Entropy-based active learning of graph neural network surrogate models for materials properties

Johannes Allotey, Keith T. Butler, and Jeyan Thiyagalingam

Published version information

Citation: J Allotey, KT Butler and J Thiyagalingam. 'Entropy-based active learning of graph neural network surrogate models for materials properties.' J Chem Phys, vol. 155, no. 17 (2021): 174116.

DOI: 10.1063/5.0065694

Entropy-based Active Learning of Graph Neural Network Surrogate Models for Materials Properties

Johannes Allotey
School of Physics, University of Bristol, BS8 1TL, UK

Keith T. Butler* and Jeyan Thiyagalingam
Scientific Machine Learning Research Group,
Scientific Computing Department,
Rutherford Appleton Laboratory,
Science and Technology Facilities Council,
Didcot, OX11 0DQ, UK

Graph neural networks, trained on experimental or calculated data are becoming an increasingly important tool in computational materials science. Networks, once trained, are able to make highly accurate predictions at a fraction of the cost of experiments or first-principles calculations of comparable accuracy. However these networks typically rely on large databases of labelled experiments to train the model. In scenarios where data is scarce or expensive to obtain this can be prohibitive. By building a neural network that provides a confidence on the predicted properties, we are able to develop an active learning scheme that can reduce the amount of labelled data required, by identifying the areas of chemical space where the model is most uncertain. We present a scheme for coupling a graph neural network with a Gaussian process to featurise solid-state materials and predict properties *including* a measure of confidence in the prediction. We then demonstrate that this scheme can be used in an active learning context to speed up the training of the model, by selecting the optimal next experiment for obtaining a data label. Our active learning scheme can double the rate at which the performance of the model on a test data set improves with additional data compared to choosing the next sample at random. This type of uncertainty quantification and active learning has the potential to open up new areas of materials science, where data are scarce and expensive to obtain, to the transformative power of graph neural networks.

I. INTRODUCTION

Machine learning (ML) has become an important tool in almost every modern scientific discipline, and materials' science is no exception [1]. In particular, the proliferation of data in recent years has given rise to the advent of deep learning approaches, where neural networks with hundreds of thousands or even millions of parameters learn to infer trends from relatively unstructured data [2–6]. In materials' science, network architectures based on graphs, Graph Neural Networks (GNNs), have proved to be successful both for molecules and condensed matter systems[2, 3, 6]. GNNs allow the encoding of domain knowledge about connections (bonds) into the topography of the underlying neural network architecture. Despite the great success and promise of GNNs for materials' science applications a number of questions remain open. In this paper we address two of these questions (i) how much can we trust the results of a GNN on a previously unseen sample, and (ii) how can we train a GNN if data is scarce or labelled data is expensive to obtain?

Uncertainty quantification (UQ) of machine learning models, and neural networks in particular, is currently an area of intense research. Conventional neural networks return a single output value per property that they are trained on (classes or values in a classification or a regression scenario). These values contain neither information about how confident the network is, nor how the output varies with variation of the input. In a recent high-profile example, a network trained to classify COVID-19 infection from lung X-rays classified a cat as COVID-19

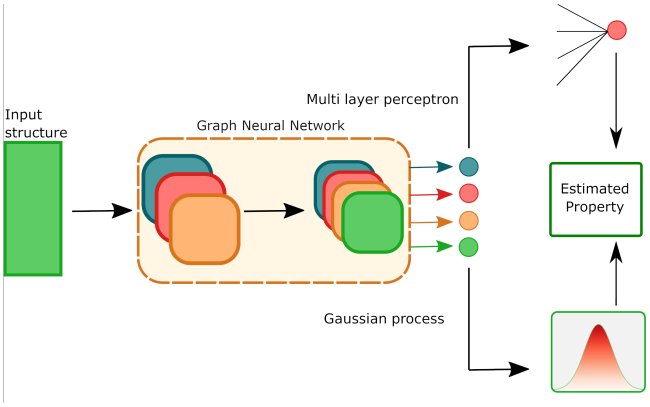


FIG. 1: The structure of the convolution fed Gaussian process (CFGP). The graph network part of MEGNET learns representations for materials, in the standard model these are fed to a multi-layer perceptron for function approximation, in the CFGP these are used as input for a Gaussian process for function approximation.

* keith.butler@stfc.ac.uk

infected lung[7]. In this case an estimation of certainty would have been very beneficial.

Bayesian neural networks (BNNs) provide NNs with a principled UQ[8]. However, BNNs are not scalable to most practical problems. Although various approximations, eg Monte Carlo dropout and mean field variational inference, [9, 10] improve scalability of BNNs, this comes at the cost of UQ reliability. Deep ensembles provide generally good estimates of UQ but are not *a priori* reliable.

Alternative machine learning methods, rooted in Bayesian statistics, such as Gaussian processes do provide principled, and scalable UQ in many scenarios [11]. GPs have proved very successful in materials modelling with a range of studies recently employing them for applications such as calculating energies and dynamics [12, 13]. The major barrier in applying GPs to practical problems is the necessity to construct meaningful representations (or features) of the data as an input to the GP. Several elegant examples, such as the SOAP descriptor, have been proposed, but still need to be derived for each new system[14].

One avenue to addressing the problem of feature engineering is to rely on representation learning or feature learning, an area of machine learning that focuses on automatically discovering the features that best describe the underlying (possibly hidden) characteristics of the data. Representation learning thus helps to not only better understand a dataset, but also to produce better outcomes. Although specialised architectures can be built to extract minimal representations of data [15], even the hidden layers of simple neural networks progressively capture these representations [16].Representation learning has been instrumental in a number of areas[17, 18].

In this work we combine a GNN with a GP by performing feature learning with the GNN and using the features learned by the GNN as input to a GP, thus circumventing the need for manual feature engineering. This approach has recently been demonstrated to work well for calculating adsorption energies for catalysis [19] and builds on a number of notable recent efforts towards uncertainty quantification for machine learning of molecular and materials properties[20, 21]. We demonstrate that the recently published MEGNET architecture can be used to learn representations that can be applied in a GP to predict formation energy of a crystal.

We first show that the latent space of the MEGNET model is structured in such a way that information about the target properties can be inferred. We then demonstrate a well-calibrated GP with reliable UQ on the prediction of formation energies, we show that this can be obtained using either MEGNET or more generic graph convolutional networks as the featurising network with very similarly good performance. We use this UQ to address the issue of using GNNs in scenarios where labelled data is scarce. We use the uncertainties calculated on previously unseen examples to choose the next training point for the ML model. This so-called entropy sampling-based active learning procedure is applied to training a model for formation energy prediction and is shown to significantly outperform random sampling of the next training point.

We are confident that the methods and models presented in this paper can be applied to a range of properties and scenarios in materials' science, for example when training data is difficult to obtain such as dielectric properties, phonons or high quality electronic structures [22]. Active learning has already demonstrated transformative power in the field of molecular quantum chemistry[23], we hope to contribute towards its application in condensed matter materials.

II. LATENT SPACE AND PREDICTED PROPERTIES

We first investigate whether a latent space representation of a material can be useful as an input vector for a Gaussian process. In order to do this we have extracted the outputs from the graph network section of MEGNET . To be precise we use the values from the layer labelled "Dense 32" from [2], this network was trained for 1,000 epochs on 1,800 randomly chosen materials from the full dataset, before being used to extract the activations for the same 1,800 materials. After a material is input to the network, we extract the values of the activations in this layer, this 32-dimensional vector serves as a compressed representation of the material. Visualisation of higher dimensional data is well-known to be difficult for human cognition, thus we reduce the dimensions of the latent space for visulaisation purposes, using the t-distributed stochastic neighbour embedding (t-SNE)[24]. t-SNE is a non-linear and stochastic dimensionality reduction technique that projects the original data points (32-dimensions) to two-dimensional space.

The t-SNE-derived 2D distribution of the activation vectors for the dataset (see Dataset section for more details) are presented in Figure 2. In this figure we have coloured the points according to their formation energy. From this Figure, it is clear that there is a structure in the latent space relating to formation energy; points tend to have similar formation energies to those clustered close by. This gives us confidence that the graph network is learning a representation (or features) that contains the relevant information required for building a subsequent model for approximating the formation energy.

III. A NETWORK-FED GAUSSIAN PROCESS

Having established that there is a structure of the latent space that is related to the output property of interest, we now develop a Gaussian process (GP) model to infer the relationship. Using the notation of Tran and coworkers[19] we refer to this as a convolution-fed Gaussian process (CFGP). The GP uses the 32-dimensional vector

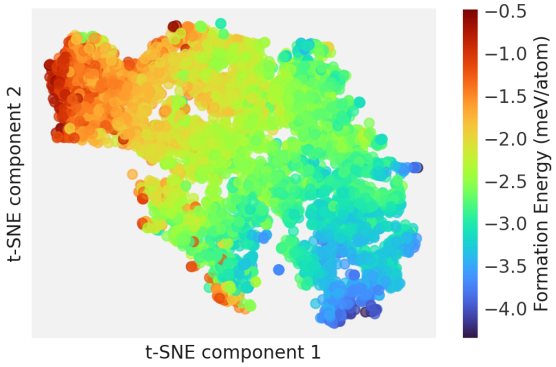


FIG. 2: A reduced dimensionality plot of the activations extracted from the first dense layer of the MEGNET model. The activations are projected onto a 2D space using the t-SNE algorithm. The points are coloured according to the formation energy of the material.

from the GNN as the input vector - full details of the GP are given in the Models section.

To characterise the similarity between samples, we use the Laplacian kernel

$$k(x, y) = a^2 \exp\left(\frac{-||x - y||}{l}\right) \quad (1)$$

where the similarity k between two data samples at points x and y depends on a function of the difference between these points and an amplitude a and length scale l of the kernel. The full details of how values for a and l are obtained are given in the Methods section.

The CFGP is trained on 7,800 materials from the original dataset. Where 1,800 materials were previously used to train the MEGNET model and the additional 6,000 materials are identified during an active learning procedure (see Methods for details). Note that during the active learning procedure the weights of the MEGNET model are not updated, only the GP is trained, however at the end of the active learning we re-optimized MEGNET and the GP on the new training set of 7,800 materials. We then test the model on 1,460 previously unseen materials to evaluate the performance.

Figure 3 plots the CFGP predicted formation energy against the DFT calculated formation energy, for a test set of 1,460 materials not used to train the model. We can observe a high degree of correlation between the predicted and ground-truth (DFT) values. The test set data has a mean absolute error (MAE) of 0.0277 eV/atom, a mean squared error of 0.002 eV/atom. This performance is comparable to the performance of other GNN-based methods as reported on the MATBENCH benchmark dataset[25], where MEGNET had and MAE of 0.0417 eV/atom. We note that the training and test sets that we apply here are somewhat more restricted than the MATBENCH data, explaining our lower MAE.We have also implemented a CFGP where the graph convolutions are performed using the graph convolution units from the [3] architecture to generate features for the Gaussian process. The results of the performance of this architecture are presented in the Supporting Information. We can see that the MAE of this model on the test data is actually a slight improvement on the MEGNET featurised model at 0.0201 eV/atom. This shows that the general principle of featurising materials using a graph network is valid regardless of the particular flavour of graph network and resonates with a recent benchmarking study which showed broadly similar performance of several flavours of materials graph neural network architectures.[26]

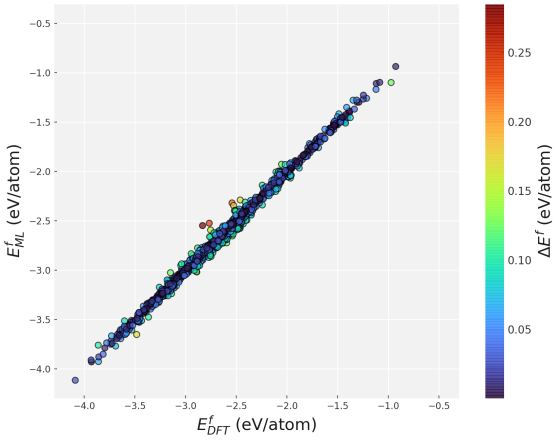


FIG. 3: The performance of the CFGP model for predicting formation energies on a test set of 1460 materials. The model has a mean absolute error of 0.0277 meV/atom, with an r^2 value of 0.99.

This demonstrates that the CFGP can perform the task of predicting properties from materials structure with comparable fidelity to GNNs employing multi-layer perceptron function approximators. The advantage of the GP is that we also obtain UQ, or degree of confidence on the prediction. In the section that follows this, we evaluate the resulting UQ performance.

IV. CALIBRATION OF UNCERTAINTY QUANTIFICATION

To assess the quality of the UQ element in our CFGP we compare the uncertainty estimates in the training set to the residuals between the predicted and true values. The work of Tran and *et al.* [19] provides an excellent, lucid introduction to benchmarking UQ methods, outlining recently proposed protocols [27, 28]. We briefly outline the criteria for good UQ methods here but, highly recommend the aforementioned work of Tran *et al.*[19].

A good uncertainty estimate should be well-calibrated,

sharp, and disperse. In our context, well-calibrated means that the residuals generally fall within the range of the error estimates (in our case twice the standard deviation of the GP). Sharpness implies that the error estimates are not unnecessarily large; conservative large error-estimates give good calibration, but are not necessarily useful for assessing how the model will perform. Finally disperse means that the error-estimates cover a range of values, so that the model does not simply predict a uniform estimate for all samples.

In Figure 4 we show the performance of the resulting UQ. To provide more clarity, the upper plot shows a randomly selected set of 100 samples from the test set. Specifically, the CFGP energies are plotted against the DFT energies with error bars equal to twice the standard deviation (2σ) of the GP; in a standard distribution, 95% of the data should fall within 2σ. We can observe that, within the sample set, the great majority of points fall within 2σ of the DFT energy. In the lower panel we plot the residuals against σ. We use the the same data as in Figure 3 for 1,460 test materials, not used in training. This plot demonstrates that validation samples with larger residuals generally do have larger uncertainty estimates. We also colour the points by the difference between the residual and 2σ, indicating which points fall outside the error estimates.

We note in Figure 4 that there are a few materials that stand out with particularly high errors and also with a large discrepancy between the UQ and the error in the prediction. Among these, three particularly poor predictions are labelled in the figure, we note that all three systems have phosphate motifs. The phosphate motif is relatively unusual in the context of an inorganic materials database. In addition two of the materials, $ZrP_2(HO_3)_2$ and $SnPO_3F$ have an apical oxygen in the phosphate replaced with P-H and P-F bonds respectively. These kinds of bonds are unusual and are therefore probably not present in large amounts in the training set. The scarcity of these motifs can explain why the model fails to predict them with good accuracy. We suggest that our UQ here is possibly over-cautious, but does generally identify well cases where the results of the model are likely to be unreliable. The calibration of the UQ could potentially be improved by exploring alternative kernels for the GP or using different layers of MEGNET as input vectors, this will be the subject of further investigations, for now we assess how well the model can perform in an active learning procedure.

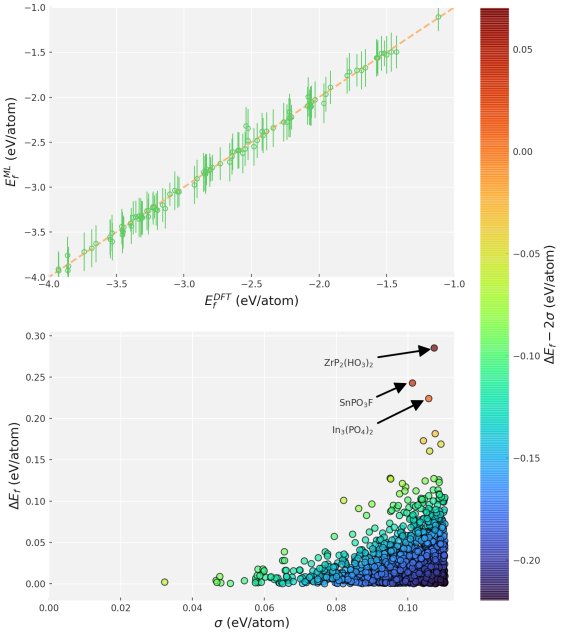


FIG. 4: The effectiveness of uncertainty quantification by the Gaussian process. Top plot shows 100 randomly selected points, the CFGP energy is plotted against the DFT energy, the error-bars are set to twice the standard deviation obtained from the CFGP. Bottom plot shows the residuals (difference between DFT energy and CFGP energy) versus the standard deviation obtained from the CFGP.

V. ACTIVE LEARNING

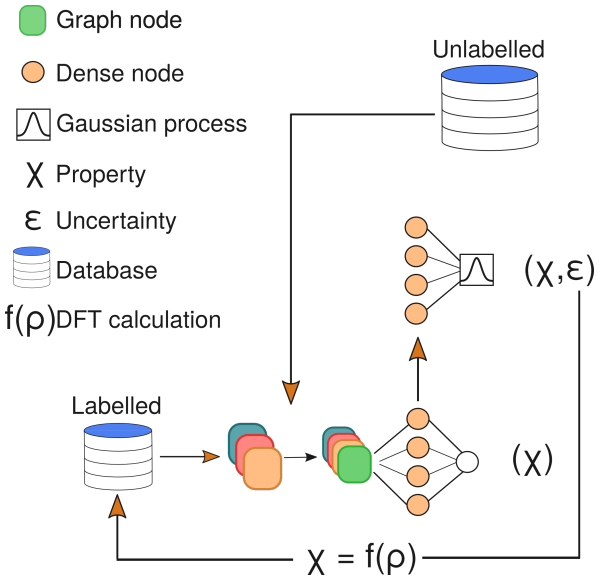
We now explore how the UQ demonstrated in the previous section can be further leveraged for active learning. In active learning, the ML model not only learns the relationships in the underlying data from labelled samples, but also chooses which unlabelled sample should be labelled next in order to best improve the performance and generality of the model. There are a number of techniques that can be used to select the points to be labelled next [29]. Recently an approach based on query by committee has been used to accelerate training for a quantum chemistry dataset [23].

In our case, we apply an entropy-based sampling approach, which chooses the unlabelled sample with the highest uncertainty and uses that as the next sample to be labelled. The uncertainty represents the entropy of the samples from an information theory perspective. Hence, opting to label the sample with greatest uncertainty next maximises the entropy reduction in the sampling process. In the context of the GP, a point with high uncertainty is likely to be found in an area of feature space that is more sparsely labelled than other areas, thus choosing this point for labelling favours a more representative sampling of all regions of feature space.

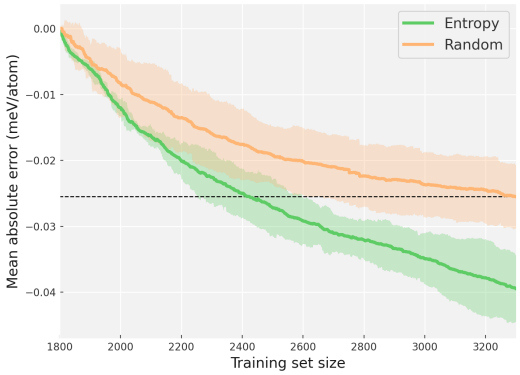
A schematic illustration of our process is shown in the upper panel of Figure 5. First, a trained MEGNET is passed the labelled data and the activations of the dense layer are used to feed a GP for training (a CFGP). Then the unlabelled data is passed through the CFGP pro-

viding predictions with uncertainties. Here, the sample with the greatest uncertainty is chosen to be labelled. Then the new larger labelled dataset is used to re-train the Gaussian process. The overall process continues until some predefined end point. In our case we started with 1,800 labelled materials from the dataset described in the Dataset section, we then run through 1,500 cycles of the active learning procedure evaluating the performance of the CFGP at each cycle model on all of the remaining materials whose labels were not used in training (i.e. the full set of materials minus the training set at that cycle).

The results of the entropy-based active learning process are presented in Figure 5 lower panel. For comparison we have also performed a sampling of new materials based on random selection from the unlabelled data. We repeated the random and entropy-based procedures 4 times each to get some indication of the range of possible results; means and 2σ values are presented in Figure 5. We can see that the maximum entropy-based sampling approach significantly out-performs the random sampling-based approach. By the end of the 1,500 cycles of active learning, the model from the maximum entropy based approach outperforms the model from the random approach by $\sim 43\%$. The enhancement in performance achieved by the random selection process after 1,500 additional samples is achieved by the entropy based sampling in fewer than half the samples; 740.



(a) Schematic representation of our active learning process



(b) Performance of our active learning process versus random sampling

FIG. 5: The process and effectiveness of entropy sampling based active learning. Upper; the active learning procedure, labelled data are passed through MEGNET and activations at the dense layer extracted to train a Gaussian process, unlabelled data are then passed through the same process and predictions and uncertainties obtained, samples with the greatest uncertainty are then labelled and the procedure loops back around. Lower; the average and twice the standard deviation of mean absolute error on the test set of the CFGP with different sampling methods; the mean and standard deviation were obtained across 5 runs of each procedure. The entropy-based sampling method chooses the next sample to label based on the uncertainty in prediction of the unlabelled samples, the random sampling method chooses at random.

VI. DISCUSSION

We have demonstrated that Gaussian process trained on the outputs of a GNN, a CFGP, performs well for prediction, uncertainty quantification and active learning for the formation energy of materials. The CFGP model that we presented here has uses in a very wide range of materials science applications. In fact, a similar architecture was recently demonstrated to perform very well for prediction and uncertainty quantification in absorption energy prediction for catalysis[19]. This approach is particularly appealing as it combines the theoretical rigour of the Gaussian processes-based approaches, with the flexibility of the GNN approach. GNNs can, in principle, represent any materials or molecular system, with minimal user-defined input required, and therefore this approach should be easily adaptable to other domains of interest with minimal feature engineering .

The ability to place reliable confidence bounds on predicted properties is crucial for the application of ML methods and building trust in the outputs of models. For example, in the context of virtual high-throughput screening ML surrogate models are often applied as a filtering step [30–34]. Access to reliable UQ in this context is very attractive as it can be used in combination with threshold values to filter out prospective candidates. Reliable UQ can also reduce the number of false positives passing through the procedure or the number of false negatives being rejected, and thus could be used on-the-fly to assess edge cases at a higher level of fidelity.

The potential application of active learning in materi-

als and molecular science is being increasingly recognised [35, 36]. In applications such as deriving new inter-atomic potentials[37, 38] or searching through materials space for optimal catalysts[39], the ability to optimise the choice of the next experiment is important. Active learning is particularly appealing in scenarios where data is scarce and expensive to obtain. For energy materials design properties such as high-quality electronic structure [40], dielectric constants[22, 41–43], effective masses[44] and defect properties[45, 46] are some high-profile examples where data-driven approaches are possible, but obtaining large sets of labelled data can be prohibitive to applying deep learning approaches. In these scenarios the availability of a method for reducing the number of samples required to train accurate, and reliable models is rather critical to the application of deep learning approaches in materials' science.

We note that the UQ, while capturing most errors within 2σ does have a tendency to overestimate uncertainty in cases where the error is low; as evidenced by the many clue points in the lower right hand side of Figure 4 (lower plot). This suggests that out UQ is not as disperse as it would ideally be. There are several approaches that could be taken to improve this performance, for example we have not fully explored the effect of different kernels in the GP. Additionally in our procedure the GP and GNN are separately trained; our early results in a follow-on study give a hint that end-to-end training of the GP and GNN in a single pass may lead to improved UQ performance of the overall model.

We note that there is definite room for improvement in the active learning strategy applied here. As demonstrated in the large errors on a family of phosphate materials, our approach may not be sampling the chemical space as effectively as possible. Improving the calibration of the UQ (as discussed above) would be beneficial in active learning. There is scope for introducing chemical and statistical heuristics in order to improve the coverage. As an example, one could sample several new materials to label per active learning cycle, but ensure that each one comes from a distinct region of the latent space by enforcing a cut-off (possibly based on the kernel length of the GP). Additionally choice of kernel for the GP may affect how well calibrated the UQ is, this is an ongoing area of research. We also note that the active learning procedure was run for a further 4,500 cycles after the process described in the section Active Learning. However the advantage of using the entropy-based sampling diminishes as the process continues because the probability of randomly choosing useful material for training approaches that of actively choosing good materials by the entropy-based method as we sample larger portions of the available additional data points (in our case 7,195 materials).

Our choice of active learning strategy (entropy-based selection), is by no means the only possible choice. When we have access to the posterior uncertainty we could also use least confidence or margin sampling, if we were dealing with a classification problem. These kinds of procedure can easily be implemented with the CFGP described in this work. Other approaches such as expected model change are also possible, where one would calculate the gradient in the loss function for a range of possible labels for un-labelled samples and select those that result in the greatest change in the model - although these approaches can be computationally expensive. We could also use the uncertainties and estimates on the unlabelled data to perform Bayesian optimisation, in which case we would seek to identify the best sample, rather than obtain the best model, such an approach (albeit based on different types of models) has recently been employed for global structure determination[47, 48] and property optimisation[49].

It is clear that the degree of advantage to be gained by applying the active learning approach will depend on the size of the non-labelled data and will likely increase as the unlabelled space increases. However, our approach already demonstrates the true potential of active learning using an example where size of the unlabelled dataset is relatively small. With the number of unexplored inorganic materials estimated in the region of 10^{11} materials[30], active learning will have an important role to play in navigating this space.

VII. CONCLUSIONS

We have presented an active learning approach based on convolution-fed Gaussian processes, which is capable of greatly enhancing the learning rate of deep neural networks for predicting the properties of materials, when compared to random sampling. Our approach uses the fact that the latent space of a graph neural network model provides a compressed representation of a 3D crystal in a space with meaningful correlations to the composition-structure-property relationship which the GNN has been trained to reproduce. We use this compressed representation to feed a Gaussian process model, which provide property estimates and uncertainty quantification (UQ) on those estimates. We demonstrate that the estimates obtained are competitive with state-of-the-art GNNs and that the UQ obtained is well calibrated and sharp. We then take advantage of this UQ to develop an active learning procedure where the training dataset is augmented on-the-fly sampling from unlabelled data based on the UQ obtained from the model. We show that this active learning procedure results in a much more rapid improvement in model performance compared to sampling the unlabelled data randomly. The methods presented in this paper will help increasing the confidence in the output of ML models by providing reliable estimates of (un)certainty. Furthermore, the active learning procedure can be extremely useful for training models in scenarios where labelled data is difficult or expensive to obtain. We hope that our methods can help to accelerate the application of machine learning for materials design.

VIII. DATASET

The dataset for the formation energy models consists of 10455 oxide materials and their DFT calculated energies from the Materials Project database[50] the exact dataset is available to download and can be used in association with our open repository of this code [51].

The dataset is initially split into three parts - 1800 samples for training MEGNET , 7,195 samples available for updating the model during active learning and 1,460 samples kept completely separate for model evaluation purposes.

The data used in this study are available from reference [52].

IX. MODELS

Our models in this paper are based on the MEGNET GNN architecture [2]. MEGNET consists of an embedding layer, followed by a flexible number of graph neural network units. The output from the graphs is then converted to a consistent sized vector using the SET2SET transformer[53]. This vector is then fed into a multi-layer perceptron feed-forward neural network to approximate the property of interest. In the CFGP, the vector is fed into a GP for function approximation.

The MEGNET model is initially trained on 1,800 randomly selected materials from the database. The model is trained for 1,000 epochs, until the validation error has equilibrated. The performance is shown in Figure **??**.

Using the MEGNET model trained as above, we build input vectors for the GP. Input vectors are obtained by passing an input to MEGNET and extracting the activations from the "Dense 32" layer. The GP uses these vectors as input. The GP we employ uses the Laplacian kernel of Equation 1. We have implemented the GP using TensorFlow Probability [54]. The hyper-parameters a and l are obtained by optimising the negative log likelihood of GP on the training dataset of 1,800 materials. This process can be rather sensitive to the initial values of the hyper-parameters, starting from 1.0 for both a and l, we obtain optimal values of 0.4281 and 2.3997 respectively.

An alternative baseline CFGP was also trained using the [3] architecture graph convolutions, rather than the MEGNET blocks. In this architecture we include a single feed-forward layer at the end of the graph convolutions to reduce the size of the output vector from 128 to 32 D - in order to mimic the featurisitaion performed using the MEGNET model.

During active learning one new material is chosen at each cycle and added to the training set for the GP (note we do not re-train the MEGNET model or the hyper-parameters of the GP at each step during the active learning). The GP is re-trained at each cycle of the active learning and the process was repeated for 6,000 cycles. At each cycle the performance of the GP on 1,460 independent samples, never used in any of the training steps is assessed.

For reference - the MEGNET model took around 5 hrs to train; the hyper-parameter optimisation of the GP took 10,000 steps and 12 hrs and the active learning for 6,000 cycles took 7 hrs, all on a NVIDIA Quadro P5000 GPU.

The code for the models, and trained model architectures are available at [51] and [52], respectively.

ACKNOWLEDGEMENTS

This work was partially supported by Wave 1 of The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1, particularly the "AI for Science" theme within that grant and The Alan Turing Institute. The ML models were trained using computing resources provided by STFC Scientific Computing Department's SCARF cluster and the PEARL cluster.

DATA ACCESS STATEMENT

All of the training data, trained neural networks and code for generating the training data for this study are openly available at `https://zenodo.org/record/4922828#.YMHksBIo-xI`.

A git repository containing the code used to build and train the neural networks, as well as notebooks to recreate them are available at `https://github.com/keeeto/gp-net`

AUTHOR CONTRIBUTIONS

KTB conceived, planned and steered the project. JA built, trained and applied the neural networks. KTB, JT and JA wrote the manuscript together. JT was involved in conception and establishing of the project and facilitated the work in the paper.

[1] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, Machine learning for molecular and materials science, Nature **559**, 547 (2018).

[2] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, Graph networks as a universal machine learning framework for molecules and crystals, Chem. Mater. **31**, 3564 (2019).

[3] T. Xie and J. C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, Phys. Rev. Lett. **120**, 145301 (2018).

[4] C. W. Park and C. Wolverton, Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery, Phys. Rev. Mater. **4**, 063801 (2020).

[5] J. Lee and R. Asahi, Transfer learning for materials informatics using crystal graph convolutional neural network, Comp. Mater. Sci. **190**, 110314 (2021).

[6] A. Raza, A. Sturluson, C. M. Simon, and X. Fern, Message passing neural networks for partial charge assignment to metal–organic frameworks, J. Phys. Chem. C **124**, 19070 (2020).

[7] A. Mallick, C. Dwivedi, B. Kailkhura, G. Joshi, and Y. Han, *Sample Efficient Uncertainty Estimation using Probabilistic Neighborhood Component Analysis*, Tech. Rep. (Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States), 2020).

[8] D. J. MacKay, A practical bayesian framework for backpropagation networks, Neural computation **4**, 448 (1992).

[9] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, Weight uncertainty in neural networks, arXiv preprint arXiv:1505.05424 (2015).

[10] Y. Gal and Z. Ghahramani, Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in *international conference on machine learning* (2016) pp. 1050–1059.

[11] I. Bilionis and N. Zabaras, Bayesian uncertainty propagation using gaussian processes, in *Handbook of Uncertainty Quantification*, edited by R. Ghanem, D. Higdon, and H. Owhadi (Springer International Publishing, Cham, 2017) pp. 555–599.

[12] N. Raimbault, A. Grisafi, M. Ceriotti, and M. Rossi, Using gaussian process regression to simulate the vibrational raman spectra of molecular crystals, New J. Phys. **21**, 105001 (2019).

[13] R. Meyer and A. W. Hauser, Geometry optimization using gaussian process regression in internal coordinate systems, J. Chem. Phys. **152**, 084112 (2020).

[14] F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, Physics-inspired structural representations for molecules and materials, arXiv preprint arXiv:2101.04673 (2021).

[15] M. Tschannen, O. Bachem, and M. Lucic, Recent advances in autoencoder-based representation learning, arXiv preprint arXiv:1812.05069 (2018).

[16] Y. Bengio, A. Courville, and P. Vincent, Representation learning: A review and new perspectives, IEEE transactions on pattern analysis and machine intelligence **35**, 1798 (2013).

[17] S. Gidaris, P. Singh, and N. Komodakis, Unsupervised representation learning by predicting image rotations, arXiv preprint arXiv:1803.07728 (2018).

[18] A. Kopf and M. Claassen, Latent representation learning in biology and translational medicine, Patterns **2**, 100198 (2021).

[19] K. Tran, W. Neiswanger, J. Yoon, Q. Zhang, E. Xing, and Z. W. Ulissi, Methods for comparing uncertainty quantifications for material property predictions, Machine Learning: Science and Technology **1**, 025006 (2020).

[20] J. P. Janet, C. Duan, T. Yang, A. Nandy, and H. J. Kulik, A quantitative uncertainty metric controls error in neural network-driven chemical discovery, Chem. Sci. **10**, 7913 (2019).

[21] G. Scalia, C. A. Grambow, B. Pernici, Y.-P. Li, and W. H. Green, Evaluating scalable uncertainty estimation methods for dnn-based molecular property prediction, arXiv preprint arXiv:1910.03127 (2019).

[22] K. Morita, D. W. Davies, K. T. Butler, and A. Walsh, Modelling the dielectric constants of crystals using machine learning, arXiv preprint arXiv:2005.05831 (2020).

[23] J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. E. Roitberg, Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning, Nat. Commun. **10**, 1 (2019).

[24] G. Hinton and S. T. Roweis, Stochastic neighbor embedding, in *NIPS*, Vol. 15 (Citeseer, 2002) pp. 833–840.

[25] A. Dunn, Q. Wang, A. Ganose, D. Dopp, and A. Jain, Benchmarking materials property prediction methods: The matbench test set and automatminer reference algorithm, arXiv preprint arXiv:2005.00707 (2020).

[26] V. Fung, J. Zhang, E. Juarez, and B. G. Sumpter, Benchmarking graph neural networks for materials chemistry, npj Computational Materials **7**, 1 (2021).

[27] V. Kuleshov, N. Fenner, and S. Ermon, Accurate uncertainties for deep learning using calibrated regression, arXiv preprint arXiv:1807.00263 (2018).

[28] D. Levi, L. Gispan, N. Giladi, and E. Fetaya, Evaluating and calibrating uncertainty prediction in regression tasks, arXiv preprint arXiv:1905.11659 (2019).

[29] Z. Del Rosario, M. Rupp, Y. Kim, E. Antono, and J. Ling, Assessing the frontier: Active learning, model accuracy, and multi-objective candidate discovery and optimization, J. Chem. Phys. **153**, 024112 (2020).

[30] D. W. Davies, K. T. Butler, A. J. Jackson, A. Morris, J. M. Frost, J. M. Skelton, and A. Walsh, Computational screening of all stoichiometric inorganic materials, Chem **1**, 617 (2016).

[31] D. W. Davies, K. T. Butler, and A. Walsh, Data-driven discovery of photoactive quaternary oxides using first-principles machine learning, Chem. Mater. **31**, 7221 (2019).

[32] C. Nyshadham, M. Rupp, B. Bekker, A. V. Shapeev, T. Mueller, C. W. Rosenbrock, G. Csányi, D. W. Wingate, and G. L. Hart, Machine-learned multi-system surrogate models for materials prediction, npj Comp. Mater. **5**, 1 (2019).

[33] N. S. Bobbitt and R. Q. Snurr, Molecular modelling and machine learning for high-throughput screening of metal-organic frameworks for hydrogen storage, Mol. Sim. **45**, 1069 (2019).

[34] Z. Li, S. Wang, W. S. Chin, L. E. Achenie, and H. Xin, High-throughput screening of bimetallic catalysts enabled by machine learning, J. Mater. Chem. A **5**, 24131 (2017).

[35] T. Lookman, P. V. Balachandran, D. Xue, and R. Yuan, Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design, npj Comp. Mater. **5**, 1 (2019).

[36] Y. Zhang *et al.*, Bayesian semi-supervised learning for uncertainty-calibrated prediction of molecular properties and active learning, Chem. Sci. **10**, 8154 (2019).

[37] E. V. Podryabinkin and A. V. Shapeev, Active learning of linearly parametrized interatomic potentials, Comp. Mater. Sci. **140**, 171 (2017).

[38] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, Less is more: Sampling chemical space with active learning, J. Chem. Phys. **148**, 241733 (2018).

[39] K. Tran and Z. W. Ulissi, Active learning across intermetallics to guide discovery of electrocatalysts for co 2 reduction and h 2 evolution, Nature Catalysis **1**, 696 (2018).

[40] G. Pilania, J. E. Gubernatis, and T. Lookman, Multi-fidelity machine learning models for accurate bandgap predictions of solids, Comp. Mater. Sci. **129**, 156 (2017).

[41] I. Petousis, D. Mrdjenovich, E. Ballouz, M. Liu, D. Winston, W. Chen, T. Graf, T. D. Schladt, K. A. Persson, and F. B. Prinz, High-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials, Scientific data **4**, 1 (2017).

[42] Y. Noda, M. Otake, and M. Nakayama, Descriptors for dielectric constants of perovskite-type oxides by materials informatics with first-principles density functional theory, Science and technology of advanced materials **21**, 92 (2020).

[43] S. A. Tawfik, O. Isayev, M. J. Spencer, and D. A. Winkler, Predicting thermal properties of crystals using machine learning, Adv. Theo. Sim. **3**, 1900208 (2020).

[44] D. W. Davies, C. N. Savory, J. M. Frost, D. O. Scanlon, B. J. Morgan, and A. Walsh, Descriptors for electron and hole charge carriers in metal oxides, J. Phys. Chem. Letters **11**, 438 (2019).

[45] V. Sharma, P. Kumar, P. Dev, and G. Pilania, Machine learning substitutional defect formation energies in abo3 perovskites, J. Appl. Phys. **128**, 034902 (2020).

[46] G. H. Gu, J. Noh, I. Kim, and Y. Jung, Machine learning for renewable energy materials, J. Mater. Chem. A **7**, 17096 (2019).

[47] M. K. Bisbo and B. Hammer, Efficient global structure optimization with a machine-learned surrogate model, Physical review letters **124**, 086102 (2020).

[48] M. S. Jørgensen, U. F. Larsen, K. W. Jacobsen, and B. Hammer, Exploration versus exploitation in global atomistic structure optimization, The Journal of Physical Chemistry A **122**, 1504 (2018).

[49] D. Bash, Y. Cai, V. Chellappan, S. L. Wong, X. Yang, P. Kumar, J. D. Tan, A. Abutaha, J. J. Cheng, Y.-F. Lim, *et al.*, Multi-fidelity high-throughput optimization of electrical conductivity in p3ht-cnt composites, Advanced Functional Materials , 2102606 (2021).

[50] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. a. Persson, The Materials Project: A materials genome approach to accelerating materials innovation, APL Materials **1**, 011002 (2013).

[51] J. Allotey and K. T. Butler, Gp-net, `https://github.com/keeeto/gp-net` (2021).

[52] J. Allotey and K. T. Butler, Data and models for: Entropy based active learning of graph neural networks for materials properties, `https://zenodo.org/record/4922828#.YMHksBIo-xI` (2021).

[53] O. Vinyals, S. Bengio, and M. Kudlur, Order matters: Sequence to sequence for sets, arXiv preprint arXiv:1511.06391 (2015).

[54] J. V. Dillon, I. Langmore, D. Tran, E. Brevdo, S. Vasudevan, D. Moore, B. Patton, A. Alemi, M. Hoffman, and R. A. Saurous, Tensorflow distributions, arXiv preprint arXiv:1711.10604 (2017).