HAR_Locator: a novel protein subcellular location prediction model of immunohistochemistry images based on hybrid attention modules and residual units

Kai Zou[1], Simeng Wang[1], Ziqian Wang[1], Zhihai Zhang[1] and Fan Yang[1,2]*

[1]School of Communications and Electronics, Jiangxi Science and Technology Normal University, Nanchang, China, [2]Artificial Intelligence and Bioinformation Cognition Laboratory, Jiangxi Science and Technology Normal University, Nanchang, China

Introduction: Proteins located in subcellular compartments have played an indispensable role in the physiological function of eukaryotic organisms. The pattern of protein subcellular localization is conducive to understanding the mechanism and function of proteins, contributing to investigating pathological changes of cells, and providing technical support for targeted drug research on human diseases. Automated systems based on featurization or representation learning and classifier design have attracted interest in predicting the subcellular location of proteins due to a considerable rise in proteins. However, large-scale, fine-grained protein microscopic images are prone to trapping and losing feature information in the general deep learning models, and the shallow features derived from statistical methods have weak supervision abilities.

Methods: In this work, a novel model called HAR_Locator was developed to predict the subcellular location of proteins by concatenating multi-view abstract features and shallow features, whose advanced advantages are summarized in the following three protocols. Firstly, to get discriminative abstract feature information on protein subcellular location, an abstract feature extractor called HARnet based on Hybrid Attention modules and Residual units was proposed to relieve gradient dispersion and focus on protein-target regions. Secondly, it not only improves the supervision ability of image information but also enhances the generalization ability of the HAR_Locator through concatenating abstract features and shallow features. Finally, a multi-category multi-classifier decision system based on an Artificial Neural Network (ANN) was introduced to obtain the final output results of samples by fitting the most representative result from five subset predictors.

Results: To evaluate the model, a collection of 6,778 immunohistochemistry (IHC) images from the Human Protein Atlas (HPA) database was used to present experimental results, and the accuracy, precision, and recall evaluation indicators were significantly increased to 84.73%, 84.77%, and 84.70%, respectively, compared with baseline predictors.

KEYWORDS

hybrid attention modules, residual units, multi-view abstract features, protein subcellular location prediction, immunohistochemistry images

1 Introduction

Proteins are important biomacromolecules at the eukaryotic cellular level and are delivered to appropriate subcellular compartments where they interact with other biomolecules. Especially, several diseases have been reported to be significantly associated with the subcellular location of protein expression. For instance, Nucelolin in several subcellular locations, such as nucleolus, nucleoplasm, cytoplasm, and cell membrane impacts cancer development and therapy (Berger et al., 2015) and the loss of BRCA1 in nuclear or cytoplasmic is observed as a marker of breast tumor aggressiveness (Madjd et al., 2011). Therefore, understanding the subcellular locations of proteins is conducive to analyzing the functional principle of proteins, comprehending the complex physiological reaction process, and finding out the cancer biomarker (Kumar et al., 2014; Thul et al., 2017; Cheng et al., 2019; Shen et al., 2021). The classical solution that wet-experiment observation has been used to identify protein subcellular location, which exposes low-efficiency, time-consuming and labor-intensive with an increasing number of proteins. A resolution to these issues is highly desired, with one intriguing option being automated high-performance predictors, which have been explored to speed up the study of protein subcellular location (Yu et al., 2006; Chung, 2010; Du, 2017).

At present, protein expression patterns that are used to study protein subcellular location prediction systems fall into two types: amino acid molecular sequence and microscopic image. The former represents the similarity of subcellular locations between proteins by quantifying the intermolecular correlation of one-dimensional amino acid sequences (Shi et al., 2007; Nair and Rost, 2009; Shen and Chou, 2009; Cheng et al., 2018; Sun and Du, 2021). The latter expresses the dependability of subcellular locations between unknown and known proteins by applying advanced image processing technology to extract two-dimensional image properties (Chen and Murphy, 2006; Newberg and Murphy, 2008; Coelho et al., 2013; Xu et al., 2013). By comparison, sequence-based systems have higher accuracy but cannot reflect changes in the biochemical environment of tissues; On the other hand, some methods founded on microscopic images provided researchers with several image views of protein regions, such as shape, outline, texture, and cell distribution information, which helps to capture changes in the physiological environment to screen pathological tissues and cancer biomarkers (Uhlen et al., 2017). Particularly, with the development of machine learning and deep learning, considerable progress has been inspired by using hand-crafted features obtained from featurization or abstract features derived from representation learning in image-based models (Bengio et al., 2013; Shao et al., 2017; Xu et al., 2018).

With respect to featurization or hand-crafted features, shallow features are introduced to describe an image's global or local numerical information by statistical methods. For instance, Subcellular Location Features (SLFs) were employed to describe the shallow features of microscope images at the global level, including morphological features, Zernike moment features, Haralick texture features, and wavelet features (Newberg and Murphy, 2008). Zernike moment features, obtaining descriptive features by applying orthogonal Zernike polynomials to a unit circle of the set of complex functions, have been adopted to express rotational invariance properties on images (Boland and Murphy, 2001; Huang and Murphy, 2004; Chebira et al., 2007; Chen et al., 2007). Haralick features were adopted to quantitatively describe inertia and isotropy of intuitive pattern of protein subcellular location relying on omni-directional Gray-Level Co-occurrence Matrix (GLCM) (Xu et al., 2013). In addition, DNA distribution information, which means protein and nuclear object overlap and distance, was deployed to supplement global information since each protein image was accompanied by nuclear information (Bengio et al., 2013; Liu et al., 2019; Xue et al., 2020; Su et al., 2021; Ullah et al., 2021). An image intensity coding strategy was utilized to quantize frequency features in the frequency domain space of image wavelet transforms, which was conducive to releasing sparsity problems of immunohistochemistry (IHC) images and strengthening discriminative ability (Yang et al., 2019). To further decrypt images from multi-view, local-level information was powerfully presented to supplement global information. Local Binary Patterns (LBP), Local Ternary Patterns (LTP), and Local Quinary Patterns (LQP) are grounded on the statistic of the histogram between the center and surrounding pixels to express local texture information (Xu et al., 2013; Yang et al., 2014). Speeded-Up Robust Features (SURF) were derived from local operator features by detecting interesting points using an approximate Gaussian blob detector in both space and scale (Coelho et al., 2013). The structural relationship among cellular components as effective prior information has been considered advanced in the protein subcellular image feature space by combining Haralick and LBP features (Shao et al., 2016). Although the above image feature properties have been effectively validated, the inherently weak supervisory properties and poor distinctness of shallow features have been limiting the further improvement of model performance.

Unlike handcrafted features such as time domain features and frequency transformation features, the representation learning that learning representation of IHC images based on deep learning makes it easier to extract more supervisory and representational information (Bengio et al., 2013). With the evolution of deep learning, predictors based on Convolutional Neural Networks (CNNs) map image feature vectors into high-dimensional space through numerous nonlinear activation functions to obtain more robust representations and produce impressive performance in many fields, such as Face Recognition, Image Recognition, and Object Detection (Szegedy et al., 2013; Simonyan and Zisserman, 2014; Sun et al., 2015). Meanwhile, various abstract features from classical deep learning models trained in a fully supervised setting have consistently proved effective on generic vision tasks, such as Person Re-identification and Human Activity Recognition (Donahue et al., 2014; Sani et al., 2017; Nie et al., 2019). Consequently, feature maps in the last or penultimate layer of pre-train CNNs were extracted and incorporated into shallow features to enhance the supervisory and distinctness of protein subcellular location in the IHC images (Shao et al., 2017; Liu et al., 2019; Xue et al., 2020; Su et al., 2021; Ullah et al., 2021). It can be explained that abstract features from deep learning models describe abstract morphological local regions, edges, corners, outlines, and other digital image characteristics and serve as a helpful supplement to texture, inertia, isotropy, and the spatial ratio of shallow features. An 11-layer neural network trained in

yeast cells describes basic digital image characteristics and spate subcellular localization classes with an increasing depth of layer; besides, abstract features from the neural network have been proven useful for predicting the subcellular localization of fluorescent proteins (Pärnamaa and Parts, 2017). Abstract features obtained from a deep CNN were organized with histomorphologic information to recognize lesional coordinates of cancer tissue images (Faust et al., 2018). Abstract features from CNNs do improve the supervisory and discriminative capabilities of shallow features in digital images, but they do not account for the poor information richness and abstraction attributed to some reasons: firstly, general convolution and pooling operations are difficult to focus on the protein-target regions in bio-images; secondly, large-scale bio-images are prone to get stuck in information degradation resulting in poor performance.

In this work, a predictor named HAR_Locator based on the Hybrid Attention modules and Residual units was proposed to predict protein subcellular location in IHC images. The advancement of HAR_Locator has several attractive attributes: firstly, the features extractor known as HARnet is developed based on Hybrid Attention modules and Residual units to effectively highlight discriminating abstract features, convey the gradient information of IHC image in the network, and prevent information loss (He et al., 2016); secondly, the multi-view abstract features from different layers of HARnet were concatenated with shallow features obtained from statistical methods to improve the supervised ability of features; thirdly, the Binary Relevance (BR) and the Stepwise Discriminant Analysis (SDA) were adopted to fit feature space from concatenation; finally, a multi-category multi-classifier decision system based on ANN was utilized to output decision results from multiple confidence levels of multiple classifiers. In addition, a benchmark dataset of 6778 IHC images from HPA, including 59 proteins, was collected to verify the effectiveness of the HAR_Locator. The experimental results show that the HAR_Locator reaches 84.73%, 84.77%, and 84.70% respectively in accuracy, precision, and recall, and is significantly improved compared with other baseline predictors.

2 Materials and methods

2.1 The benchmark dataset

In this study, a benchmark dataset with a total number of 6,778 IHC images was collected from HPA (https://www.proteinatlas.org/), including 5,725 high-level stain expressions with Enhance label reliability images and 1,053 high-level stain expressions with Support label reliability images. The HPA database was created in 2005 with the goal of providing researchers with information on the expression and localization of proteins in human tissues or cells. Researchers can freely access three types of protein images, namely, immunohistochemistry (IHC) images, immunofluorescence (IF) images, and pathology (PA) images, which respectively reflect the protein information at the tissue, cell, and pathology levels. As IHC images are widely used in clinical applications and screening cancer biomarkers, this work selects IHC images as the research target. There are seven subcellular locations of proteins: Cytosol, Endoplasmic Reticulum (ER), Golgi

apparatus, Nucleoli, Mitochondria, Centrosome, and Vesicles. The dataset is shown in Table 1. Hereinto, the high-level stain expression means the protein channel with the best staining in IHC images; similarly, weaker staining levels include Medium, Low, and Not detected. Enhanced label reliability refers to the annotation of proteins being validated by one or several antibodies, and proteins with the Enhance annotation are not reported in contradiction with the existing annotation in the HPA database by published literature. Support label reliability is not validated by several antibodies like Enhanced, but the annotation of subcellular localization is described in other literature. Expect for mentioned two reliability levels, there are two other lower levels of evaluation, i.e., Improve and Uncertain.

2.2 The HAR_Locator constructed by concatenating multi-view abstract and shallow features

The algorithm framework of HAR_Locator is shown in Figure 1, and it consists of four protocols: A, getting interesting regions by preprocessing technologies for the subsequent feature extraction; B, extracting shallow features by statistical methods and abstract multi-view features from HARnet; C, the establishment of multiple classifiers through SDA and BR classifier; D, getting decision result by the ANN. The details are covered in the next section.

2.2.1 The preprocessing of IHC images

IHC images from HPA were photographed at bright field images of tissue level using an RGB camera, reflecting purple DNA in nuclei and brown protein in subcellular locations after DNA and proteins with the corresponding chemical reagent staining, whose size would be roughly $3,000 \times 3,000$ resolutions. In order to eliminate badly stained images, the empirical threshold filtering method was employed to delete those IHC images with bad staining quality (Newberg and Murphy, 2008). There are six images with poor staining, and 6,772 images were left after deletion. After that, IHC images were unmixed into protein and DNA channels by Linear Spectral Separation (LIN) (Newberg and Murphy, 2008), as shown in Figure 1A. LIN was employed to transform the IHC from RGB to HSV space for calculating the statistic histogram of hue value. The original IHC image was unmixed out of protein and DNA channels based on the color conversion matrix from the two peaks of the histogram. Moreover, to remove invalid border information from IHC images, the canny operator with two scale factors is used to obtain the protein region, and then it is mapped back to the original IHC image (Bao et al., 2005). After mentioned preprocessing stags, the protein and DNA channels were adopted to get shallow features by statistical methods and the new images with 512×512 resolutions were fed into HARnet to gain abstract features. Details are as described later.

2.2.2 Global and local shallow feature operators acting on protein and DNA channels

As a classical quantitative representation method of IHC images, SLFs based on the statistical method have been proven advanced in describing global and local information (Newberg

TABLE 1 The data volume of protein subcellular location in the dataset.

Item	Subcellular location	Number of images
Class0	Cyto.	999
Class1	ER	996
Class2	Golgi	1000
Class3	Nucl.	1000
Class4	Mito.	1000
Class5	Cent.	788
Class6	Vesi.	995

Notes: Cyto., cytosol; ER., endoplasmic reticulum; Golgi, Golgi apparatus; Nucl., nucleoli; Mito., mitochondria; Cent., centrosome; Vesi., vesicles.

and Murphy, 2008; Xu et al., 2013). The well-known Haralick features are leveraged to describe the inertia and isotropy of intuitive patterns of protein subcellular location from a global perspective (Newberg and Murphy, 2008). Specifically, 836-dimension Haralick features of the protein channel were calculated by discrete wavelet transform using the Daubechies filter, which was extracted by calculating texture feature components in the horizon, vertical, and two diagonal directions. Furthermore, 4-dimensional DNA spatial distribution features were obtained by calculating the DNA occupancy ratio in the protein and DNA channels. Moreover, the LBP algorithm is used to capture the histogram statistical

information of protein channels based on the image coding strategy of the central pixel and peripheral pixels in the local mask, including 256-dimensional features (Xu et al., 2013). Finally, the 1096-dimension (836 + 4+256 = 1096) shallow features that combine the above features are fed into the BR classifier to construct HAR_Locator[S] (Boutell et al., 2004). Above mentioned process can be shown in Figures 1B, C

2.2.3 Multi-view abstract features derived from different layers of HARnet

Numerous papers have reported that abstract features derived from deep learning models are an effective supplement to shallow features for improving classification accuracy (Shao et al., 2017; Liu et al., 2019; Xue et al., 2020; Su et al., 2021; Ullah et al., 2021); most abstract features are from the last or penultimate layer of CNNs. Furthermore, abstract features from different depths of deep learning models also expressed infusive performance (Long et al., 2020). However, some problems pose challenges for complex, fine-grained, and large-sized IHC images (SeyedJafari and Hunger, 2017), such as information loss and feature degradation with deeper depth of CNNs, which demands a treatment for capturing more robust abstract features. In this work, a deep feature extraction network named HARnet based on hybrid attention modules and residual units was designed to capture multi-view abstract features in different layers for releasing the problems mentioned. The extraction of deep features is composed of two steps: in step 1, the HARnet is trained in an end-to-end training fashion and its
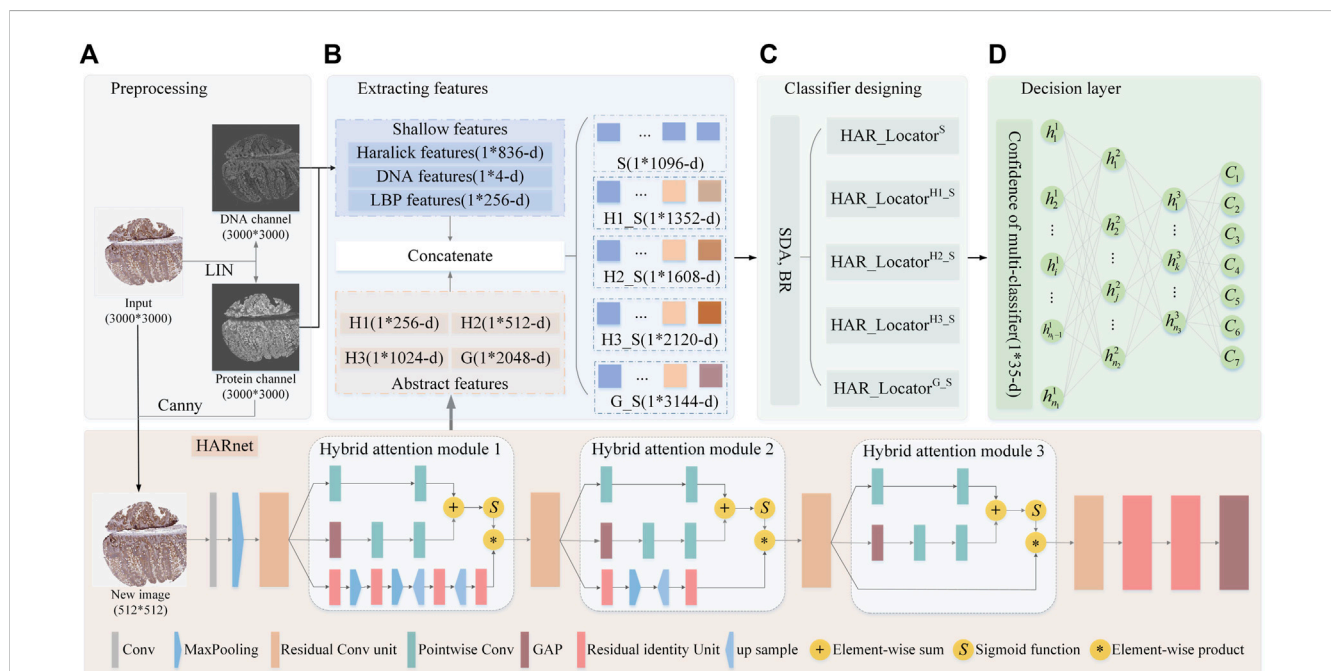

FIGURE 1
The flowchart of HAR_Locator proposed in this work. **(A)** the preprocessing of IHC images; **(B)** extracting shallow and abstract features; **(C)** fitting multiple BR classifiers in integrated features space; **(D)** outputting decision result by ANN. Abbreviation definitions: H1, Hybrid attention module 1; H2, Hybrid attention module 2; H3, Hybrid attention module 3; G, the output of the last Global Average Pooling (GAP) layer; HAR_Locator[H1–S], the predictor was constructed by concatenating H1 features with shallow features; HAR_Locator[H2–S], the predictor was constructed by concatenating H2 features with shallow features; HAR_Locator[H3–S], the predictor was constructed by concatenating H3 features with shallow features; HAR_Locator[G–S], the predictor was constructed by concatenating G features with shallow features.

architecture is shown at the bottom of Figure 1; in step 2, an IHC was described by the output of the four modules of HARnet, as part of which the basic image characteristics and abstract category properties were highlighted with an increasing depth of HARnet, i.e., three hybrid attention modules and the last Global Average Pooling (GAP) layer. The details about HARnet are described below.

2.2.4 The HARnet based on hybrid attention modules and residual units

In natural image computer vision tasks, by stacking multiple convolutions, activation, and pooling layers, CNNs with multiple nonlinear functions can capture abstract image information after iterative parameter optimization (Ji et al., 2012; Chan et al., 2015; Mezgec and KoroušićSeljak, 2017). However, simple stacked convolutional networks are prone to sticking in gradient dispersion, network degradation, and poor performance due to complexity and fine-grained IHC images. Hence, the HARnet based on the **H**ybrid **A**ttention modules and **R**esidual unit was first developed to extract abstract features, and its properties can be summarized as follows: firstly, to effectively capture discriminant features of IHC images, the hybrid attention modules that fuse bottom-up top-down feedforward structure and multi-scale channel attention are introduced to focus on protein-target regions (Wang et al., 2017; Dai et al., 2021); secondly, the backbone network of HARnet was superimposed multiple residual units, and the gradient information can be preferably transmitted (He et al., 2016); finally, in the last layer of HARnet, the GAP layer rather than Fully Connection(FC) layer is employed, which not only releases the training burden and time but also avoids overfitting problems. The abstract features came from the three hybrid attention modules and the last GAP layer. The details of HARnet are displayed at the bottom of Figure 1.

The hybrid attention module is composed of three branches, among which the top two represent global information of IHC feature maps at multi-scale channel attention, the bottom one amplifies local region information through a fusion of bottom-up top-down feedforward structure, and the equation of hybrid attention module 2 (H2) is illustrated in Eqs 1–4. The key ideas of multi-scale channel attention are as follows: firstly, the local context $P(X)$ from pointwise convolution of the top pipeline was executed to express pixel-level discriminant features by varying spatial size; secondly, the global context $G(X)$ from global average pooling of every channel was described global view in whole feature maps; thirdly, the pixel-level local context $P(X)$ and the global view of the global context $G(X)$ enrich the feature information in the hybrid attention modules, as shown in Eqs 1, 2. Furthermore, local region features $B(X)$ in the bottom branch of H2 can be broken down into two steps: in step 1, residual units R are used to remove irrelevant information and further refine image abstract information, and the response of the receptive field is enhanced by performing maxpooling layer M_1 in a 3×3 local mask; step 2, after reaching the lowest resolution of feature maps, bilinear interpolation was executed to achieve the original resolution of feature maps. The sum of local context $P(X)$ and global context $G(X)$ is then normalized to (0, 1) through the Sigmoid function and the result is used as the weight to obtain the local discriminant features of $B(X)$, as shown in Eqs 3, 4. The hybrid attention modules

1 and 3 (H1 and H3) have a structure that is similar to that of H2, with the primary exception being that H1 includes an additional bottom-up top-down feedforward layer in its bottom branch, and H3 has the identity map in its bottom branch. The process of refining and modifying of all attention modules are also displayed in Figure 1.

$$P(X) = p_2^1\left(p_1^1(X)\right) \tag{1}$$

$$G(X) = p_2^2\left(p_1^2(g(X))\right) \tag{2}$$

$$B(X) = R_2\left(I_1\left(M_1\left(R_1(X)\right)\right)\right) \tag{3}$$

$$X' = \sigma(P(X) + G(X)) {}^\star B(X) \tag{4}$$

Where X is the input, $X \in R^{C*H*W}$, X' is the output of the hybrid attention module 2, $P(X)$ is the output of top branch of the hybrid attention module 2 by pointwise convolution, $G(X)$ is the output of second branch of the hybrid attention module 2, $B(X)$ is the output of bottom branch of the hybrid attention module 2, the filter size of p_1^1 and p_1^2 is $(\frac{c}{r}, H, W)$, r is 4, the filter size of p_2^1 and p_2^2 is $(C, 1, 1)$, g is GAP layer, M_1 is MaxPooling, I_1 is interpolation function, σ is sigmoid function, R_1 and R_2 is Residual identity unit.

2.2.5 Designing multi-classifier of HAR_Locator via SDA and BR

After the above processing, the integrated features were concatenated by combing abstract and shallow features. To avoid irrelevant information or redundant features, SDA is employed to select a more discriminative feature subset, and following that, the subset feature is fed into BR. In SDA, Wilks'λ statistical method was employed to judge iteratively discriminative features in the original feature space (Huang et al., 2003). The BR classifier uses seven One-vs-Rest (OvR) Support Vector Machine (SVM) classifiers, which are effective at determining class probability (Boutell et al., 2004). There are five subset classifiers in HAR_Locator: the HAR_Locator$^{H_i–S}$ ($i = 1, 2, 3$) and HAR_Locator$^{G–S}$ are trained by concatenating shallow and abstract features extracted from three hybrid attention modules, and the last GAP layer, the HAR_Locator^S is trained by shallow feature. After that, five classifiers were generated, that is, five sets of 1*7 confidence vectors from them were output to express a sample, just as in the section of Figures 1C, D.

2.2.6 Multi-category multi-classifier decision system of HAR_Locator based on ANN

Taking up the above multi-classifier, an effective decision system is also helpful to further improve output results from the multi-classifier. In previous work, the output confidence was the mean of all classifier output probabilities, where the largest was the output label (Newberg and Murphy, 2008). However, this approach would weaken the representation of sample confidence. Based on the previous work, the ANN is designed with three hidden layers to get decision results; the hidden neurons of n_1, n_2, and n_3 are 256, 128, and 64 respectively. The ReLU activation function and Softmax function were adopted in this ANN (Cao et al., 2018). In the experiment, the output confidence of five predictors was concatenated into a 1*35-D vector to fit the network parameters. Finally, the test dataset was used to assess the performance of the HAR_Locator, as shown in the section of Figure 1D.

TABLE 2 Comparison of HAR_Locator with baseline protein location predictors.

Model	Accuracy (%)	Precision (%)	Recall (%)
Newberg and Murphy (2008)	57.95	57.14	57.31
Xu et al. (2013)	65.95	64.77	65.59
HAR_Locator	**84.73**	**84.77**	**84.70**

3 Results

In this section, the 10-fold cross-validating strategy is used to assess HAR_Locator and compare its performance to other predictors. The training process of HARnet is completed in 300 iterations using GPU parallel computing architecture and the Tensorflow-gpu2.4.0 deep learning framework. The BR classifiers were executed by Matlab software. Furthermore, the initial learning rate of the training process of HARnet is 0.001, which is multiplied by 0.1 around 60 epochs. Adam was utilized to optimize the parameters of HARnet.

3.1 HAR_Locator outperforms other baseline predictors

The baseline predictors were compared with HAR_Locator in performance evaluation indices like accuracy, precision, and recall (Newberg and Murphy, 2008; Xu et al., 2013; Jiao and Du, 2016), and the results showed that HAR_Locator ranks first among them. This was done to objectively and thoroughly verify the performance of HAR_Locator. The experimental results are shown in Table 2. From these three indexes, the predictor built by Newberg et al. through adopting Haralick features with different values of 8.0%, 7.63%, and 8.28% is inferior to the predictor built by Xu et al. through concatenating LBP features. However, the result of HAR_Locator was over 18.78%, 20.0%, and 19.11% higher than the method proposed by Xu et al. HAR_Locator significantly improves prediction performance, the main advances include the following. Firstly, the feature space composition of each image in HAR_Locator consists of two parts, namely, abstract features extracted from some modules of HARnet and shallow features based on the statistical method. After concatenating, the supervision ability of feature maps is significantly enhanced. Secondly, the HARnet built by hybrid attention modules and residual units was imported to extract abstract features, and the gradient information of IHC images can be effectively transmitted. The discriminative regions of feature maps in HARnet can be retained and enhanced, so abstract features with more supervisory and discriminative information have a better supplementary for shallow features and improve the performance of baseline predictors. Finally, a multi-category multi-classifier decision system is constructed to obtain the final output results of samples by fitting the most representative results in five basic predictors through ANN; it further improves the performance of HAR_Locator. It can be demonstrated that the HAR_locator has better experimental performance than baseline predictors based on shallow features by concatenating multi-view abstract features

obtained from HARnet with shallow features derived from statistical methods.

3.2 HAR_Locator outshined other deep networks and derived models

The performance of HARnet was confirmed when using 512*512 revolution IHC images as input and achieved the best performance in the mainstream network, such as InceptionV3, Resnet56, Densenet121, and MobilenetV3 (He et al., 2016; Szegedy et al., 2016; Huang et al., 2017; Hu et al., 2018; Howard et al., 2019). The experimental results of various CNNs were presented in Figure 2. The scatter plot of Figure 2A shows that HARnet outperformed other mainstream CNNs with 67.18% overall accuracy; multiple accuracies of different protein subcellular locations were higher than those of other CNNs, such as Golgi apparatus, Centrosome, Vesicles, and Cytosol. However, InceptionV3 achieved the last results in terms of accuracy. Among the protein subcellular locations involved, the prediction accuracy of the Nucleoli and Mitochondria exceeded that of the other subcellular locations in all models. Additionally, HARnet's accuracy in the Nucleoli and Mitochondria was 81.5%, placing it second and third, respectively, among the CNNs mentioned. Correspondingly, the Receiver Operating Characteristic curve (ROC) was visualized to show fluctuations of various mainstream CNNs stimulated by different thresholds in Figure 2B. It can be seen that HARnet ranks first with an Area Under Curve (AUC) of 0.90. This shows that HARnet has the highest permutation ratio of positive samples to negative samples and the highest true validity of the test. In model structure, the advantages of HARnet can be summarized in the following aspects for IHC images: firstly, using a smaller filter size for the convolution kernel instead of stacking simply general convolution operation would conducive to capturing fine details and keeping feature information richness; secondly, the GAP layer rather than the FC layer was adapted to map three-dimensional deep features into a one-dimensional feature vector and sent into the Softmax layer, which is prone to overcome overfitting problems due to substantially increased training parameters. Expect for these, Resnet_SE refers to the network in which the attention module is replaced by the Squeeze Excitation (SE) module from the HARnet (Hu et al., 2018). The results of Resnet56, Resnet_SE, and HARnet show that the network with added attention modules has better performance than the models without attention modules, but the hybrid attention module utilized in HARnet allows more efficient acquisition for discriminative features of protein subcellular location patterns in IHC images.
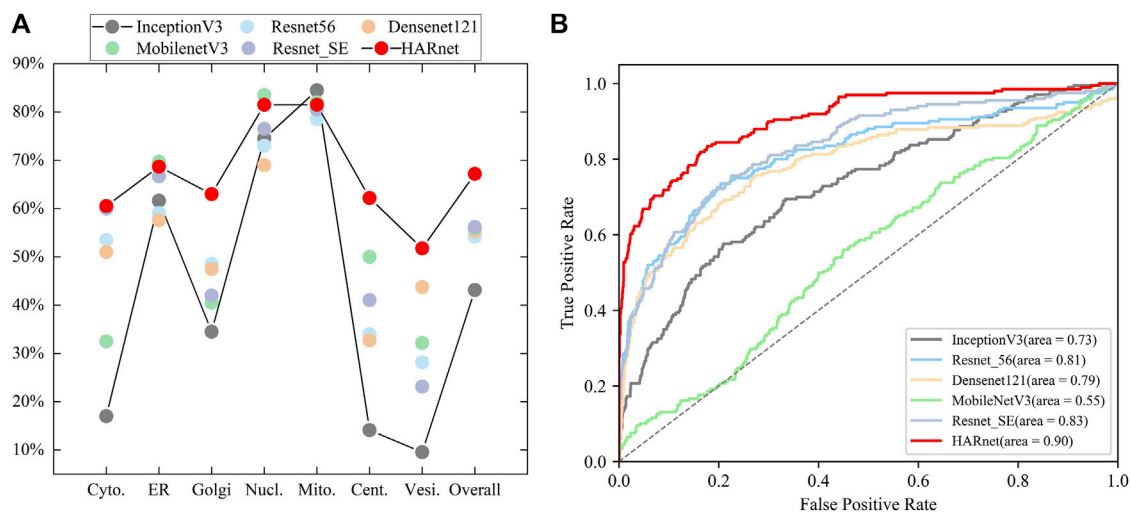
FIGURE 2
Visualization of performance evaluation in mainstream CNNs; **(A)** Single-class and overall predictive accuracy of protein subcellular location in different CNNs; **(B)** Receiver Operating Characteristic curve and Area Under Curve of various CNNs.

TABLE 3 Comparison of results between the handcraft predictors models and sub-classifier of HAR_Locator.
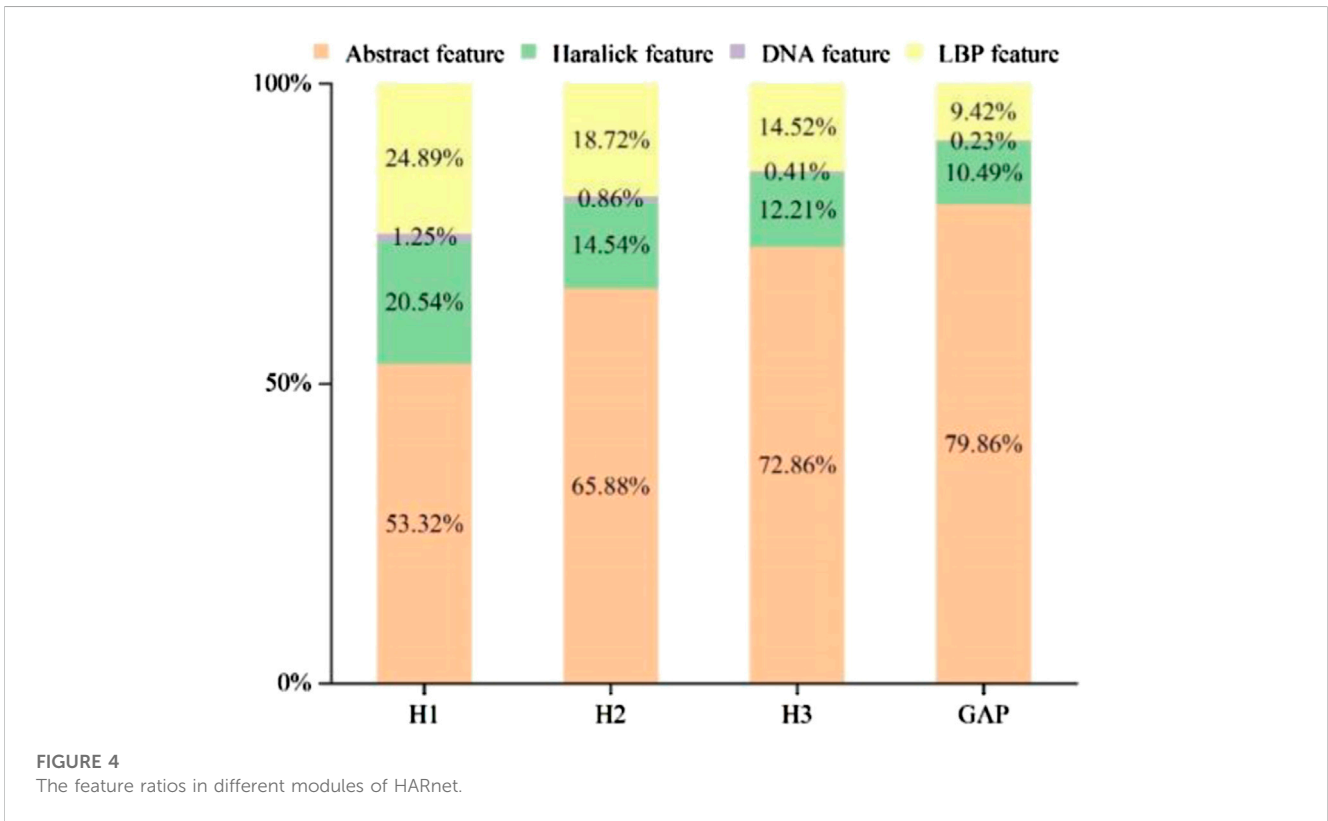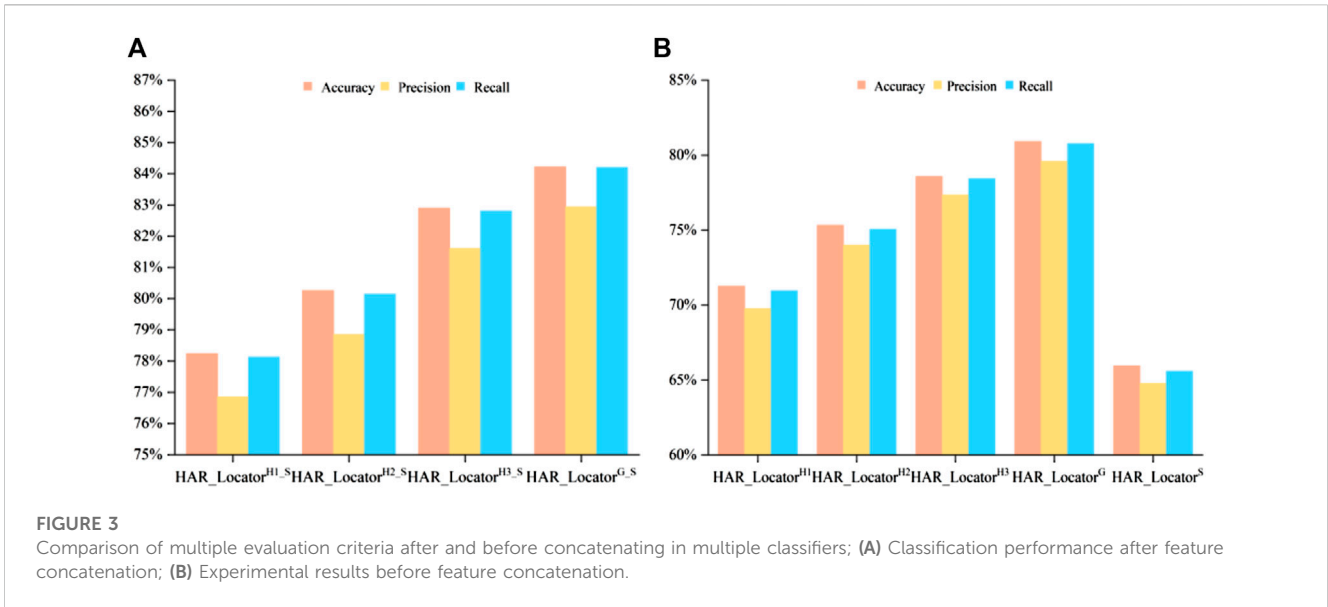
Predictor	Accuracy (%)	Precision (%)	Recall (%)
Resnet56^G_BR	59.68	59.27	59.36
Resnet56^{G-S}_BR	74.56	73.22	74.38
Resnet_SE^G_BR	57.84	57.39	57.18
Resnet_SE^{G-S}_BR	72.67	71.56	72.47
HAR_Locator^{A3-S}	82.90	81.61	82.81
HAR_Locator^{G-S}	84.22	82.94	84.20
HAR_Locator	**84.48**	**84.43**	**84.42**

In addition, some handcrafted predictors were constructed by feature fusion and BR classifier to recognize protein subcellular location. For example, the Resnet_SE^G_BR in Table 3 is built from the following pipelines: firstly, similar to HARnet, the Resnet_SE is the backbone network built by residual units and embedded with SE modules; secondly, the G feature maps were derived from the GAP layer of trained Resnet_SE; finally, the Resnet_SE^G_BR is successfully constructed by feeding the features into the SDA feature reduction dimension and BR classifier. The Resnet56^G_BR is similar to the Resnet_SE^G_BR except that the Resnet56^G_BR consists of residual units only. Then, the Resnet56^{G-S}_BR and the Resnet_SE^{G-S}_BR are akin to the Resnet_SE^G_BR; however, they do so by introducing shallow features. The derived models are inferior to the HAR_Locator with different experimental indices from Table 3. Some conclusions can be summarized as follows: firstly, it can be informed that the predictor trained by abstract features and shallow features can further improve its performance; secondly, the Resnet_SE achieved the best performance expected for HARnet in mentioned CNNs, and the derived

models are similar to these experimental results; finally, the HARnet based on hybrid attention modules and residual units can present more effective feature maps, and the advanced HAR_Locator was proven.

3.3 Comparison of abstract features and shallow features before and after concatenating

Some digital image characteristics of abstract features in HARnet, such as morphological local regions, edges, corners, and outlines, were collected to supplement the properties of shallow features. A crucial fact is that HAR_Locator performance may unquestionably be enhanced by concatenating shallow and deep features. In this part, multi-view features from HARnet were adopted to investigate the feature prediction effect under different module depths, with the results shown in Figure 3. It can be seen that the prediction performance obtained only by shallow features is inferior to that obtained by concatenating abstract and shallow features. For example, HAR_Locator^S can reach 65.95%, 64.77%, and 65.59% in accuracy, precision, and recall, while HAR_Locator^{G-S} constructed by connecting with GAP layer abstract features, can reach 84.22%, 82.94%, and 84.02%. The latter significantly improves the experimental results. The advanced performance mainly attributes to the following four aspects: firstly, the abstract features obtained from HARnet underwent multiple nonlinear function mappings and hybrid attention modules, which purposefully highlighted the subcellular location properties of the protein-target regions. Secondly, shallow features can be described by basic digital image characteristics, such as texture, inertia, isotropy, and spatial ratio, while abstract features express the structural components of the feature maps of the protein-target regions from different layers of HARnet and enrich the information

FIGURE 3
Comparison of multiple evaluation criteria after and before concatenating in multiple classifiers; **(A)** Classification performance after feature concatenation; **(B)** Experimental results before feature concatenation.

FIGURE 4
The feature ratios in different modules of HARnet.

richness of the protein subcellular location. Thirdly, as HARnet's depth increases, feature maps transition from describing fundamental features of digital images to describing abstract features of the categories to which IHC images belong. The feature maps at different depths represent protein-target regions from different views and express the different generalization abilities of the abstract features at different layers of HARnet. Finally, the weakness caused by the poor supervisory and discriminative properties of shallow features can be addressed by combining shallow features with abstract features.

4 Discussion

4.1 Analyzing the composition of subset feature spaces after SDA

A few ratios of various feature components were determined, as shown in Figure 4, to better understand the effect of the feature following SDA. Abstract features, Haralick features, LBP features, and DNA features all played a role in the integrated feature space. The figure respectively shows the feature selection distribution ratios
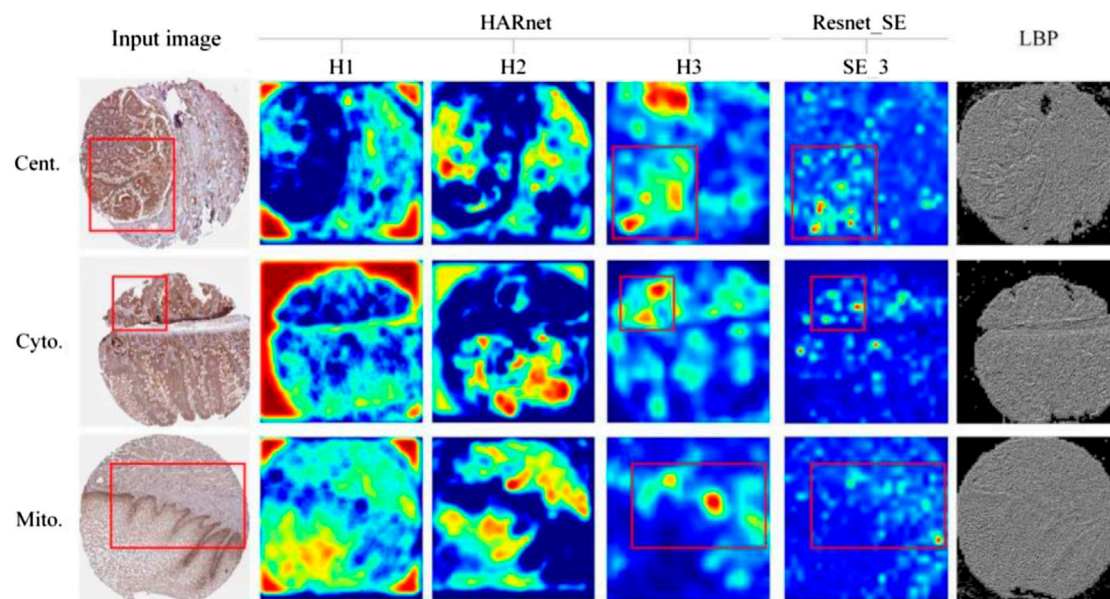
FIGURE 5
Feature maps visualization in different modules.

of HAR_Locator$^{H_i-S}$ (i = 1, 2, 3) and HAR_locator^{G-S} after SDA through 10-fold cross validation. For instance, the 3144-dimension feature of HAR_Locator^{G-S} was composed of 2048-dimension abstract features derived from the GAP layer of HARnet, 836-dimension Haralick features, 256-dimension LBP features, and 4-dimension DNA features. It reveals that shallow features make up 20.14% of the total, while abstract features account for 79.86%. In keeping with the aforementioned experimental results, we also discovered that the fraction of abstract characteristics increased as the depth of HARnet increased. Compared with the result of Table 2, the performance of some predictors constructed by shallow features, i.e., the automated framework proposed by Newberg et al. and the iLocator proposed by Xu et al., are far inferior to HAR_Locator. The complementarity between abstract features acquired by HARnet and shallow features can effectively enrich the information richness of protein subcellular location patterns and improve the model's performance. Specifically, feature maps represent the various spatial information from the IHC image in various HARnet layers. The abstract representation and identifiability of feature maps are also significantly improved with increasing HARnet depth, which enhances the HAR_Locator's generalization capabilities.

4.2 Investigating feature maps under different modules

The outputs of a few attention modules and LBP were represented in Figure 5 to help more easily comprehend the specifics of feature maps under various modules. H1, H2, and H3 are the outputs of three hybrid attention modules of HARnet, SE_3 is the output of Resnet_SE in the third module of the Resnet_SE, and the last column is texture image under LBP. As can be seen

from Figure 5, each row of the visualization represents the subcellular location pattern of one protein IHC image under different modules, namely, Centrosome, Cytosol, and Mitochondria. The second to fourth columns show that the protein-target regions in the feature map are gradually highlighted with the deepening of HARnet. Specifically, the H1 feature maps tend to express the contours and edges of protein-target regions; the H2 feature maps gradually started to focus on protein-target regions, but it was not inaccurate; and the H3 feature maps correctly identified the protein-target regions and displayed the target abstract morphology and protein highlight properties. As can be seen in the red box of Figure 5, the feature maps of SE_3 in Resnet_SE can highlight some protein-target regions, but they are weaker than H3 from HARnet. Meanwhile, the protein-target regions exhibited by the texture features in LBP are ambiguous. These investigations show that the high-level abstract features derived from HARnet are a potential addition to shallow features and can capture abstract information for boosting model performance. Also noteworthy is the fact that H3 typically captures local abstract information of the IHC image, whereas H2 typically captures global concrete information of the image. This further demonstrates how the integrated feature space collected from the HARnet's multi-view layers may be mutually complementary and enhance the HAR_Locator's robustness and generalizability.

5 Conclusion

In this study, an exact and effective model called HAR_Locator has been developed for predicting protein subcellular location. Concerning the complex and fine-grained IHC images, an integrated feature space made up of multi-view abstract features

from HARnet and shallow features derived from statistical methods was used to improve information richness, supervision, and discriminant, which is helpful to increase performance. The HARnet assembled by hybrid attention modules and residual units was designed to spotlight discriminative regions of protein-target subcellular location patterns in IHC images, and aim to capture more robust abstract features from different layers; moreover, multiple sub-classifiers constructed by different depth abstract features and shallow features were adopted to output the probability that IHC images belong to the subcellular location; finally, a decision system based on ANN has been embraced to produce a nondestructive decision result. The experimental results reveal that HAR_Locator can achieve 84.73% prediction accuracy in the benchmark dataset from HPA, which is better than other baseline models' performance. HAR_Locator participated in multi-view feature maps of HARnet and significantly improved feature richness and discriminant, in contrast to other baseline models based on shallow features and a last or penultimate abstract feature. The effectiveness of the combination of the hybrid attention modules and residual units has been verified by quantitative and qualitative analysis. Taken together, it shows that HAR_Locator is effective for accurately analyzing protein subcellular location patterns. Naturally, there is also a crucial problem that necessitates consideration. The remaining feature dimensions of abstract features after SDA were decreased substantially compared with the original abstract feature dimension, which indicates that the original feature space has great redundancy. The subsequent research strategy for protein subcellular pattern analysis will therefore involve screening more discriminant feature maps from various layers.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

FY and KZ contributed to the conception and design of the study, and KZ carried out experiments and conducted evaluations. SW and ZW helped perform statistical analysis. KZ wrote the first draft of the manuscript. KZ, ZW, and ZZ contributed to the visualization of these graphs. FY supervised the project. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bao, P., Zhang, L., and Wu, X. (2005). Canny edge detection enhancement by scale multiplication. *IEEE Trans. pattern analysis Mach. Intell.* 27 (9), 1485–1490. doi:10.1109/TPAMI.2005.173

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: a review and new perspectives. *IEEE Trans. pattern analysis Mach. Intell.* 35 (8), 1798–1828. doi:10.1109/TPAMI.2013.50

Berger, C. M., Gaume, X., and Bouvet, P. (2015). The roles of nucleolin subcellular localization in cancer. *Biochimie* 113, 78–85. doi:10.1016/j.biochi.2015.03.023

Boland, M. V., and Murphy, R. F. (2001). A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. *Bioinformatics* 17 (12), 1213–1223. doi:10.1093/bioinformatics/17.12.1213

Boutell, M. R., Luo, J., Shen, X., and Brown, C. M. (2004). Learning multi-label scene classification. *Pattern Recognit.* 37 (9), 1757–1771. doi:10.1016/j.patcog.2004.03.009

Cao, Z., Pan, X., Yang, Y., Huang, Y., and Shen, H. B. (2018). The lncLocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier. *Bioinformatics* 34 (13), 2185–2194. doi:10.1093/bioinformatics/bty085

Chan, T. H., Jia, K., Gao, S., Lu, J., Zeng, Z., and Ma, Y. (2015). PCANet: a simple deep learning baseline for image classification?[J]. *IEEE Trans. image Process.* 24 (12), 5017–5032. doi:10.1109/TIP.2015.2475625

Chebira, A., Barbotin, Y., Jackson, C., Merryman, T., Srinivasa, G., Murphy, R. F., et al. (2007). A multiresolution approach to automated classification of protein subcellular location images. *BMC Bioinforma.* 8, 210. doi:10.1186/1471-2105-8-210

Chen, S. C., and Murphy, R. F. (2006). A graphical model approach to automated classification of protein subcellular location patterns in multi-cell images[J]. *BMC Bioinforma.* 7 (1), 1–13. doi:10.1186/1471-2105-7-90

Chen, S-C., Zhao, T., Gordon, G. J., and Murphy, R. F. (2007). Automated image analysis of protein localization in budding yeast. *Bioinformatics* 23 (13), i66–i71. doi:10.1093/bioinformatics/btm206

Cheng, X., Lin, W. Z., Xiao, X., and Chou, K. C. (2019). pLoc_bal-mAnimal: predict subcellular localization of animal proteins by balancing training dataset and PseAAC. *Bioinformatics* 35 (3), 398–406. doi:10.1093/bioinformatics/bty628

Cheng, X., Xiao, X., and Chou, K. C. (2018). pLoc-mHum: predict subcellular localization of multi-location human proteins via general PseAAC to winnow out the crucial GO information. *Bioinformatics* 34 (9), 1448–1456. doi:10.1093/bioinformatics/btx711

Chung, M. J. P. (2010). Subcellular proteomics today. *Proteomics* 10 (22), 3933. doi:10.1002/pmic.201090098

Coelho, L. P., Kangas, J. D., Naik, A. W., Osuna-Highley, E., Glory-Afshar, E., Fuhrman, M., et al. (2013). Determining the subcellular location of new proteins from microscope images using local features. *Bioinformatics* 29 (18), 2343–2349. doi:10.1093/bioinformatics/btt392

Dai, Y., Gieseke, F., Oehmcke, S., Wu, Y., and Barnard, K. (2021). "Attentional feature fusion[C]," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 03-08 January 2021 (IEEE), 3560–3569.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., et al. (2014). Decaf: aa deep convolutional activation feature for generic visual recognition[C]. *Int. Conf. Mach. Learn.* 32, 647–655. doi:10.48550/arXiv.1310.1531

Du, P. F. (2017). Predicting protein submitochondrial locations: tthe 10th anniversary. *Curr. Genomics* 18 (4), 316–321. doi:10.2174/1389202918666170228143256

Faust, K., Xie, Q., Han, D., Goyle, K., Volynskaya, Z., Djuric, U., et al. (2018). Visualizing histopathologic deep learning classification and anomaly detection using nonlinear feature space dimensionality reduction[J]. *BMC Bioinforma.* 19 (1), 1–15. doi:10.1186/s12859-018-2184-4

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition[C]," in Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27-30 June 2016 (IEEE), 770–778.

Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., et al. (2019). "Searching for mobilenetv3[C]," in Proceedings of the IEEE/CVF international conference on computer vision (IEEE), 1314–1324.

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks[C]," in Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 18-23 June 2018 (IEEE), 7132–7141.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks[C]," in Proceedings of the IEEE conference on computer vision and pattern recognition (USA: IEEE), 4700–4708.

Huang, K., and Murphy, R. F. (2004). Boosting accuracy of automated classification of fluorescence microscope images for location proteomics[J]. *BMC Bioinforma.* 5 (1), 1–19. doi:10.1186/1471-2105-5-78

Huang, K., Velliste, M., and Murphy, R. F. (2003). Feature reduction for improved recognition of subcellular location patterns in fluorescence microscope images[C]. *SPIE* 4962, 307–318. doi:10.1117/12.477903

Ji, S., Xu, W., Yang, M., and Yu, K. (2012). 3D convolutional neural networks for human action recognition. *IEEE Trans. pattern analysis Mach. Intell.* 35 (1), 221–231. doi:10.1109/TPAMI.2012.59

Jiao, Y., and Du, P. (2016). Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quant. Biol.* 4, 320–330. doi:10.1007/s40484-016-0081-2

Kumar, A., Rao, A., Bhavani, S., Newberg, J. Y., and Murphy, R. F. (2014). Automated analysis of immunohistochemistry images identifies candidate location biomarkers for cancers. *Proc. Natl. Acad. Sci.* 111 (51), 18249–18254. doi:10.1073/pnas.1415120112

Liu, G. H., Zhang, B. W., Qian, G., Wang, B., Mao, B., and Bichindaritz, I. (2019). Bioimage-based prediction of protein subcellular location in human tissue with ensemble features and deep networks. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 17 (6), 1966–1980. doi:10.1109/TCBB.2019.2917429

Long, W., Yang, Y., and Shen, H. B. (2020). ImPLoc: a multi-instance deep learning model for the prediction of protein subcellular localization based on immunohistochemistry images. *Bioinformatics* 36 (7), 2244–2250. doi:10.1093/bioinformatics/btz909

Madjd, Z., Karimi, A., Molanae, S., and Asadi-Lari, M. (2011). BRCA1 protein expression level and CD44(+)Phenotype in breast cancer patients. *Cell J. (Yakhteh)* 13 (3), 155–162.

Mezgec, S., and Koroušić Seljak, B. (2017). NutriNet: a deep learning food and drink image recognition system for dietary assessment. *Nutrients* 9 (7), 657. doi:10.3390/nu9070657

Nair, R., and Rost, B. (2009). Sequence conserved for subcellular localization. *Protein Sci.* 11 (12), 2836–2847. doi:10.1110/ps.0207402

Newberg, J., and Murphy, R. F. (2008). A framework for the automated analysis of subcellular patterns in human protein atlas images. *J. proteome Res.* 7 (6), 2300–2308. doi:10.1021/pr7007626

Nie, J., Huang, L., Zhang, W., Wei, G., and Wei, Z. (2019). Deep feature ranking for Person Re-identification. *IEEE Access* 7, 15007–15017. doi:10.1109/access.2019.2894347

Pärnamaa, T., and Parts, L. (2017). Accurate classification of protein subcellular localization from high-throughput microscopy images using deep learning. *G3 Genes, Genomes, Genet.* 7 (5), 1385–1392. doi:10.1534/g3.116.033654

Sani, S., Massie, S., Wiratunga, N., and Cooper, K. (2017). Learning deep and shallow features for human activity recognition[C]. in Proceedings 10, Knowledge Science, Engineering and Management: 10th International Conference, KSEM 2017, Melbourne, VIC, Australia, August 19-20, 2017 (Berlin, Germany: Springer International Publishing), 469–482.

Seyed Jafari, S. M., and Hunger, R. (2017). IHC optical density score: aa new practical method for quantitative immunohistochemistry image analysis. *Appl. Immunohistochem. Mol. Morphol.* 25 (1), e12–e13. doi:10.1097/PAI.0000000000000370

Shao, W., Ding, Y., Shen, H. B., and Zhang, D. (2017). Deep model-based feature extraction for predicting protein subcellular localizations from bio-images. *Front. Comput. Sci.* 11 (2), 243–252. doi:10.1007/s11704-017-6538-2

Shao, W., Liu, M., and Zhang, D. (2016). Human cell structure-driven model construction for predicting protein subcellular location from biological images. *Bioinformatics* 32 (1), 114–121. doi:10.1093/bioinformatics/btv521

Shen, H. B., and Chou, K. C. (2009). A top-down approach to enhance the power of predicting human protein subcellular localization: hum-mPLoc 2.0. *Anal. Biochem.* 394 (2), 269–274. doi:10.1016/j.ab.2009.07.046

Shen, Z. A., Luo, T., Zhou, Y. K., Yu, H., and Du, P. F. (2021). NPI-GNN: ppredicting ncRNA-protein interactions with deep graph neural networks. *Briefings Bioinforma.* 22 (5), bbab051. doi:10.1093/bib/bbab051

Shi, J., Zhang, S. W., Pan, Q., and Zhou, G. P. (2007). "Amino acid composition distribution: aa novel sequence representation for prediction on protein subcellular localization[C]," in 2007 1st International Conference on Bioinformatics and Biomedical Engineering, Wuhan, China, 06-08 July 2007 (IEEE), 115–118.

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556. Available at: https://doi.org/10.48550/arXiv.1409.1556 (Accessed April 10, 2015).

Su, R., He, L., Liu, T., Liu, X., and Wei, L. (2021). Protein subcellular localization based on deep image features and criterion learning strategy. *Briefings Bioinforma.* 22 (4), bbaa313. doi:10.1093/bib/bbaa313

Sun, J., and Du, P. F. (2021). Predicting protein subchloroplast locations: tthe 10th anniversary[J]. *Front. Comput. Sci.* 15, 1–11. doi:10.1007/s11704-020-9507-0

Sun, Y., Liang, D., Wang, X., and Tang, X. (2015). Deepid3: face recognition with very deep neural networks[J]. arXiv preprint arXiv:1502.00873. Available at: https://doi.org/10.48550/arXiv.1502.00873 (Accessed February 3, 2015).

Szegedy, C., Toshev, A., and Erhan, D. (2013). "Deep neural networks for object detection[J]," in Advances in neural information processing systems (Berlin, Germany: Springer), 26.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the inception architecture for computer vision [J]," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Melbourne, VIC, Australia: IEEE), 2818–2826. doi:10.1109/CVPR.2016.308

Thul, P. J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., et al. (2017). A subcellular map of the human proteome. *Science* 356 (6340), eaal3321. doi:10.1126/science.aal3321

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., et al. (2017). A pathology atlas of the human cancer transcriptome. *Science* 357 (6352), eaan2507. doi:10.1126/science.aan2507

Ullah, M., Han, K., Hadi, F., Xu, J., Song, J., and Yu, D. J. (2021). PScL-HDeep: image-based prediction of protein subcellular location in human tissue using ensemble learning of handcrafted and deep learned features with two-layer feature selection. *Briefings Bioinforma.* 22 (6), bbab278. doi:10.1093/bib/bbab278

Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., et al. (2017). "Residual attention network for image classification[C]," in Proceedings of the IEEE conference on computer vision and pattern recognition (United States: IEEE), 3156–3164.

Xu, Y. Y., Yang, F., Zhang, Y., and Shen, H. B. (2013). An image-based multi-label human protein subcellular localization predictor (iLocator) reveals protein mislocalizations in cancer tissues. *Bioinformatics* 29 (16), 2032–2040. doi:10.1093/bioinformatics/btt320

Xu, Y. Y., Yao, L. X., and Shen, H. B. (2018). Bioimage-based protein subcellular location prediction: aa comprehensive review. *Front. Comput. Sci.* 12 (1), 26–39. doi:10.1007/s11704-016-6309-5

Xue, Z. Z., Wu, Y., Gao, Q. Z., Zhao, L., and Xu, Y. Y. (2020). Automated classification of protein subcellular localization in immunohistochemistry images to reveal biomarkers in colon cancer[J]. *BMC Bioinforma.* 21 (1), 1–15. doi:10.1186/s12859-020-03731-y

Yang, F., Liu, Y., Wang, Y., Zin, Z., and Yang, Z. (2019). MIC_Locator: aa novel image-based protein subcellular location multi-label prediction model based on multi-scale monogenic signal representation and intensity encoding strategy[J]. *BMC Bioinforma.* 20 (1), 1–21. doi:10.1186/s12859-019-3136-3

Yang, F., Xu, Y. Y., Wang, S. T., and Shen, H. B. (2014). Image-based classification of protein subcellular location patterns in human reproductive tissue by ensemble learning global and local features. *Neurocomputing* 131, 113–123. doi:10.1016/j.neucom.2013.10.034

Yu, C. S., Chen, Y. C., Lu, C. H., and Hwang, J. K. (2006). Prediction of protein subcellular localization. *Proteins Struct. Funct. Bioinforma.* 64 (3), 643–651. doi:10.1002/prot.21018