



PhosTransfer: A Deep Transfer Learning Framework for Kinase-Specific Phosphorylation Site Prediction in Hierarchy

Ying Xu¹(✉), Campbell Wilson², André Leier³, Tatiana T. Marquez-Lago³, James Whisstock², and Jiangning Song²

¹ IBM Research Australia, Melbourne, Australia
ying.jo.xu@au1.ibm.com

² Monash University, Melbourne, Australia
{campbell.wilson,james.whisstock,jiangning.song}@monash.edu

³ University of Alabama at Birmingham, Birmingham, USA
{aleier,tmarquezlago}@uabmc.edu

Abstract. Machine learning algorithms have been widely used for predicting kinase-specific phosphorylation sites. However, the scarcity of training data for specific kinases makes it difficult to train effective models for predicting their phosphorylation sites. In this paper, we propose a deep transfer learning framework, PHOSTRANSFER, for improving kinase-specific phosphorylation site prediction. It banks on the hierarchical information encoded in the kinase classification tree (KCT) which involves four levels: kinase groups, families, subfamilies and protein kinases (PKs). With PHOSTRANSFER, predictive models associated with tree nodes at higher levels, which are trained with more sufficient training data, can be transferred and reused as feature extractors for predictive models of tree nodes at a lower level. Our results indicate that models with deep transfer learning out-performed those without transfer learning for 73 out of 79 tested PKs. The positive effect of deep transfer learning is better demonstrated in the prediction of phosphosites for kinase nodes with less training data. These improved performances are further validated and explained by the visualisation of vector representations generated from hidden layers pre-trained at different KCT levels.

Keywords: Phosphorylation site prediction · Hierarchical representation · Transfer learning

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-47436-2_29) contains supplementary material, which is available to authorized users.

1 Introduction

Phosphorylation is the most common post-translational modification. It plays an important role in gene expression and cellular processes. The identification of phosphorylation sites in substrate sequences represents an important step toward a deeper understanding of cell signaling processes. Dozens of computational tools have been developed to automatically identify phosphorylation sites from protein sequences [9,12,15,19,23]. They are generally categorised into two types, general phosphorylation site (GPS) prediction and kinase-specific phosphorylation site (KPS) prediction. In GPS prediction, any Serine(S), Threonine(T) and Tyrosine(Y) site is classified either as a phosphorylation site or not, irrespective of the specific kinase that catalyses the phosphorylation. However, due to structural differences among different kinases, their target substrates have to meet kinase-specific requirements, including residue patterns. With KPS prediction, such patterns are taken into account.

It is challenging to structurally characterise kinases for the purpose of KPS prediction. Yet, kinases are classified into groups, families and subfamilies according to the sequential pattern of their catalytic domains, resulting in a kinase classification tree (KCT) [14]. In GPS 2.0, this KCT was used as heuristic for KPS prediction for the first time [22]. It provided a new hierarchical perspective for investigating kinase-specific phosphorylation. However, due to the lack of phosphosites annotated for specific kinases, few methods approached KPS prediction in such a hierarchical manner. A more recently work MUSIT-DEEP [18] implemented the idea of transfer models in GPS prediction for KPS prediction. But it did not explore the effect of transfer learning using heuristic from the hierarchy of the KCT.

In this paper, we propose a deep transfer learning framework, PHOSTRANSFER, for KPS prediction. With PHOSTRANSFER, we observed improved performance for kinases with limited annotated phosphosites. We also analysed the factors that affect its prediction performance and visualised the vector representations generated by PHOSTRANSFER at different tree levels. A benchmark is constructed and released for hierarchical KPS prediction. The source codes for PHOSTRANSFER is available at: <https://github.com/yxu132/PhosTransfer>

2 Methodology

The lack of annotated phosphosites is a common issue in building model for predicting KPS. This issue is even more problematic for deep learning models that in general involve more parameters and therefore are more likely to suffer from overfitting. In this section, we introduce PHOSTRANSFER in which deep learning models are trained and transferred for predicting sites that are phosphorylated by kinases at each level of the KCT.

2.1 Deep Transfer Learning in Hierarchy

According to Manning et al. [14], there are 8 major *kinase groups*, each of which has multiple *kinase families* and *subfamilies* with individual *PKs* clustered to

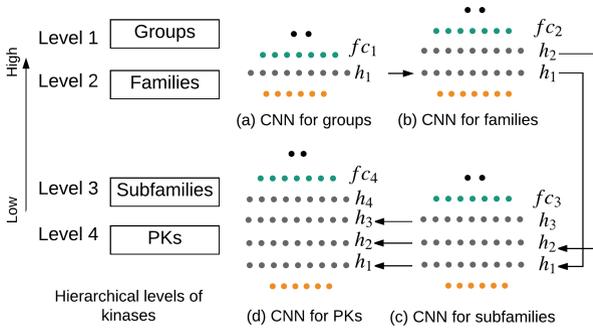


Fig. 1. Deep transfer learning for KPS prediction.

different subfamilies, forming a KCT of four levels. For simplicity, we refer to any tree node in the KCT as a *kinase node*. GPS 2.0 [22] proposed to use the annotated phosphosites of kinase nodes at lower levels to train models for their ancestor kinase nodes at higher levels. In this case, the training data is reused in a bottom-up manner. In contrast, we adopted the general idea of transfer learning [16] to transfer the knowledge learned for kinase nodes at higher levels to descendant kinase nodes at lower levels, in a top-down manner.

When implemented with deep convolutional neural network (CNN)¹, the transferable knowledge refers to the hidden layers that are learned to extract high abstractive features in the source task, which are latter reused or fine-tuned for target tasks [1]. Generally speaking, source tasks and target tasks are usually related tasks where the former has more training data and its model can be properly trained while the latter has limited training data and potentially suffers from overfitting during the process of training.

For KPS prediction, we transfer hidden layers trained for kinase nodes at higher levels to those at lower levels. We map the level t in the KCT to the hidden layer i in the generic deep CNN as in Fig. 1. Let x be the input of the deep CNN and h_i the convolutional filter at the i -th hidden layer, the binary output y_i of KPS prediction for kinase nodes at the t -th level is represented as,

$$y_t = \sigma(W_t h_t^{\theta_t}(h_{t-1}(\dots(h_1(x)))) + b_t); \theta_t \tag{1}$$

where $h_1 \dots h_{t-1}$ are convolutional filters pre-trained by phosphosites of kinase nodes at level $1 \dots t - 1$, respectively, $h_t^{\theta_t}$ is the convolutional filter for kinase nodes at level t whose parameters θ_t are to be learned, σ is the activation function at the output layer, and W_t and b_t are the weight and bias parameters for the fully connected layer, respectively. Note that only parameter θ_t , W_t and b_t are trainable for h_t while parameters for $h_1 \dots h_{t-1}$ are pre-trained and fixed.

¹ We used a sliding window of size w to extract the neighbouring residues of the target residue r_i . The local sequence segment is represented as $r_{i-w}, \dots, r_i, \dots, r_{i+w}$ with the length of $L = 2w + 1$. Here, 1-dimensional CNN is used.

2.2 Level-by-Level Representation Extraction

We trained the above deep transfer learning framework for KPS prediction using a level-by-level strategy.

Level 1 (Kinase group). For each of the 8 kinase groups in the KCT, we trained a CNN with their respective training data sets. Each of these CNNs has a single hidden layer h_1^g where $g \in \{AGC, CAMK, \dots, CMGC, STE, \text{ and } TK\}$.

Level 2 (Kinase family). For any kinase family f in group g , we reused the hidden layer h_1^g (fixed) for feature extraction and top-up a second hidden layer h_2^f (trainable) for kinase family f . For example, when $g = AGC$ kinase family $f \in \{DMPK, GRK, NDK, PKA, \dots\}$.

Level 3 (Kinase subfamily). For any subfamily s in kinase family f , we reused the hidden layer h_1^g and h_2^f (fixed) for feature extraction and top-up a third hidden layer h_3^s (trainable) for kinase subfamily s . For example, when $g = AGC$ and $f = PKC$ subfamily $s \in \{Alpha, \dots\}$.

Level 4 (Protein kinase). For PK k in kinase subfamily s , we reused the hidden layer h_1^g , h_2^f and h_3^s (fixed) and top-up a fourth hidden layer h_4^k (trainable) for protein kinase k . For instance, when $g = AGC$, $f = PKC$ and $s = Alpha$, protein kinase $k \in \{PKC\alpha, PKC\beta\}$.

In order to explore the large amount of S/T/Y phosphorylation sites that are not specifically annotated for any kinase, we added an extra level on top of the level *kinase group* in the KCT, namely, the **Level 0 (AA type)**, for which the single hidden layer h_0^{aa} is inserted and trained as the feature extractor for subsequent hidden layers h_1^g , h_2^f , h_3^s and h_4^k . Here, the amino acid type $AA \in \{S/T, Y\}$. Among the 8 kinase groups, only the model of TK is trained based on h_0^Y while the models of other groups are trained based on $h_0^{S/T}$.

2.3 Feature Vectors

Previous studies have highlighted multiple factors that are relevant to GPS/KPS prediction [9, 12, 15, 23]. In this study, we construct the feature vector V by combining the following three feature categories: a) Evolutionary-based features. We incorporated two evolutionary features calculated from the weighted observed percentage (WOP), namely the Shannon entropy and the relative entropy [9]. b) Structural-based features. We incorporated two structural properties of proteins, including protein secondary structures generated using the PSIPRED tool [6] and disordered protein states generated using DISOPRED3 [13]. c) Physicochemical properties. We incorporated Taylor’s overlapping properties and the average cumulative hydrophobicity described in PhosphoSVM [9].

The above six features were concatenated to form a 26-dimensional feature vector for each S/T/Y site.

3 Datasets

We constructed datasets by combining phosphorylation sites from UNIPROT [7] and PHOSPHO.ELM (v9) [8]. First, we downloaded 555,594 reviewed proteins from UNIPROT and extracted all the proteins that had at least one phosphosite annotation, resulting in a total of 14,458 proteins. We then collected triple-record annotations (protein identification, site position, kinase) from UNIPROT for these 14,458 proteins and removed 2,155 that were labeled as ‘by similarity’. The resulting 56,772 triple-record annotations contained 43,785 S sites, 10,397 T sites, and 4,711 Y sites, among which 7,021, 2,515 and 2,066 were annotated for specific kinases, respectively. We performed similar steps for PHOSPHO.ELM, resulting in 43,027 S sites, 9,556 T sites, and 4,723 Y sites, among which 2,961, 943, and 1,031 sites were kinase-specific, respectively.

The combined annotations are cross-referenced to the hierarchical structure introduced in Table S1 of the GPS 2.0 paper [22]. We removed kinases that had less than 15 triple-record annotations and obtained consolidated phosphorylation sites for 8 kinase groups, 50 families, 52 subfamilies and 69 PKs. In addition, to fully explore the annotated phosphorylation sites, we also include S/T and Y sites (even if they are not annotated for specific kinases), resulting in an extra *amino-acid* (AA) level on top of the *group* level. Finally, we constructed the training and independent test sets for each of the 179 groups/families/subfamilies/PKs by randomly partitioning the datasets using a size ratio of 4:1. Please refer to Table S1–S8 in supplementary materials for details dataset statistics.

4 Experiments

4.1 Experimental Settings

In order to investigate the effect of using deep transfer learning for KPS prediction, we conducted a comparison among multiple models for each kinase node in the KCT. Figure 2 demonstrates a partial tree surrounding the tree path A-B-C-D-E. For model 3', 5'–6', 8'–10', and 12'–15', there are more than one hidden layer and deep transfer learning was applied. For example, in model 13', the first hidden layer h_2^f was trained using phosphosites of family C (transferred from model 4'), based on which the second hidden layer h_3^s was trained using phosphosites of subfamily D (transferred from model 8'). Based on these two pre-trained hidden layers, the third hidden layer h_4^k was trained using phosphosites of protein kinase E. Therefore, Eq. 1 is represented as $y_4 = \sigma(W_4 h_4^k (h_3^s (h_2^f (x))) + b_4); \theta_4$.

Among the 15 compared models, the last hidden layers of models 11'–15' were trained using phosphosites of protein kinase E, which we refer to as the *direct models* of E. Since the models of its ancestors (e.g. A, B, C, and D) can also be applied to predict phosphorylation sites of protein kinase E, model 1'–10' are referred to as *indirect models* of protein kinase E.

In phosphorylation site prediction, datasets are strongly unbalanced. Therefore, we evaluated the prediction performance of PHOSTRANSFER using the area under the ROC curve (AUC), the Matthew’s coefficients of correlation (MCC) [4] and the balanced accuracy (BACC) [5].

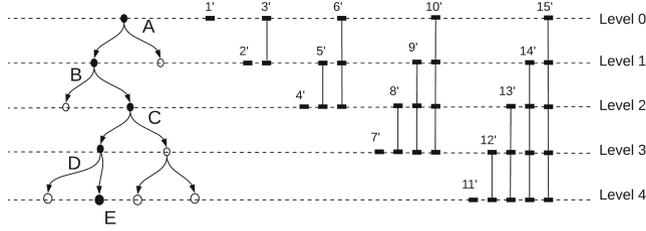


Fig. 2. Compared models for kinase nodes in hierarchies of the KCT.

4.2 Results for PKs in Different Groups

We first compared the prediction performance of direct models 11'–15' for the 69 PKs in 8 different groups. Heat maps in Fig. 3 correspond to the the normalised AUC scores of the five models, with 1, 2, 3, 4 and 5 hidden layers respectively, for each of the PKs. Better performance is in lighter colour.

Results demonstrated that for most PKs in kinase groups *AGC*, *CMGC* and *Atypical*, prediction performance was improved with the increase of hidden layers. It indicates that the hidden layers transferred from models of ancestor nodes play a positive role in improving the prediction performance. For most PKs in the group *Other*, the best performance was achieved by model 14' while the performance of model 15', which also included h_0^{aa} , ($aa = S/T$) as the first hidden layer, was inferior. This demonstrates that the phosphosites of S/T sites negatively affected the prediction performance. Considering that the PKs in group *Other* are structurally different from PKs in other groups, it makes sense that the annotated S/T sites, among which most were from other groups, did not help in improving the performance.

In group *CAMK*, *STE* and *CK1*, better performance was achieved by among model 12'–15' for most PKs, demonstrating the positive effect of deep transfer learning. However, for kinases LKB1, CHK1 and DAPK3, the best performance was achieved by model 11' that was trained solely on the phosphosites of PKs themselves. Especially for kinase LKB1, the prediction performance decreased with the increase of the number of hidden layers. It indicates that phosphosites of LKB1 may be well distinguished from phosphosites of others.

For most PKs in kinase group *TK*, the best prediction performance was achieved by models 14'. According to the normalised results, the pre-trained layer h_3^s and h_2^f played little positive effect in improving the performance. The prediction performance was only improved in model 14' when the hidden layer h_1^g trained for group *TK* was added. At the same time, adding the hidden layer h_0^{aa} ($aa = Y$) trained on amino acid Y in model 15' played a negative effect in improving the performance, which can be explained by the diverse local sequential patterns of Y phosphosites [2].

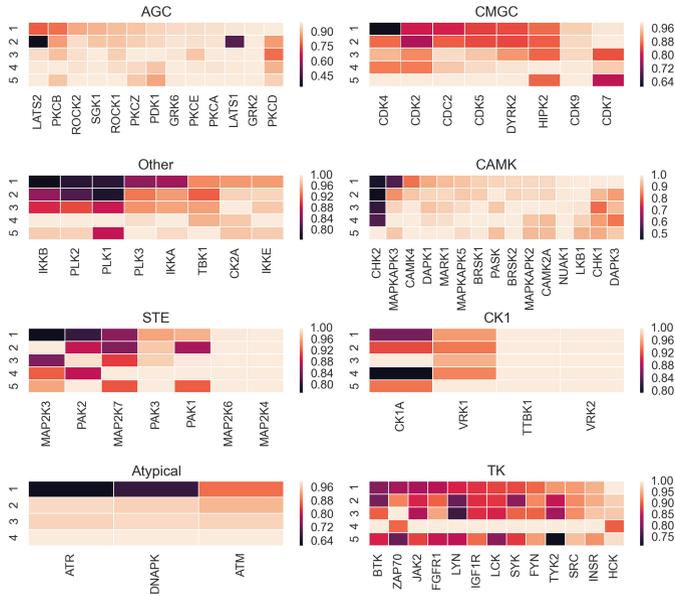


Fig. 3. Prediction performance of models with 1–5 hidden layers for PKs.

4.3 Results for Kinases with Insufficient Annotations

Transfer learning was designed in part to improve the performance for prediction tasks with insufficient training data. Among all the tested 179 kinase nodes, 84% had no more than 200 annotated phosphosites, 45% had no more than 50 annotated phosphosites and 10% had no more than 20 annotated phosphosites. Here, we investigated the performance improvements due to deep transfer learning with respect to the number of phosphosites for each kinase node.

Given the prediction performance of various kinase nodes were different even without deep transfer learning, we defined the performance improvement rate (*PIR*): Let M and M' denote the model with and without deep transfer learning respectively, the measurement *PIR* is defined as

$$PIR(M, M') = \frac{s_M - s_{M'}}{s_{M'}} \quad (2)$$

where s_M and $s'_{M'}$ represent the prediction performance of model M and M' , respectively. Here, the AUC was used as the performance evaluation score.

According to Fig. 2, models 1', 4', 7' and 11' are based on the single-layer feed-forward neural network (SLFN), to which no deep transfer learning was applied. All other models are deep transfer learning models with different numbers of pre-trained hidden layers. Therefore, models 1', 4', 7' and 11' correspond to model M' while all others correspond to model M . For each tested kinase node n , we then calculated the *PIR* between the best performing deep transfer learning model and the corresponding SLFN-based model, and plot the calculated *PIR*

score with respect to the number of phosphosites of n . Figure 4 depicts the scatter plot, with each point representing a node’s relationship between performance improvement rate and number of annotated phosphosites.

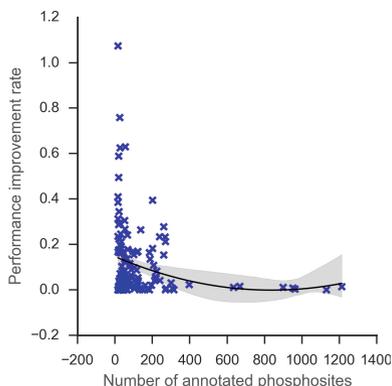


Fig. 4. The correlation between PIR and the number of annotated phosphosites.

According to the results, the prediction performance for kinases with insufficient annotated sites are more likely to improve when deep transfer learning was applied. While for kinases with more than 400 annotated sites, the prediction performance was improved by no more than 10% compared to the performance of the SLFN-based models. However, there is no guarantee for kinases with less annotated sites to have their prediction model’s performance improved, when deep transfer learning is applied. Other factors, such as the local sequential patterns, may also affect the *PIR* of deep transfer learning models. Nevertheless, the results confirmed the positive effect of deep transfer learning in predicting phosphorylation sites for kinases with insufficient annotations.

4.4 Visualisation of Layer-by-Layer Feature Extraction

In PHOSTRANSFER, hidden layers pre-trained based on phosphosites of kinase nodes in higher levels of the KCT are used as the feature extractors for models of kinase nodes in lower levels. Therefore, in model 15’, the first four hidden layers h_0^{aa} , h_1^g , h_2^f , and h_3^s were pre-trained in model 1’, 3’, 6’ and 10’, respectively. In order to evaluate these hidden layers as the feature extractor, we generated the vector representations from h_0^{aa} , h_1^g , h_2^f , h_3^s and h_4^k in model 1’, 3’, 6’, 10’ and 15’, respectively, for phosphosites of five PKs including CDK2, CDK5, GRK2, PLK1 and SRC. Figure 5 demonstrated the scattered plot of the vector representations generated by different hidden layers (feature extractors), where each vector representation was mapped to a 2-dimensional vector using t-SNE.

In Fig. 5 (a), no hidden layer was used and the distribution of phosphosites of all five PKs overlap with each other. In Fig. 5 (b), hidden layers pre-trained at

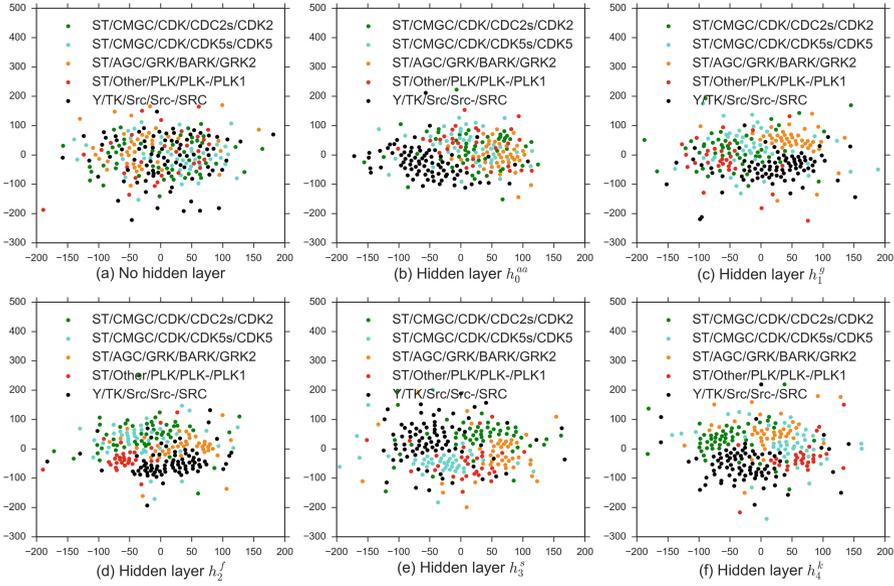


Fig. 5. t-SNE plot of vector representations generated from hidden layers.

amino-acid level was applied, and the phosphosites of SRC was clustered to the left side of the vector space, while the distributions of phosphosites of other four kinases still overlap. Here, SRC is the only PK that catalyzes the Y sites. It indicates that the hidden layer h_0^{ST} trained with S/T sites and the hidden layer h_0^Y trained with Y sites are capable of distinguishing phosphosites of kinases catalyze S/T sites and Y sites, respectively. In Fig. 5 (c), the hidden layers were trained at the *group* level, and the phosphosites of GRK2, PLK1, and CDK2/CDK5 were separated from each other. Here, GRK2, PLK1 and CDK2/CDK5 belong to group AGC, Other and CMGC, respectively. It indicates that the hidden layers pre-trained at the group level are capable of distinguishing phosphosites of kinases from different groups. In Fig. 5 (d), the distribution of phosphosites of CDK2 and CDK5 still overlap with each other, which is consistent with the KCT where both CDK2 and CDK5 belong to the same kinase family CDK. In Fig. 5 (e), the distribution of phosphosites of CDK2 and CDK5 were separated from each other, which is consistent with the KCT where CDK2 and CDK5 were classified to have different subfamily CDK2s and CDK5s. Finally, in Fig. 5 (f), phosphosites of all five kinases were clustered into five groups, which corresponds to the five PKs, respectively.

4.5 Comparison with Baseline Methods

We further compared the prediction performance of PHOSTRANSFER to 5 baseline phosphorylation site prediction methods, among which GPS 3.0 [21–23] uses hierarchical clustering to perform both GPS and KPS prediction based

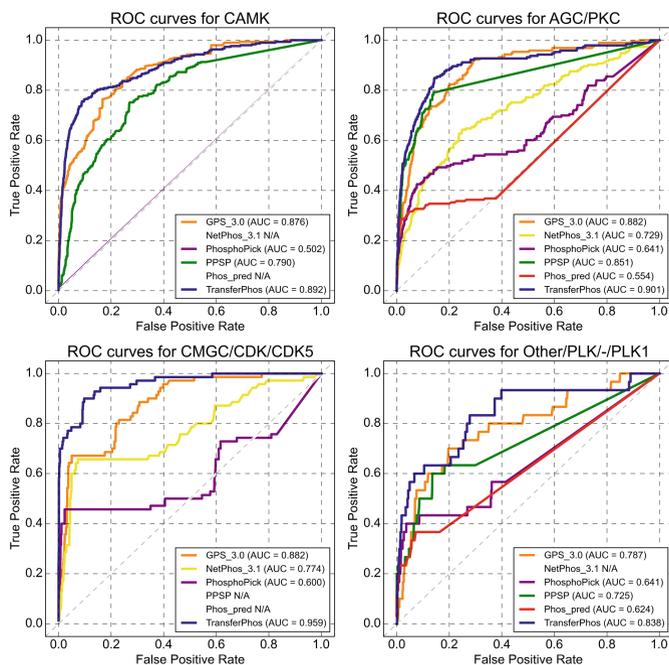


Fig. 6. Prediction performance comparison among multiple KPS prediction methods.

on similarities between local sequences around phosphosites; KINASEPHOS 2.0 [19] perform both GPS and KPS prediction based on HMM profile [10] and the coupling pattern of the surrounding sequence segment using the support vector machines; NETPHOS 3.1 [2, 3] used a neural network to combine sequential and structural motifs in a unified prediction model to predict KPS; PHOSPHOPICK [15] incorporated the protein interaction networks for KPS prediction; and PPSP [20] and PHOS_PRED [11] predicts KPS using Bayesian decision theory and random forest, respectively.

Considering the inconsistent sets of kinases that are available in the different prediction methods, we selected the group CAMK, family AGC/PKC, subfamily CMGC/CDK/CDK5, and PK Other/PLK/-/PLK1 as the representative kinase nodes from each level of the KCT. Figure 6 demonstrates the ROC plots and corresponding AUC scores of different prediction methods for each of the representative kinase nodes. According to the results below, PHOSTRANSFER achieved the best performance for the four representative kinase nodes in different levels of the KCT, when compared to that of the baseline methods.

5 Conclusions

In this study, we introduced the deep transfer learning framework PHOSTRANSFER for KPS prediction. This framework was inspired by the hierarchical

classification system of kinases and transfer learning. The basic idea is that phosphorylation sites of protein kinases within the same subfamily, family and group are likely to share similar local sequential and structural patterns, therefore models trained for higher level kinase nodes, which have more sufficient training data, can be transferred (reused) as feature extractors for lower level kinase nodes. When combined with deep learning, this idea is implemented in form of convolutional neural networks with multiple hidden layers, where each layer was trained individually based on the annotated phosphosites of kinase groups, families, subfamilies and PKs that are on the same tree path.

According to our investigation, the improved performance achieved by PHOSTRANSFER is affected by the following factors. First, the performance improvement rate of PHOSTRANSFER is related to the number of annotated phosphosites of the kinase node itself. For protein kinases with sufficient training data (more than 400 annotated phosphosites), the prediction performance was improved no more than 10% when PHOSTRANSFER was applied. This indicates that the application of PHOSTRANSFER does address the issue of overfitting during the process of model training for kinase nodes with insufficient training data. Second, the number of phosphosites of the group itself can affect the prediction performance for families, subfamilies and PKs within this group. The basic idea of PHOSTRANSFER is that the models trained for kinase nodes (especially, kinase groups) with more sufficient training data can be reused as feature extractors for kinase nodes with insufficient training data. However, if the kinase group itself does not have enough annotated phosphosites, this idea may not work properly. This explains the unsatisfactory performance of PHOSTRANSFER in groups STE and Atypical, which had the least annotated phosphosites among all kinase groups. Third, the prediction performance of PHOSTRANSFER is negatively correlated to the motif diversity of the phosphorylation sites in the training data. According to [2], PROSITE motifs could recognise only 10% of annotated Y phosphorylation sites [17], which may explain the unsatisfactory performance of PHOSTRANSFER for kinases in group TK.

References

1. Bengio, Y.: Deep learning of representations for unsupervised and transfer learning. In: Proceedings of ICML Workshop on Unsupervised and Transfer Learning, pp. 17–36 (2012)
2. Blom, N., Gammeltoft, S., Brunak, S.: Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**(5), 1351–1362 (1999)
3. Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., Brunak, S.: Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**(6), 1633–1649 (2004)
4. Boughorbel, S., Jarray, F., El-Anbari, M.: Optimal classifier for imbalanced data using Matthews correlation coefficient metric. *PLoS One* **12**(6), e0177678 (2017)
5. Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M.: The balanced accuracy and its posterior distribution. In: 2010 20th International Conference on Pattern Recognition, pp. 3121–3124. IEEE (2010)

6. Buchan, D.W., Minneci, F., Nugent, T.C., Bryson, K., Jones, D.T.: Scalable web services for the psipred protein analysis workbench. *Nucleic Acids Res.* **41**(W1), W349–W357 (2013)
7. Consortium, U.: UniProt: a hub for protein information. *Nucleic Acids Res.* **43**(D1), D204–D212 (2014)
8. Dinkel, H., et al.: Phospho. ELM: a database of phosphorylation sites-update 2011. *Nucleic Acids Res.* **39**(Suppl. 1), D261–D267 (2010)
9. Dou, Y., Yao, B., Zhang, C.: PhosphoSVM: prediction of phosphorylation sites by integrating various protein sequence attributes with a support vector machine. *Amino Acids* **46**(6), 1459–1469 (2014)
10. Eddy, S.R.: Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**(10), e1002195 (2011)
11. Fan, W., Xu, X., Shen, Y., Feng, H., Li, A., Wang, M.: Prediction of protein kinase-specific phosphorylation sites in hierarchical structure using functional information and random forest. *Amino Acids* **46**(4), 1069–1078 (2014). <https://doi.org/10.1007/s00726-014-1669-3>
12. Gao, J., Thelen, J.J., Dunker, A.K., Xu, D.: Musite, a tool for global prediction of general and kinase-specific phosphorylation sites. *Mol. Cell. Proteomics* **9**(12), 2586–2600 (2010)
13. Jones, D.T., Cozzetto, D.: DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**(6), 857–863 (2014)
14. Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S.: The protein kinase complement of the human genome. *Science* **298**(5600), 1912–1934 (2002)
15. Patrick, R., Lê Cao, K.A., Kobe, B., Bodén, M.: PhosphoPICK: modelling cellular context to map kinase-substrate phosphorylation events. *Bioinformatics* **31**(3), 382–389 (2014)
16. Pratt, L.Y.: Discriminability-based transfer between neural networks. In: *Advances in Neural Information Processing Systems*, pp. 204–211 (1993)
17. Sigrist, C.J., et al.: PROSITE: a documented database using patterns and profiles as motif descriptors. *Briefings Bioinform.* **3**(3), 265–274 (2002)
18. Wang, D., et al.: MusiteDeep: a deep-learning framework for general and kinase-specific phosphorylation site prediction. *Bioinformatics* **33**(24), 3909–3916 (2017)
19. Wong, Y.H., et al.: KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Res.* **35**(Suppl. 2), W588–W594 (2007)
20. Xue, Y., Li, A., Wang, L., Feng, H., Yao, X.: PPSP: prediction of PK-specific phosphorylation site with Bayesian decision theory. *BMC Bioinform.* **7**(1), 163 (2006)
21. Xue, Y., et al.: GPS 2.1: enhanced prediction of kinase-specific phosphorylation sites with an algorithm of motif length selection. *Protein Eng. Des. Sel.* **24**(3), 255–260 (2010)
22. Xue, Y., Ren, J., Gao, X., Jin, C., Wen, L., Yao, X.: GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol. Cell. Proteomics* **7**(9), 1598–1608 (2008)
23. Zhou, F.F., Xue, Y., Chen, G.L., Yao, X.: GPS: a novel group-based phosphorylation predicting and scoring method. *Biochem. Biophys. Res. Commun.* **325**(4), 1443–1448 (2004)