RESEARCH

Combining graph neural networks and transformers for few-shot nuclear receptor binding activity prediction

Luis H. M. Torres[1*], Joel P. Arrais[1] and Bernardete Ribeiro[1]

Abstract

Nuclear receptors (NRs) play a crucial role as biological targets in drug discovery. However, determining which compounds can act as endocrine disruptors and modulate the function of NRs with a reduced amount of candidate drugs is a challenging task. Moreover, the computational methods for NR-binding activity prediction mostly focus on a single receptor at a time, which may limit their effectiveness. Hence, the transfer of learned knowledge among multiple NRs can improve the performance of molecular predictors and lead to the development of more effective drugs. In this research, we integrate graph neural networks (GNNs) and Transformers to introduce a few-shot GNN-Transformer, Meta-GTNRP to predict the binding activity of compounds using the combined information of different NRs and identify potential NR-modulators with limited data. The Meta-GTNRP model captures the local information in graph-structured data and preserves the global-semantic structure of molecular graph embeddings for NR-binding activity prediction. Furthermore, a few-shot meta-learning approach is proposed to optimize model parameters for different NR-binding tasks and leverage the complementarity among multiple NR-specific tasks to predict binding activity of compounds for each NR with just a few labeled molecules. Experiments with a compound database containing annotations on the binding activity for 11 NRs shows that Meta-GTNRP outperforms other graph-based approaches. The data and code are available at: https://github.com/ltorres97/Meta-GTNRP.

Scientific contribution

The proposed few-shot GNN-Transformer model, Meta-GTNRP captures the local structure of molecular graphs and preserves the global-semantic information of graph embeddings to predict the NR-binding activity of compounds with limited available data; A few-shot meta-learning framework adapts model parameters across NR-specific tasks for different NRs in a joint learning procedure to predict the binding activity of compounds for each NR with just a few labeled molecules in highly imbalanced data scenarios; Meta-GTNRP is a data-efficient approach that combines the strengths of GNNs and Transformers to predict the NR-binding properties of compounds through an optimized meta-learning procedure and deliver robust results valuable to identify potential NR-based drug candidates.

Keywords Graph Neural Network, Transformer, Few-shot Learning, Meta-Learning, Nuclear Receptor Binding Activity Prediction, Drug Discovery

*Correspondence:
Luis H. M. Torres
luistorres@dei.uc.pt
[1] Department of Informatics Engineering, Univ Coimbra, Centre for Informatics and Systems of the University of Coimbra, Coimbra 3030-790, Portugal

Introduction

Nuclear receptors (NRs) are a family of transcription factors that play a crucial role in regulating various biological processes including cell growth, development, and metabolism [1, 2]. Their biochemical significance has promoted to a great deal of research in the fields of

Torres *et al. Journal of Cheminformatics*   (2024) 16:109

Page 2 of 19

toxicology and medicinal chemistry, with many drug discovery projects using machine learning (ML) to select compounds for the development of NR-based drugs. Nonetheless, discovering novel NR-modulators with high affinity and specificity is difficult due to structural similarities and shared domains among multiple NRs [3–5].

Ongoing research in computational toxicology is focused on developing in silico methods to modulate the activity for a group of NRs or the selectivity among specific NRs. Several ML models of chemical activity against multiple NRs have begun to emerge to predict various NR-modulators with the potential to target various diseases [6]. However, there is limited research on the impact of using distinct types of NRs to design quantitative structure-activity relationship (QSAR) models for a given target receptor. In addition, current computational approaches are centered around a single receptor, and there hasn't been an attempt to transfer the learned knowledge across multiple NRs [7].

One potential approach for NR-modulation with QSAR is multi-task learning. In the case of NRs, multi-task learning can be applied to train a model across multiple related tasks and evaluate the NR-activity with different ligands, or to infer the effects of NR-ligands in different tissues [8, 9]. In drug discovery, where high-quality labeled information is limited, meta-learning is particularly useful as it allows to learn across few-shot tasks for different molecular properties and improve generalization with few labeled compounds [10–14]. Hence, multi-task meta-learning can improve the accuracy of QSAR models to predict the activity of compounds on specific biological targets with limited data [15–19].

Compounds can be represented using molecule graphs, with nodes representing the atoms, and edges describing the chemical bonds [20]. Graph neural networks (GNNs) update node-edge embeddings in graph-structured data with neighborhood aggregation to output a graph-level embedding useful for molecular property discovery [21, 22]. However, standard GNNs only aggregate local dependencies and are incapable of capturing the broader aspects of node-edge connections significant for compound classification. Alternatively, Transformers have been developed to tackle this issue by learning long-range dependencies while maintaining the global structure of molecular embeddings. These models attend to multiple positions to preserve global-semantic information in molecule embeddings and generalize for different molecular properties [23, 24]. Vision Transformers (ViT) extend standard Transformer attention to propagate sequences of visual tokens and obtain improved performances on image classification tasks [25]. Recent advancements in ViT approaches have derived multiple hybrid architectures that combine them with different neural network models [26]. Nevertheless, the potential of ViT networks is still to be revealed in molecule representation learning when inferring the NR-binding activity of chemical compounds for NR-based drug discovery.

To address this challenge, a novel few-shot GNN-Transformer, Meta-GTNRP is introduced for NR-binding activity prediction using limited labeled compounds. The proposed approach considers compounds as graph-structured data encoding the local-to-global context of molecule structures for NR-binding activity prediction. In addition, a meta-learning approach is proposed to optimize model parameters in multiple few-shot tasks and predict their specific NR-binding properties with limited data. In this research, we make use of a Nuclear Receptor Activity (NURA) database [27] to describe the experimentally-derived binding, agonist and antagonist activities against various human NRs. Multiple experiments with NR-activity data demonstrate that Meta-GTNRP achieves an improved performance on NR-activity tasks over the conventional graph-based approaches.

Related work
Few-shot learning for NR-binding activity prediction
Few-shot learning (FSL) is a meta-learning approach that focuses on generalizing with reduced amounts of supervised information. FSL has found recent applications in compound discovery by predicting clinically-relevant properties using limited high quality data. Here, the goal of FSL is to adapt model parameters for different molecular tasks (meta-training) and use them to predict important molecular property tasks using small sets of labeled compounds (meta-testing) [28, 29]. FSL methods can be classified in two categories: metric-based models [30] learn a distance metric that captures the relationship between task-specific support sets and separate query sets, enabling effective transfer of knowledge for different few-shot tasks. On the other hand, optimization-based methods [31] adapt model parameters within different tasks represented by task-specific support sets and generalize for novel representations in separate query sets using few gradient steps. In this research, we introduce an optimization-based meta-learning approach to learn across different NR-tasks and generalize to new NR-binding meta-testing tasks. In meta-training, few-shot models are trained using NR-specific support sets to adapt model parameters for different NR-specific tasks by computing gradients and losses in disjoint query sets of molecules. In meta-testing, these parameters are used to infer the NR-binding properties of compounds for new NR tasks using limited data.

Torres *et al. Journal of Cheminformatics*   (2024) 16:109

Page 3 of 19

Graph representation learning

Representing molecules using graph-structured data can more accurately depict the relationships among atoms important to predict NR-binding properties [32]. GNNs work with molecule graphs to encode molecules and use a set of nodes to represent the atoms and a set of edges to describe chemical bonds between atoms. Through a message-passing approach, GNNs aggregate node-edge information to compute molecule graph embeddings, capturing the molecule's overall structure in a multi-dimensional graph embedding space. More specifically, graph convolutional networks (GCNs) [33] incorporate a convolutional operation which aggregates local information and updates node-edge embeddings, similar to convolutional filters used in convolutional layers. An alternative technique, called GraphSAGE [34], utilizes a node-centric inductive training approach to learn the node embeddings in large molecule graph structures for unseen graph features. In addition, graph isomorphism networks (GIN) [35] are powerful GNN frameworks that extend the Weisfeiler-Lehman (WL) isomorphism test, demonstrating impressive results on different downstream applications. In a pioneering study, Hu et al. [36] have pre-trained GNNs to learn local information and obtain improved performances across various chemical property tasks. Based on this method, Guo et al. [37] proposed a novel meta-learning approach that allows GNNs to fast adapt across tasks using task-specific weights to meet self-supervised objectives in molecular property discovery.

Transformer networks

Transformer networks, introduced by Vaswani et al. [38], are natural language processing (NLP) models that leverage self-attention to learn from sequential data and retain its global structure. The attention mechanism can be complemented using feed-forward (FF) layers, making it a commonly used method for various NLP tasks. Vision Transformers (ViT) introduce a new application of Transformers to generalize in image classification tasks. Dosovitskiy et al. [39] develop this novel approach, to outperform the conventional convolutional networks by treating inputs as sequences of non-overlapping image tokens known as patches. ViT blocks include multi-head self-attention layers and FF networks which model long-range dependencies among patches for computer vision tasks [40, 41]. In molecule discovery, the application ViT networks has not been extensively studied. In this research, we develop a few-shot graph-based ViT architecture which combines the local context of molecule graphs with global-semantic information captured by attention operations

to effectively predict the NR-binding properties using reduced amounts of labeled compounds.

Nuclear receptor data

In this work, data is collected using a public compound repository known as NURA (NUclear Receptor Activity) database [27], which includes public information on the activity of 15,247 compounds on 11 human NRs. The database contains information of compounds collected from sources such as ChEMBL25 [42], BindingDB [43], NR-DBIND [44] and Tox21 [45] to express the chemical structure using SMILES strings (Simplified Molecular Input Line Entry System) [46]. In this study, molecules are represented using molecular graphs obtained from SMILES using the RDKit.Chem library [47] which are pre-processed, so that SMILES are canonicalized and duplicates are removed. This curated dataset refers to the in vitro bioactivity data of compounds on 11 nuclear receptors (NRs), selected based on their biological significance and availability on public databases including the following: androgen receptor (AR), estrogen receptor α (ERA), estrogen receptor β (ERB), progesterone receptor (PR), glucocorticoid receptor (GR), peroxisome proliferator-activated receptor α (PPARA), peroxisome proliferator-activated receptor γ (PPARG), peroxisome proliferator-activated receptor δ (PPARD), pregnane X receptor (PXR), retinoid X receptor (RXR) and farnesoid X receptor (FXR). Thus, the dataset comprises measurements of bioactivity against 11 NRs such as binding activity, agonist activity and antagonist activity. In each case, compounds are assigned an activity label, given their experimental bioactivities against specific NRs: (1) "active", if bioactivity is equal to or lower than 10,000 nM (positive); (2) "weakly active", for bioactivity between 10,000 and 100,000 nM (positive); (3) "inactive", for bioactivity values greater than 100,000 nM (negative); (4) "inconclusive", for compounds having conflicting labels for all 3 cases; (5) "missing", for compounds having missing information for at least one case. In our experiments, we merge both "active" and "weakly active" into a positive label and "inactive" into a negative label for binding (BIN), agonist (AGO) and antagonist (ANT) activity classification tasks and compounds with "inconclusive" and "missing" labels are excluded. Table 1 below reports the distribution of compounds for all 11 NR activity labels and for all NR binding (BIN), agonist (AGO) and antagonist (ANT) activity classification tasks.

Methods

Graph neural network module (GNN)

Molecular graphs are graph-structured representations of atoms and their connections via chemical bonds within a molecule. Molecular graphs are denoted by $G = (V, E)$,

Table 1 Distribution of positive and negative samples for binding, agonist and antagonist activity labels for all 11 nuclear receptors

NR	Binding (BIN)		Agonism (AGO)		Antagonism (ANT)	
	Positive	Negative	Positive	Negative	Positive	Negative
AR	1523	5130	634	5578	1167	4942
ERA	1464	4861	937	5060	684	5160
ERB	1225	5554	334	5744	453	5133
PR	1305	5040	376	5670	1289	4400
GR	1900	5228	778	5384	847	4577
PPARA	1352	15	1084	14	18	1
PPARG	1904	5458	1510	5223	241	5249
PPARD	782	5742	689	5663	52	5561
PXR	1327	3866	1319	3866	10	0
RXR	1006	4569	263	4549	1	3
FXR	658	5272	457	5349	267	4829

with V the set of nodes v (atoms) and E the set of edges e (chemical bonds). Edges are defined by $e = (v, u)$, where v and u are nodes interconnected in a neighborhood $N(v)$. Graph neural networks (GNNs) use a neighborhood aggregation function to update node embeddings h_v and build graph embedding representations h_G used in molecule classification. In this research, a GIN with $L_{GIN} = 5$ layers is proposed as an embedding network to detect the local dependencies in molecular graphs G and compute graph embeddings h_G. The GIN performs AGGREGATE and COMBINE steps as a sum of node and edge features. Node embeddings h_v are updated for each message-passing iteration l by

$$m_{N(v)}^l = AGGREGATE^l(\{h_u^{l-1}, \forall u \in N(v)\}, \{h_e^{l-1} : e = (v,u)\}) \tag{1}$$

$$h_v^l = \sigma(MLP^l(COMBINE^l(h_v^{l-1}, m_{N(v)}^l))) \tag{2}$$

with $m_{N(v)}$ the "message" propagated throughout GNN layers, h_u^l the embeddings for neighboring nodes, and h_e^l the embedding for the edge between nodes u and v. After node-edge aggregation, multiple message-passing iterations l update node embeddings h_v^l using prior representations of that node h_v^{l-1} and embeddings of its neighboring nodes h_u^{l-1} with $u \in N(v)$. The UPDATE step applies multi-layer perceptrons MLP to introduce non-linearity followed by non-linear activation $\sigma = ReLU$

$$h_v^l = ReLU(MLP^l(\sum_{u \in N(v) \cup v} h_u^{l-1} + \sum_{e=(v,u):u \in N(v) \cup v} h_e^{l-1})). \tag{3}$$

At the final layer $L_{GIN} = 5$, a READOUT step pools node embeddings to produce a graph-level embedding h_G. This graph embedding is obtained by averaging node representations h_v using a mean-pooling operation, $h_G = mean(\{h_v^{L_{GIN}} : v \in V\})$. Input node-edge embedding features (h_v^0, h_e^0) are described by multiple atom-bond attributes including atom type (AT), atom chirality (AC) with $h_v^0 = \{v_{AT}, v_{AC}\}$, and bond type (BT), bond direction (BD) with $h_e^0 = \{e_{BT}, e_{BD}\}$. Pre-trained GNNs of Hu et al. (2020) [36] are leveraged to pre-train the GIN model for better initialization. In this setting, we consider 5 GIN layers and an embedding size of 300. A schematic of the GNN-Transformer architecture is presented in Fig. 1.[1]

Transformer prediction module (TR)

In our research, we investigate how to combine Transformers and GNNs to better discriminate the global-semantic context and long-range dependencies within molecule graph embeddings h_G for NR-binding activity prediction. A Transformer network with $L_T = 5$ blocks is introduced to convert graph embeddings h_G into token embeddings h_T. This prediction module operates as a vision Transformer (ViT) [35, 48] that accepts graph embedding h_G transformed into sequences of patches considering a space of dimension $D = N \times P^2$ where N is the number of patch tokens and P the size of individual patch tokens. The Transformer accepts embeddings x converted into sequences of patches x_p

$$T(x) = [x_p^1, x_p^2, \ldots, x_p^N] \tag{4}$$

---

[1] In the figure, individual nodes are represented by blue circles, while neighboring nodes are illustrated by black circles. In addition, blue and white squares denote node and graph embeddings h_v and h_G. The AGGREGATE, COMBINE, UPDATE steps are performed simultaneously for all nodes $v \in V$ within the graph. Here, we consider graph operations for $L_{GIN} = 5$ GIN layers, and a READOUT step is performed at the final GIN layer. The vision Transformer (ViT) computes token embeddings h_T using graph embeddings h_G of size 300.
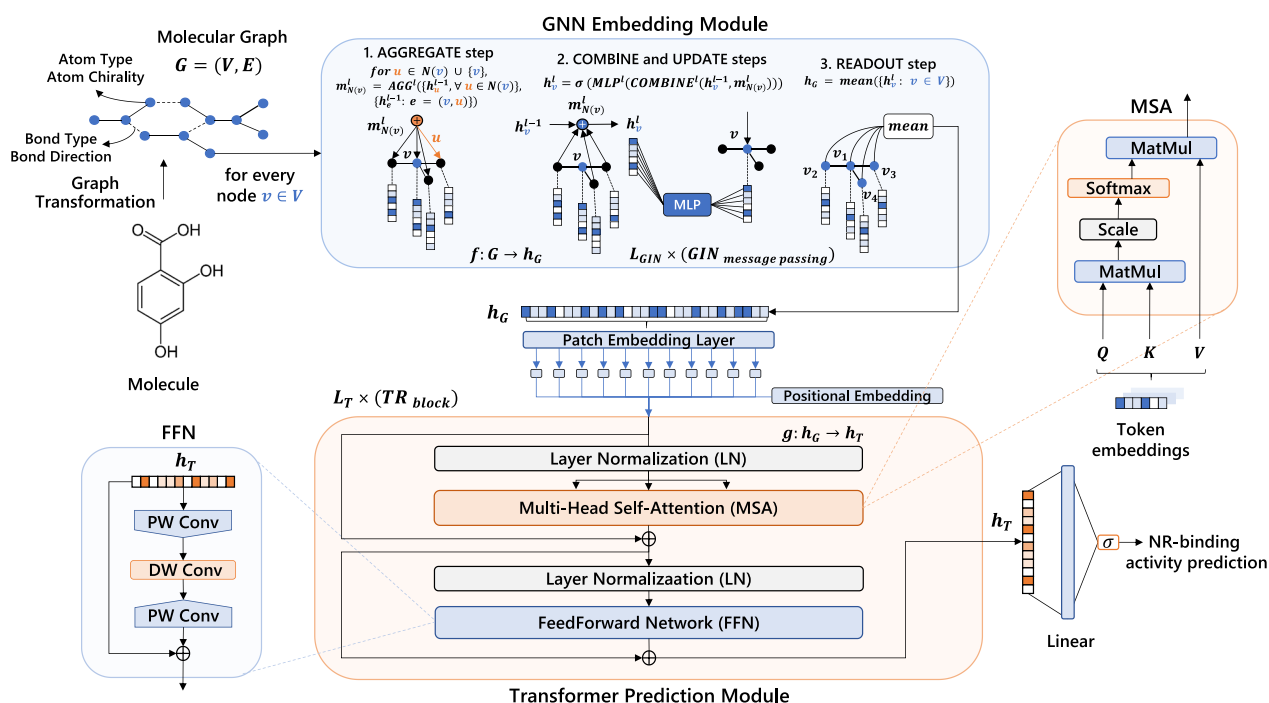
Fig. 1 Graphical depiction of the two-module GNN-Transformer architecture, Meta-GTNRP

where x_p^i denotes the i-th patch vector. Specifically, the Transformer converts graph embeddings h_G into $N = (\lfloor \frac{300}{P} \rfloor)^2$ patch tokens of size P. Token embeddings $h_T(x) = T(x).K$ are produced by a linear projection $K \in \mathbb{R}^{P^2 \times D}$ of patch vectors x_p^i in a Transformer dimension D. The Transformer propagates token embeddings h_T in MSA layers. MSA takes three inputs: queries q, keys k and values v stacked in matrices (Q, K, V) to calculate a dot-product attention between queries q in Q and keys k in K. MSA considers a softmax operation to obtain the attention weights for values v in V. In addition, MSA layers include H projection heads and attention values are calculated by

$$MSA(Q, K, V) = CONCAT(head_1, \ldots, head_H)W \quad (5)$$

$$\begin{aligned} head_j &= Attention(QW_j^Q, KW_j^K, VW_j^V) \\ &= softmax(\frac{QW_j^Q(KW_j^K)^T}{\sqrt{D}})VW_j^V \end{aligned} \quad (6)$$

where (W_j^Q, W_j^K, W_j^V) are the projection matrices of (Q, K, V) for each attention head j. Transformer blocks use *MSA* layers followed by feed-forward networks (*FFN*). *FFN* include a point-wise (*PW*) convolutional operation to undersample Q and K and model the local context of token embeddings more efficiently. Then, a convolution operation is applied to the Q, K and V

matrices using a depth-wise (*DW*) separable convolution with kernel size $s = 3$ followed by batch normalization and a *PW* convolution operation. The *MSA* and *FFN* layers are preceded by a layer normalization (*LN*) followed by residual connections. The individual patch tokens x_p are propagated across multiple Transformer blocks l to obtain the token embedding representations h_T given by

$$h_T^0 = [x_p^1 K, x_p^2 K, x_p^3 K, \ldots, x_p^N K] + K_{pos} \quad (7)$$

$$h_T^{l*} = MSA(LN(h_T^{l-1})) + h_T^{l-1} \quad (8)$$

$$h_T^l = FFN(LN(h_T^{l*})) + h_T^{l*} \quad (9)$$

$$y = LN(h_T^{L_T}) \quad (10)$$

where $l = \{1, \ldots, L_T\}$, h_T^l are the deep token embedding representations, K are the linear projections of individual patch tokens and positional embeddings with $K \in \mathbb{R}^{P^2 \times D}$ and $K_{pos} \in \mathbb{R}^{(N+1) \times D}$ and where y is an output vector. Then, a *MLP* followed by a sigmoid activation function is applied to the the output *cls* token to predict a molecular label for different NR-binding activity prediction tasks (with an output value $\in \{0, 1\}$). The Transformer prediction module of Meta-GTNRP has the main hyper-parameters displayed in Table 2.

Torres *et al. Journal of Cheminformatics*   (2024) 16:109

Page 6 of 19

Table 2 Main hyper-parameters of the Transformer module

Hyper-parameter	Value
Input dim	300
Patch size (P)	30
Number of patches (N)	100
Number of attention heads	5
Linear transform dim (D)	128
MLP dim	256
Conv kernel size	3
Number of blocks (L_T)	5
Batch size	10
Pooling	Mean pool
Learning rate	$1e^{-5}$
Optimizer	Adam

Few-shot meta-learning framework for NR-binding activity prediction

In this research, a few-shot meta-learning framework built upon two distinct neural network modules is introduced to learn complementary information across few-shot tasks for NR-binding activity prediction. This strategy leverages the relationship among different NRs by the means of integrating information of NR-specific predictive tasks with a joint learning procedure. This framework is composed of two main components: a GNN and a Transformer (TR) modules. Both of these meta-models update model parameters for different few-shot tasks (meta-training) for 10 different NRs using random support sets for training and query sets with remaining samples for evaluation. Then, these parameters are leveraged to infer the binding activity of compounds against a 1 new specific NR (meta-testing) [30, 31]. In this framework, molecules are organized across meta-training tasks to optimize the model parameters by evaluating the binding activity of compounds for 10 different NRs. Then, the parameters obtained in meta-training are leveraged to infer the binding activity of compounds for 1 new NR, in a new meta-testing task. The main objective is to predict the binding activity of compounds for 1 specific NR using the NR-binding information of other 10 NR-binding tasks, so that $\{f_\theta(G), g_{\theta*}(h_G)\} : S \Rightarrow \{0, 1\} \in Y$, where S is the chemical space of molecule graphs G, h_G is the output graph embedding space of a GNN f_θ, $g_{\theta*}$ is a Transformer (TR), and Y the labels for each individual NR. A GIN model f_θ with parameters θ and a TR model

$g_{\theta*}$ with parameters $\theta*$ learn across different few-shot tasks $t \in \{1 \ldots, N_{NR-train}\}$ for each individual NR. For each meta-task, meta-models f_θ and $g_{\theta*}$ are trained using random support sets S_t of molecule graphs $G_{S_{t_i}}$ and evaluated using query sets Q_t of graphs $G_{Q_{t_i}}$.

In meta-training, support sets with size k are randomly sampled as the input for a GNN $f(\theta)$ and TR $g(\theta*)$ models to obtain the support losses $\mathcal{L}_t^{GNN}$, $\mathcal{L}_t^{TR}$ for each individual NR across meta-training tasks $t \in \{1 \ldots, N_{NR-train}\}$ with $N_{NR-train} = 10$. Support losses are used to iteratively update parameters $\theta \rightarrow \theta'$, $\theta* \rightarrow \theta*'$. Then, the GNN and TR models compute query losses $\mathcal{L}_t^{GNN'}, \mathcal{L}_t^{TR'}$ with the remaining n samples for a specific task. At this stage, to update parameters $\theta, \theta*$ we consider few gradient steps

$$\theta_t = \theta - \alpha \nabla_\theta \mathcal{L}_t^{GNN}(\theta) \tag{11}$$

$$\theta_t^* = \theta^* - \alpha^* \nabla_{\theta*} \mathcal{L}_t^{TR}(\theta^*) \tag{12}$$

where α and α^* are the step sizes for gradient descent updates. Then, in meta-testing, support sets with k random samples is obtained for 1 new NR-specific test task $t = N_{NR-train} + N_{NR-test}$ and parameters θ, θ^* are initialized using model parameters obtained in meta-training, $\theta \rightarrow \theta'$, $\theta^* \rightarrow \theta^{*'}$ with $N_{NR-test} = 1$. Then, GNN and TR models are evaluated using query sets of new molecules with the remainder of the samples n for this test task, for predicting the binding activity of compounds for 1 specific NR with just a few labeled compounds. In this meta-learning framework, the NR data is divided into a set of meta-training and meta-testing tasks for different NRs. During the meta-training phase, a set of meta-training tasks for 10 different NRs are performed. For each task, a random support set of size $(k+, k-)$ ($k+$ positive and $k-$ negative samples) is sampled for training, and a disjoint query set is sampled for evaluation. More specifically, we compute the gradient of the loss with respect to Meta-GTNRP parameters using just a few examples from that task and update model parameters such that it performs well on the query set of this task. The updated parameters obtained in meta-training are then used to initialize Meta-GTNRP to predict a query set of new compounds of a new NR meta-testing task with limited data. This meta-learning framework is illustrated in Fig. 2 with the algorithm shown below.

Algorithm 1 Meta-GTNRP: Few-shot GNN-Transformer Meta-Learning Framework

---

Require: Support data: (S_t, Y_t), Query data: (Q_t, Y'_t), $\alpha, \beta, \alpha^*, \beta^*$: update step sizes
   $\theta \leftarrow$ Pre-trained GNN
   while *not done* **do**
      Sample a batch of tasks $t \sim \rho(T)$
      for all t **do**
         Sample k examples $\{G_{S_{t_1}}, G_{S_{t_2}}, ... G_{S_{t_k}}\} \in S_t$
         for $i = 1$ to k **do**
            $y_{t_i}, h_{G_{S_{t_i}}} = \text{GNN}(G_{S_{t_i}}, \theta)$
         end for
         $\mathcal{L}_t^{GNN} \leftarrow \{y_{t_1}, y_{t_2}, ..., y_{t_k}\}$
         for $i = 1$ to k **do**
            $y_{t_i}, h_{T_{S_{t_i}}} = \text{TR}(h_{G_{S_{t_i}}}, \theta^*)$
         end for
         $\mathcal{L}_t^{TR} \leftarrow \{y_{t_1}, y_{t_2}, ..., y_{t_k}\}$
         $\theta' = \theta - \alpha \nabla \mathcal{L}_t^{GNN}$
         $\theta^{*'} = \theta^* - \alpha^* \nabla \mathcal{L}_t^{TR}$
         Sample n examples $\{G_{Q_{t_1}}, G_{Q_{t_2}}, ... G_{Q_{t_n}}\} \in Q_t$
         for $j = 1$ to n **do**
            $y'_{t_j}, h_{G_{Q_{t_j}}} = \text{GNN}(G_{Q_{t_j}}, \theta')$
         end for
         $\mathcal{L}_t^{GNN'} \leftarrow \{y'_{t_1}, y'_{t_2}, ..., y'_{t_n}\}$
         for $j = 1$ to n **do**
            $y'_{t_j}, h_{T_{Q_{t_j}}} = \text{TR}(h_{G_{Q_{t_j}}}, \theta^{*'})$
         end for
         $\mathcal{L}_t^{TR'} \leftarrow \{y'_{t_1}, y'_{t_2}, ..., y'_{t_n}\}$
      end for
      $\theta \leftarrow \theta \beta \nabla_\theta \sum_{t \sim \rho(T)} \mathcal{L}_i^{GNN'}$
      $\theta^* \leftarrow \theta^* \beta^* \nabla_{\theta^*} \sum_{t \sim \rho(T)} \mathcal{L}_i^{TR'}$
   end while

Loss function for NR-binding activity prediction

The loss function for the GNN and Transformer models, $\mathcal{L}^{GNN}$ and $\mathcal{L}^{TR}$ is a binary cross-entropy loss. However, to answer the issue of class imbalance in NR-binding activity prediction, a weighted loss significantly penalizes the misclassifications with rare-class instances. Hence, the cross-entropy loss includes a weight w for a minority class and is formalized by

$$\mathcal{L} = -\frac{1}{n} \sum_{i=1}^{n} w \, y_i \, log(y'_i) + (1 - y_i) \, log(1 - y'_i) \quad (13)$$

where y' are the predictions and y the binding activity labels with n the number of datapoints. Since we observe different positive–negative ratios for individual NR tasks, the value w is obtained by exploring different values within an appropriate range. Thus, we establish a value of $w = 5$ due to a task-specific variability among NR-specific predictive tasks.

Baseline methods

The proposed few-shot GNN-Transformer model, Meta-GTNRP is compared with other 3 types of GNNs:

1. GIN - top-performing GNN method that applies the Weisfeiler-Lehman (WL) isomorphism test to aggregate important parts of the node's neighborhood [35];
2. GCN - updates node embeddings using convolution for neighborhood aggregation similar to convolutional filters found in convolutional networks [33];

Torres *et al. Journal of Cheminformatics*   (2024) 16:109

Page 8 of 19
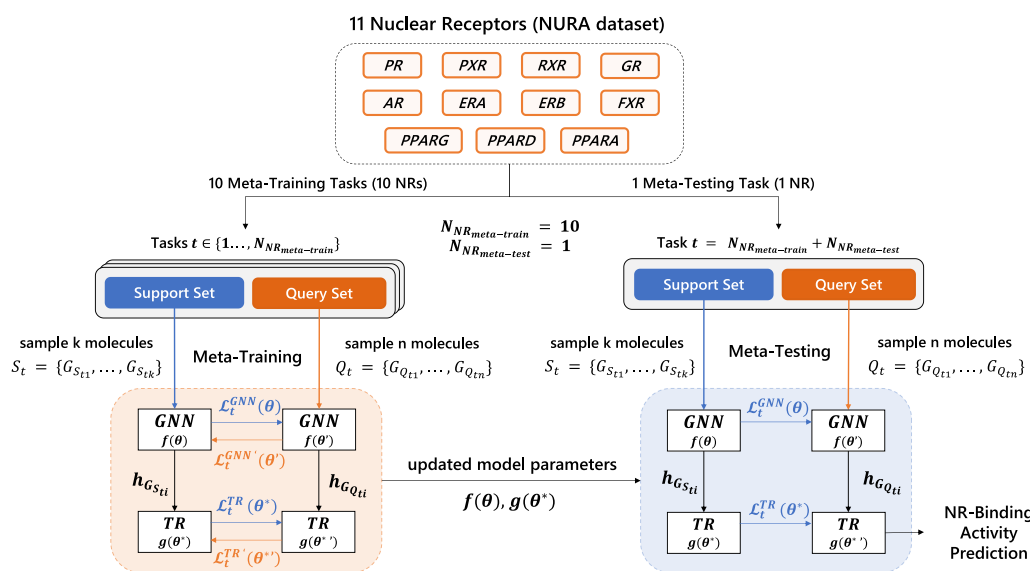
Fig. 2 Schematic of the GNN-Transformer meta-learning framework for NR-binding activity prediction. This framework is composed by two neural networks a GNN f and a Transformer (TR) g with parameters θ and θ', respectively. In meta-training, few-shot tasks t include random support sets S_t with positive samples k_+, negative samples k_- provided for training. Then, the remaining n data points are used as query sets Q_t for evaluation. The GNN and TR models, f and g consider 10 few-shot tasks for 10 different NRs in meta-training. Then, in meta-testing, the updated parameters obtained are used to initialize the Meta-GTNRP model to predict the NR-binding activity of new compounds in query sets of 1 new NR task using random support sets of size (k_+, k_-) for k-shot experiments

3. GraphSAGE - uses an inductive framework based on a node-centric training method to update node features within unseen graph representations [34].

These GNN baselines are obtained by the removal of the Transformer component of Meta-GTNRP and are also trained and evaluated using a few-shot meta-learning approach to ensure a comparable performance across few-shot tasks in the 5-shot and 10-shot settings. These GNN models also use pre-trained models of Hu et al. [36] for improved initialization and are optimized using a standard binary cross-entropy loss function.

The graph-based baselines and Meta-GTNRP are implemented in Python 3.9.16 and PyTorch 1.13.0 with CUDA 11.6, along with functions in Scikit-learn 1.2.2, Numpy 1.22.3, Pandas 1.5.3 and RDKit 2022.03.5. The best GNN and Meta-GTNRP models are selected at the epoch giving the best ROC-AUC score in the query set of the final meta-testing task and we allow it to run for at most 1000 epochs. Additionally, we consider update steps of 5 for meta-training and 10 for meta-testing. In addition, we consider GNN models with a total of 5 message-passing layers and a graph embedding dimension of 300.

The GNN baseline models are trained and evaluated using a few-shot meta-learning approach [31] in the 5-shot and 10-shot settings. Similarly to the Meta-GTNRP model, for NR-binding activity prediction tasks, we consider 10 meta-training tasks for 10 different NRs and 1 final meta-testing task for the remaining NR. For NR-agonist and NR-antagonist activity prediction tasks, we also consider the single-task (ST) models considering 1 meta-training task and 1 final meta-testing task. This experimental setup ensures a comparable performance between Meta-GTNRP and the GNN baselines.

Nuclear receptor binding activity experiments
In this study, we evaluate the binary classification of compounds across few-shot tasks for NR-binding activity prediction. For a total of 11 NRs, the proposed Meta-GTNRP model considers 10 meta-training tasks for 10 different NRs and is evaluated on 1 final meta-testing task for a specific NR with limited available data (see Fig. 2). Specifically, we calculate seven performance metrics on the query set of a test task for each of 11 different NRs: Sensitivity (Sn), Specificity (Sp), Precision (Pr), Accuracy (Acc), ROC-AUC score, and F1 score (F1s). ROC-AUC is the main performance metric and computes the area under the receiver operating characteristic curve to evaluate the performance in imbalanced scenarios for NR-binding activity prediction. Here, we conduct 5-shot and 10-shot experiments with random support

sets of size (5+, 5−) and (10+, 10−) for each individual NR, respectively. Experiments are repeated 30 times, using random support sets each time, to obtain a robust estimate of performance for each metric. In Table 3, we report the mean and standard deviations of ROC-AUC results obtained by the Meta-GTNRP model considering 10 tasks in meta-training for 10 different NRs across 30 experiments with (5+, 5−) (5-shot) and (10+, 10−) (10-shot) random support sets and evaluated on 1 meta-testing task for 1 specific NR. In **bold**, we show the best ROC-AUC results for each NR-specific test task. The △ column shows the difference in performance results of the proposed model and graph-based baselines. In the Supplementary Material, figures show scatter plots overlaid by boxplots show the ROC-AUC scores and standard deviations obtained across 30 experiments with 5-shot and 10-shot random supports sets for NR-binding activity prediction tasks on 11 different NRs. The full performance results for each metric are shown in Tables also displayed in the Supplementary Material section.

Transfer-learning agonist and antagonist activity experiments

In our experiments, we also test the ability of the Meta-GTNRP model to transfer the knowledge of a single NR-binding task for one individual NR to predict the agonist or antagonist activity of compounds on that specific NR (see Fig. 3). The main goal is to evaluate the generalization power of the Meta-GTRNP model by the means of single-task (ST) models obtained using the NR-binding information to determine which compounds have agonist and antagonist activity for individual NRs. Here, we conduct few-shot experiments to transfer the learning of ST models trained on NR-binding activity information to predict the NR-specific agonist and antagonist activity with limited data.

In this case, we conduct 5-shot and 10-shot experiments for ST models trained on a single NR-binding activity task to predict the agonist or antagonist activity for that specific NR on a single NR agonist or antagonist activity test task. These experiments are repeated 30 times, using random support sets each time. In Tables 4

Table 3 Average ROC-AUC scores obtained across 30 experiments with random support sets of size (5+, 5−) (5-shot) and (10+, 10−) (10-shot) by Meta-GTNRP and few-shot GNN baselines in NR-binding activity prediction for 11 different NRs

Binding activity (BIN)					
NR	GIN	GCN	GraphSAGE	Meta-GTNRP (GIN+TR)	△
5-shot (5+, 5−)					
PR	0.8812 ± 0.0110	0.8634 ± 0.0353	0.8831 ± 0.0332	$\mathbf{0.9409 \pm 0.0007}$	$+0.0578$
PXR	0.6080 ± 0.0147	0.6649 ± 0.0412	0.6486 ± 0.0387	$\mathbf{0.6662 \pm 0.0036}$	$+0.0013$
RXR	0.6128 ± 0.0354	0.8194 ± 0.0323	$\mathbf{0.8537 \pm 0.0189}$	0.7630 ± 0.0439	-0.0907
GR	0.8313 ± 0.0256	0.8363 ± 0.0348	0.8352 ± 0.0273	$\mathbf{0.9082 \pm 0.0014}$	$+0.0719$
AR	0.7652 ± 0.0253	0.8277 ± 0.0469	0.8101 ± 0.0525	$\mathbf{0.8745 \pm 0.0023}$	$+0.0468$
ERA	0.8759 ± 0.0143	0.8474 ± 0.0342	0.8442 ± 0.0170	$\mathbf{0.9012 \pm 0.0045}$	$+0.0253$
ERB	0.9273 ± 0.0081	0.8832 ± 0.0235	0.8717 ± 0.0170	$\mathbf{0.9308 \pm 0.0032}$	$+0.0035$
FXR	0.7452 ± 0.0170	0.7937 ± 0.0260	0.7995 ± 0.0389	$\mathbf{0.8605 \pm 0.0037}$	$+0.0610$
PPARD	0.9136 ± 0.0154	0.8853 ± 0.0292	0.9298 ± 0.0203	$\mathbf{0.9315 \pm 0.0043}$	$+0.0017$
PPARG	0.7901 ± 0.0279	0.8226 ± 0.0454	0.8610 ± 0.0250	$\mathbf{0.8936 \pm 0.0044}$	$+0.0326$
PPARA	0.5771 ± 0.0408	0.6648 ± 0.0595	$\mathbf{0.7115 \pm 0.0798}$	0.6238 ± 0.0563	-0.0877
10-shot (10+, 10−)					
PR	0.9143 ± 0.0148	0.8887 ± 0.0221	0.9062 ± 0.0124	$\mathbf{0.9458 \pm 0.0016}$	$+0.0315$
PXR	0.6328 ± 0.0263	0.6573 ± 0.0372	0.6630 ± 0.0375	$\mathbf{0.7046 \pm 0.0029}$	$+0.0416$
RXR	0.7978 ± 0.0414	0.7832 ± 0.0375	$\mathbf{0.8598 \pm 0.0160}$	0.8392 ± 0.0186	-0.0206
GR	0.8713 ± 0.0185	0.8463 ± 0.0410	0.8646 ± 0.0128	$\mathbf{0.8967 \pm 0.0035}$	$+0.0254$
AR	0.8356 ± 0.0305	0.8238 ± 0.0530	0.8452 ± 0.0287	$\mathbf{0.8912 \pm 0.0033}$	$+0.0460$
ERA	0.8910 ± 0.0136	0.8493 ± 0.0196	0.8589 ± 0.0125	$\mathbf{0.9013 \pm 0.0053}$	$+0.0103$
ERB	0.9255 ± 0.0108	0.8631 ± 0.0455	0.8896 ± 0.0145	$\mathbf{0.9318 \pm 0.0039}$	$+0.0063$
FXR	0.7997 ± 0.0214	0.7657 ± 0.0605	0.8162 ± 0.0221	$\mathbf{0.8686 \pm 0.0032}$	$+0.0524$
PPARD	$\mathbf{0.9247 \pm 0.0221}$	0.8686 ± 0.0748	0.9157 ± 0.0281	0.8999 ± 0.0093	-0.0257
PPARG	0.8395 ± 0.0191	0.8171 ± 0.0481	0.8763 ± 0.0166	$\mathbf{0.8904 \pm 0.0073}$	$+0.0141$
PPARA	0.6957 ± 0.1088	$\mathbf{0.7401 \pm 0.1004}$	0.6892 ± 0.1224	0.6068 ± 0.0931	-0.1333

Bold indicate the highest ROC-AUC scores for each NR task

and 5, we report the mean and standard deviations of ROC-AUC results for the agonist and antagonist activity predictions made by ST models across 30 experiments with $(5+, 5-)$ (5-shot) and $(10+, 10-)$ (10-shot) random support sets, considering 1 NR-specific binding task in meta-training and evaluated on 1 agonist or antagonist meta-testing task for that specific NR, respectively. In bold, we present the best ROC-AUC results for each individual NR-specific test task. The △ column shows the difference in performance results of the proposed model and graph-based baselines. In the Supplementary Material, figures show scatter plots overlaid by boxplots show ROC-AUC scores and standard deviations obtained across 30 experiments with 5-shot and and 10-shot random supports sets in agonist and antagonist activity prediction on 11 different NRs. The full performance results using different metrics are shown in Tables also displayed in the Supplementary Material.

Statistical significance analysis of performance results

In this work, ROC-AUC (Receiver Operating Characteristic-Area Under the Curve) scores evaluate the performance of the Meta-GTNRP model and graph-based baselines. The ROC-AUC score is a widely used metric to measure the ability to distinguish between positive and negative samples and, particularly useful to measure performance in limited imbalanced data. However, simply reporting the ROC-AUC scores alone may not be sufficient to draw conclusions about the significance of the performance difference between different models.

To address this issue, we performed a statistical significance analysis to determine whether the differences in the ROC-AUC scores between the Meta-GTNRP model and GNN baselines are statistically significant or merely due to chance. This analysis compared the ROC-AUC results obtained using a statistical significance test. The p-values calculated indicate the probability of observing such a difference by chance alone, and a threshold level of significance ($\alpha = 0.05$) is used to determine whether the difference is statistically significant. The first step was to compute a normality test to determine if ROC-AUC scores for each pair of performance results are normally distributed. We used a normality test provided by the SciPy library [49] based on the D'Agostino-Pearson omnibus test [50, 51], which combines skewness and kurtosis measurements to provide a p-value that indicates the likelihood that the data is normally distributed. The null hypothesis for this test is that the data is normally distributed. Hence, if the p-value is less than the significance level ($p < 0.05$), we can reject the null hypothesis and conclude that the data is not normally distributed. Otherwise, we fail to reject the null hypothesis, indicating that ROC-AUC results are more likely to follow a normal distribution.

Next, we assessed the descriptive statistic of the normality test to evaluate whether the variance among both distributions was the same. We found a difference in
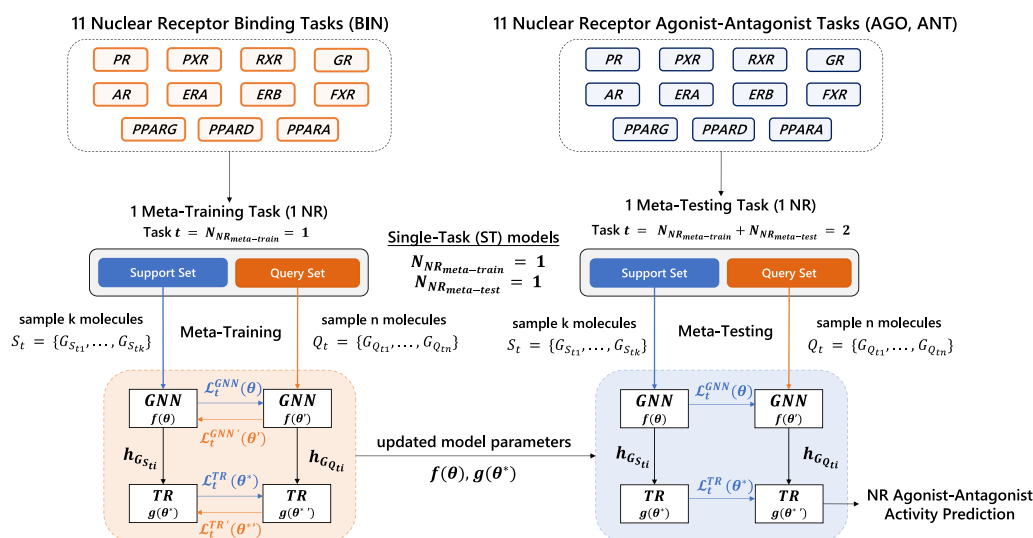
Fig. 3 Schematic of the meta-learning framework for NR-agonist and NR-antagonist activity prediction. In this experiment, we consider 2 few-shot tasks, 1 meta-training task with NR-binding information and 1 meta-testing task for NR-agonist or NR-antagonist activity prediction. In our proposed framework, both GNN and Transformer models, *f* and *g* consider a NR-binding task in meta-training to build single-task (ST) models evaluated on the corresponding NR-specific agonist or antagonist test task. The model performance is assessed on the query set of a new NR-agonist or NR-antagonist meta-testing task using a random support set of size (k_+, k_-) for *k*-shot experiments

Table 4 Average ROC-AUC scores obtained across 30 experiments with random support sets of size $(5+, 5-)$ (5-shot) and $(10+, 10-)$ (10-shot) by the single-task (ST) Meta-GTNRP models and single-task (ST) GNN baselines considering 1 NR binding task in meta-training and 1 NR agonist task in meta-testing

Agonist activity (AGO)					
NR	GIN	GCN	GraphSAGE	Meta-GTNRP (GIN+TR)	Δ
5-shot $(5+, 5-)$					
PR	0.9556 ± 0.0091	0.9086 ± 0.0318	0.9188 ± 0.0210	$\mathbf{0.9617 \pm 0.0005}$	$+0.0061$
PXR	0.8370 ± 0.0140	0.7800 ± 0.0184	0.7854 ± 0.0199	$\mathbf{0.8636 \pm 0.0008}$	$+0.0266$
RXR	$\mathbf{0.7290 \pm 0.0100}$	0.6440 ± 0.0310	0.6772 ± 0.0110	0.7236 ± 0.0047	-0.0054
GR	$\mathbf{0.9513 \pm 0.0009}$	0.9164 ± 0.0221	0.9220 ± 0.0176	0.9478 ± 0.0006	-0.0035
AR	$\mathbf{0.8622 \pm 0.0020}$	0.8090 ± 0.0419	0.8176 ± 0.0289	0.8606 ± 0.0015	-0.0016
ERA	0.7213 ± 0.0071	0.6982 ± 0.0128	0.7093 ± 0.0116	$\mathbf{0.7278 \pm 0.0025}$	$+0.0065$
ERB	0.9134 ± 0.0069	0.8714 ± 0.0381	0.8804 ± 0.0233	$\mathbf{0.9205 \pm 0.0013}$	$+0.0071$
FXR	0.9190 ± 0.0075	0.8620 ± 0.0317	0.8885 ± 0.0078	$\mathbf{0.9303 \pm 0.0021}$	$+0.0113$
PPARD	$\mathbf{0.9515 \pm 0.0007}$	0.9154 ± 0.0214	0.9324 ± 0.0085	0.9492 ± 0.0007	-0.0023
PPARG	$\mathbf{0.9433 \pm 0.0012}$	0.8975 ± 0.0155	0.9114 ± 0.0109	0.9403 ± 0.0011	-0.0030
PPARA	0.9605 ± 0.0268	0.9401 ± 0.1010	0.9563 ± 0.0248	$\mathbf{0.9622 \pm 0.0415}$	$+0.0017$
10-shot $(10+, 10-)$					
PR	0.9474 ± 0.0094	0.8263 ± 0.1612	0.9112 ± 0.0272	$\mathbf{0.9543 \pm 0.0018}$	$+0.0069$
PXR	0.8362 ± 0.0079	0.7737 ± 0.0334	0.7955 ± 0.0094	$\mathbf{0.8589 \pm 0.0018}$	$+0.0227$
RXR	$\mathbf{0.7349 \pm 0.0160}$	0.6522 ± 0.0384	0.6771 ± 0.0151	0.7274 ± 0.0061	-0.0075
GR	$\mathbf{0.9518 \pm 0.0011}$	0.9030 ± 0.0339	0.9173 ± 0.0199	0.9460 ± 0.0008	-0.0058
AR	$\mathbf{0.8614 \pm 0.0018}$	0.7839 ± 0.0672	0.8197 ± 0.0267	0.8574 ± 0.0025	-0.0040
ERA	0.7257 ± 0.0068	0.6809 ± 0.0233	0.7123 ± 0.0131	$\mathbf{0.7388 \pm 0.0022}$	$+0.0131$
ERB	0.9182 ± 0.0078	0.8252 ± 0.0985	0.8908 ± 0.0129	$\mathbf{0.9213 \pm 0.0017}$	$+0.0031$
FXR	$\mathbf{0.9261 \pm 0.0094}$	0.8623 ± 0.0257	0.8823 ± 0.0119	0.9208 ± 0.0030	-0.0053
PPARD	$\mathbf{0.9549 \pm 0.0026}$	0.8984 ± 0.0951	0.9255 ± 0.0115	0.9488 ± 0.0008	-0.0061
PPARG	$\mathbf{0.9432 \pm 0.0013}$	0.8643 ± 0.0749	0.9086 ± 0.0092	0.9410 ± 0.0009	-0.0022
PPARA	0.9611 ± 0.0582	0.6728 ± 0.1404	0.8899 ± 0.0989	$\mathbf{0.9680 \pm 0.0808}$	$+0.0069$

Bold indicate the highest ROC-AUC scores for each NR task

variance between both distributions for all model results. To test the statistical significance between each pair of results, we used a modified version of the Student t-test if both distributions are more likely to be normal. This version of the t-test, known as Welch's t-test, is used when there is unequal variance between the two distributions being compared. If the distributions are unlikely to follow a normal distribution, we apply the Mann–Whitney U non-parametric test. In the statistical significance test, p-values are calculated considering the hypothesis:

> **H0**: Performance results are likely drawn from the same distribution;
> **H1**: Performance results are likely drawn from different distributions (reject H0).

The calculated p-values are used to assess the statistical significance of the mean differences between the two distributions, considering a significance level of $\alpha = 0.05$. If the $p < 0.05$, we reject the null hypothesis (H0) and conclude

that there is evidence to support the alternative hypothesis (H1), indicating that the observed result is statistically significant. In Table 6, we report the significance analysis of ROC-AUC scores obtained across 30 experiments by Meta-GTNRP with respect to GNN baselines in the 5-shot and 10-shot settings for the NR-binding, NR-agonist and NR-antagonist activity prediction tasks. The results show that all the p-values are lower than the significance level, leading us to reject the null hypothesis and conclude that the ROC-AUC results of Meta-GTNRP are statistically significant when compared with the GNN baselines.

This analysis enabled us to determine the significance of the performance differences between Meta-GTNRP and GNN baselines, allowing us to draw important conclusions about the effectiveness of the Meta-GTNRP model in NR-activity prediction tasks. In addition, the statistical significance analysis provided valuable insights for comparing the performance of few-shot models, offering robust evidence to support the validity of our findings.

Torres *et al. Journal of Cheminformatics*   (2024) 16:109

Page 12 of 19

Table 5 Average ROC-AUC scores obtained across 30 experiments with random support sets of size (5+, 5−) (5-shot) and (10+, 10−) (10-shot) by the single-task (ST) Meta-GTNRP models and single-task (ST) GNN baselines considering a 1 NR binding task in meta-training and 1 NR antagonist task in meta-testing

Antagonist activity (ANT)

NR	GIN	GCN	GraphSAGE	Meta-GTNRP (GIN+TR)	Δ
5-shot (5+, 5−)					
PR	0.7738 ± 0.0061	0.7522 ± 0.0267	0.7682 ± 0.0188	**0.7933 ± 0.0038**	+0.0195
GR	0.8739 ± 0.0035	0.8256 ± 0.0245	0.8463 ± 0.0152	**0.8802 ± 0.0016**	+0.0063
AR	0.8037 ± 0.0058	0.7724 ± 0.0220	0.7817 ± 0.0181	**0.8173 ± 0.0043**	+0.0136
ERA	0.7326 ± 0.0079	0.7179 ± 0.0178	0.7199 ± 0.0196	**0.7596 ± 0.0076**	+0.0270
ERB	0.7255 ± 0.0084	0.6916 ± 0.0155	0.7091 ± 0.0139	**0.7303 ± 0.0059**	+0.0048
FXR	0.7199 ± 0.0161	0.7113 ± 0.0335	0.7165 ± 0.0238	**0.7280 ± 0.0077**	+0.0081
PPARD	**0.6876 ± 0.0273**	0.6303 ± 0.0376	0.6330 ± 0.0278	0.6571 ± 0.0212	−0.0305
PPARG	**0.6669 ± 0.0128**	0.6227 ± 0.0292	0.6183 ± 0.0192	0.6563 ± 0.0093	−0.0106
10-shot (10+, 10−)					
PR	0.7854 ± 0.0117	0.7514 ± 0.0227	0.7689 ± 0.0152	**0.7923 ± 0.0068**	+0.0069
GR	0.8755 ± 0.0063	0.8067 ± 0.0618	0.8375 ± 0.0127	**0.8813 ± 0.0014**	+0.0058
AR	0.8091 ± 0.0063	0.7501 ± 0.0291	0.7802 ± 0.0211	**0.8105 ± 0.0029**	+0.0014
ERA	0.7408 ± 0.0134	0.6874 ± 0.0515	0.7297 ± 0.0128	**0.7526 ± 0.0066**	+0.0118
ERB	0.7326 ± 0.0085	0.6787 ± 0.0520	0.7192 ± 0.0168	**0.7422 ± 0.0052**	+0.0096
FXR	0.7271 ± 0.0128	0.6885 ± 0.0340	0.7227 ± 0.0238	**0.7331 ± 0.0089**	+0.0060
PPARD	**0.7133 ± 0.0293**	0.6691 ± 0.0348	0.6366 ± 0.0480	0.6885 ± 0.0295	−0.0248
PPARG	**0.7233 ± 0.0172**	0.6072 ± 0.0287	0.6345 ± 0.0149	0.6920 ± 0.0088	−0.0313

PXR, RXR and PPARA have insufficient data to obtain conclusive ROC-AUC results in the 5-shot and 10-shot settings

Bold indicate the highest ROC-AUC scores for each NR task

Discussion

In this work, we introduce a few-shot GNN-Transformer, Meta-GTNRP to predict the NR-binding properties across 11 different NR tasks using limited available data. The proposed few-shot two-module meta-learning framework combines information of 10 NR-specific meta-training tasks to predict the NR-binding activity of compounds for 1 new NR in a final meta-testing task. It is demonstrated that Meta-GTNRP achieves a superior performance in NR-binding activity prediction over the standard GNN methods.

In Table 3, we report the average ROC-AUC scores obtained across 30 experiments with (5+, 5−) (5-shot) and (10+, 10−) (10-shot) random support sets for models considering 10 meta-training NR tasks in meta to predict the binding activity of compounds for 1 remaining NR meta-testing task. The results show that the proposed Meta-GTNRP model outperforms the GNN baseline methods (GIN, GCN, GraphSAGE) for the majority of NR test tasks in the 5-shot and 10-shot settings. The proposed approach achieves a superior performance taking into account the class imbalance scenarios across tasks and the lack of labeled information for each individual NR. In general, the Meta-GTNRP model shows the best ROC-AUC results and high Sn and Sp values, thus

correctly predicting a high proportion of both active and non-active binders. The standard deviations reported also indicate a smaller variance, which ensures a stable performance and more robust results across 5-shot and 10-shot experiments (see Figures in the Supplementary Material). Inversely, GIN, GCN and GraphSAGE produce unstable performances and high-variance predictions across NR test tasks, which means that they may generalize well in some cases, but collapse for the majority of NR-binding experiments.

In our experiments, we also test the ability of the Meta-GTNRP model considering a single NR-binding task in meta-training to predict the NR agonist and antagonist activity results for a specific NR. The goal is to evaluate the performance of single-task (ST) models for each individual NR and compare the performance in NR agonist and antagonist activity prediction with the standard GNN baselines. There are relevant differences in performance yielded by ST-models for each NR in NR-agonist and NR-antagonist predictive tasks. Nonetheless, in 5-shot and 10-shot experiments, the ROC-AUC scores in Table 4 show that the Meta-GTNRP model exhibits higher and more robust results in NR-agonist activity prediction for most NR test tasks when compared with the graph-based baselines. Similarly, Meta-GTNRP also

Table 6 *p*-value results of the statistical significance test for 5-shot and 10-shot experiments in NR-binding (BIN), NR-agonist (AGO) and NR-antagonist (ANT) activity prediction tasks

NR	Binding activity (BIN)			Agonist activity (AGO)			Antagonist activity (ANT)		
	Meta-GTNRP								
	GIN	GCN	GraphSAGE	GIN	GCN	GraphSAGE	GIN	GCN	GraphSAGE
5-shot (5+, 5−)									
PR	$2.6770e-23$	$9.2155e-13$	$2.0132e-10$	$1.2168e-09$	$2.7862e-11$	$2.7862e-11$	$1.8126e-19$	$3.0161e-11$	$9.2787e-10$
PXR	$1.7833e-20$	$1.7632e-02$	$1.8875e-02$	$2.7263e-11$	$2.9785e-11$	$2.9766e-11$	–	–	–
RXR	$1.2914e-20$	$3.0939e-06$	$7.8179e-13$	$1.0821e-02$	$1.0815e-14$	$3.7008e-23$	–	–	–
GR	$3.1254e-16$	$3.8332e-12$	$2.9953e-11$	$3.6642e-21$	$2.7565e-11$	$2.7547e-11$	$4.2228e-11$	$3.0047e-11$	$3.0010e-11$
AR	$1.3406e-22$	$1.0879e-10$	$2.9991e-11$	$1.9951e-04$	$3.0123e-11$	$3.0123e-11$	$2.8950e-14$	$3.3239e-12$	$3.0085e-11$
ERA	$7.3051e-11$	$3.0066e-11$	$3.0047e-11$	$3.4432e-05$	$3.0104e-11$	$1.1843e-09$	$1.7275e-19$	$2.0412e-14$	$3.0123e-11$
ERB	$3.6243e-02$	$3.0066e-11$	$1.9035e-18$	$2.8250e-04$	$2.9841e-11$	$2.9766e-11$	$1.5539e-02$	$4.1490e-15$	$2.4780e-10$
FXR	$2.0997e-27$	$1.1764e-14$	$2.3701e-10$	$3.4482e-09$	$1.2652e-12$	$2.9860e-11$	$1.2996e-02$	$1.2062e-02$	$1.6185e-02$
PPARD	$7.1381e-07$	$4.0696e-11$	$4.8388e-02$	$9.2552e-10$	$2.9617e-11$	$2.9543e-11$	$1.2073e-05$	$1.4178e-03$	$4.0686e-04$
PPARG	$4.2933e-19$	$3.0161e-11$	$3.6829e-11$	$5.4190e-10$	$2.9561e-11$	$2.9506e-11$	$6.0748e-04$	$2.4901e-08$	$2.4595e-12$
PPARA	$5.6073e-04$	$8.1525e-03$	$9.1655e-06$	$1.0758e-02$	$2.1905e-02$	$3.0340e-03$	–	–	–
10-shot (10+, 10−)									
PR	$1.6004e-12$	$2.9822e-11$	$2.9859e-11$	$4.6075e-04$	$3.0047e-11$	$3.0047e-11$	$7.7693e-03$	$4.7821e-11$	$2.1317e-09$
PXR	$2.7126e-15$	$1.1776e-07$	$1.3261e-06$	$3.0010e-11$	$3.0029e-11$	$3.0010e-11$	–	–	–
RXR	$1.5926e-07$	$4.7701e-09$	$1.0267e-06$	$2.2382e-02$	$3.0161e-11$	$3.0732e-19$	–	–	–
GR	$2.7440e-08$	$2.2112e-07$	$1.0078e-14$	$5.9087e-29$	$2.9487e-11$	$2.9468e-11$	$2.5832e-05$	$2.9935e-11$	$2.9935e-11$
AR	$5.9845e-11$	$2.9972e-11$	$3.3155e-11$	$9.3443e-09$	$3.0085e-11$	$3.0066e-11$	$6.6613e-03$	$2.9898e-12$	$3.0123e-11$
ERA	$4.6858e-04$	$1.8965e-15$	$3.0142e-11$	$7.5240e-12$	$3.0123e-11$	$3.0104e-11$	$1.0052e-04$	$3.0142e-11$	$5.2512e-11$
ERB	$5.0692e-03$	$3.0142e-11$	$1.1401e-16$	$4.2498e-02$	$2.9972e-11$	$2.9953e-11$	$3.9184e-06$	$3.0142e-11$	$4.6032e-10$
FXR	$2.5994e-17$	$2.9972e-11$	$8.8757e-14$	$6.0097e-03$	$3.0161e-11$	$6.8654e-18$	$3.6582e-02$	$6.0280e-08$	$3.1330e-02$
PPARD	$1.6262e-06$	$2.4376e-03$	$1.5376e-04$	$1.1017e-13$	$2.9137e-11$	$2.9100e-11$	$1.8108e-03$	$1.9491e-02$	$7.1099e-06$
PPARG	$5.3717e-16$	$3.0161e-11$	$6.7283e-06$	$2.1728e-09$	$2.9822e-11$	$2.9822e-11$	$2.7335e-11$	$4.5529e-17$	$6.2620e-23$
PPARA	$1.2437e-03$	$1.6889e-06$	$4.8922e-03$	$3.8543e-07$	$4.0779e-13$	$1.4713e-03$	–	–	–

performs better in NR-antagonist activity prediction for most NR tasks across 5-shot and 10-shot experiments as shown in Table 5. It is important to note that individual NRs that include higher class imbalance or limited data are more likely expected to negatively affect predictive performance.

This study addresses key challenges related to class imbalance and limited data when predicting the activity of compounds for specific NRs. Class imbalance, where certain classes are underrepresented, biases predictive models toward the majority class, leading to misclassification and high-variance predictions in the minority class. This is particularly problematic in drug discovery. Limited data aggravates this issue, providing insufficient samples for effective learning, which results in poor generalization, especially for NRs with just a few labeled compounds.

When quantifying the class imbalance across different NRs, we found significant disparities. For instances, PPARA reveals an extreme class imbalance with only 15 negative samples compared to 1352 positive samples for binding activity prediction tasks or 14 negative samples compared to 1084 positive samples for agonist activity prediction tasks. Similarly, RXR is heavily imbalanced in agonist activity prediction tasks with only 263 positives and 4549 negative samples. This imbalance results in lower sensitivity and higher false-negative rates for these NRs, causing predictive models to underperform on the minority class.

Moreover, limited available data for specific NRs, such as PXR and RXR, also poses a significant challenge for predictive models. PXR, for instance, has only 1327 positive samples and 3866 negative samples for binding activity predictive tasks, leading to potential overfitting issues and reduced predictive performance. Similarly, RXR also suffers from limited data with only 1006 positive and 4569 negative samples for binding activity prediction tasks, which can hinder the ability of predictive models to generalize more effectively.

Torres *et al. Journal of Cheminformatics*   (2024) 16:109

Page 14 of 19

Meta-GTNRP mitigates the challenges of low-data and class imbalance by adopting a few-shot meta-learning approach based on model-agnostic meta-learning (MAML) [31]. In meta-training, this strategy learns across few-shot tasks (k-shot) for different NRs using random support sets of ($k+, k-$) samples for training and a disjoint query set with the remaining samples for evaluation. Consequently, by considering support sets in a balanced manner with the same number of positive ($k+$) and negative ($k-$) samples, with each class is equally represented, Meta-GTNRP learns to handle imbalanced data more effectively. In addition, support sets include a small number of representative k samples for the model to learn from, making it well-suited for situations with limited available data.

Therefore, Meta-GTNRP applies a few-shot meta-learning approach to address the challenges posed by class imbalance and limited data in NR activity prediction. This meta-learning strategy facilitates the transfer of learned knowledge across NR-specific few-shot tasks in meta-training and the updated parameters are used to initialize Meta-GTNRP and generalize to new compounds for new NR tasks in meta-testing. This approach improves the performance of Meta-GTNRP, when there is limited data available and high class imbalance, without having a significant negative impact on predictive performance.

The performance of Meta-GTNRP on PPARA is notably distinct due to several challenges: the extreme class imbalance and limited data for PPARA hinders the ability to learn the minority class dependencies. In addition, the complexity introduced by the Transformer component exacerbates performance issues on such small and imbalanced data. PPARA's unique ligand-binding domain, which interacts with diverse ligands, adds complexity to this prediction task, as the model struggles to capture this diversity with limited data. To improve performance on PPARA, strategies such as data augmentation, a PPARA-specific weighted loss function, or reducing model complexity could be considered. However, these adaptations may negatively impact the performance on other NR tasks, as the model might not generalize well to data with different distributions. Thus, it is essential to balance optimizing for PPARA with maintaining performance across all NRs.

The complexity of Meta-GTNRP also poses challenges for scalability, particularly with larger datasets and more diverse NR activity prediction tasks. Future research will focus on optimizing Meta-GTNRP for scalability by exploring techniques to improve efficiency and better manage computational resources. This will potentially make the model more adaptable to new NR-based drug discovery applications. Another potential limitation is the sensitivity of Meta-GTNRP to hyperparameter settings, requiring additional fine-tuning for broader generalization to new NRs. Consequently, employing alternative hyperparameter tuning methods and conducting sensitivity analyses will be crucial for achieving robust and generalizable performance across new and diverse NR activity prediction tasks.

Despite these challenges, Meta-GTNRP has the potential to transfer the learned knowledge among multiple NRs to help in the discovery of compounds that target other NRs involved in different biological processes. Meta-GTNRP can help researchers to accelerate drug discovery, making it more efficient to identify NR-modulators with limited data, which is crucial for developing therapies for multiple diseases. This makes Meta-GTNRP a useful tool in the field of computational drug discovery, offering new opportunities for the identification of NR-based drug candidates.

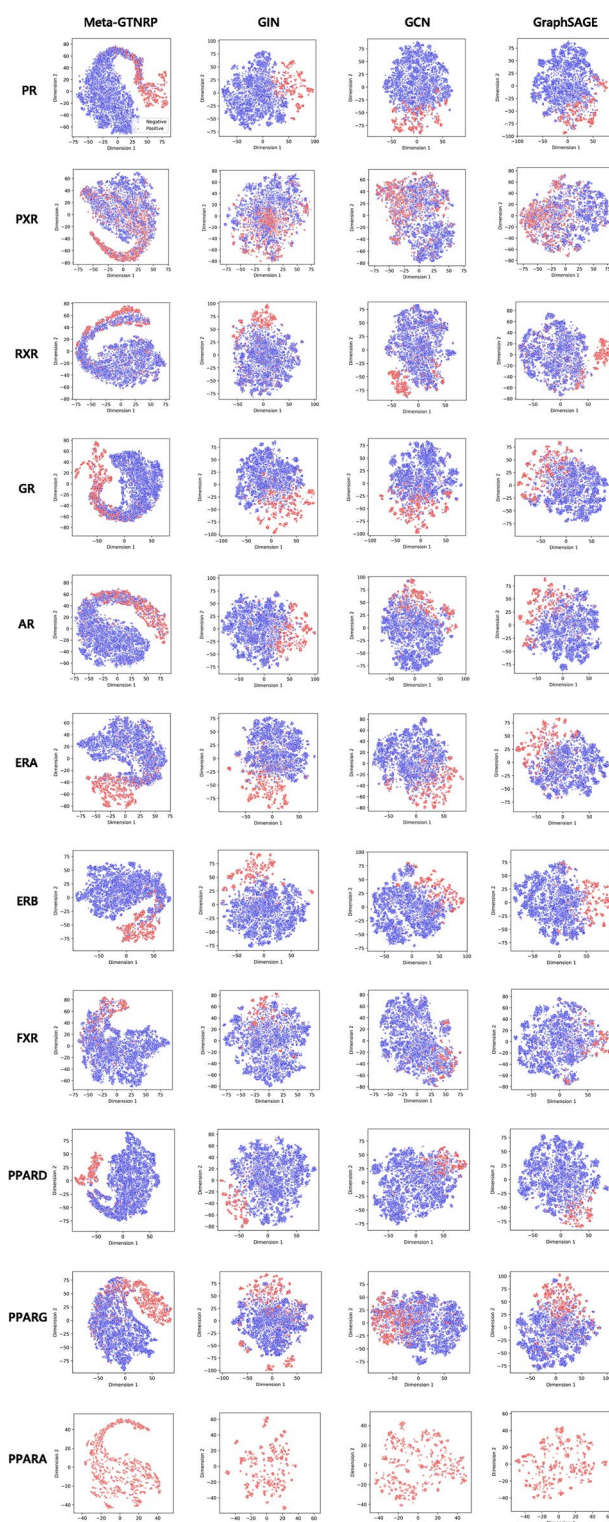## t-SNE visualization experiments in NR-binding activity prediction

To better show the effectiveness of our proposed approach in NR-binding activity prediction over graph-based baselines, we visualize the token embeddings h_T computed by Meta-GTNRP and graph embeddings h_G obtained by the GNN baselines across each one of 11 NR-binding tasks for 5-shot experiments. Therefore, we computed the t-distributed stochastic neighbor embeddings (t-SNE) [52] implemented in Scikit-learn with the following parameters: n_components = 2, perplexity = 50 and learning rate = 300 for Meta-GTNRP and standard GNN methods. The t-SNE cluster plots are displayed in Fig. 4 (5-shot) for 11 different NRs, where red dots denote positive samples and blue dots describe negative samples.

In Fig. 4, positive and negative compounds predicted by the baselines GIN, GCN and GraphSAGE are mixed up together, indicating that they have limited ability to distinguish between active and non-active binders for different NRs, respectively. Conversely, the Meta-GTNRP model obtains well-defined clusters of non-active binders progressively separating from active binders in the low-dimensional feature space for most NR-binding activity tasks. In addition, Meta-GTNRP shows clusters of negative datapoints (blue dots) closer to each other, well-separated from positive datapoints (red dots) with some overlapping to express a sense of global connectivity among active and non-active binders. Hence, for most NR-binding tasks, it is clear that Meta-GTNRP outperforms the GNN baselines in discriminating both positive and negative samples for NR-binding activity prediction.

Torres *et al. Journal of Cheminformatics*   (2024) 16:109

Page 15 of 19

Analysis of structural alerts in NR-binding activity prediction

Structural alerts (SA) are molecular substructures which help to identify key molecular fragments and functional groups with an important role in NR-binding activity. These structural fragments are often used to indicate biological activity, but can also be used to illustrate a possible mode of action for a certain compound. Therefore, the combination of structural alerts and predictive models can offer a robust solution to understand the prediction through a more interpretable approach. In this case, we analyse the results obtained for the structural alerts to identify the key substructures responsible for NR-binding activity. In this experiment, we obtain different types of molecular substructures identified using predictions of Meta-GTNRP for each specific NR considering 10 other different NR tasks in meta-training. The significant molecular substructures are identified using Bio-alerts [53], a Python package for the derivation of SAs using bioactivity data. The probability of a substructure to be a structural alert is given by the probability density function of the binomial distribution and *p*-values are calculated to assess the statistical significance using the predictions of Meta-GTNRP for each NR. The threshold frequency is set to 0.70, and the other parameters were set to default ($p - value \leq 0.05, nb \geq 50$). In Fig. 5, we show the main substructures obtained using the predictions of Meta-GTNRP for each specific NR significant in NR-binding activity 5-shot experiments. The structural features that were developed for the workflow are summarized below for each NR studied.

The structural alerts (SAs) identified for various nuclear receptors (NRs) reveal both shared and unique molecular features that influence the ligand NR-binding and NR activation. For PR, GR, and AR, several key structural elements are shared, reflecting their steroid-based nature. In PR, the 3-keto groups of most pregnane-based ligands are crucial for the ligand-receptor binding. These groups establish vital hydrogen bonds (H-bonds) with amino acid residues in the PR ligand-binding domain (LBD), making them essential for the PR activity. The removal or replacement of this 3-keto group significantly reduces the NR-binding activity [54–57]. Similarly, in GR, ketone groups play a central role, forming key H-bonds with the arginine and glutamine residues in the LBD, which enhances the NR affinity. The structural alerts for GR ligands also emphasize the importance of backbone ring structures with oxygen or nitrogen groups that contribute to more stable ligand-receptor interactions [58]. For AR, the 3-keto group and OH groups are critical for the androgenic activity, facilitating interactions with key amino acid residues such as the T877 AR side chains in the AR LBD [59]. Additionally, nonsteroidal ligands such

Fig. 4 t-SNE visualizations of token embeddings h_T generated by Meta-GTNRP and graph embeddings h_G obtained by GNN baselines for 5-shot experiments. In this figure, blue dots denote negative samples and red dots represent positive samples for each NR-binding activity task

Torres *et al. Journal of Cheminformatics*   (2024) 16:109

Page 16 of 19

Fig. 5 Analysis of structural alerts (SA) and representative molecular structures in NR-binding activity experiments. Each subfigure shows the significant substructures to determine NR-binding activity for each specific NR obtained using the predictions of the Meta-GTNRP model for 5-shot experiments

as quinolones, hydantoin, and bicalutamide derivatives can also bind effectively to the AR, allowing for flexible structural modifications used to develop potential drug candidates for the treatment of androgen-sensitive prostate cancer [60, 61].

PXR and RXR ligands exhibit distinct features. PXR ligands are characterized by scaffold ring structures and oxygen functional groups, particularly ketones, which are essential for ligand-receptor interactions via H-bonding within the PXR ligand-binding pocket [58, 62]. In contrast, RXR ligands are generally lipophilic and include

functional groups like double-bond oxygens or carboxylic acids that interact with arginine and serine residues in the RXR LBD [63]. These ligands typically contain aromatic or aliphatic ring structures, such as the cyclohexene group in retinoic acid, which aligns with the structural alerts identified for this specific NR [58].

ER ligands share similarities with those of RXR. Both ERA and ERB ligands rely on ring structures and oxygen or nitrogen functional groups to establish hydrogen bonds within the ER LBD, which are crucial for an effective ligand-receptor interaction [64]. These shared

Torres *et al. Journal of Cheminformatics*   (2024) 16:109

Page 17 of 19

structural patterns across different NR families highlight the conserved features required for the NR activation.

FXR, however, displays more diverse structural alerts. While hydrogen bonding with arginine and histidine residues via carboxylic groups is a common feature, the SAs for FXR also show functional groups like nitrogen, sulfur, and halogens connected to aromatic and aliphatic rings [58, 65]. This diversity suggests that FXR can accommodate a broader range of chemical structures compared to other NRs.

Lastly, PPAR ligands represent a distinct class of NRs. The three PPAR isoforms (PPARA, PPARD, and PPARG) prefer diaromatic scaffolds with specific functional groups tailored to each PPAR isoform activity. Unlike the steroidal NRs, the PPAR ligands often lack a steroid backbone but still maintain structural motifs necessary for NR activation. For example, the PPARG agonists are used to manage insulin resistance, while PPARA and PPARD primarily regulate glucose metabolism. In addition, fatty acid- and retinoid-like ligands with moderate PPAR affinity are also observed among the identified SAs [58, 66].

The structural alerts and their representative structures obtained for different NRs identified key molecular fragments and functional groups playing significant roles in NR-binding activity. These substructures are critical for understanding the biological activity of compounds and their potential modulator properties for specific NRs. The most significant molecular substructures reveal both similarities and differences in the structural alerts across various NRs, highlighting distinct and common features that influence the NR-binding activity.

It is important to note that the frequency of ligand structures appearing across NRs can be attributed to several factors inherent to the original data and the biological nature of the NRs. Certain substructures, such as the 3-keto groups in PR, GR, and AR, appear frequently due to their critical role in establishing stable interactions through H-bonding within the NR ligand-binding domains. Conversely, the diversity of structural alerts for FXR suggests that its ligand-binding domain can accommodate a broader range of chemical structures. This variability of structural alerts is indicative of potential selective NR modulation and highlights the adaptability of these NRs to different chemical environments.

Conclusion

Nuclear receptors (NRs) are important biological targets that modulate the binding activity of *drug-like* compounds. In this work, the goal is to take into account the individual contribution of different NRs and leverage their complementarity to predict the NR-binding properties of compounds with high sensitivity and high specificity with imbalanced and limited data, which is crucial in drug discovery.

In this paper, we propose a few-shot GNN-Transformer, Meta-GTNRP to capture local information of molecular graphs and preserve the global structure of graph embeddings using a two-module meta-learning framework for NR-binding activity prediction with limited data. This few-shot learning strategy combines the information of 11 individual predictive tasks for 11 different NRs in a joint learning procedure to predict the binding, agonist and antagonist activity with just a few labeled compounds in highly imbalanced scenarios. The results yielded by Meta-GTNRP provide strong evidence that meta-learning is a data-efficient approach to model the NR-binding activity of compounds across few-shot tasks, when there is a limited data available, without having a negative impact on predictive performance. The ROC-AUC results show that Meta-GTNRP generalizes well to new NR tasks with a smaller variance, showing a superior performance over the standard graph-based methods. Hence, the proposed Meta-GTNRP framework is an effective method to predict the NR-binding properties of compounds through an optimized meta-learning procedure, delivering faster and more robust results with just a few labeled compounds. This approach can be used to identify potential NR-based drug candidates with limited available data, making Meta-GTNRP a valuable tool to accelerate the process of drug discovery and development.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13321-024-00902-4.

> Supplementary Material 1.

Author contributions
Luis H. M. Torres: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. Joel P. Arrais: Writing - review & editing, Supervision Bernardete Ribeiro: Writing - review & editing, Supervision.

Data availability
The dataset and code scripts supporting this article are available at: https://github.com/ltorres97/Meta-GTNRP.

Declarations

Torres *et al. Journal of Cheminformatics*   (2024) 16:109

Page 18 of 19

References

1. Francis GA, Fayard E, Picard F, Auwerx J (2003) Nuclear receptors and the control of metabolism. Ann Rev Physiol. https://doi.org/10.1146/annurev.physiol.65.092101.142528
2. Huang P, Chandra V, Rastinejad F (2009) Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. Ann Rev Physiol. https://doi.org/10.1146/annurev-physiol-021909-135917
3. Khandelwal A, Krasowski MD, Reschly EJ, Sinz MW, Swaan PW, Ekins S (2008) Machine learning methods and docking for predicting human pregnane x receptor activation. Chem Res Toxicol. https://doi.org/10.1021/tx800102e
4. Kleinstreuer NC, Ceger P, Watt ED, Martin M, Houck K, Browne P, Thomas RS, Casey WM, Dix DJ, Allen D, Sakamuru S, Xia M, Huang R, Judson R (2017) Development and validation of a computational model for androgen receptor activity. Chem Res Toxicol. https://doi.org/10.1021/acs.chemrestox.6b00347
5. Mansouri K, Kleinstreuer N, Abdelaziz AM, Alberga D, Alves VM, Andersson PL, Andrade CH, Bai F, Balabin I, Ballabio D, Benfenati E, Bhhatarai B, Boyer S, Chen J, Consonni V, Farag S, Fourches D, García-Sosa AT, Gramatica P, Grisoni F, Grulke CM, Hong H, Horvath D, Hu X, Huang R, Jeliazkova N, Li J, Li X, Liu H, Manganelli S, Mangiatordi GF, Maran U, Marcou G, Martin T, Muratov E, Nguyen DT, Nicolotti O, Nikolov NG, Norinder U, Papa E, Petitjean M, Piir G, Pogodin P, Poroikov V, Qiao X, Richard AM, Roncaglioni A, Ruiz P, Rupakheti C, Sakkiah S, Sangion A, Schramm KW, Selvaraj C, Shah I, Sild S, Sun L, Taboureau O, Tang Y, Tetko IV, Todeschini R, Tong W, Trisciuzzi D, Tropsha A, Driessche GVD, Varnek A, Wang Z, Wedebye EB, Williams AJ, Xie H, Zakharov AV, Zheng Z, Judson RS (2020) Compara: collaborative modeling project for androgen receptor activity. Environ Health Perspect. https://doi.org/10.1289/EHP5580
6. Ramaprasad ASE, Smith MT, McCoy D, Hubbard AE, La Merrill MA, Durkin KA (2022) Predicting the binding of small molecules to nuclear receptors using machine learning. Brief Bioinform. https://doi.org/10.1093/bib/bbac114
7. Wang J, Lou C, Liu G, Wu Z, Tang Y (2022) Profiling prediction of nuclear receptor modulators with multi-task deep learning methods: toward the virtual screening. Brief Bioinform. https://doi.org/10.1093/bib/bbac351
8. Zhang Y, Yang Q (2018) An overview of multi-task learning. Natl Sci Rev 5:30–43. https://doi.org/10.1093/nsr/nwx105
9. Valsecchi C, Collarile M, Grisoni F, Todeschini R, Ballabio D, Consonni V (2022) Predicting molecular activity on nuclear receptors by multitask neural networks. J Chemom. https://doi.org/10.1002/cem.3325
10. Altae-Tran H, Ramsundar B, Pappu AS, Pande V (2017) Low data drug discovery with one-shot learning. ACS Centr Sci 3:283–293. https://doi.org/10.1021/acscentsci.6b00367
11. Ma J, Fong SH, Luo Y, Bakkenist CJ, Shen JP, Mourragui S, Wessels LFA, Hafner M, Sharan R, Peng J, Ideker T (2021) Few-shot learning creates predictive models of drug response that translate from high-throughput screens to individual patients. Nat Cancer 2:233–244. https://doi.org/10.1038/s43018-020-00169-2
12. Wang J, Zheng S, Chen J, Yang Y (2021) Meta learning for low-resource molecular optimization. J Chem Inform Model 61:1627–1636. https://doi.org/10.1021/acs.jcim.0c01416
13. Torres LHM, Ribeiro B, Arrais JP (2023) NFew-shot learning with transformers via graph embeddings for molecular property prediction. Expert Syst Appl 225. https://doi.org/10.1016/j.eswa.2023.120005
14. Torres LHM, Ribeiro B, Arrais JP (2024) Multi-scale cross-attention transformer via graph embeddings for few-shot molecular property prediction. Appl Soft Comput 153. https://doi.org/10.1016/j.asoc.2024.111268
15. Simões RS, Maltarollo VG, Oliveira PR, Honorio KM (2018) Transfer and multi-task learning in QSAR modeling: advances and challenges. Front Pharmacol. https://doi.org/10.3389/fphar.2018.00074
16. Cai T, Xie L, Zhang S, Chen M, He D, Badkul A, Liu Y, Namballa HK, Dorogan M, Harding WW, Mura C, Bourne PE, Xie L (2023) End-to-end sequence-structure-function meta-learning predicts genome-wide chemical-protein interactions for dark proteins. PLoS Comput Biol. https://doi.org/10.1371/journal.pcbi.1010851
17. Ren Q, Qu N, Sun J, Zhou J, Liu J, Ni L, Tong X, Zhang Z, Kong X, Wen Y, Wang Y, Wang D, Luo X, Zhang S, Zheng M, Li X (2024) Kinomemeta: meta-learning enhanced kinome-wide polypharmacology profiling. Brief Bioinform. https://doi.org/10.1093/bib/bbad461
18. Li Z, Qu N, Zhou J, Sun J, Ren Q, Meng J, Wang G, Wang R, Liu J, Chen Y, Zhang S, Zheng M, Li X (2024) KinomeMETA: a web platform for kinome-wide polypharmacology profiling with meta-learning. Nucleic Acids Res 52(W1):489–497. https://doi.org/10.1093/nar/gkae380
19. Gao Y, Gao Y, Fan Y, Zhu C, Wei Z, Zhou C, Chuai G, Chen Q, Zhang H, Liu Q (2023) Pan-peptide meta learning for t-cell receptor-antigen binding recognition. Nat Mach Intell 5:236–249. https://doi.org/10.1038/s42256-023-00619-3
20. Kearnes S, McCloskey K, Berndl M, Pande V, Riley P (2016) Molecular graph convolutions: moving beyond fingerprints. J Comput Aided Mol Des. https://doi.org/10.1007/s10822-016-9938-8
21. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE (2017) Neural message passing for quantum chemistry. Int Conf Mach Learn 3:2053–2070. https://doi.org/10.48550/arXiv.1704.01212
22. Yang K, Swanso K, Jin W, Coley C, Eiden P, Gao H, Guzman-Perez A, Hopper T, Kelley B, Mathea M, Palmer A, Settels V, Jaakkola T, Jensen K, Barzilay R (2019) Analyzing learned molecular representations for property prediction. J Chem Inform Model 59:3370–3388. https://doi.org/10.1021/acs.jcim.9b00237
23. Maziarka Tabor J (2019) Molecule-augmented attention transformer. NeurIPS Workshop Mach Learn Phys Sci. 10.48550/arXiv.2002.08264
24. Mialon G, Chen D, Selosse M, Mairal J (2021) Graphit: Encoding graph structure in transformers. 10.48550/arxiv.2106.05667
25. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N (2020) An image is worth 16x16 words: Transformers for image recognition at scale . 10.48550/arxiv.2010.11929
26. Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, Tang Y, Xiao A, Xu C, Xu Y, Yang Z, Zhang Y, Tao D (2022) A survey on vision transformer. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2022.3152247
27. Valsecchi C, Grisoni F, Motta S, Bonati L, Ballabio D (2020) Nura: a curated dataset of nuclear receptor modulators. Toxicol Appl Pharmacol. https://doi.org/10.1016/j.taap.2020.115244
28. Wang Y, Yao Q, Kwok JT, Ni LM (2020) Generalizing from a few examples: a survey on few-shot learning. ACM Comput Surv. https://doi.org/10.1145/3386252
29. Sun Q, Liu Y, Chua T.S, Schiele B (2019). Meta-transfer learning for few-shot learning. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition. pp. 403–412 . https://doi.org/10.1109/CVPR.2019.00049
30. Vinyals O, Blundell C, Lillicrap T, Kavukcuoglu K, Wierstra D (2016) Matching networks for one shot learning. https://doi.org/10.48550/arXiv.1606.04080††
31. Finn C, Abbeel P, Levine S (2017) Model-agnostic meta-learning for fast adaptation of deep networks. Int Conf Mach Learn 3:1856–1868. https://doi.org/10.48550/arXiv.1703.03400
32. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A (2018) Machine learning for molecular and materials science. Nature 559:547–555. https://doi.org/10.1038/s41586-018-0337-2
33. Defferrard M, Bresson X, Vandergheynst P (2016) Convolutional neural networks on graphs with fast localized spectral filtering. Adv Neural Inform Process Syst. https://doi.org/10.48550/arXiv.1606.09375
34. Hamilton WL, Ying R, Leskovec J (2017) Inductive representation learning on large graphs. Adv Neural Inform Process Syst. https://doi.org/10.48550/arXiv.1706.02216
35. Xu K, Jegelka S, Hu W, Leskovec J (2019). How powerful are graph neural networks? In: 7th International conference on learning representations, ICLR 2019. https://doi.org/10.48550/arXiv.1810.00826 10.48550/arXiv.1810.00826
36. Hu W, Liu B, Gomes J, Zitnik M, Liang P, Pande V, Leskovec J (2019) Strategies for pre-training graph neural networks. https://doi.org/10.48550/arxiv.1905.12265
37. Guo Z, Zhang C, Yu W, Herr J, Wiest O, Jiang M, Chawla NV (2021) Few-shot graph learning for molecular property prediction. https://doi.org/10.1145/3442381.3450112

38. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A.N, Kaiser Polosukhin I (2017) Attention is all you need. https://doi.org/10.48550/arXiv.1706.03762

39. Dosovitskiy A, Beye L, Kolesnikov, A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N (2020) An image is worth 16x16 words: transformers for image recognition at scale . https://doi.org/10.48550/arxiv.2010.11929

40. Beyer L, Zhai X, Kolesnikov A (2022) Better plain vit baselines for imagenet-1k. https://doi.org/10.48550/arxiv.2205.01580

41. Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jégou H (2020) Training data-efficient image transformers and distillation through attention. Int Conf Mach Learn 139:10347

42. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, Mutowo P, Atkinson F, Bellis LJ, Cibrián-Uhalte E, Davies M, Dedma N, Karlsson A, Magariños MP, Overington JP, Papadatos G, Smit I, Leach AR (2016) The ChEMBL database in 2017. Nucleic Acids Res 45(1):945–954. https://doi.org/10.1093/nar/gkw1074

43. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J (2015) BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res 44(D1):1045–1053. https://doi.org/10.1093/nar/gkv1072

44. Réau M, Lagarde N, Zagury JF, Montes M (2019) Nuclear receptors database including negative data (nr-dbind): a database dedicated to nuclear receptors binding data including negative data and pharmacological profile. J Med Chem 62:2894–2904. https://doi.org/10.1021/acs.jmedchem.8b01105

45. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, Leswing K, Pande V (2018) Moleculenet: a benchmark for molecular machine learning. Chem Sci 9:513–530. https://doi.org/10.1039/c7sc02664a

46. Weininger D (1988) Smiles, a chemical language and information system: 1: introduction to methodology and encoding rules. J Chem Inform Comput Sci 28:31–36. https://doi.org/10.1021/ci00057a005

47. Landrum G (2021) Rdkit: Open-source cheminformatics software

48. Li Y, Zhang K, Cao J, Timofte R, Gool L.V (2021) Localvit: Bringing locality to vision transformers. 10.48550/arXiv.2104.05707 arXiv:2104.05707 [cs.CV]

49. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, Mulbregt P, Vijaykumar A, Bardelli AP, Rothberg A, Hilboll A, Kloeckner A, Scopatz A, Lee A, Rokem A, Woods CN, Fulton C, Masson C, Häggström C, Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik DV, Olivetti E, Martin E, Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold GL, Allen GE, Lee GR, Audren H, Probst I, Dietrich JP, Silterra J, Webber JT, Slavič J, Nothman J, Buchner J, Kulick J, Schönberger JL, Miranda Cardoso JV, Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Trit K, Thoma M, Newville M, Kümmerer M, Bolingbroke M, Tartre M, Pak M, Smith NJ, Nowaczyk N, Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT, Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T, Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T, Krauss T, Upadhyay U, Halchenko YO, Vàzquez-Baeza Y (2020) Scipy 1.0: fundamental algorithms for scientific computing in python. Nat Methods 17:261–272. https://doi.org/10.1038/s41592-019-0686-2

50. D'Agostino RB (1971) An omnibus test of normality for moderate and large size samples. Biometrika 58:341–348. https://doi.org/10.1093/biomet/58.2.341

51. Pearson ES, D'agostino RB, Bowman KO (1977) Tests for departure from normality: comparison of powers. Biometrika 64:231–246. https://doi.org/10.1093/biomet/64.2.231

52. Maaten LVD, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579

53. Cortes-Ciriano I (2016) Bioalerts: a python library for the derivation of structural alerts from bioactivity and toxicity data sets. J Cheminform. https://doi.org/10.1186/s13321-016-0125-7

54. Levina IS, Kuznetsov YV, Shchelkunova TA, Zavarzin IV (2021) Selective ligands of membrane progesterone receptors as a key to studying their biological functions in vitro and in vivo. J Steroid Biochem Mol Biol. https://doi.org/10.1016/j.jsbmb.2021.105827

55. Williams SP, Sigler PB (1998) Atomic structure of progesterone complexed with its receptor. Nature. https://doi.org/10.1038/30775

56. Madauss KP, Deng SJ, Austin RJH, Lambert MH, McLay I, Pritchard J, Short SA, Stewart EL, Uings IJ, Williams SP (2004) Progesterone receptor ligand binding pocket flexibility: crystal structures of the norethindrone and mometasone furoate complexes. J Med Chem. https://doi.org/10.1021/jm030640n

57. Polikarpova AV, Maslakova AA, Levina IS, Kulikova LE, Kuznetsov YV, Guseva AA, Shchelkunova TA, Zavarzin IV, Smirnova OV (2017) Selection of progesterone derivatives specific to membrane progesterone receptors. Biochemistry (Moscow). https://doi.org/10.1134/S0006297917020055

58. Mellor CL, Steinmetz FP, Cronin MTD (2016) Using molecular initiating events to develop a structural alert based screening workflow for nuclear receptor ligands associated with hepatic steatosis. Chem Res Toxicol. https://doi.org/10.1021/acs.chemrestox.5b00480

59. Tan ME, Li J, Xu HE, Melcher K, Yong EL (2015) Androgen receptor: structure, role in prostate cancer and drug discovery. Acta Pharmacol Sin. https://doi.org/10.1038/aps.2014.18

60. Gao W, Bohl CE, Dalton JT (2005) Chemistry and structural biology of androgen receptor. Chem Rev. https://doi.org/10.1021/cr020456u

61. Nendza M, Wenzel A, Müller M, Lewin G, Simetska N, Stock F, Arning J (2016) Screening for potential endocrine disruptors in fish: evidence from structural alerts and in vitro and in vivo toxicological assays. Environ Sci Eur. https://doi.org/10.1186/s12302-016-0094-5

62. Pavek P (2016) Pregnane x receptor (pxr)-mediated gene repression and cross-talk of pxr with other nuclear receptors via coactivator interactions. Front Pharmacol. https://doi.org/10.3389/fphar.2016.00456

63. Dawson MI, Xia Z (2012) The retinoid x receptors and their ligands. Biochim Biophys Acta Mol Cell Biol Lipids. https://doi.org/10.1016/j.bbalip.2011.09.014

64. Paterni I, Granchi C, Katzenellenbogen JA, Minutolo F (2014) Estrogen receptors alpha (erα) and beta (erβ): Subtype-selective ligands and clinical potential. Steroids. https://doi.org/10.1016/j.steroids.2014.06.012

65. Xu W, Yu J (2017) Chapter 20-obesity and hepatocellular carcinoma. In: Muriel P (ed) Liver pathophysiology. Academic Press, Boston. https://doi.org/10.1016/B978-0-12-804274-8.00020-5

66. Chiazza F, Collino M (2016) Peroxisome proliferator-activated receptors (PPARs) in glucose control. Mol Nutr Diabetes. https://doi.org/10.1016/B978-0-12-801585-8.00009-9

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.