Position Paper

Towards green machine learning: challenges, opportunities, and developments

Witold Pedrycz

Department of Electrical and Computer Engineering, University of Alberta, Edmonton T6G 2R3, Canada.

Correspondence to: Prof. Witold Pedrycz, Department of Electrical and Computer Engineering, University of Alberta, 116 St. and 85 Ave., Edmonton T6G 2R3, Canada. E-mail: wpedrycz@ualberta.ca

Abstract

Machine Learning has assumed a prominent position in the plethora of design and analysis of intelligent systems. Learning is the holy grail of Machine Learning, and with rapidly growing complexity and the size of the constructed networks (the trend which is profoundly visible in deep learning architectures), the overwhelming computing is staggering. The return on investment clearly diminishes: even a very limited improvement in performance (commonly expressed as a classification rate or prediction error) does call for intensive computing because of learning a large number of parameters. The recent developments in green Artificial Intelligence (or better to say, green Machine Learning) has identified and emphasized a genuine need for a holistic multicriteria assessment of the design practices of Machine Learning architectures by involving computing overhead, interpretability, robustness, and identifying sound trade-offs present in these problems. We discuss a realization of green Machine Learning and advocate how Granular Computing contributes to the augmentation of the existing technology. In particular, some paradigms that exhibit a sound potential to support the sustainability of Machine Learning such as federated learning and transfer learning, are identified, critically evaluated, and cast into some general perspective.

Keywords: Green machine learning, granular computing, sustainable computing, federated learning, transfer learning, knowledge distillation

INTRODUCTORY COMMENTS

Recently, the technology of Machine Learning (ML) has enjoyed unprecedented growth and high visibility. In particular, ML is visible through a large number of paradigms of learning and underlying architectures and a truly remarkable wealth of applications encountered in numerous areas such as computer vision, natural language processing (NLP), and computer games, among others applications. There are two dominant features that are omnipresent across the general ML area:

(i) enormous computing overhead. It comes as a result of exploring advanced architectures of deep learning structures and intensive hyperparameter adjustments. The required computing burden calls for intensive optimization schemes and high-performance computing environments. The numbers are startling: an overall increase in computing effort was 300,000 times during 2012-2019, with training costs doubling every few months[1]. The most visible trend is encountered in natural language processing (NLP word embedding approaches, namely, ELMO-BERT, openGPT-2, and XLNet, among others).

(ii) a dominant quest for the higher performance of ML models (such as a classification rate or a low prediction error being commonly studied) or the lowest loss function being regarded as the predominant objective to be met.
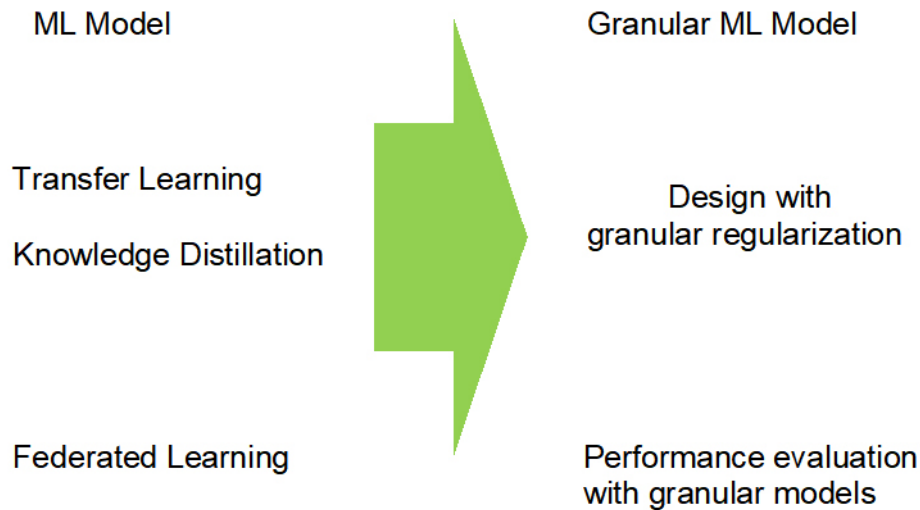
Few years ago, the area was already cast in a broader and holistic perspective giving rise to Green AI[1], or better name, green ML. The ultimate goal of green ML is to raise awareness about the need for a comprehensive design and efficiency of evaluation of ML constructs by identifying possible sound and meaningful trade-offs among an array of carefully selected objectives. In contrast to so-called red ML guided by a single performance index, green ML stresses a need for a multi-dimensional and holistic way of assessment of the quality and usage of the ML architectures. This, in turn, helps construct a better learning paradigm and understand trade-offs among crucial criteria and anticipate some sound design practices leading to environmental sustainability. An interesting development within the setting of green ML is reported in[2], where the role of Gaussian processes was identified.

In this position note, we elaborate on the agenda of green ML by focusing on a set of well-thought-out efficiency criteria and linking them to the associated characteristics of ML architectures. Several existing directions of ML of particular interest to support green ML are discussed by highlighting their essential aspects. It is also advocated that an innovative way of realizing the objectives of green ML is accomplished by engaging the conceptual ad algorithmic framework of Granular Computing. The concept of information granule being viewed as a fundamental vehicle to support abstraction to move to a higher level of generalization is emphasized. We argue and provide a number of compelling reasons behind the statement that information granules play a pivotal role in the augmentation of federated learning and knowledge transfer as well as knowledge distillation. All these learning paradigms, put in certain perspective, offer support to realize the focal objectives of green ML in one way or another.

A bird's eye view of these augmented pursuits to be discussed in the sequel is displayed in Figure 1.

As a preamble of this study, we advocate that Granular Computing becomes instrumental to:

(i)facilitate and augment learning mechanisms of federated learning, transfer learning, and knowledge distillation.

ML Model   Granular ML Model

Transfer Learning

Knowledge Distillation   Design with
                                            granular regularization

Federated Learning   Performance evaluation
                                            with granular models

Figure 1. An overall agenda of green ML: from developments in ML to green ML constructs augmented by concepts and algorithms of Granular Computing.

(ii)deliver sound and practically viable evaluation mechanisms of assessment of the quality of the ML models so that a sound trade-off among many conflicting criteria could be accomplished.

While the study does not delve into technical details, we highlight the main concepts and point at the general algorithmic developments so that an overall picture becomes painted completely and convincingly.

In Section 2, we devote our discussion to the key objectives identified and stressed in green ML. Some interesting paradigms of transfer learning and knowledge distillation, whose emergence is also vital to the realization of the focal points of green ML are elaborated in Section 3. Section 4 covers the key ideas of federated learning. Information granules and an environment of Granular Computing are discussed in Section 5 followed by their direct usage in the design of granular models. We show how the quantification of the quality of results of ML arises in the form of granular constructs, say predictors. In the sequel, the discussed federated learning and transfer learning are expanded with the incorporation of information granules either as a vehicle to quantify the relevance of the ML model or to facilitate the learning by optimizing a certain loss function augmented by the component of granular regularization (Section 7 and 8).

The discussions are presented at a certain level of generality to focus on the essence of the underlying principles; however, at the same time, a level of formal rigor is retained. This could help the reader to explore some related technical details encountered in the existing literature.

GREEN ML: A NEW TREND IN THE ML AND AI

The term Green AI (Green ML) has emerged in recent years as a tangible and timely manifestation of a growing awareness of the enormous and rapidly increasing number of computations, especially those encountered in the realm of deep learning architectures[1]. Deep learning assumed a central position in the design practices of commonly encountered classifiers, and predictors envisioned in challenging tasks of computer vision and NLP processing. The dominant trend is witnessed today: more than ever before, nowadays, accuracy is a holy grail of any classifier. Any improvement in classification rate is reached by

developing more advanced and structurally complex models. This comes hand in hand with a large number of parameters to be estimated. The accuracy is paid by involving a steep increase in computing costs. The straightforward observation holds: computational power buys stronger results. Model performance and model complexity are strongly linked: exponentially larger models lead to a linear gain in performance[1]. Notably, the return on investment starts to sharply diminish: a struggle for an improvement in a fraction of the percentage of accuracy comes with a very steep price in terms of the training and validation time along with an intensive effort going into the validation process and time-consuming tweaking of the hyperparameters. It is also needless to say that with the complexity of the model, the amount of required data is also growing at a rapid pace. To put all efforts in a general perspective, it is worth stressing that no matter which advanced design strategy is pursued and which architecture has been selected, there is a bound on the classification performance implied by the Bayesian classifier in pointing out the accuracy limit one can theoretically achieve.

The essence of Green AI is on the efficiency of a slew of evaluation criteria but taking into consideration economic, environmental, and social costs and subsequently assessing the relevancy of the overall design within such a multicriteria decision-making process. To achieve efficiency, attention needs to be paid to carbon emission, which comes with electricity usage (carbon footprint), elapsed real time, and floating point operations. These two determine the amount of time required to design the ML architecture. It also associates with the carbon footprint, as clearly demonstrated in[3-5].

There are several other crucial factors that are contributors to the efficiency of ML development processes. The key question is: could a slight increase in the classification error be acceptable? There is no universally acceptable answer to this question, as there are several factors that are instrumental and have to be taken into account:

-reduction of computing overhead associated with the increased accuracy. A general visible tendency is that increasing the already high level of accuracy requires a substantial increase in the ML size and a significant increase in the learning time.

-the interpretability requirement of ML architectures. The recent developments in the area of explainable AI(XAI)[6,7] call for ML models that are transparent and comprehended by the end user. In particular, XAI delivers critical abilities to provide "what if" analysis and deliver a sound explanation mechanism behind the result produced by the ML model. A slight reduction in accuracy could not lead to any deterioration of interpretability. The term explainable system is not precisely defined. It is human-centric, so it depends on the end-user and her preferences. For instance, from the perspective of ML, linear regression models come with some degree of explainability; however, the formula, albeit simple, might not be comprehended. In this sense, a number of XAI methodologies, approaches, and detailed constructs in some cases could offer some quasi-explanation mechanisms.

-instead, a desirable feature of the ML model is its robustness. One should avoid situations where a slight change in the data yields a large departure from the previous findings. This effect speaks loudly to the volatility and brittleness of the findings (model), and this is a phenomenon that has to be avoided or at least limited.

-level of abstraction of results is inherently associated with interpretability. In the plethora of ML architectures and paradigms, there are several pursuits whose usefulness can be identified to support the agenda of green ML. Those of interest fall under the umbrella of transfer learning and knowledge distillation. Also of interest is the paradigm of federated learning.

Granular Computing and information granules contributing to the agenda of green ML exhibit a tangible potential in several key ways: (i) supporting a certain level of information granularity (abstraction) that is central to facilitating the requests for interpretability and explainability; (ii) helping to quantify the credibility of ML constructs (and in this way also contributing to the formation of a sound trade-off between accuracy and computing overhead as well as supporting the quantification of the brittleness of the models); and (iii) facilitating the realization of privacy by hiding data and communicate them at the higher level of abstraction (lower level of information granularity).

TRANSFER LEARNING AND KNOWLEDGE DISTILLATION

Transfer learning has been an intensively studied paradigm of learning[8,9]. The essence of this learning is to extract previously acquired knowledge (model) and apply it to a new environment. In the context of green ML, this pursuit is about cutting down on computing expenses via reuse and a prudent adaptation of the already acquired knowledge. The commonly used terminology here relates to the source domain D_s in which the knowledge has been acquired and the target domain D_t in which the already acquired knowledge (model) is used. Along with the domains associated are the corresponding tasks, T_s and T_d, which denoteclassification or prediction, respectively. In sum, the structures involved in transfer learning are pairs (D_s, T_s) and (D_t, T_s). More formally, we deal with the domains $D_s = \{F_s, P()\}$ and $D_t = \{F_t, P()\}$ (where P is a probability function expressed on F_s or F_t) being described by the features (attributes) and associated tasks $T_s = \{Y_s, f_s(.)\}$ and $T_t = \{Y_t, f_t(.)\}$ where the reuse needs are to be carefully arranged as we have $D_s \neq D_t$ and /or $T_s \neq T_t$

There are several highly motivating and convincing factors behind the use of transfer learning, especially (i) the existence of small data available in the target domain (which does not support designing advanced models equipped with a large number of parameters); (ii) enhancements of robustness of the ML model; and (iii) elimination of cold start problem, in particular, visible in recommender systems. Some detailed considerations are covered in[10,11]. Transfer learning is also referred to as knowledge reuse, learning by analogy, domain adaptation, and pre-training, among others, and as such, it has been in existence for a long time. Conceptually, this type of learning can be illustrated in Figure 2. The challenge is how to transfer knowledge efficiently as the domains and tasks differ and how their "distance" (or resemblance) has to be accommodated in the transfer mechanism; say, how much the already accumulated knowledge can be transferred and in which way one can quantify the departure from the source to the target domain.

Knowledge distillation can be regarded as a certain variant of transfer learning; however, the key motivating factor comes with the nature of the computing environment. In the source domain of the design, there is a highly advanced ML model (say, one composed of thousands of decision trees) where the computing expenditures are huge. This model is referred to as a teacher. Deploying the model to a new environment (in which a handheld device is available) is not feasible and a compact, more economical version of the teacher model is required, usually referred to as a student model. Knowledge distillation also comes under the name of model compression, and as such, it can be thought of as a meaningful way to support green ML.

FEDERATED LEARNING - COPING WITH DATA ISLANDS AND PRIVACY ISSUES

Federated learning is a paradigm of learning carried out in the presence of data that cannot be shared and are distributed across a number of local devices (data islands)[12-16]. This challenges the generic scheme of learning and calls for a paradigm shift. The essence is to build a model at the level of a server, Figure 3,
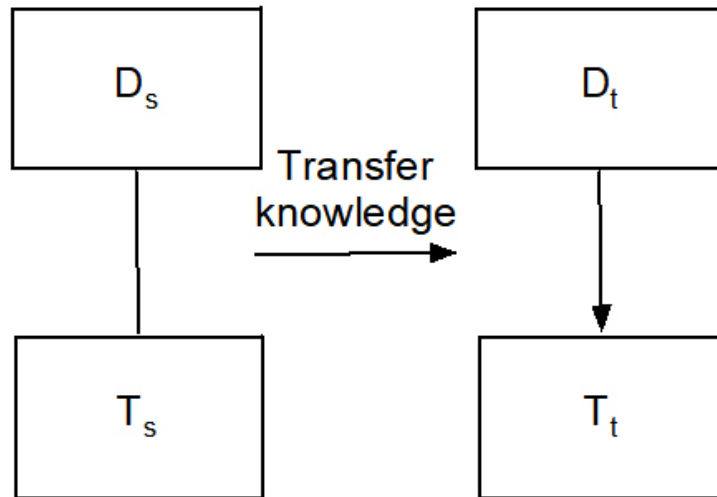
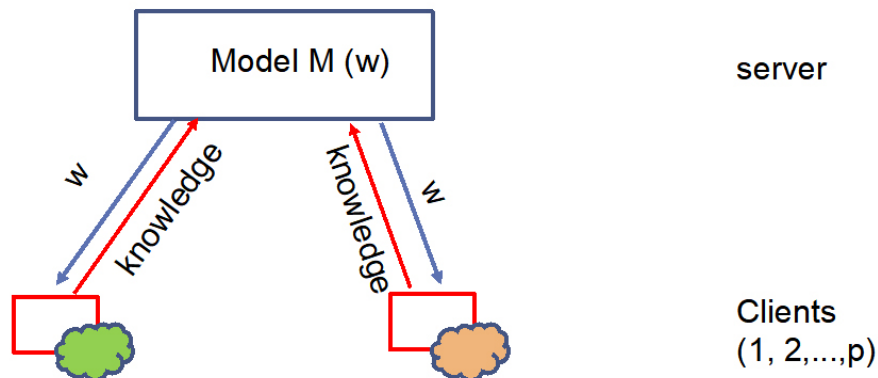**Figure 2.** From source to target domain: realizing transfer knowledge.

Figure 3. Federated learning: building a holistic model in the presence of data islands.

where knowledge rather than data is used to update the model (its parameters). Federated learning addresses a number of challenges: (i) meeting requirements of privacy and security; (ii) coping with unreliable and limited communication links and legal requirements. A number of applications include education, healthcare, the Internet of Things (IoT), and smart transportation.

The two dominant modes of federated learning are driven by the format of knowledge being supplied to the server. They involve averaged federated learning (each client provides parameters of its own model) and gradient-based federated learning (where each client delivers the gradient of the loss function computed for the weights provided by the server), with the generic expressions coming in the following form

-averaged federated learning

$$w = \frac{1}{p} \sum_{i=1}^{p} w_i \tag{1}$$

-gradient-based federated learning

$$w(iter+1)=w(iter)-\alpha\sum_{ii=1}^{p} g[ii] \tag{2}$$

Where *iter* stands for the iteration index of the learning of the weights. The parameters of the model (weights of the neural network, coefficients of the regression model, etc.) are denoted as *w* while the gradient of the minimized loss function provided by the ii^{th} client is given as $g[ii]$. Likewise, w_1, w_2, ..., w_p are the parameters of the models produced by the individual clients. There have been numerous studies concerning ways of arranging efficient communication among the clients and the server to achieve efficient update learning schemes and allow for fair, non-dominated communication between the parties. The key challenge arising at the level of the server is to take into consideration the diversified nature of data (as coming from numerous clients) and assess the performance of the constructed ML model; this quest impacts a way of communication with the individual client and helps quantify its contribution in the updates of the model (for instance, through the successive modifications of the gradient of the loss function provided by the individual client). A prudent assessment of the performance of the model built by the server and monitoring of its changes, and guiding contributions of the clients become essential to the successful realization of federated learning.

INFORMATION GRANULES AND GRANULAR COMPUTING

The concept of information granules and their formal settings, along with the computing, gives rise to the area of Granular Computing[17-20].

Information granules are regarded as pieces of knowledge resulting in an abstraction of data, exhibiting well-defined semantics and forming functional modules in further interpretable system modeling (granular models). Information granules are intuitively appealing constructs. A flood of experimental evidence (data) is structured in the form of some meaningful, semantically sound entities.

Generally, by information granules, one regards a collection of elements drawn together by their closeness (resemblance, proximity, functionality, etc.) and articulated in terms of some useful spatial, temporal, or functional relationships.

The essential items delivered by information granules support various levels of abstraction by hiding unnecessary details and supporting interpretability and explainability (as noted in the context of XAI). The level of abstraction also supports the stability of findings in XAI.

Granular Computing is a knowledge-based environment supporting the design and processing of information granules. Granular Computing is about representing, constructing, processing, and communicating information granules. Information granules are formalized as sets, fuzzy sets, rough sets, and probabilities, just to point at several among available alternatives.

In sum, the three requirements that facilitate the realization of green ML involve the sequence of linkages:

Information granules - level of abstraction - computing overhead

GRANULAR MODELS

No matter how complex they could be, numeric models constructed on the basis of numeric data cannot be ideal, viz., the optimized loss function cannot achieve zero. This implies that no matter how much costs are associated with the design and deployment of the ML model, there is limited credibility one has to associate with the results of ML models. This problem is of particular importance when it comes to the use of ML models in prediction or classification tasks. How much confidence/trust could we associate with the obtained numeric results? How to quantify their credibility? We advocate that the quality of the result can be achieved by admitting that the result is an information granule rather than a single numeric outcome. In this way, the role of granular models built over the already constructed mode becomes of genuine interest.

In a succinct way, the overall process of proceeding with the designed model to its granular counterpart can be carried out as follows. First, a numeric model expressed as $y = M(x; a)$ is designed in a supervised mode on the basis of pairs of input-output data $(x_k, target_k)$, $k=1,2, ..., N$. Here x stands for the vector of input variables, a denotes a vector of parameters of the model to be estimated, and the target is the output data.

In the sequel, its granular augmentation $M \xrightarrow{G} G(M)$ is formed where the aspect of information granularity is accommodated across the parameters of the model (which leads to granular output Y)

$$y=M(x;a) \xrightarrow{G} Y=M(x;G(a))= M(x; A) \tag{3}$$

or the granular output is being formed

$$M \xrightarrow{G} Y =G(M(x;a)) \tag{4}$$

The key role is being played by the adjustable level of information granularity ε which elevates the numeric parameter of the model to its granular (interval-valued) counterpart. Two transformations involving interval information granules are sought with this regard[5,21]

$$a \xrightarrow{\epsilon} [\min(a_i(1+\varepsilon),a_i(1-\varepsilon)), \max(a_i(1+\varepsilon),a_i(1-\varepsilon))], \varepsilon \in [0,1] \tag{5}$$
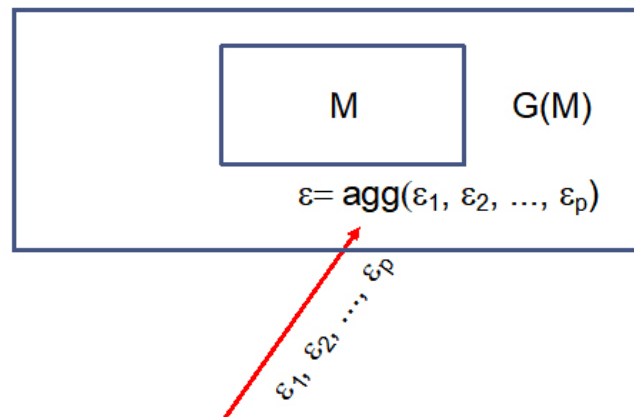$$a \xrightarrow{\epsilon} [\min(a_i(1+\varepsilon),a_i/(1+\varepsilon)), \max(a_i(1+\varepsilon),a_i/(1-\varepsilon))], \varepsilon \geq 0 \tag{6}$$

The value of the level of information granularity ε is determined through the optimization of the granular results confronted with the numeric data. In the optimization, a product of coverage *cov* and specifity *sp* (which are essential descriptors of information granules). The optimization yields a certain compromise between these two conflicting criteria of coverage and specificity. With the increase of the values of ε, the coverage increases but results in lower values of specificity.

GRANULAR FEDERATED LEARNING

In light of the *non-i.i.d* nature of data (one cannot claim in federated learning that the data islands satisfy the assumption of independent and identically distributed random variables) supplied by individual clients,

Figure 4. Building a granular model in federated learning through an aggregation of levels of information granularity.

it is unlikely the model constructed by the server fits all of them. The careful quantification of the resulting model is essential in many ways. In particular, it helps monitor contributions provided by the individual clients (e.g., by weighting the gradient of the loss function) and finally provides a certain figure of merit to the constructed model. The quality is assessed by elevating the numeric model to its counterpart, where the level of information granularity is viewed as a measure of the performance of the model, which directly manifests in the granular character of the generated results.

The model M designed in the federated learning is confronted with local data D_{ii} located at ii^{th} client. This results in its *granular* counterpart $G(M)|_{Dii}$. Such a granular model is designed as outlined in Section 6. In the same way, we proceed with all remaining data islands, obtaining the granular characterizations of the model $G(M)|_{D_1} G(M)|_{D_2} .... G(M)|_{Dp}$, where p stands for the number of data islands.

Note that $G(M)|_{Dii}$ is characterized by the level of information granularity ε_{ii}. This level is the result of the maximization of the product of coverage and specificity, namely $\varepsilon_{ii} = \arg \max(cov*sp)$. The corresponding levels of information granularity $\varepsilon_1, \varepsilon_2, ..., \varepsilon_p$ are aggregated, giving rise to some value of ε, which is used to form a granular model $G(M)$ and deliver granular results. Various aggregation schemes *agg* could be sought, namely logic-based operators, a family of generalized averaging operators, among other alternatives.

The levels of information granularity could also be involved in the successive steps of learning by impacting contributions from the clients. For instance, the gradient-based scheme of learning (2) is modified as follows

$$w(iter+1)=w(iter)-\alpha\sum_{ii=1}^{p} g(\epsilon_{ii})\gamma[ii] \qquad (7)$$

Here $\gamma(\varepsilon_{ii})$ is a function of the level of information granularity impacting the contribution of the corresponding gradient to the calculations of the overall gradient, as shown in Figure 4.
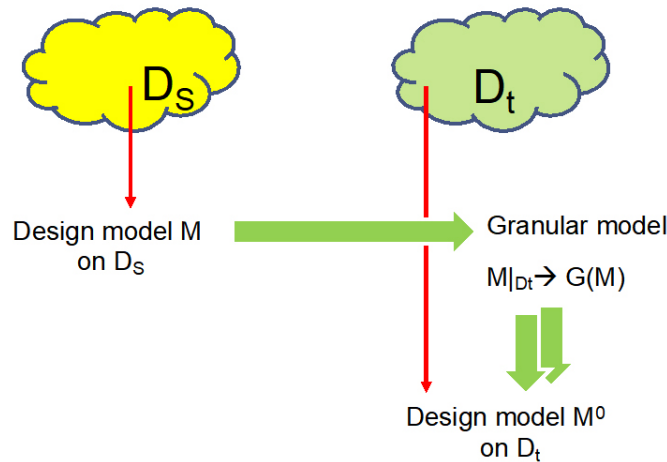
Figure 5. Design of model in *Dt* through granular regularization.

KNOWLEDGE TRANSFER AND KNOWLEDGE DISTILLATION WITH MECHANISMS OF GRANULAR COMPUTING

Transferring knowledge from the source to the target domain implies inherently a granular nature of the results of the model when it operates in the target domain. The level of information granularity entails a quantification of the results produced in the modified domain. As shown in Figure 5, two types of approaches are sought:

(i)passive approach. The model constructed in the source domain is transferred to D_t without any changes, but the results are elevated to the granular level. The level of information granularity serves as a measure of usefulness of the model in the new environment.

(ii)active approach. A new model $M^0(x; w)$ is constructed in D_t, and in its design, one uses a loss function whose important part comes in the form of the guidance provided by the granular model $G(M)$ transferred to D_t.

The design of the model M^0 is driven by the augmented loss function Q. Its essential component is a so-called granular regularization that guides the construction of the model. The loss function is expressed as follows

$$Q = \sum_{D_t} \left\| target_k - M^0(x_k, w) \right\| + \alpha \sum_{D_t} [1 - cov(M^0(x_k, w), G(M(x_k, w)))] * sp(G(M(x_k, w))) \tag{8}$$

where α is a weight coefficient emphasizing a contribution of the granular model $G(M)$. This second term of Q plays an essential role in supporting the reuse of knowledge already acquired for D_s. The value of the hyperparameter supplies a certain level of flexibility in the construction of the model. In the sequel, the parameters of M^0 are then updated by following a standard gradient-based iterative scheme minimizing Q in the form

$$w = w - \beta \nabla_w Q \tag{9}$$

Where β is the learning rate of the learning scheme. As data in D_t could be far smaller than those in D_s as well as M^o is more concise than M, the reuse of knowledge conveyed by M supports the agenda of green ML both in terms of the reduced computing overhead and enhanced interpretability (note that M^o is far simpler than M, so it is easier to comprehend and interpret).

Knowledge distillation is one of the interesting realizations of transfer learning being motivated by real-world scenarios[22]. While complex models come with high performance (say, low classification error), they require enormous computing and memory resources. Not only is the design highly demanding, but such models (namely, a model composed of thousands of decision trees) can only be deployed in an environment equipped with enough computing capabilities to realize any prediction or classification tasks. Considering that the environment one intends to deploy the designed model is very limited such as a typical handheld device, a new model has to be designed, and in its development, one can be guided by the complex mode. The underlying development process is referred to as knowledge distillation. In fact, it is an example of knowledge transfer. The terminology commonly used refers to the originally constructed model M as a teacher model, whereas the new model is a student model M^o and an overall process is also known as model compression. The loss function is the same as given by (8).

CONCLUSION

Undoubtedly, *red* ML has provided a large number of advanced solutions to complex problems of computer vision and NLP and thus plays an important role in the delivery of solutions to complex real-world problems. Green ML triggers another important wave of thoughtful considerations with an ultimate pursuit of the agenda of intelligent systems to build smart environments and green computing. The objectives of green ML stress the importance of the holistic perception of the problem and the delivery of successful solutions. While there are several interesting and promising developments, some of which are outlined here, there is still a long way to go and new exciting developments of the existing paradigms and challenging problems require innovative approaches to problem-solving. One promising alternative is to posit the problem in the framework of multi-objective group decision-making considerations where various, quite often conflicting views could be reconciled when building efficient solutions.

In the general context of advanced modeling, as it becomes predominantly visible in ML, in its design, it is worth keeping in mind that ...*just as the ability to devise simple but evocative models is the signature of great scientists, so over-elaboration and overparameterization is often the mark of mediocrity*[23] .

DECLARATIONS
Author's contributions
The author contributed solely to the article.

Availability of data and materials
Not applicable.

Financial support and sponsorship
None.

Conflicts of interest
Not applicable.

Ethical approval and consent to participate
Not applicable

Consent for publication
Not applicable.

Copyright
© The Author(s) 2022.

REFERENCES

1.   Schwartz R, Dodge J, Smith NA, Etzioni O. Green AI. *Commun ACM* 2020;63:54-63. DOI
2.   Candelieri A, Perego R, Archetti F. Green machine learning via augmented Gaussian processes and multi-information source optimization. *Soft Comput* 2021;25:12591-603. DOI
3.   Castanyer RC, Martínez-Fernández S, Franch X. Which design decisions in AI-enabled mobile applications contribute to greener AI? Available from: https://arxiv.org/abs/2109.15284 [Last accessed on 30 Dec 2022].
4.   Tornede T, Tornede A, Hanselle J, Wever M, Mohr F, Hüllermeier E. Towards green automated machine learning: status quo and future directions. Available from: https://arxiv.org/abs/2111.05850 [Last accessed on 30 Dec 2022].
5.   Vale Z, Gomes L, Ramos D, Faria P. Green computing: a realistic evaluation of energy consumption for building load forecasting computation. *J Smart Environ Green Comput* 2022;2:34-45. DOI
6.   Barredo Arrieta A, Díaz-rodríguez N, Del Ser J, et al. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion* 2020;58:82-115. DOI
7.   der Waa J, Nieuwburg E, Cremers A, Neerincx M. Evaluating XAI: a comparison of rule-based and example-based explanations. *Artif Intell* 2021;291:103404. DOI
8.   Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010;22:1345-59. DOI
9.   Pan SJ, Zheng VW, Yang Q, Hu DH, Transfer learning for wifi-based indoor localization. Available from: https://www.aaai.org/Papers/Workshops/2008/WS-08-13/WS08-13-008.pdf [Last accessed on 30 Dec 2022].
10. Lu J, Zuo H, Zhang G. Fuzzy Multiple-Source Transfer Learning. *IEEE Trans Fuzzy Syst* 2020;28:3418-31. DOI
11. Zuo H, Zhang G, Pedrycz W, Behbood V, Lu J. Fuzzy regression transfer learning in takagi-sugeno fuzzy models. *IEEE Trans Fuzzy Syst* 2017;25:1795-807. DOI
12. Abdulrahman S, Tout H, Ould-slimane H, Mourad A, Talhi C, Guizani M. A Survey on federated learning: the journey from centralized to distributed on-site learning and beyond. *IEEE Internet Things J* 2021;8:5476-97. DOI
13. Hu X, Shen Y, Pedrycz W, Wang X, Gacek A, Liu B. Identification of fuzzy rule-based models with collaborative fuzzy clustering. *IEEE Trans Cybern* 2022;52:6406-19. DOI PubMed
14. Kairouz P, Mcmahan HB, Avent B, et al. Advances and open problems in federated learning. *FNT Mach Learn* 2021;14:1-210. DOI
15. Yang Q, Liu Y, Cheng Y, Kang Y, Chen T, Yu H. Federated learning. *Synth Lect Artif Intell Mach Learn* 2019;13:1-207. DOI
16. Federated learning. In: Yang Q, Fan L, Yu H, editors. Privacy and incentive, Springer; 2020.
17. Pedrycz W. Granular computing: analysis and design of intelligent systems. Boca Raton: CRC press; 2013. DOI
18. Pedrycz W. Granular computing for data analytics: a manifesto of human-centric computing. *IEEE/CAA J Autom Sinica* 2018;5:1025-34. DOI
19. Pedrycz W. An introduction to computing with fuzzy sets. IEEE ASSP Magazine; 2021. DOI
20. Zadeh LA. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst* 1997;90:111-27. DOI
21. Moore RE. Introduction to interval computations (Götz Alefeld and Jürgen Herzberger). *SIAM Rev* 1985;27:296-7. DOI
22. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531* 2015:2. DOI
23. Box GEP. Science and statistics. *J Am Stat Assoc* 1976;71:791-9.