltae	605
Kingsbury Experimental Formation Energy (KEFE)	Formation Energy ($eV atom^{-1}$)	Deltae	1557
Kingsbury Experimental Bandgap (KEB)	Bandgap (eV)	Bg	2432
Harvard Organic Photovoltaic Dataset (HOPV)	{b3lyp, bp86, m06, pbe0} gap (eV)	-	4854
	{b3lyp, bp86, m06, pbe0} homo (eV)	-	4854
	{b3lyp, bp86, m06, pbe0} lumo (eV)	-	4854
	{b3lyp, bp86, m06, pbe0} scharber jsc ($mAcm^{-2}$)	-	4854
	{b3lyp, bp86, m06, pbe0} scharber pce (%)	-	4854
	{b3lyp, bp86, m06, pbe0} scharber voc (V)	-	4854

Supplementary Table 4: Modified target properties and their description.

Notation	Description
Max Mode	Maximum value of the Modes Array
Min Mode	Minimum value of the Modes Array
P Effective Masses 300K Average	Average of the P type Effective Masses at 300K
N Effective Masses 300K Average	Average of the N type Effective Masses at 300K
Elastic Tensor C11	Elastic Tensor (ET) at (1,1) of the ET array
Elastic Tensor C12	Elastic Tensor (ET) at (1,2) of the ET array
Elastic Tensor C13	Elastic Tensor (ET) at (1,3) of the ET array
Elastic Tensor C22	Elastic Tensor (ET) at (2,2) of the ET array
Elastic Tensor C33	Elastic Tensor (ET) at (3,3) of the ET array
Elastic Tensor C44	Elastic Tensor (ET) at (4,4) of the ET array
Elastic Tensor C55	Elastic Tensor (ET) at (5,5) of the ET array
Elastic Tensor C66	Elastic Tensor (ET) at (6,6) of the ET array

Supplementary Table 5: The table shows the type of transfer learning (TL) model used on the test set to get the results shown in Table 1.

Dataset	Property	Best Performing Model
JARVIS-3D	BgOptb	Fine-Tuning
	Deltae	Fine-Tuning
	KLU	Fine-Tuning
	Ehull	Fine-Tuning
	Encut	Feature Extraction
	MagOszi	Fine-Tuning
	MagOut	Fine-Tuning
	Epsx	Fine-Tuning
	Epsy	Fine-Tuning
	Epsz	Fine-Tuning
	PPf	Fine-Tuning
	NPf	Fine-Tuning
	NSb	Fine-Tuning
	PSb	Feature Extraction
	NEM300K	Feature Extraction
	PEM300K	Feature Extraction
	ETC33	Fine-Tuning
	ETC22	Fine-Tuning
	ETC11	Fine-Tuning
	Bulk	Feature Extraction
	ETC13	Fine-Tuning
	ETC12	Fine-Tuning
	Poisson	Feature Extraction
	Shear	Fine-Tuning
	BgMbj	Fine-Tuning
	ETC44	Fine-Tuning
	AvgME	Feature Extraction
	AvgHE	Feature Extraction
	ETC55	Fine-Tuning
	ETC66	Fine-Tuning
	Mepsx	Fine-Tuning
	Mepsy	Fine-Tuning
	Mepsz	Fine-Tuning
	MaxM	Fine-Tuning
	MinM	Fine-Tuning
	MaxEFG	Fine-Tuning
	Spillage	Fine-Tuning
	SLME	Fine-Tuning
	MaxIrM	Feature Extraction
	PMEij	Feature Extraction
	PMDi	Feature Extraction
	PMDiEl	Feature Extraction
	MinIrM	Feature Extraction
	PMDiIo	Feature Extraction
	PMDij	Feature Extraction
	Exfoli	Feature Extraction

Supplementary Table 6: The table shows the type of TL model used on the test set to get the results shown in Table 2.

Dataset	**Property**	**Best Performing Model**
Flla	Deltae	Feature Extraction
	Deltae pa	Fine-Tuning
	Ehull	Feature Extraction
DC	Vol	Fine-Tuning
	Bg	Feature Extraction
	N	Feature Extraction
	PolyElec	Feature Extraction
	PolyTotal	Feature Extraction
PT	Vol	Feature Extraction
	Eij	Feature Extraction

Supplementary Table 7: The table shows the type of TL model used on the test set to get the results shown in Table 3.

Dataset	Property	Best Performing Model
JARVIS-2D	BgOptb	Feature Extraction
	Deltae	Feature Extraction
	KLU	Feature Extraction
	MagOut	Feature Extraction
	Encut	Fine-Tuning
	MagOszi	Feature Extraction
	Epsx	Feature Extraction
	Epsy	Feature Extraction
	Epsz	Feature Extraction
	PPf	Fine-Tuning
	NPf	Feature Extraction
	Exfoli	Feature Extraction
	NSb	Fine-Tuning
	PSb	Feature Extraction
	Spillage	Feature Extraction
	PEM300K	Feature Extraction
	NEM300K	Feature Extraction
	Mepsz	Feature Extraction
	Mepsx	Feature Extraction
	SLME	Feature Extraction
	Mepsy	Feature Extraction
	ETC11	Fine-Tuning
	ETC22	Fine-Tuning
	BgMbj	Feature Extraction
	ETC12	Fine-Tuning
	ETC44	Feature Extraction
	MaxM	Fine-Tuning
	MinM	Feature Extraction
	ETC55	Feature Extraction
	ETC66	Feature Extraction
	ETC13	Feature Extraction
	ETC33	Feature Extraction

Supplementary Table 8: The table shows the type of TL model used on the test set to get the results shown in Table 4.

Dataset	Property	Best Performing Model
HOPV	b3lyp gap	Feature Extraction
	b3lyp homo	Feature Extraction
	b3lyp lumo	Feature Extraction
	b3lyp scharber jsc	Fine-Tuning
	b3lyp scharber pce	Feature Extraction
	b3lyp scharber voc	Fine-Tuning
	bp86 gap	Feature Extraction
	bp86 homo	Feature Extraction
	bp86 lumo	Feature Extraction
	bp86 scharber jsc	Fine-Tuning
	bp86 scharber pce	Fine-Tuning
	bp86 scharber voc	Fine-Tuning
	m06 gap	Feature Extraction
	m06 homo	Feature Extraction
	m06 lumo	Feature Extraction
	m06 scharber jsc	Fine-Tuning
	m06 scharber pce	Fine-Tuning
	m06 scharber voc	Fine-Tuning
	pbe0 gap	Feature Extraction
	pbe0 homo	Feature Extraction
	pbe0 lumo	Feature Extraction
	pbe0 scharber jsc	Fine-Tuning
	pbe0 scharber pce	Feature Extraction
	pbe0 scharber voc	Fine-Tuning

Supplementary Table 9: The table shows the type of TL model used on the test set to get the results shown in Table 4.

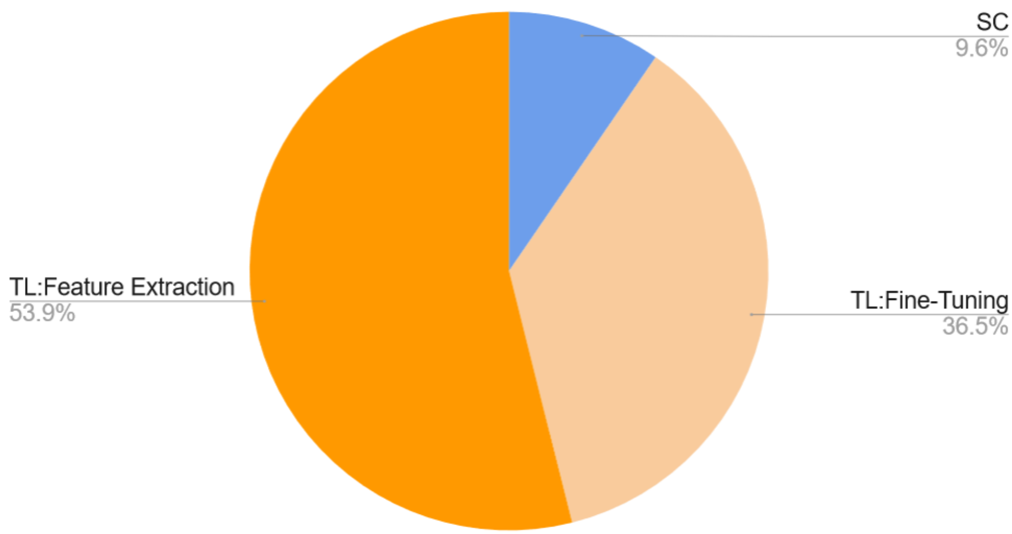
Dataset	Property	Best Performing Model
EFE	Deltae	Feature Extraction
KEFE	Deltae	Fine-Tuning
KEB	Bg	Feature Extraction

Supplementary Table 10: The table shows the mean ± standard deviation of MAE for different target properties from 9-fold cross-validation for SC and the proposed TL model.

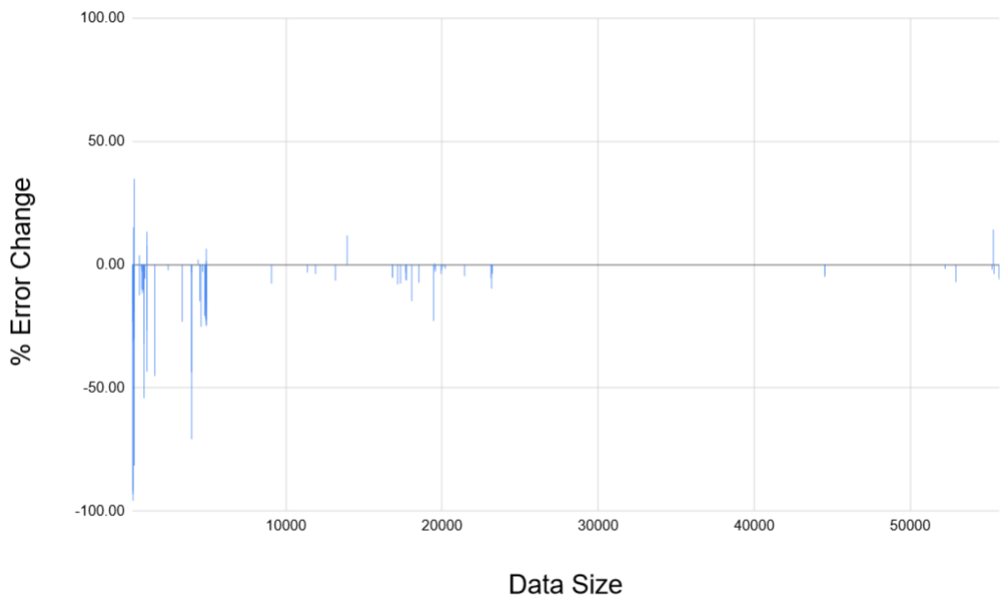
Dataset	**Property**	**SC**	**TL**
Jarvis3D	Deltae	0.0329±0.0009	0.0301±0.0007
	BgOptb	0.1349±0.0035	0.1208±0.0026
Jarvis2D	Deltae	0.1152±0.0176	0.0660±0.0124
	BgOptb	0.4844±0.0585	0.4106±0.0492
Experimental	Deltae	0.0973±0.0118	0.0534±0.0069
	Bg	0.4098±0.0765	0.3339±0.0677

Supplementary Table 11: The table shows the test MAE of the SC model and proposed TL model when the source and target model is trained on single and multiple materials properties. Here the source model is trained on the MP dataset, and the target model is trained on the JARVIS dataset.

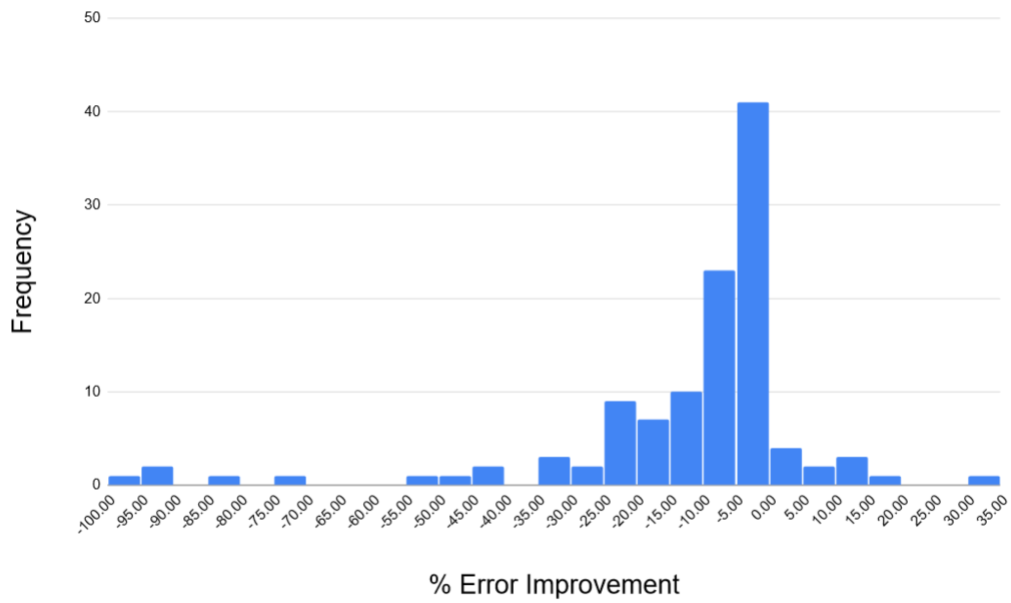
Target Property	**Single Property**			**Multiple Properties (Deltae + Bandgap)**	
	SC	**TL (Source Models)**		**SC**	**TL**
		Deltae	**Bandgap**		
Deltae	0.0343	0.0325	0.0419	0.0513	0.0539
Bandgap	0.1341	0.1258	0.1190	0.1335	0.1190



Supplementary Figure 1: Pie chart representing the number of cases for which each model performed the best.



Supplementary Figure 2: Percent improvement of the TL model against the SC model as a function of dataset size.



Supplementary Figure 3: Histogram representing percent improvement of the TL model against the SC model.