

## Structural bioinformatics

# BionoiNet: ligand-binding site classification with off-the-shelf deep neural network

Wentao Shi<sup>1</sup>, Jeffrey M. Lemoine<sup>2</sup>, Abd-El-Monsif A. Shawky<sup>2,3</sup>, Manali Singha<sup>2</sup>, Limeng Pu<sup>4</sup>, Shuangyan Yang<sup>1</sup>, J. Ramanujam<sup>1,4</sup> and Michal Brylinski<sup>2,4,\*</sup> 

<sup>1</sup>Division of Electrical and Computer Engineering, Louisiana State University, Baton Rouge, LA 70803, USA, <sup>2</sup>Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA, <sup>3</sup>Department of Cell Biology, National Research Centre, 12622 Giza, Egypt and <sup>4</sup>Center for Computation & Technology, Louisiana State University, Baton Rouge, LA 70803, USA

\*To whom correspondence should be addressed.

Associate Editor: Arne Elofsson

Received on November 4, 2019; revised on January 27, 2020; editorial decision on January 31, 2020; accepted on February 5, 2020

## Abstract

**Motivation:** Fast and accurate classification of ligand-binding sites in proteins with respect to the class of binding molecules is invaluable not only to the automatic functional annotation of large datasets of protein structures but also to projects in protein evolution, protein engineering and drug development. Deep learning techniques, which have already been successfully applied to address challenging problems across various fields, are inherently suitable to classify ligand-binding pockets. Our goal is to demonstrate that off-the-shelf deep learning models can be employed with minimum development effort to recognize nucleotide- and heme-binding sites with a comparable accuracy to highly specialized, voxel-based methods.

**Results:** We developed BionoiNet, a new deep learning-based framework implementing a popular ResNet model for image classification. BionoiNet first transforms the molecular structures of ligand-binding sites to 2D Voronoi diagrams, which are then used as the input to a pretrained convolutional neural network classifier. The ResNet model generalizes well to unseen data achieving the accuracy of 85.6% for nucleotide- and 91.3% for heme-binding pockets. BionoiNet also computes significance scores of pocket atoms, called BionoiScores, to provide meaningful insights into their interactions with ligand molecules. BionoiNet is a lightweight alternative to computationally expensive 3D architectures.

**Availability and implementation:** BionoiNet is implemented in Python with the source code freely available at: <https://github.com/CSBG-LSU/BionoiNet>.

**Contact:** [michal@brylinski.org](mailto:michal@brylinski.org)

**Supplementary information:** [Supplementary data](#) are available at *Bioinformatics* online.

## 1 Introduction

Most proteins perform their molecular functions by interacting with other biomolecules such as nucleic acids, hormones, other proteins, peptides, neurotransmitters and metabolites. Binding sites for low molecular weight molecules, called ligands, are typically concave surfaces presenting specific amino acids in a certain conformation to bind organic compounds. A comprehensive characterization of ligand-binding sites across proteomes is essential to elucidate the functions of proteins, study the molecular mechanisms of disease, and develop new pharmacological agents with improved selectivity profiles. Structural bioinformatics offers a wide collection of tools to facilitate these tasks. For instance, numerous algorithms are available to identify ligand-binding pockets in protein structures, including Fpocket (Le Guilloux *et al.*, 2009), COACH (Yang *et al.*, 2013), eFindSite (Brylinski and Feinstein, 2013) and 3DLigandSite (Wass

*et al.*, 2010), to mention a few. Other methods commonly used in computer-aided drug discovery analyze the distinctive features of binding sites, such as their druggability (Kana and Brylinski, 2019; Schmidtke and Barril, 2010), composition (Khazanov and Carlson, 2013; Skolnick *et al.*, 2015), the location of interaction hot spots (Brenke *et al.*, 2009; Ngan *et al.*, 2012) and conformational dynamics (Araki *et al.*, 2018; Stank *et al.*, 2016). One of the outstanding challenges in structural bioinformatics is to accurately classify ligand-binding sites with respect to the type of small molecules binding to these regions on protein surface.

The last decade has witnessed tremendous advances in artificial intelligence leading to solutions to many challenging problems. Examples include computer vision (Szegedy *et al.*, 2016), natural language processing (Lipton *et al.*, 2015) and other research areas (Li and Dong, 2014; Najafabadi *et al.*, 2015). Deep learning computational models, usually referred to as deep neural networks (DNNs)

are the most popular algorithms. DNNs can learn intrinsic relationship between input data and their labels under proper regularizations and are capable of generalizing to unseen data (Neyshabur et al., 2017). These models approximate sophisticated functions and effectively extract informative features from raw data such as images and speech signals (LeCun et al., 2015). The performance of various DNNs in the highly competitive ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015) is continuously improving. AlexNet (Krizhevsky et al., 2012) was one of the first DNNs to win ILSVRC in 2012 with a top-5 error of 16.4%, which was 10.8% lower than the second ranked algorithm. GoogLeNet (Szegedy et al., 2015) was the winner of 2014 ILSVRC with a top-5 error of 6.7%. Finally, ResNet (He et al., 2016) was the top performer in 2015 ILSVRC achieving a top-5 error of only 3.6% in. A unique feature of ResNet is identity mapping across layers allowing gradients to take shortcuts when passing through layers during training. These shortcuts allow bottom layers to learn more efficiently, which greatly speeds up the learning process and makes training very deep architectures feasible.

Many deep learning-based methods employ custom architectures and models depending on the type of input data, for instance, the voxel representation of spatial data requires 3D convolutional neural network (CNN) frameworks (Kamnitsas et al., 2017; Pu et al., 2019; Qureshi et al., 2019; Skalic et al., 2019). This prerequisite not only makes the development of new tools laborious but also puts responsibilities on developers, who are often domain scientists, to continuously maintain, debug, and improve their codes. An alternative approach is to use a pretrained, off-the-shelf DNN architecture, such as ResNet (He et al., 2016), OxfordNet (Simonyan and Zisserman, 2015) or GoogLeNet (Szegedy et al., 2015). In this case, the input data need to be transformed to be compatible with the model and the DNN parameters should be fine-tuned. The major advantage of this strategy is that it allows domain scientists to use state-of-the-art models yielding the best performance at significantly reduced development efforts. In this spirit, we developed Bionoi, a new method to represent the 3D structures of ligand-binding sites as 2D images. We demonstrate that this representation contains sufficient spatial physicochemical information to ensure a high accuracy of binding pocket classification with an off-the-shelf deep learning model. We call this algorithm BionoiNet because it employs ResNet, a state-of-art DNN proven to be successful in image recognition, to classify ligand-binding sites.

## 2 Materials and methods

### 2.1 Datasets

The primary dataset used in this study, TOUGH-C1, was compiled previously to evaluate the performance of DeepDrug3D (Pu et al., 2019). It comprises non-redundant sets of 1553 nucleotide-binding pockets and 596 heme-binding pockets obtained by clustering PDB structures at 80% sequence identity. The control dataset contains 1946 pockets selected from TOUGH-M1, which was used previously to benchmark several pocket matching algorithms (Govindaraj and Brylinski, 2018). Control pockets bind ligands chemically different from ATP and heme with the Tanimoto coefficient (Kawabata, 2011) of  $\leq 0.5$ . Further, control proteins share  $\leq 40\%$  sequence identity with nucleotide- and heme-binding proteins and have different structures with the template modeling score (Zhang and Skolnick, 2004) of  $\leq 0.5$ . Binding residues in nucleotide-, heme-binding and control proteins were identified with the ligand-protein contacts (LPCs) software (Sobolev et al., 1999). These datasets are combined for three binary classification tasks, nucleotide-heme, nucleotide-control and heme-control. In addition to these binding pocket datasets, we use 5000 images of cats and 5000 images of dogs from Kaggle, Google's resource for data science and machine learning (www.kaggle.com). The cat-dog dataset is employed to benchmark BionoiNet and compare the results to those obtained for the nucleotide-heme dataset. Moreover, the images of cats and dogs are more intuitive to understand than the Voronoi diagrams of binding sites making it easier to analyze and interpret the trained DNN

models. The original Kaggle images are of different sizes, therefore, we resized them to  $256 \times 256$  pixels.

### 2.2 Flowchart of BionoiNet

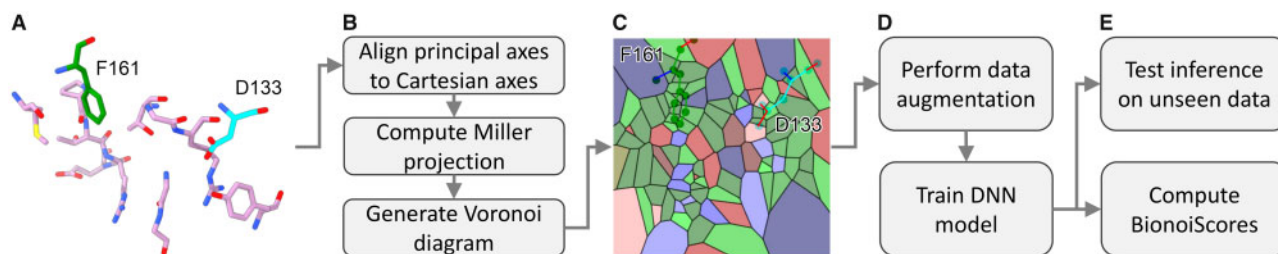
Diagram of the sequence of functions in BionoiNet is presented in Figure 1. Ligand-binding sites extracted from protein structures (Fig. 1A) are subjected to a series of transformations (Fig. 1B). First, principal axes are calculated from the coordinates of residue atoms and aligned to Cartesian axes in order to ensure that the subsequent projections are invariant under different initial orientations of pockets. Next, the Miller cylindrical projection, originally devised to portray the Earth in two dimensions (Miller, 1942), is computed to generate the 2D coordinates of binding site atoms. Given that binding sites typically have irregular structures, this particular projection was found to maintain relative distances between atoms so that the spatial arrangement of atoms is preserved after projection. Furthermore, the Miller projection avoids eclipsing atoms, i.e. projecting atoms from the opposite sides of a pocket next to one another on the 2D plane. The projected atoms are used as seeds for a Voronoi tessellation (Aurenhammer, 1991). The basic properties of Voronoi diagrams are explained in the Supplementary Information.

Cells in the resulting Voronoi diagrams can be colored either by atom/residue type or according to physicochemical properties. Figure 1C shows an example of a binding site represented by a Voronoi diagram colored by atom type with two selected residues marked as sticks. Indeed, the spatial arrangement of atoms is preserved after the projection of the 3D coordinates of residues onto a 2D plane. We found that coloring by the hydrophobicity of binding residues according to the Kyte-Doolittle index (Kyte and Doolittle, 1982) yields a high classification performance of the model, therefore this coloring scheme is used by default in BionoiNet. Ligand-binding sites are represented with this scheme as images providing the spatial physicochemical information to the DNN classifier. In order to mitigate the effects of overfitting, the dataset is augmented according to a procedure described in the Supplementary Information. A DNN model is trained in a supervised manner using the true labels of the data (Fig. 1D). Finally, the trained DNN model is applied to classify unseen data and to compute atom significance scores, called BionoiScores (Fig. 1E).

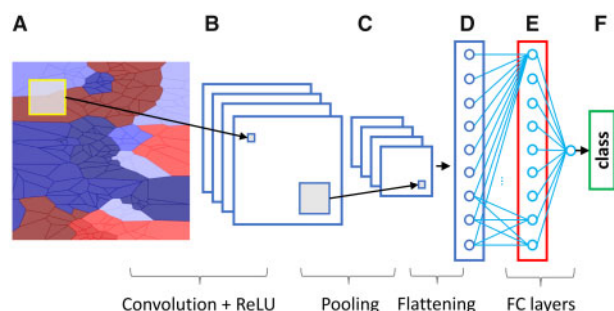
### 2.3 Convolutional neural network for image classification

CNNs, which evolved from traditional neural networks, consist of convolutional layers, pooling layers and fully connected layers (LeCun et al., 1998). Neurons in a convolutional layer are locally connected to the input feature maps and share sets of trainable parameters, namely filters. Figure 2 illustrates a simple CNN working on a Voronoi diagram of a ligand-binding site to generate a classification result. The function of convolutional layers is to scan the input image (Fig. 2A) to extract features (Fig. 2B). BionoiNet employs a rectified linear unit (ReLU) as the activation function for its convolutional layers (Jarrett et al., 2009). Pooling layers are usually added after convolutional layers to reduce the data resolution by down-sampling the spatial dimension of feature maps (Fig. 2C). Feature maps are then flattened into a vector (Fig. 2D). Fully connected layers are usually employed as the final layers (Fig. 2D) performing the high-level induction in order to generate the output (Fig. 2E).

In BionoiNet, all layers of the ResNet-18 CNN (He et al., 2016), except the last fully connected layer, are used as a module for feature extraction. A fully connected neuron followed by a sigmoid function takes the output of this feature extractor to compute classification results. The ResNet-18 feature extractor is pretrained on the ImageNet dataset, which improves both the convergence rate and the model performance by initializing the model at a good position in the parameter space. Although ImageNet and the dataset of binding sites belong to different data domains, previous work demonstrated that the lower layers of a CNN (those closer to the input image) learn similar low-level features, such as edges, even on different datasets (Zeiler and Fergus, 2014). Therefore, low-level features



**Fig. 1.** Flowchart of BionoiNet. (A) The example of a ligand binding site taken from heat resistant RNA dependent ATPase with two residues colored in green (F161) and cyan (D133). (B) A series of transformations needed to generate Voronoi diagrams. (C) The example of an atom type-based Voronoi diagram constructed for the binding site in A with individual cells colored by atom type (nitrogen—blue, carbon—green, oxygen—red) with the color intensity representing different hybridization states. Residues F161 and D133 are shown as sticks. (D) Data augmentation followed by model training and cross-validation. (E) The performance of the trained model is tested against the unseen data and significance scores for individual atoms are calculated



**Fig. 2.** Simplified convolutional neural network extracting features from ligand-binding sites. (A) An input image showing the Voronoi representation of a ligand-binding site. (B) Four feature maps constructed with four filters followed by ReLU activation. (C) Latent feature space after a pooling operation. (D) Feature maps flattened into a vector. (E) A fully connected layer performing high-level induction followed by a single fully connected neuron with a sigmoid activation function. (F) Classification results

learned by ImageNet are still useful for the CNN model to learn from the dataset of binding sites. In addition, Voronoi diagrams are composed of simple polygons that can be treated as the combinations of edges in different directions. The knowledge of these edges is presumably already learned during the pretraining of ImageNet.

## 2.4 Saliency maps and atom significance scores

In addition to pocket classification, BionoiNet computes BionoiScores, significance scores for all atoms in the ligand-binding sites, from the saliency maps of input images. These atom scores together with saliency maps provide meaningful insights and guidance for the study of ligand-binding sites in proteins. The algorithm to calculate saliency maps was originally invented for weakly supervised object localization (Simonyan *et al.*, 2014). These maps are computed by applying the gradient-based visualization, which is the generalization of deconvolutional networks (Zeiler and Fergus, 2014), to the class score. If  $I_0$  is an input image and the corresponding class score  $S_c$  is computed by feedforwarding, then the derivative of the class score with respect to the input is computed by one round of backpropagation:

$$w = \frac{\partial S_c}{\partial I} |_{I_0} \quad (1)$$

The value of the derivative  $w$  indicates how much the class score  $S_c$  changes when the values of the corresponding pixels change. For example, if the derivative has a high magnitude at position  $i, j$  with the color channel  $c$ , a small change of the pixel at  $(i, j, c)$  will result a significant change of the class score  $S_c$ , indicating that this pixel is important for the classification result. To find the important locations of an input image, a saliency map  $M$  is computed as the maximum magnitudes among three color channels of the derivative:

$$M_{ij} = \max_c w_{(i, j, c)} \quad (2)$$

A saliency map is basically a single-channel image, in which each pixel is assigned a value representing the attention of the CNN model to that pixel. A BionoiScore for each atom is computed by averaging the values of all pixels inside the corresponding polygon of a Voronoi cell.

## 3 Experimental setup

### 3.1 Setup of CNN training in BionoiNet

All layers of the ResNet-18 CNN, except the final fully connected layer, are used as the feature extractor, which is already pretrained on the ImageNet dataset. During training against the binding pocket data, the feature extractor was fine-tuned, i.e. its parameters were updated by the optimization algorithm. The batch size was set to 32. The loss function is calculated as the average binary cross-entropy of a mini-batch of training data:

$$l(x, y) = \text{mean}(l_1, l_2, \dots, l_N) \quad (3)$$

$$l_n = -[w_p y_n \times \log x_n + (1 - y_n) \times \log(1 - x_n)] \quad (4)$$

where  $N$  is the number of data in a mini-batch,  $x_n$  is the output of CNN,  $y_n$  is the value of the label and  $w_p$  is the class weight assigned to positive data which is the ratio of the number of negative instances to the number of positive instances. The class weight reduces the effect of imbalanced class numbers by assigning higher weights to the minority class in the loss function. Adam (Kingma and Ba, 2015) was used as the optimization algorithm with the learning rate set to 0.0003. The L2 weight decay was set to 0.0001 for nucleotide-control and nucleotide-heme, and 0.0003 for heme-control datasets. The optimization algorithm ran 3 epochs for cat-dog, and 8 epochs for nucleotide-heme, nucleotide-control and heme-control datasets.

Individual CNN models were trained with the above configuration to perform four binary classification tasks, cat-dog, nucleotide-heme, control-nucleotide and control-heme. Both the five-fold cross-validation and the one-fold training were conducted. The five-fold cross validation was used to assess the classification performance of the model, whereas the one-fold training was employed to test the inference on unseen data and to generate BionoiScores. The dataset was first clustered with BLASTClust (Altschul *et al.*, 1990) at the sequence identity threshold of 20%. Next, the resulting clusters were randomly assigned to five equal-sized portions creating a homology-reduced dataset in which training, testing and validation instances share no more than 20% sequence identity. In the five-fold cross validation, each portion was used as a validation set for each fold. For the one-fold training, the dataset was randomly split into three subsets, 80% for training, 10% for validation and 10% for testing. A model yielding the highest performance against the validation set was selected to assess how well it generalizes to unseen data.

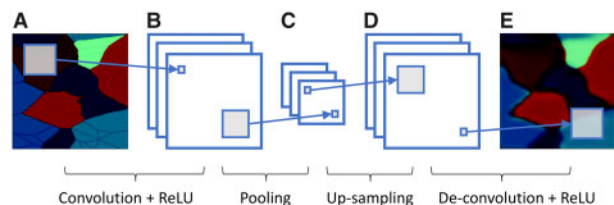


Fig. 3. Simplified stacked convolutional autoencoder. (A) An input image showing the Voronoi representation of a ligand-binding. (B) Three feature maps constructed with three filters followed by ReLU activation. (C) Latent feature space after a pooling operation. (D) Three feature maps generated from the latent feature space by an up-sampling operation. (E) An image reconstructed by a transposed convolution layer

### 3.2 Other machine learning-based classifiers

Three baseline algorithms are used to evaluate the performance of BionoiNet against nucleotide-control and heme-control datasets, a multilayer perceptron taking flattened pixels as input (Flat/MLP) and two convolutional autoencoder-based approaches (Masci *et al.*, 2011) employing random forest (Autoencoder/RF) and multilayer perceptron (Autoencoder/MLP) classifiers. A simplified convolutional autoencoder is shown in Figure 3. An autoencoder has two parts, the encoder and the decoder. The encoder takes an image as input (Fig. 3A), generates feature maps (Fig. 3B), and the corresponding latent feature vector (Fig. 3C). The decoder takes the latent feature vector generated by the encoder and attempts to recover the original image by learning a decoding function (Fig. 3D and E). The convolutional autoencoder was trained on the Voronoi representations of nucleotide-, heme-binding and control pockets in an unsupervised manner. The mean squared error over all pixels was employed as the loss function. The autoencoder was trained with the Adam optimizer for 30 epochs with 0.001 learning rate, and 0.0001 weight decay. The learning rate was decayed by the factor of 0.5 at epochs 10 and 20.

A RF and an MLP were trained on feature vectors generated by the autoencoder. Baseline classifier architectures are listed in Supplementary Table S1. For instance, for the first layer of the encoder utilizes 16  $3 \times 3$  filters, one-pixel stride and padding, leaky ReLU activation function and  $2 \times 2$  max pooling function. Leaky ReLU (Xu *et al.*, 2015) was used as the activation function for the hidden layers. Dropout (Srivastava *et al.*, 2014) with a probability of 0.5 was applied to the fully connected layers in MLP to reduce overfitting. The loss function in MLP was the same as in CNN. MLP was trained with the Adam optimizer for 100 epochs with 0.001 learning rate. RF has 1000 estimators, the minimum number of samples required to split a node is 0.0005 times total number of samples, the minimum number of samples per leaf is 0.0002 times total number of samples, and the impurity is measured with the entropy. Input images for Flat/MLP were resized to  $64 \times 64$  for memory efficiency. The loss function in Flat/MLP was the same as in CNN. Flat/MLP was trained by the Adam optimizer for 100 epochs with 0.0003 learning rate and the L2 weight decay was set to 0.0001.

## 4 Results

### 4.1 Classification performance on cat-dog and heme-nucleotide datasets

BionoiNet is first tested against a standard Kaggle dataset using evaluation metrics described in the Supplementary Information. Table 1 shows that the cross-validated performance against the cat-dog dataset is very high, confirming that the model was correctly set up and trained. The performance of BionoiNet on the heme-nucleotide dataset is slightly lower, e.g. the accuracy is 0.944 versus 0.991 for the cat-dog dataset. This can be expected because pretraining on ImageNet gives the ResNet-18 feature extractor an extensive prior knowledge of cats and dogs, therefore, the model performs extremely well on the cat-dog task. A slight decrease in the

Table 1. Performance of BionoiNet on cat-dog and nucleotide-heme datasets

Dataset	ACC	PPV	TPR	MCC	AUC
Cat-dog	0.991	0.990	0.990	0.980	0.995
Nucleotide-heme	0.944	0.973	0.950	0.865	0.982

Note: Accuracy (ACC), precision (PPV), recall (TPR), the Matthews correlation coefficient (MCC) and the area under the ROC curve (AUC) are calculated from 10-fold cross-validation.

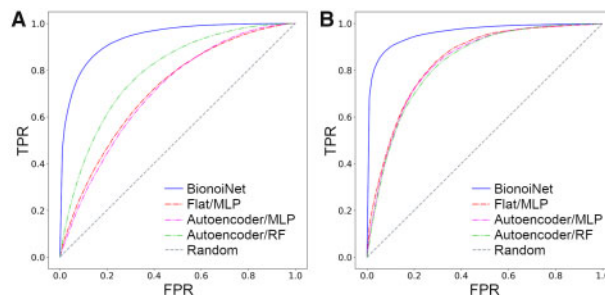


Fig. 4. Cross-validated performance of several algorithms to classify pockets. Averaged ROC curves are shown for (A) nucleotide-control and (B) heme-control datasets. BionoiNet is compared to a multi-layer perceptron with flattened pixels as input (Flat/MLP) and two autoencoders employing random forest (RF) and MLP classifiers. The diagonal represents the performance of a random classifier. FPR, false positive rate; TPR, true positive rate

performance of BionoiNet is due to dissimilarities between ImageNet and the dataset of ligand-binding sites. The prior knowledge of ImageNet is simply less helpful to classify protein pockets because those two datasets belong to different data domains. Nonetheless, the overall performance of BionoiNet against the nucleotide-heme dataset is still very high, indicating that ResNet-18 effectively learned distinct features of nucleotide- and heme-binding pockets.

### 4.2 Classification performance on nucleotide-control and heme-control datasets

Next, we evaluate the classification performance of BionoiNet for more challenging tasks, the recognition of nucleotide- and heme-binding sites (positives) against a large and diverse dataset of control pockets binding a variety of ligands (negatives). Since these tasks contain imbalanced classes, 1553 nucleotide-binding, 596 heme-binding and 1946 control instances, and the identification of positive instances is more important, precision, recall and the Matthews correlation coefficient (MCC) are the most useful metrics to evaluate the classification performance. Figure 4 and Table 2 show that BionoiNet outperforms baseline classifiers on both datasets in terms of these metrics. For the nucleotide-control dataset, it yields a precision of 0.837 and a recall of 0.843 showing a high capability to correctly identify positive instances. In contrast, the precision of all baseline classifiers is significantly lower, ranging from 0.591 for Autoencoder/MLP to 0.673 for Autoencoder/RF. BionoiNet also achieves the highest MCC of 0.712, which measures the overall classification performance. In addition, the corresponding receiver operating characteristics (ROC) plots for the nucleotide-control dataset are presented in Fig. 4A. BionoiNet has the best area under the curve (AUC) of 0.935, whereas Autoencoder/RF outperformed both MLP-based algorithms. The performance of BionoiNet is compared to those of baseline classifiers on the heme-control dataset in Figure 4B and Table 2. BionoiNet clearly outperformed baseline classifiers with an accuracy of 0.913, a recall of 0.882, a precision of 0.777 and an MCC of 0.771. ROC plots for the heme-control dataset are shown in Figure 4B. BionoiNet yields the best ROC curve with an



**Table 2.** Performance of BionoiNet and baseline classifiers on nucleotide-control and heme-control datasets

Algorithm	Nucleotide-control					Heme-control				
	ACC	PPV	TPR	MCC	AUC	ACC	PPV	TPR	MCC	AUC
BionoiNet	0.856	0.837	0.843	0.712	0.935	0.913	0.777	0.882	0.771	0.960
Flat/MLP	0.655	0.594	0.706	0.320	0.715	0.747	0.478	0.822	0.472	0.845
Autoencoder/RF	0.721	0.673	0.723	0.441	0.790	0.795	0.557	0.597	0.441	0.833
Autoencoder/MLP	0.650	0.591	0.687	0.306	0.708	0.744	0.475	0.815	0.465	0.839

Note: Accuracy (ACC), precision (PPV), recall (TPR), the Matthews correlation coefficient (MCC) and the area under the ROC curve (AUC) are calculated from 10-fold cross-validation. Baseline classifiers are a multilayer perceptron with flattened pixels as input (Flat/MLP) and two autoencoders employing machine learning with random forest (RF) and MLP.

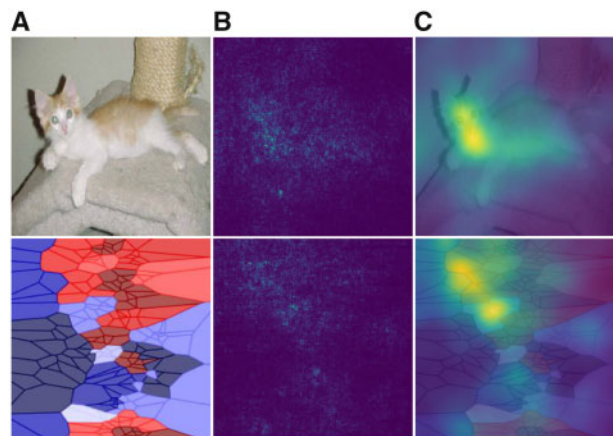


Fig. 5. Examples of saliency and heat maps generated with BionoiNet. (A) Input images, (B) saliency maps and (C) heat maps are shown for a picture of a cat (first row) and the Voronoi representation of a nucleotide-binding site (second row)

AUC of 0.960, which is higher than AUC values obtained for other classifiers.

Regardless of the classification task, BionoiNet outperforms Flat/MLP taking image pixels directly as input. CNN-based models, such as BionoiNet, employ filters with shared parameters on different locations of an image in order to extract local invariant features more easily and accurately, leading to better performance. BionoiNet is also more accurate than both autoencoder-based methods. The difference between the CNN- and autoencoder-based algorithms is that the former are trained in an end-to-end style, whereas the latter have two separate phases, the training of the autoencoder and the training of the actual classifier. Although the classifier could acquire some knowledge from data labels during training, the autoencoder was trained in an unsupervised manner, thus, it was not provided any labels during training. In contrast, since the CNN model was trained with an end-to-end supervised protocol, convolutional filters were able to learn more information from data labels during training. Another interesting observation is that all classifiers yield better results on the heme-control dataset compared to the nucleotide-control dataset. Many heme-binding pockets have a similar shape because the porphyrin ring of heme is rigid, while nucleotide-binding pockets are generally much more diverse in shape and physicochemical properties (Pu *et al.*, 2019). Since the spatial arrangement of atoms is preserved in the Voronoi diagrams of binding sites, classifiers can learn these patterns to yield better prediction accuracy on the heme-control dataset than on the nucleotide-control dataset. No discernible patterns were found to explain individual misclassifications.

### 4.3 BionoiScores for binding pockets

BionoiNet calculates BionoiScores from saliency maps constructed for input images. Two examples of saliency maps generated by BionoiNet are presented in Figure 5 for a picture of a cat (first row)

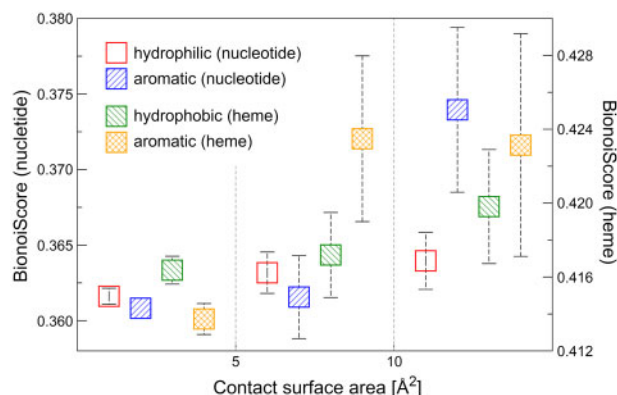


Fig. 6. Distribution of BionoiScores over nucleotide- and heme-binding residues. Binding residues are divided into three groups depending on the contact surface area with a ligand,  $<5 \text{ \AA}^2$ ,  $5\text{--}10 \text{ \AA}^2$  and  $>10 \text{ \AA}^2$ . For nucleotide-binding pockets, mean  $\pm$  standard deviation values are calculated separately for residues forming hydrophilic and aromatic interactions with nucleotides. For heme-binding pockets, these values are calculated for residues forming hydrophobic and aromatic interactions with heme molecules

and the Voronoi representation of a ligand-binding site (second row). Figure 5A shows the original images with the corresponding saliency maps presented in Figure 5B. Figure 5C shows heat maps constructed by overlaying saliency maps processed through a Gaussian filter on the original images. Heat maps provide better visualization by highlighting those regions that are important for image classification. It can be seen that the prediction for a cat (the first row of Fig. 5C) is triggered by different semantic regions of the image and the model learned to localize the common visual patterns, such as the animal face. According to the model, the remaining ‘cold’ regions covering a scratching post in the background are irrelevant for the classification.

The same principle can be applied to the Voronoi representations of ligand-binding sites. As shown in the second row of Figure 5, the saliency map (Fig. 5B) and the heat map (Fig. 5C) are generated by BionoiNet to highlight those locations that are important for pocket classification. We first identified binding residues forming the key interactions with nucleotide and heme molecules in our dataset. Here, we consider hydrophilic and aromatic interactions in nucleotide-protein complexes (Mao *et al.*, 2004) and hydrophobic and aromatic interactions in heme-binding proteins (Li *et al.*, 2011). Using LPC, we divided these residues into three groups depending on the contact surface area with a ligand,  $<5 \text{ \AA}^2$ ,  $5\text{--}10 \text{ \AA}^2$  and  $>10 \text{ \AA}^2$ , assuming that binding residues more strongly interact with a ligand and when the corresponding contact surface area is high. Figure 6 shows the distribution of BionoiScores over nucleotide- and heme-binding residues with respect to the interaction type and the contact surface group assignment. Interestingly, the average BionoiScores are low for residues assigned to the first group ( $<5 \text{ \AA}^2$ ) forming rather weak interactions with a ligand. BionoiScores start to be higher in the second group ( $5\text{--}10 \text{ \AA}^2$ ), particularly for those residues forming aromatic interactions with heme molecules. Residues with the

highest contact surface area in the third group ( $>10 \text{ \AA}^2$ ) have the highest BionoiScores. Note that BionoiNet computes BionoiScores from the conformation and physicochemical properties of pockets without any information on binding ligands. Two representative examples selected from the nucleotide-heme dataset are described in the [Supplementary Information](#). Overall, the analysis of the strongly discriminative regions of input images indicates that the attention of BionoiNet is on animal faces for the cat-dog dataset and on residues forming key interactions with ligands for the nucleotide-heme dataset.

## 5 Discussion

In this communication, we describe BionoiNet, a new deep learning-based framework to classify ligand-binding site structures based on their 2D representations. We demonstrate that BionoiNet significantly outperforms traditional machine learning methods providing a lightweight alternative to more sophisticated 3D CNN models, such as DeepDrug3D ([Pu et al., 2019](#)). This recently developed deep learning-based method first creates 3D voxel representations of ligand-binding sites and then performs binary classification with 3D CNN. DeepDrug3D was demonstrated to be more accurate than many other methods, including volume- and shape-based approaches, a classifier employing the histogram of gradients with principal component analysis (HOG/PCA), pocket matching with G-LoSA ([Lee and Im, 2016](#)), molecular docking with Vina ([Trott and Olson, 2010](#)) and sequence signature detection with ScanProsite ([de Castro et al., 2006](#)). Although BionoiNet also significantly outperformed all these baseline methods, its performance is slightly lower than that of DeepDrug3D. For instance, the AUC of DeepDrug3D against the nucleotide-control (heme-control) dataset is 0.986 (0.987), whereas the corresponding AUC of BionoiNet is 0.935 (0.960). It is important to note that both DeepDrug3D and BionoiNet have high precision and recall values at the same time indicating that these approaches effectively detect positive instances. The classification performance of DeepDrug3D is slightly higher because its 3D voxel representation of ligand-binding sites with 14 channels contains more information than 2D Voronoi diagrams with a single channel.

Nonetheless, BionoiNet has four major advantages over 3D methods. First, 2D images are more efficient in terms of computing time, memory and storage compared to 3D voxels. For example, calculating a 3D voxel requires 10–30 min on a single core and produces a binary file whose size is 3.7 MB. In contrast, Voronoi images are only about 14 kB in size and can be generated in a fraction of second, which is highly advantageous when working with large datasets. Second, off-the-shelf DNN architectures, such as ResNet ([He et al., 2016](#)), OxfordNet ([Simonyan and Zisserman, 2015](#)) and GoogLeNet ([Szegedy et al., 2015](#)), can directly be applied to classify binding sites eliminating the necessity to develop highly-specialized deep learning frameworks. Third, users can initialize those off-the-shelf models with publicly available pretrained parameters significantly reducing training time. Forth, the convolutional autoencoder implemented as part of the BionoiNet software outputs fixed-size feature vectors effectively encoding ligand-binding sites regardless of their size. Similar to ProtVec ([Asgari and Mofrad, 2015](#)) and Mol2vec ([Jaeger et al., 2018](#)) generating feature vectors for protein sequences and ligand chemical structures, BionoiNet constructs feature vectors specifically for binding pockets, which can then be used in other machine learning-based projects that require this kind of data. Although the current version is trained to recognize nucleotide- and heme-binding sites, BionoiNet will be extended to other ligand types by employing various data augmentation techniques to account for fewer structures currently available for certain complexes.

## Acknowledgements

Portions of this research were conducted with computing resources provided by Louisiana State University.

## Funding

This work has been supported in part by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM119524, by the US National Science Foundation award CCF-1619303, the Louisiana Board of Regents contract LEQSF(2016-19)-RD-B-03 and by the Center for Computation and Technology, Louisiana State University.

*Conflict of Interest:* none declared.

## References

- Altschul, S.F. et al. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Araki, M. et al. (2018) Improving the accuracy of protein-ligand binding mode prediction using a molecular dynamics-based pocket generation approach. *J. Comput. Chem.*, **39**, 2679–2689.
- Asgari, E. and Mofrad, M.R. (2015) Continuous distributed representation of biological sequences for deep proteomics and genomics. *PLoS One*, **10**, e0141287.
- Aurenhammer, F. (1991) Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Comput. Surv.*, **23**, 345–405.
- Brenke, R. et al. (2009) Fragment-based identification of druggable ‘hot spots’ of proteins using Fourier domain correlation techniques. *Bioinformatics*, **25**, 621–627.
- Brylinski, M. and Feinstein, W.P. (2013) eFindSite: improved prediction of ligand and binding sites in protein models using meta-threading, machine learning and auxiliary ligands. *J. Comput. Aided Mol. Des.*, **27**, 551–567.
- de Castro, E. et al. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.*, **34**, W362–365.
- Govindaraj, R.G. and Brylinski, M. (2018) Comparative assessment of strategies to identify similar ligand-binding pockets in proteins. *BMC Bioinformatics*, **19**, 91.
- He, K. et al. (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA, pp. 770–778.
- Jaeger, S. et al. (2018) Mol2vec: unsupervised machine learning approach with chemical intuition. *J. Chem. Inf. Model.*, **58**, 27–35.
- Jarrett, K. et al. (2009) What is the best multi-stage architecture for object recognition? In: *2009 IEEE 12th International Conference on Computer Vision*. Kyoto, Japan, pp. 2146–2153.
- Kamnitsas, K. et al. (2017) Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.*, **36**, 61–78.
- Kana, O. and Brylinski, M. (2019) Elucidating the druggability of the human proteome with eFindSite. *J. Comput. Aided Mol. Des.*, **33**, 509–519.
- Kawabata, T. (2011) Build-up algorithm for atomic correspondence between chemical structures. *J. Chem. Inf. Model.*, **51**, 1775–1787.
- Khazanov, N.A. and Carlson, H.A. (2013) Exploring the composition of protein-ligand binding sites on a large scale. *PLoS Comput. Biol.*, **9**, e1003321.
- Kingma, D.P. and Ba, J. (2015) Adam: a method for stochastic optimization. In: *Proceedings of 3rd International Conference on Learning Representations*. San Diego, CA, USA.
- Krizhevsky, A. et al. (2012) Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*. pp. 1097–1105.
- Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.*, **157**, 105–132.
- Le Guilloux, V. et al. (2009) Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics*, **10**, 168.
- LeCun, Y. et al. (2015) Deep learning. *Nature*, **521**, 436–444.
- LeCun, Y. et al. (1998) Gradient-based learning applied to document recognition. *Proc. IEEE*, **86**, 2278–2324.
- Lee, H.S. and Im, W. (2016) G-LoSA: an efficient computational tool for local structure-centric biological studies and drug design. *Prot. Sci.*, **25**, 865–876.
- Li, D. and Dong, Y. (2014) Deep learning: methods and applications. In: Li, D. and Dong, Y. (eds) *Found Trends Signal Process.* Now Publishers Inc, Hanover, MA, USA, Vol. 7. pp. 197–387.
- Li, T. et al. (2011) Structural analysis of heme proteins: implications for design and prediction. *BMC Struct. Biol.*, **11**, 13.
- Lipton, Z.C. et al. (2015) A critical review of recurrent neural networks for sequence learning. arXiv 2015.1506.00019.

- Mao, L. *et al.* (2004) Molecular determinants for ATP-binding in proteins: a data mining and quantum chemical analysis. *J. Mol. Biol.*, **336**, 787–807.
- Masci, J. *et al.* (2011) Stacked convolutional auto-encoders for hierarchical feature extraction. In: *International Conference on Artificial Neural Networks*. Springer, pp. 52–59.
- Miller, O.M. (1942) Notes on a cylindrical world map projection. *Geograph. Rev.*, **32**, 424–430.
- Najafabadi, M.M. *et al.* (2015) Deep learning applications and challenges in big data analytics. *J. Big Data*, **2**, 1.
- Neyshabur, B. *et al.* (2017) Exploring generalization in deep learning. In: *Advances in Neural Information Processing Systems*, pp. 5947–5956.
- Ngan, C.H. *et al.* (2012) FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic Acids Res.*, **40**, W271–W275.
- Pu, L. *et al.* (2019) DeepDrug3D: classification of ligand-binding pockets in proteins with a convolutional neural network. *PLoS Comput. Biol.*, **15**, e1006718.
- Qureshi, M.N.I. *et al.* (2019) 3D-CNN based discrimination of schizophrenia using resting-state fMRI. *Artif. Intell. Med.*, **98**, 10–17.
- Russakovsky, O. *et al.* (2015) Imagenet large scale visual recognition challenge. *Int. J. Comp. Vision*, **115**, 211–252.
- Schmidtke, P. and Barril, X. (2010) Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J. Med. Chem.*, **53**, 5858–5867.
- Simonyan, K. *et al.* (2014) Deep inside convolutional networks: visualising image classification models and saliency maps. In: *Proceedings of 2nd International Conference on Learning Representations*. Banff, AB, Canada.
- Simonyan, K. and Zisserman, A. (2015) Very deep convolutional networks for large-scale image recognition. In: *Proceedings of 3rd International Conference on Learning Representations*. San Diego, CA, USA.
- Skalic, M. *et al.* (2019) LigVoxel: inpainting binding pockets using 3D-convolutional neural networks. *Bioinformatics*, **35**, 243–250.
- Skolnick, J. *et al.* (2015) Implications of the small number of distinct ligand binding pockets in proteins for drug discovery, evolution and biochemical function. *Bioorg. Med. Chem. Lett.*, **25**, 1163–1170.
- Sobolev, V. *et al.* (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics*, **15**, 327–332.
- Srivastava, N. *et al.* (2014) Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, **15**, 1929–1958.
- Stank, A. *et al.* (2016) Protein binding pocket dynamics. *Acc. Chem. Res.*, **49**, 809–815.
- Szegedy, C. *et al.* (2015) Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA, USA, pp. 1–9.
- Szegedy, C. *et al.* (2016) Rethinking the inception architecture for computer vision. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA, pp. 2818–2826.
- Trott, O. and Olson, A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, **31**, 455–461.
- Wass, M.N. *et al.* (2010) 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res.*, **38**, W469–W473.
- Xu, B. *et al.* (2015) Empirical evaluation of rectified activations in convolutional network. In: *International Conference on Machine Learning, Deep Learning Workshop*. Lille, France.
- Yang, J. *et al.* (2013) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, **29**, 2588–2595.
- Zeiler, M.D. and Fergus, R. (2014) Visualizing and understanding convolutional networks. In: *European Conference on Computer Vision*. Zurich, Switzerland, pp. 818–833.
- Zhang, Y. and Skolnick, J. (2004) Scoring function for automated assessment of protein structure template quality. *Proteins*, **57**, 702–710.